






ARTICLE



<https://doi.org/10.1057/s41599-020-0500-x>

OPEN

# Gender gaps in urban mobility

Laetitia Gauvin<sup>1,8</sup><sup>✉</sup>, Michele Tizzoni <sup>1,8</sup>, Simone Piaggese<sup>1,2</sup>, Andrew Young<sup>3</sup>, Natalia Adler<sup>4</sup>, Stefaan Verhulst<sup>3</sup>, Leo Ferres <sup>5,6</sup> & Ciro Cattuto<sup>1,7</sup>

Mobile phone data have been extensively used to study urban mobility. However, studies based on gender-disaggregated large-scale data are still lacking, limiting our understanding of gendered aspects of urban mobility and our ability to design policies for gender equality. Here we study urban mobility from a gendered perspective, combining commercial and open datasets for the city of Santiago, Chile. We analyze call detail records for a large cohort of anonymized mobile phone users and reveal a gender gap in mobility: women visit fewer unique locations than men, and distribute their time less equally among such locations. Mapping this mobility gap over administrative divisions, we observe that a wider gap is associated with lower income and lack of public and private transportation options. Our results uncover a complex interplay between gendered mobility patterns, socio-economic factors and urban affordances, calling for further research and providing insights for policy-makers and urban planners.

<sup>1</sup>ISI Foundation, 10144 Turin, Italy. <sup>2</sup>Doctoral School in Data Science and Computation, University of Bologna, 40126 Bologna, Italy. <sup>3</sup>The Governance Lab, New York University, New York, NY 10003, USA. <sup>4</sup>United Nations International Children's Emergency Fund (UNICEF), New York, NY 10017, USA. <sup>5</sup>Data Science Institute, Universidad del Desarrollo, Santiago, Chile. <sup>6</sup>Telefónica R&D, Santiago, Chile. <sup>7</sup>University of Torino, 10124 Torino, TO, Italy. <sup>8</sup>These authors contributed equally: Laetitia Gauvin, Michele Tizzoni. ✉email: [laetitia.gauvin@isi.it](mailto:laetitia.gauvin@isi.it)

## Introduction

Cities, and how they are designed, are not gender-neutral. Consider daily mobility: simply moving around provides for different experiences depending on whether you are a woman or a man. Insecurity and the fear of physical or sexual violence in public spaces and when using public transportation are key factors that limit the everyday movement of women and girls (Loukaitou-Sideris, 2014). The location of bus stops, or how well-lit streets are, can also greatly affect women's movement. Women and girls also engage more in multi-purpose, multi-stop trips ('trip chaining') in order to do household chores, as well as other gender differentiated roles (Brown et al., 2014). Women-headed households are often more prominent in urban settings and they tend to work more in low-paid or informal jobs than men, with limited access to transportation subsidies (Tacoli, 2012).

At the same time, urbanization offers many possibilities to reduce gender gaps through a wealth of new opportunities. However, urbanization also increases inequalities by, for example, reinforcing geographical segregation, especially in developing-world settings (Chant, 2013). Mobility is a critical factor in reducing such segregation. Investigating the role of gender in urban mobility is key to better understanding of whether women and young girls can fully benefit from opportunities offered by cities, and in the process realize their human rights.

While several cities around the globe are starting to pay more attention to women's experiences, the unique mobility needs of women and girls are rarely taken into account in urban and transportation planning. One reason involves entrenched gendered power hierarchies, present in most societies (Uteng and Cresswell, 2016). Yet another reason involves the absence of robust data about the lives of women and girls - especially as it relates to daily mobility. As pointed out by the World Bank and by initiatives such as Data2X, studies on pressing issues such as fighting poverty and hunger or epidemics, among many others, have suffered from a lack of gender disaggregated data, effectively assuming that the problems of different genders are equivalent (Bank, 2011; Buvinic et al., 2014). Collapsing gender disparities prevents relevant organizations and policy-makers from getting a full picture of reality and ultimately limits the possibility of intervening to bridge the gendered gap that does exist. Non-existent and substandard data on gender differences have many drawbacks and effectively means that urban planning is often gender-blind (Buvinic and Levine, 2016). If there is no understanding on how women travel around differently, certain transportation options that could support women and girls may be overlooked.

Most traditional mobility studies are derived from surveys based on relatively few observations over a limited time span, or with low spatial resolution or other errors (e.g., due to self-reporting) (Groves, 2006; Couper, 2000; Clarke et al., 1981). More importantly, the question of how observed differences in mobility can be explained by innate sex-related differences, such as physical differences, or by gendered socially constructed factors, such as household roles, remains highly debated (Rosenbloom, 2004; Hanson, 2010). Also, long-term trends of gender differences in mobility are intertwined with global demographic and socio-economic trends and are therefore hard to capture. This is particularly true in urban areas, where the population is continuously and rapidly growing; according to the United Nations, the world's urban population is projected to account for almost 70% of total global population by 2050 (United Nations, 2018). Mobility is a complex issue and no single dataset or approach is sufficient to unpack its multidimensionality and offer insights on the way forward for decision-makers. In addition, the Data2X report on Big Data and the Well Being of Girls points out that much of the

data that could provide new insights on these issues is collected by corporations, and is therefore often not available to researchers and public policymakers (Vaitla et al., 2017).

Nowadays, with the pervasiveness of mobile devices, it has become possible to achieve large-scale urban sensing and explore the mobility of individuals at unprecedented scale (Blondel et al., 2015; Naboulsi et al., 2016). In the past 10 years, many studies have successfully used Call Details Records (CDRs) to extract and analyze human mobility patterns (González et al., 2009; Song et al., 2010; Calabrese et al., 2013; Beiró et al., 2018; Graells-Garrido et al., 2017). However, even if gender differences in mobility patterns derived from mobile phone data have been sometimes investigated, they have been mostly considered as somewhat peripheral (Song et al., 2010) or comparatively small-scale (Pssylla et al., 2017).

In this work, we study urban mobility from a gendered perspective in the greater metropolitan area of Santiago, Chile. With almost 7 million inhabitants, Santiago's metro is one of the largest metropolitan areas of South America, and like many cities across the continent, Santiago continues to expand and sprawl (Lovera, 2015). Its population already accounts for about 40% of the national population and it is projected to grow steadily in the next decades (Puertas et al., 2014) posing many challenges to urban planners and policy makers, especially related to the design and adaptation of new and existing transportation infrastructures to serve the complex mobility needs of its inhabitants.

Our study had two main objectives. First, to assess and quantify gender disparities in the mobility patterns of Santiago residents; and, second, to identify socio-demographic factors and the availability of transport options that are associated with mobility inequalities. To achieve these objectives, we analyzed the mobility traces extracted from Call Detail Records (CDRs) of a large cohort of anonymized mobile phone users disaggregated by sex (male or female) over a period of 3 months. We then mapped indicators of mobility differences between males and females to 51 *comunas* (Spanish for municipalities) of the Santiago Metropolitan Region and we investigated the association between mobility inequalities and socio-demographic indicators in different areas of the city, as well as their relationship with the Santiago transportation network structure. Finally, we added a "semantic layer" to the mobility patterns of Santiago residents by identifying specific points of interests that are more frequently present along women's or men's trajectories in the urban space, thus demonstrating how our approach can identify specific gendered mobility needs.

## Background

Previous literature has investigated the complex relation between gender and mobility, with particular attention to gender differences in transport mode choices. An important study by Robin Law (Law, 1999) provides an extensive overview of the research on gender and mobility that has been conducted since the early 70s and that spans several disciplines, ranging from social sciences to geography, to environmental studies. Over the past three decades, feminist scholars have provided significant evidence supporting the claim that "how people move (where, how fast, how often) is demonstrably gendered" (Cresswell and Uteng, 2016). In general, several empirical studies conducted in both developed and developing countries have found that women tend to travel shorter distances, to chain trips, to spend more time traveling and to prefer public transport and taxi services to cars more than men (Ng and Acker, 2018).

The large majority of studies on women and transport have solely relied on the analysis of survey data, using multinomial

logistic regressions to estimate the effect of gender on mobility variables. While surveys can provide a detailed description of individual attitudes and behaviors, they are usually costly, hard to conduct on a regular basis and limited by a small sample size. On the other hand, in the past decade, mobile phone data have been demonstrated to provide valuable, timely, fine-grained information on human mobility, from citywide to nationwide spatial scales (Blondel et al., 2015). CDRs have rapidly become a standard for mobile phone data used in human mobility studies. Although CDRs come with limitations due to their sparsity, previous studies have shown that CDRs can be used to infer origin–destination trips by purpose and time of day (Alexander et al., 2015) and to model travel demand on road networks (Çolak et al., 2016). Also, individual daily trajectories inferred from CDRs are consistent with those measured by household surveys (Schneider et al., 2013). To what extent gender inequalities in urban mobility can be investigated using mobile phone data, remains an open research question.

Our study seeks to understand gender differences in the mobility of mobile phone users, making two main assumptions: (i) the sex recorded in a subscriber profile corresponds to the sex of the person making the calls; (ii) the mobility inferred from CDRs is representative of a user mobility behavior. Both assumptions are reasonable in the context of urban Chile, a country that is characterized by a high-income economy according to the definition of the World Bank. Mobile phone technology is widely adopted in Chile, with 27.5 million mobile connections (153% penetration) and a unique subscriber penetration level above 90%, the highest in Latin America (GSMA, 2020). Also, previous studies have analyzed mobility patterns in Chile, and more specifically in Santiago, showing that the analysis of CDRs can provide insights into the daily movements of the residents of Santiago (Graells-Garrido et al., 2018).

## Methods

**Mobile phone data and metadata.** The mobile phone data set includes three months (May, June, July 2016) of anonymized (hashed) CDRs enriched with gender, socioeconomic segment, and number of phone lines registered under that number for a total of 2,148,132,995 events (calls only). We filtered these events to those phone numbers that had only 1 registered line, placed at least one call per day on average (91 calls in total), had an identifiable home location, and visited more than two distinct locations over three months. This yielded a total of 418,624 unique users. The associated socioeconomic segment was known for 315,844 users. For all users, a binary gender value (F or M) was provided by the operator based on the information provided by the user at the time of subscription. In our sample, 51% of the users are females.

The socioeconomic status (GSE) of mobile phone users was classified by the phone carrier based on the GSE definition by the Chilean *Asociación de Investigadores de Mercado* (AIM) using a pay slip that cell phone users have to provide at the time of hiring the services of the telephone company. The socioeconomic groups are: Upper Class (AB), Wealthy Middle Class, (C1a), Emerging Middle Class (C1b), Typical Middle Class (C2), Medium-low Class (C3), Lower Class (D), Poores (E). Since the GSE definitions can vary year by year, we divided our population into 2 socioeconomic segments corresponding to the upper class (AB,C1a, C1b,C2) and the lower class (C3, D, E). Our approach can be interpreted as dividing the population based on a reference household income,  $I^*$ , roughly corresponding to 1 M Chilean pesos per year (1500 USD). Then, for each spatial unit under study (tower cell or *comuna*), we define the GSE ratio, that is the ratio between the resident population whose income is below the

reference one, denoted as  $P(I < I^*)$ , and the population whose income is above  $I^*$ ,  $P(I > I^*)$ .

We inferred users' mobility patterns by extracting the locations of each call placed or received, during the study period. Beside anonymization, to preserve the privacy of the users, the analysis of CDRs was carried out by spatially aggregating users' visited locations. More specifically, there are more than 1300 towers (called Base Transceiver Stations, or BTSs) in urban Santiago for which we have access to the latitude and longitude rounded at the second decimal place. Due to this rounding, multiple antennas appear as having the same coordinates and they are merged into a single tower. This leads to a grid of 726 cells regularly spaced in which the center is given by the position (rounded) of the tower, and users' trajectories were aggregated at this spatial resolution.

We assigned a home location to each user based on user's calling activity. We define a user's home to be the most visited cell tower during the time interval 7 p.m.–8 a.m., over the whole 91 days of data collection. Such approach, which can be denoted as “time constrained home detection”, has been shown to outperform several alternatives when tested on ground truth data (Vanhoof et al., 2018).

**Ethical considerations.** We are aware of the concerns related to users' privacy when conducting research on CDRs data (de Montjoye et al., 2018). This research was solely based on the analysis of anonymized data, after taking a number of precautions to ensure an appropriate protection of users' privacy and address any associated risks. The mobile phone numbers of subscribers making and receiving calls was anonymized by the mobile phone operator inside their premises, through a hashing process using the secure SHA-3 algorithm. Anonymized CDR data were never transferred outside of the operator's system. Analysis of CDR data took place on the mobile operator's systems and only the output of the analysis (aggregated population estimates and indicators) was subsequently made available to those researchers who were not based in Chile. The analysis never singled out identifiable individuals and no attempts were made to link CDR data to third party data about an individual. No reports were made from towers that display fewer than three unique phone numbers and we did not work with antennas themselves, but aggregated all towers using rounded coordinates. Moreover, there is a natural privacy-preserving mechanism inherent in the data themselves: even at the finest level of granularity (the antenna), devices are never connected to the same antenna all the time, but rotate according to antenna demand (peak vs. regular times), time-of-day (some antennas get turned off at certain times), azimuth, etc. which makes extremely hard to identify a single user based on the location data that were analyzed in this work.

**Census data.** The census data reports social, economic and demographic data about 17,574,003 people and 6,499,355 households in Chile collected by the *Instituto Nacional de Estadísticas* (INE) in 2017 for the National Census (*Censo de Población*). Data is publicly available at <https://www.censo2017.cl/>. The census questionnaire is structured into three sections with questions about houses (*Vivienda*), resident households (*Hogares*) and people belonging to each household (*Personas*). Each questionnaire includes the location of the corresponding *vivienda* at different administrative levels for the whole country. For our purposes we restricted our data analysis to the 52 *comunas* of the Santiago Metropolitan Region. For this study, in each *comuna*, we extracted the following features that are reported in Table 2:

- the *employment gender ratio* is defined as the ratio between (Employed Women/Total Women) and (Employed Men/Total Men);

- the *higher education gender ratio* is defined as the ratio between the proportion of women and men who completed a higher education level course: (High Edu Women/Total Women) and (High Edu Men/Total Men);
- the *secondary education gender ratio* is defined as the ratio between the proportion of women and men who completed a secondary education level course;
- the *primary education gender ratio* is defined as the ratio between the proportion of women and men who completed a primary education level course;
- the *general fertility rate* is defined as (Total Births/Total Women age 15–49)  $\times 1000$ ;
- *couples households* is the number of households composed by the householder and a partner without children;
- *extended household* is the number of households composed by a nuclear family plus other family members of the householder;
- *family household* is the number of households composed by the householder, a partner and their children;
- *single parent household* is the number of households composed by one householder and his/her children;
- *single person households* is the number of households composed only by the householder.

**GTFS data.** The public transportation datasets of Santiago are provided in General Transit Feed Specification (GTFS) format. GTFS is a common format to represent public transportation schedules. It is composed of a series of text files related to the routes and the stops of the public transportation network. Here we consider only the file containing the GPS coordinates of the stops which enables us to identify cells of Santiago without access to public transportation stops. The data is publicly available at <http://datos.gob.cl/dataset/33245/>.

**OpenStreetMap points of interest.** Data about point of interests (POI) were downloaded from OpenStreetMap (OSM), a collaborative project to create a free editable map of the world, through the Overpass API<sup>1</sup>. We collected OSM map features inside the CDR data boundaries, and filtered them considering only *nodes* features, i.e., points with a geographic position, stored as pairs of latitude-longitude.

We focused on those with the *amenity* tag, which are the most common type of POI, representing facilities used by visitors and residents, and we selected the POIs that have at least 50 locations or more in Santiago for a total of 28 different POI types.

In addition to the OpenStreetMap “amenity” POIs, we considered a few POI types from other data sources.

- **Malls** (“*mall*” POI type). This POI layer includes all tower locations  $l_i$  falling within the polygon describing the perimeter of mall complexes in the map of Greater Santiago (according to OpenStreetMap).
- **Subway stops** (“*metro*” POI type). This POI layer includes all metro stop locations in the Metro Area of Santiago. Data is made publicly available by the *Observatorio de Ciudades*, at the Faculty of Architecture of the Catholic University of Chile: <http://ideocuc.cl/maps/162/download>.
- **Collective taxis** (“*colectivos*” POI type). This POI layer includes locations along the routes of shared taxis in Santiago (“*colectivos*”), obtained by scraping the geographical information used to draw taxi routes on the Web site [http://www.ubicatucolectivo.cl/cliente\\_final/all\\_lines/vercion\\_1.php?id=14](http://www.ubicatucolectivo.cl/cliente_final/all_lines/vercion_1.php?id=14).

To aid with the interpretation of our results and to establish suitable baselines, we also create two synthetic POI types that we use as reference:

- **Tower clusters** (“*towers*” synthetic POI type). This POI type comprises all the locations of telephone towers (tower clusters) we use in our analysis. We know that towers are not uniformly distributed in space and that they are usually positioned following user density, hence their distribution contains information about the spatial distribution of the population in Santiago, and it is important to compare against this reference layer the results we obtain for other POI types, to rule out that the results are not purely determined by the spatial distribution of the population and of the towers.
- **Uniform grid** (“*uniform*” synthetic POI type). This POI type contains all the vertices of the two-dimensional spatial grid we use here (0.01 degree step long both the latitude and longitude coordinates), restricted to a rectangular area that covers the Greater Santiago Area. By design, this set of locations has no spatial structure (except for the size effects of the rectangular region for large distance).

In total, in our analysis we consider 33 different types of points of interest.

**Gender differences in visit patterns.** We measured the excess ratio (women to men) of visits to locations characterized by the presence of specific category of POIs in the following way.

For each POI type, we follow a standard non-parametric approach based on (multivariate) Kernel Density Estimation (Hwang et al., 1994) to compute a POI density over the entire spatial domain under study. Namely, let us consider a given POI type  $k$  comprising  $N_k$  POIs at locations  $\{\vec{x}_1^{(k)}, \vec{x}_2^{(k)}, \dots, \vec{x}_{N_k}^{(k)}\}$ , where the POI vector components are the latitudes and the longitudes of each POI. We define the density of POI  $k$  at location  $\vec{x}$  as

$$\rho_k(\vec{x}) = \frac{1}{N_k} \sum_{i=1}^{N_k} K \left[ D(\vec{x}, \vec{x}_i^{(k)}); d \right], \quad (1)$$

where  $K$  is a normalized non-negative kernel, the bandwidth  $d > 0$  is a kernel parameter that defines the spatial scale over which the density distribution is smoothed, the function  $D(\vec{x}, \vec{x}_i^{(k)})$  is the great circle distance of the positions vectors  $\vec{x}$  and  $\vec{x}_i^{(k)}$ . We use a simple isotropic Gaussian kernel defined as

$$K(D; d) = \frac{1}{d\sqrt{2\pi}} e^{-\frac{1}{2}(D/d)^2}, \quad (2)$$

and since we want to study potential gender differences at all spatial scales, we will not select a specific bandwidth value  $d$ , but rather carry out our analysis for a broad range of values, from slightly below the spatial resolution of our CDR data (hundreds of meters) to the size of the entire city (tens of kilometers).

Given a user  $u$  and the POI densities defined above, we can define a user-level POI density by averaging over all locations visited by  $u$ :

$$\rho_k^{(u)} = \left\langle \rho_k(\vec{l}) \right\rangle_{\vec{l} \in L(u)}. \quad (3)$$

That is,  $\rho_k^{(u)}$  is the density of POI type  $k$  averaged over all locations  $L$  visited by user  $u$ . Notice that this density has an implicit dependency on the bandwidth parameter  $d$  of the kernel density estimator.

Finally, we carry out averages of the user-level POI densities separately over all males and female. That is, indicating with  $U_F$  and  $U_M$  the set of all female and male users, respectively, we define:

$$\rho_k^F = \left\langle \rho_k^{(u)} \right\rangle_{u \in U_F}, \quad (4)$$

and similarly

$$\rho_k^M = \left\langle \rho_k^{(u)} \right\rangle_{u \in U_M} \tag{5}$$

Finally, we define the gender density ratio

$$r_k = \rho_k^F / \rho_k^M, \tag{6}$$

which, for each POI type  $k$ , is meant to indicate gender imbalances in visits to location associated with that specific POI type. This ratio also has an implicit dependency on the bandwidth parameter  $d$  used for kernel density estimation, hence we study  $r_k$  as a function of both the POI type  $k$  and of the smoothing distance  $d$ .

To assess statistical significance, we compare the observed value of  $r_k$  with those obtained by using two spatial null models that randomly generates synthetic POI locations with two methods:

- **Spatial sampling** Given a POI type  $k$  that comprises  $N_k$  POIs, we generate a realization of the null model by sampling (with repetition)  $N_k$  towers according to the probability  $p_i$  that a tower  $i$  is visited by any user, i.e., by the fraction of CDR records that tower (tower cluster) is associated with. Subsequently, we perturb the locations so that they no longer lie on the regular grid, by adding a uniformly distributed random variate in the range  $[-0.01, +0.01]$  to both the latitude and the longitude. For each POI type  $k$  we generate 100 realizations and compute the ratio  $r_k$  for each realization. Finally, we compute the 95% confidence interval for the distribution of  $r_k$  generated by the null model.
- **Spatial perturbation** For each POI type  $k$  we build a spatially-perturbed version of the same set of POIs by randomly adding or subtracting an offset of 0.01 from the latitude and longitude of each POI. We treat the perturbed POIs as a new POI type  $k'$  and compute the gender ratio  $r_{k'}$  for all values of  $d$ , as described above.

**Mobility metrics.** We analyzed user’s mobility computing four different mobility metrics, first defined at individual level and then aggregated by averaging over users with same gender, GSE or home location.

As a basic measure of mobility behavior, we computed the number of distinct locations visited by a user,  $N_b$  that corresponds to the number of distinct grid cells in which a user made or received at least one call over the 3 month period. Given that  $N_l$  can vary significantly between users and it is affected by fluctuations of user’s activity, we also computed the “core activity locations”,  $\hat{N}_l$ , defined as the set of locations that account for 80% of a user’s calling activity.

To quantify the diversity of individual mobility, we calculated the Shannon entropy of user’s trajectories as:

$$S = - \sum_{l \in L} p_l \ln p_l, \tag{7}$$

where  $L$  is the full set of locations visited by a user, and  $p_l$  is the the probability of observing a user in  $l$ , computed as the fraction of calls made by the user at location  $l$ . A user with high  $S$  will distribute her visits across many different locations with equal probability, while a lower  $S$  corresponds to a higher regularity of mobility patterns with a smaller set of regularly visited locations (Song et al., 2010).

Finally, we measured the radius of gyration of each user,  $r_g$ , which quantifies the characteristic distance traveled by an

individual. It is defined as:

$$r_g = \frac{1}{L} \sqrt{\sum_{i=1}^L (\mathbf{r}_i - \mathbf{r}_{cm})^2} \tag{8}$$

where  $L$  is the full set of locations visited by a user,  $\mathbf{r}_i$  is the vector of coordinates of location  $i$  and  $\mathbf{r}_{cm}$  is the vector of coordinates of the center of mass, weighted by the visiting frequency  $p_i$ .

For each mobility metric, we carry out averages of the user-level metrics separately over all males and females who live in a given location  $l$  (*comuna* or cell). That is, indicating with  $U_{F,l}$  and  $U_{M,l}$  the set of all female and male users who live in  $l$ , respectively, we define the average of the mobility metric  $x$ :

$$x_F = \left\langle x^{(u)} \right\rangle_{u \in U_{F,l}}, \tag{9}$$

and

$$x_M = \left\langle x^{(u)} \right\rangle_{u \in U_{M,l}}. \tag{10}$$

Finally, to measure the gender gap for a given mobility metric  $x$ , we measure the gender ratio:

$$R_x = x_F / x_M. \tag{11}$$

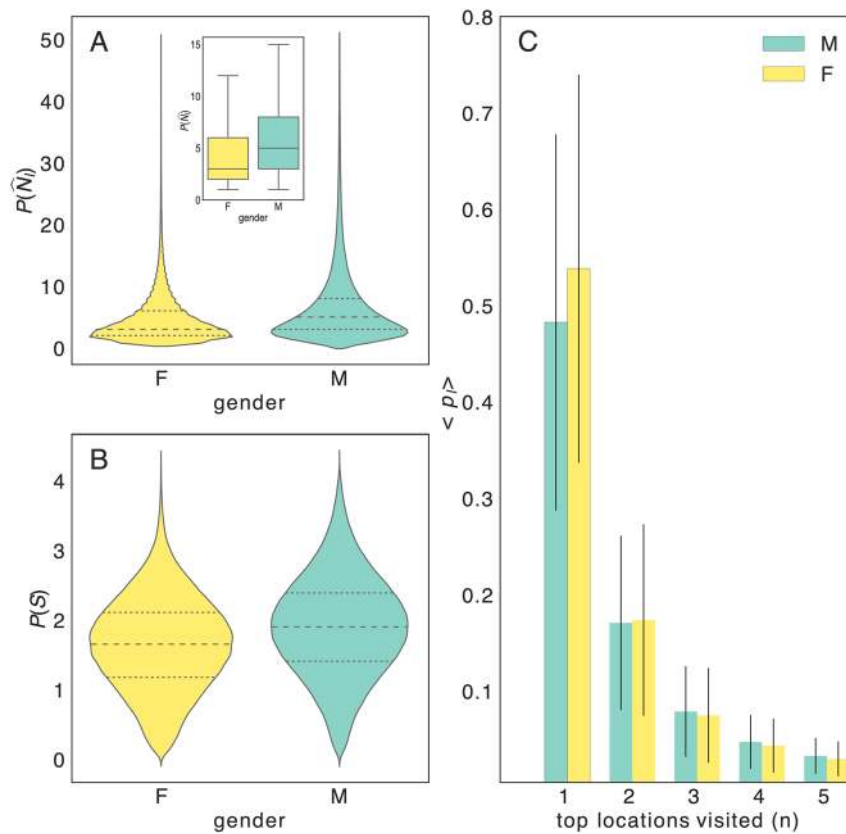
**Results**

**Gender inequalities in mobility.** We analyze the mobility patterns of 418,624 individuals extracted from about 2 billion anonymized Call Detail Records (CDRs) collected between May 1 and July 30, 2016. Our anonymized users’ sample carries information about users’ sex and socioeconomic status (see Materials and Methods for definitions). The sample is highly representative of the socio-demographic structure of the Santiago Metropolitan Region in terms of population, gender ratio and socioeconomic group distributions at the level of *comuna* (see Fig. S1 in the Supplementary Materials).

We assess gender differences in mobility by computing four mobility metrics for each individual and by evaluating the gender effect size through estimation statistics (Ho et al., 2019). We characterize the mobility behavior of Santiago residents by looking at: (i) the number of distinct locations visited by a user during the study period,  $N_b$ , (ii) the number of distinct locations that account for at least 80% of a user calling activity,  $\hat{N}_l$ , (iii) the Shannon mobility entropy,  $S$ , and (iv) the radius of gyration,  $r_g$ . To distinguish locations, the study area has been divided into 726 cells of about 1 km<sup>2</sup> regularly spaced according to the position of the cell towers as explained in the Materials and Methods. In the following, the words *location* and *cell* are used interchangeably.

We first look at gender differences in the total number of unique locations visited by the users. We observe that women travel to fewer unique locations than men. Specifically, considering the complete set of locations visited by a user, we find that—over 3 months—women have traveled about nine locations less than men, on average:  $\Delta N_b = \langle N_b \rangle_M - \langle N_b \rangle_F = 8.57$ , 95% CI [8.42, 8.72]. This means that, when compared to the average  $N_b$ , men have visited 30% more locations than women on average ( $\langle N_b \rangle_M = 37.05$ ,  $\langle N_b \rangle_F = 28.48$ ). If we only look at unique locations that belong to the core of users’ daily activity,  $\hat{N}_l$ , the difference between men and women becomes  $\Delta \hat{N}_l = 2.13$ , 95% CI [2.09, 2.16] as shown in Fig. 1a. Such a difference corresponds to a 45% increase in  $\langle \hat{N}_l \rangle_M$  with respect to  $\langle \hat{N}_l \rangle_F$ , which are equal to 6.92 and 4.79, respectively.

Another key aspect that characterizes human mobility is the average distance traveled by an individual. In our sample, the average radius of gyration of women,  $\langle r_g \rangle_F$ , is 1.09 Km (95% CI



**Fig. 1 Distributions of mobility metrics by gender.** Violin plots show the distributions by gender of the number of locations accounting for 80% of a user's activity (a) and the users' Shannon mobility entropy (b). Women visit fewer locations and their movements are characterized by a smaller entropy. Panel c shows the distributions of the mean probability of visiting the 5 most frequented locations of each user, by gender. Error bars correspond to the standard deviation of the mean. The inset of panel a shows the boxplot of  $P(\hat{N}_i)$  by gender. Whiskers correspond to the 95% reference range. Outliers are not shown.

[1.07, 1.12]) shorter than the average men's radius of gyration,  $\langle r_G \rangle_M$ . Thus, women movements tend to be more spatially localized than men, as shown by the distributions of  $r_g$  by gender (see Fig. S2 in the Supplementary Materials).

We then look at the diversity of mobility patterns, measured by the Shannon entropy. Figure 1b shows that women movements are consistently characterized by a smaller entropy compared to men,  $\Delta S = 0.26$  (95% CI [0.26, 0.27]), indicating that women distribute their trips among a few highly preferred locations, while men distribute their trips among many locations with almost equal probability. Accordingly, we observe that women can be more frequently found at their most visited location. Indeed the frequency rank plot of all visited locations for all users (Fig. 1c) shows a higher frequency of visits,  $\langle p_i \rangle$ , for women's first and second ranked locations, while the women to men ratio of  $\langle p_i \rangle$  reverses from the third ranked location onward. Mean values and 95% reference ranges of all the 4 spatial mobility metrics are reported in Table S1 of the Supplementary Materials.

Since we infer mobility patterns from users' calling activity, it is natural to ask whether the observed disparities are due to gender differences in mobile phone usage. Indeed, in our sample, women call less frequently than men, for an average of 4.73 calls/day compared to 5.21 calls/day. We answer this question in two ways (see Text A of Supplementary Materials for more details). First, we restrict our analysis to users' with the highest calling activity, i.e., those who made at least 3 calls per day on average over the 3 month period, thus increasing the number of observations for men and women in our dataset. Focusing on the most active users, of which 51% are women, the gender differences in mobility become larger than those observed with the original user

sample, with a smaller CI ( $\Delta \hat{N}_i = 2.73$ , 95% CI [2.68, 2.79]). Second, we down-sample the call activity of men by removing up to 50% of the calls made by men in our original dataset and recomputing the Shannon entropy of all users. Even after doing so, women's entropy remains significantly smaller than men's. We also check the robustness of our findings when varying the time frame of our analysis. Overall, all users' mobility metrics, entropy and number of visited locations, display a high stability (Pearson correlation  $r = 0.81$ ,  $p < 10^{-3}$ ), when measured over time windows of 9 days and considering each interval separately (see Text A of the Supplementary Materials). Also, women and men distribute their calls through the day in the same way, displaying a very similar activity pattern by hour (see Fig. S3 in the Supplementary Materials), hence we can rule out that the observed gender differences in entropy are trivially associated with gender differences in temporal activity patterns.

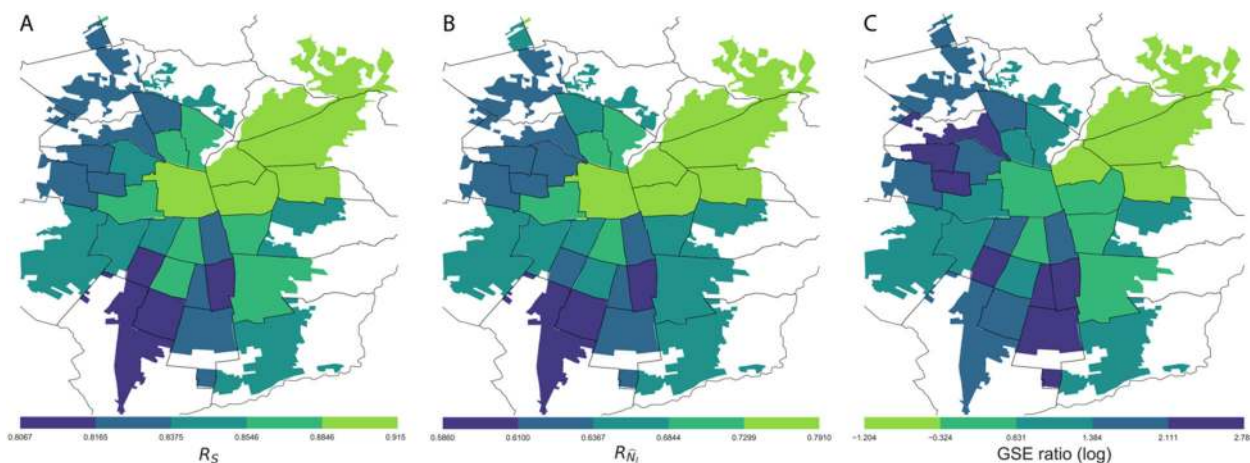
**Gendered mobility and socio-demographic indicators.** To better understand what are the factors underlying the observed gender disparities in mobility, we examine their relationship with a number of socio-demographic indicators.

First, we examine the relationship between the socioeconomic status of the users and their mobility metrics. To this aim, we evaluate the gender differences in mobility through estimation statistics after splitting the users into 5 socioeconomic groups ("Grupos Socio-Económicos" or GSE, in the Spanish acronym) that are defined by income and household status (see Materials and Methods for the exact definition of GSE and Fig. S4 in the Supplementary Materials for the distribution of users by GSE and

**Table 1 Estimation statistics of gender differences in mobility, disaggregated by users' socioeconomic group (GSE).**

	$\Delta S$	$\Delta \hat{N}_f$	$\Delta N_f$	$\Delta r_g$
ABC1	0.19 [0.17-0.20]	1.66 [1.53-1.79]	7.28 [6.72-7.82]	0.83 [0.75-0.90]
C2	0.23 [0.22-0.24]	2.08 [1.99-2.18]	8.54 [8.14-8.95]	0.94 [0.88-1]
C3	0.28 [0.27-0.29]	2.50 [2.40-2.59]	10.16 [9.78-10.53]	1.15 [1.09-1.21]
D	0.30 [0.30-0.31]	2.53 [2.45-2.61]	10.22 [9.87-10.53]	1.26 [1.21-1.32]
E	0.31 [0.29-0.33]	2.50 [2.32-2.66]	9.82 [9.16-10.48]	1.36 [1.25-1.47]

GSE are ranked from the highest (ABC1) to the lowest (E). For each metric  $X$ , the effect size  $\Delta$  is measured as  $\Delta X = (X)_M - (X)_F$ . Square brackets report 95% CIs.



**Fig. 2 Spatial patterns of gender mobility inequalities and wealth.** Choropleth maps of the metropolitan area of Santiago showing the women to men ratio of entropy (a) and of the number of locations accounting for 80% of users' activity (b) by *comuna*. Panel c shows the spatial distribution of the GSE ratio. Black lines indicate the administrative boundaries of the *comunas*. The boundary of the colored area corresponds to the urban area of Santiago as defined by the National Institute of Statistics (INE).

gender). Table 1 shows that the gender gap in all the mobility metrics widens as the socioeconomic status of the users decreases (ABC1 being the wealthiest segment and E being the most deprived). For instance,  $\Delta \hat{N}_f$  grows from 1.66 for users in the ABC1 group to 2.50 for those in the group E. In other words, poorer women tend to remain more localized than their male counterparts. Interestingly, although smaller, the gender gap in mobility is observed even among users who belong to the wealthiest class, thus indicating that a full mobility equality is not achieved even in presence of a high income.

To further investigate how gender differences in mobility vary across different socioeconomic strata of the population, we perform a spatial analysis of the gender mobility gap by first assigning a home location to each user, based on their call activity (see Materials and Methods for details). In this way, we are able to examine gender mobility disparities at two resolutions: a very fine “grid” resolution given by the cell locations and a coarser resolution corresponding to the *comunas* of the Santiago Metropolitan Region.

We measure the gender mobility gap in a given location or *comuna* by computing the women to men ratio of two mobility metrics,  $S$  and  $\hat{N}_f$ , averaged over all users who live in that location (see Materials and Methods). We denote the women to men ratios as  $R_S$  and  $R_{N_f}$ , respectively. Panels A and B of Fig. 2 show choropleth maps of the Santiago Metropolitan Area displaying the spatial variation of  $R_S$  and  $R_{N_f}$  across 30 *comunas* of the urban Santiago. The map in panel C of Fig. 2 shows the spatial distribution of wealth by *comuna*, measured by the GSE ratio, that we define as the ratio between the number of residents

belonging to the socioeconomic groups C3, D, and E, and the number residents belonging to a higher GSE (see Materials and Methods for the exact definition of GSE ratio). In general, Santiago displays a high level of segregation. The wealthiest *comunas*, characterized by a lower GSE ratio are mostly located in northwestern Santiago, while areas in the eastern and the southern outskirts are home to the poorest residents and display a high GSE ratio. A similar segregation pattern is also evident in the spatial distribution of the gender mobility gap: the gap increases significantly when moving from the wealthiest to the most deprived *comunas* of Santiago. In particular, we measure the semi-partial Pearson correlation coefficient (Fisher, 1924) between  $R_S$ ,  $R_{N_f}$  and the GSE ratio, by controlling for the variations in the call activity by gender and the differences in the sex ratio across *comunas*. Both  $R_S$  and  $R_{N_f}$  are strongly and negatively correlated with the GSE ratio, with correlation coefficients  $r = -0.59$  ( $p < 0.001$ ) and  $r = -0.53$ , ( $p < 0.001$ ), respectively. Thus as the GSE ratio decreases, in the wealthiest areas of Santiago, both  $R_S$  and  $R_{N_f}$  converge to the unit value, corresponding to gender equality in mobility (see Fig. S5 in the Supplementary Materials for the scatterplot of  $R_S$  and  $R_{N_f}$  against the GSE ratio).

We further investigate the relationship between the socio-demographic characteristics of the Santiago Metropolitan Region and the gender differences in mobility, by considering 12 census variables as predictors of the gender gap (see Materials and Methods for the complete definitions of the variables).

We select all variables available from census which represent gender disparities that may relate to the gender mobility gap, based on previous findings in the literature. In particular, we

**Table 2 Semi-partial correlation values (Pearson) between  $R_S$  and  $R_{N_i}$  and the sociodemographic features of 51 municipalities in the SMR.**

	$R_S$	$R_{N_i}$
GSE ratio (log)	-0.59***	-0.53***
HDI	0.42**	0.37**
Primary education gender ratio	0.00	-0.05
Secondary education gender ratio	0.24	0.34*
Higher education gender ratio	-0.08	-0.10
Employment gender ratio	0.51***	0.37**
General fertility rate	-0.53***	-0.40**
Couples household	0.55***	0.50***
Extended household	-0.61***	-0.57***
Family household	-0.30	-0.14
Single parent household	-0.32*	-0.32*
Single person household	0.56***	0.44**

All correlation values are computed by correcting for differences in calling activity and users distribution by gender. We report statistical significance using the following notation: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

consider measures of gender inequality in education attainment and employment, together with household structure.

We assess the magnitude and the statistical significance of correlations between census variables and the gender gap in two ways. First, we compute the semi-partial Pearson correlation coefficients between census variables and both  $R_S$  and  $R_{N_i}$  in 51 *comunas* of the SMR, for which we have at least 1,000 users, controlling for gender differences in calling activity and population distribution at the same time. Results are shown in Table 2.

Second, we assess the statistical significance of a standard Pearson correlation between the same set of variables, by comparing the observed correlation against 1000 instances of a null model where users are randomly assigned to one of the 51 *comunas*, only preserving the gender distribution of users in each *comuna* (see Table S2 in the Supplementary Materials).

Both approaches provide a consistent picture of what are the most significant socio-demographic predictors of the gender gap. The gender gap in mobility is significantly correlated with the gender gap in employment, suggesting that employment status may explain the observed differences in mobility behavior. On the other hand, gender differences in education levels, from primary to higher education, are not significantly associated with the gender gap in mobility. Childcare duties are often considered a significant cause of mobility inequalities for women. Indeed, we find that a higher fertility rate is associated with a larger gender mobility gap. If we look at the household structures in different *comunas*, we find that a higher presence of large households (i.e., including dependent relatives or children) tends to be associated with a higher inequality between women and men in terms of mobility, with women staying more put, probably to bear the brunt of childcare duties. Conversely, in those *comunas* where a larger proportion of households is formed by a single person (either a woman or a man), mobility patterns of men and women are more similar. But instead, a higher proportion of single parents living with their children—which can be reasonably thought of being mostly single mothers—is associated to a larger gender difference in mobility patterns.

**Gendered mobility and access to transport.** Access to different means of transportation plays a crucial role in determining

individual mobility patterns. We investigate the relationship between the accessibility of public transport and the presence of gender disparities in mobility, by mapping the stops of the Santiago public transportation network (trams, buses, and metro) onto the gendered mobility patterns of mobile phone users. To this aim, we characterize each cell of the SMR by the presence of at least one Transantiago stop, the public transport system that serves Santiago (see Materials and Methods). We then measure the gender effect on mobility through estimation statistics comparing two groups: the residents who have access to public transport and those who don't.

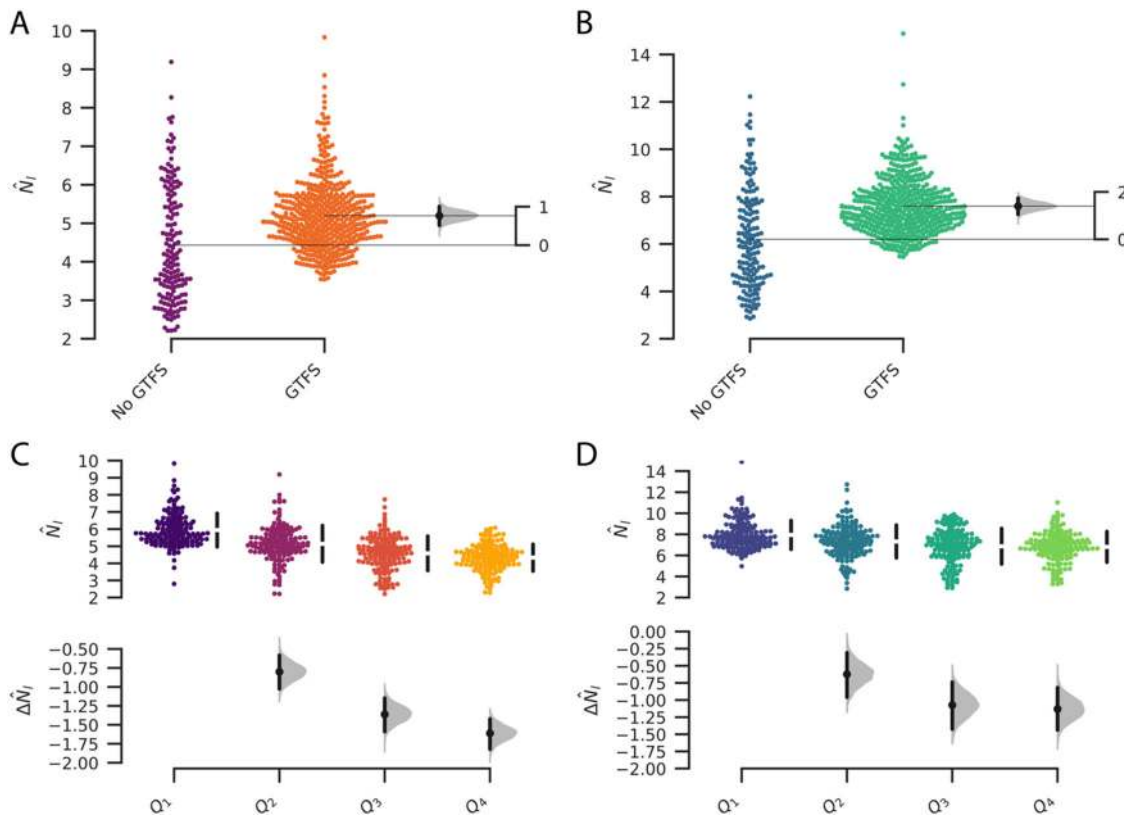
In Fig. 3 each dot corresponds to the average number of locations,  $\hat{N}_i$ , visited by the residents of a given cell. Residents of cells with access to public transport (labeled as GTFS) visit significantly more locations than those living in areas without access to public transport (no GTFS). However, while access to public transport increases the mobility of both women (panel A) and men (panel B) in terms of number of unique locations visited, it does not fill the gender gap completely. The average value of  $\hat{N}_i$  for women increases by 0.76 (95% CI [0.51, 1]) locations, when there is at least one Transantiago stop close to their home. For men, the mean value of  $\hat{N}_i$  increases by 1.39 (95% CI [1.05, 1.72]).

One would expect the availability of public transport options to mainly serve the most deprived communities, whose residents must usually rely on public services to move around the city. We investigate the interplay between the accessibility of public transport and the socioeconomic fabric of different areas of Santiago by looking at the mean value of  $\hat{N}_i$  in cells belonging to different quartiles of the GSE ratio distribution. Figure 3 shows the effect of the GSE ratio on  $\hat{N}_i$  through estimation statistics, considering only those cells with access to public transport, grouped by quartiles of GSE ratio. Access to public transport proves to be less equalizing across socioeconomic segments for women (panel C) than for men (panel D). Women who live in cells belonging to the lowest quartile of the GSE ratio (Q4) visit on average 1.53 (95% CI [1.29, 1.78]) locations less than those in the highest quartile (Q1), even when they have access to public transport. The same difference measured for men is less than 1 location (95% CI [0.51, 1]).

In some parts of the Santiago metro area residents may not have access to public transport, yet they may have access to a private vehicle, which might favor a higher mobility. We assess the impact of owning a private vehicle on gender mobility by assigning each cell of Santiago to one of two groups, defined by census data: cells where there is more than 1 car every 4 residents, and cells where the number of private vehicles does not meet such threshold (see Materials and Methods). We then quantify the impact of belonging to the higher car ownership group through estimation statistics as done for the public transport analysis (see Fig. S6 in the Supplementary Materials). As expected, a higher proportion of car owners is associated to higher values of  $\hat{N}_i$  both for women and men. When looking at only those cells where having access to a car is more likely, we notice a smaller difference in the number of locations visited by the residents living in cells belonging to the different quartiles of GSE ratio for men than women.

**Gender differences in types of visited locations.** To further investigate gender differences in mobility beyond their most general statistical features, we turn our attention to potential gender differences in visitation patterns to different types of locations within the city. To characterize locations within the city, we turn to geographic databases of Points of Interest (POIs). We





**Fig. 3 How gender differences in mobility correlate with access to public transport and socioeconomic status.** Top: estimation plots of the difference in the number of locations visited by women (a) and men (b), where each dot is a cell classified according to having access to public transport (GTFS) or not (No GTFS). Bottom: estimation plots of the difference in the number of locations visited by women (c) and men (d) where each dot is a cell ranked by quartile of GSE ratio. Each group represents cells in decreasing quartiles of income, from the top to the bottom quartile.

use POI data from the openly accessible OpenStreetMap project and 3 other data sources described in Materials and Methods.

Although we have no information on whether the presence of a user near a specific POI actually corresponds to the user engaging with that POI, or whether the POI actually motivates the visit by the user, it is nevertheless of interest to study whether the visitation patterns to certain locations are gendered, and to this end the POI data we collected are potentially valuable as proxies for important characteristics of a given location (e.g., perceived or actual safety, pedestrian friendly areas, etc.) that might be associated with gender differences in visits.

Thus, we study whether significant gender differences can be observed for visited locations that lie at or near a specific type of POI. To this aim, we compute, for each POI type,  $k$ , a POI density over the entire spatial domain under study, using a Kernel Density Estimator with varying bandwidth distance  $d$ . Then, we compute an average POI density ( $\rho_k^{F,M}$ ) for all female and male users based on their set of visited locations. That is,  $\rho_k^{F,M}$  is the density of POI type  $k$  averaged over all locations (cells) visited by all females ( $F$ ) or males ( $M$ ). Finally, we define the gender density ratio

$$r_k = \rho_k^F / \rho_k^M, \tag{12}$$

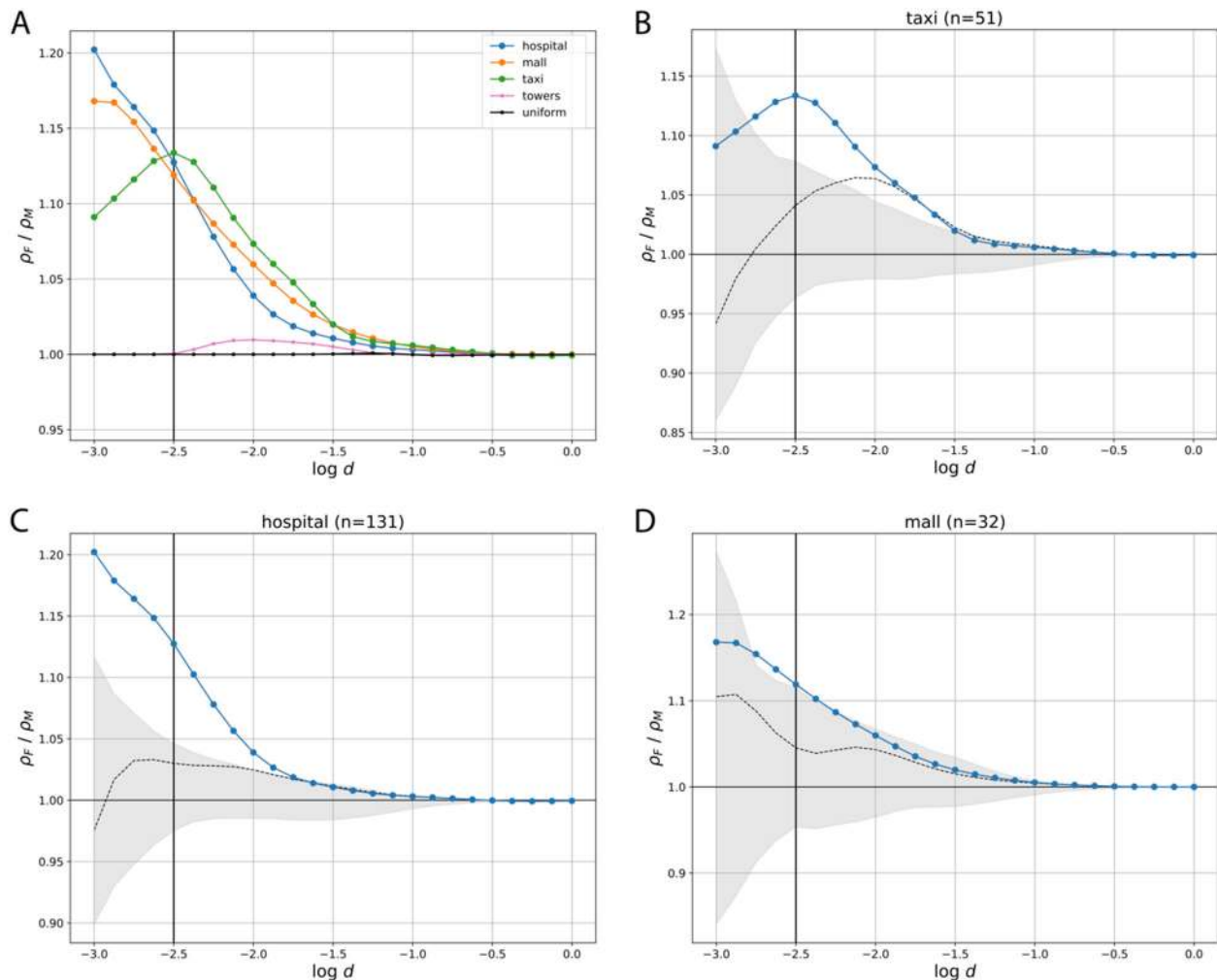
which, for each POI type  $k$ , is meant to indicate gender imbalances in visits to location associated with that specific POI type (see Materials and Methods for a complete description of the method).

Figure 4a shows the gender ratio  $r_k$  as a function of the decimal logarithm of the bandwidth  $d$  for the three POI types with the largest imbalance in  $r_k$ : “taxi”, “hospital” and “mall”. The bandwidth is expressed in decimal degrees of latitude/longitude,

to match the step of the spatial grid we use to cluster cell phone towers. Since 0.01 degrees along the North-South direction correspond to about 1.1 km of linear distance, and 0.01 degree along the East-West direction correspond to about 0.92 km, to avoid using anisotropic kernels for kernel density estimation, latitude and longitude were both rescaled so that variations of 0.01 degree in longitude or in latitude both correspond to a 1 kilometer displacement in Euclidean distance. For comparison, Fig. 4a shows  $r_k$  for two reference layers: the “towers” grid and the “uniform” grid, which are defined as uniform distributions of POIs (see Materials and Methods).

We notice how for large distances the ratio  $r_k \sim 1$ , as the density for both females and males becomes essentially a Gaussian centered in the middle of the city. Moving towards shorter bandwidth  $d$  i.e., shorter distances, gender differences appear for some, but not for all, POI types indicating an imbalance of women visiting more frequently locations near some specific POIs (Fig. S7 in the Supplementary Materials shows the values of  $r_k$  for all POI types considered).

The vertical line at  $\log_{10} d = -2.5$  in Fig. 4 corresponds to an Euclidean distance of about 300 meters, hence to a Gaussian kernel with over 90% of its mass in a disc of diameter 1 km around the kernel center, which is the distance at which the spatial resolution of the kernel density estimator approaches the spatial resolution in the position of telephone towers, i.e., our spatial resolution limit on the position of users. We also notice how the “towers” synthetic POI type exhibits some weak gender imbalance at the scale of a few kilometers, whereas the “uniform” reference layer yields  $r = 1$  for all values of  $d$ , as expected.



**Fig. 4 Gender differences in visit patterns.** Gender ratio  $\rho_F / \rho_M$  as a function of the bandwidth  $d$  (decimal logarithm of the value in decimal degrees), for a few selected POI types, including the "towers" and "uniform" reference layers (a). The vertical line corresponds to a kernel bandwidth comparable to the spatial resolution afforded by the CDR data we use. The gender ratio  $\rho_F / \rho_M$  as a function of the bandwidth  $d$  is compared to two reference null models for "taxi" (b), "hospital" (c), and "mall" (d). The shaded area indicates, for each value of the bandwidth  $d$ , the 95% confidence interval of the values  $\rho_F / \rho_M$  computed on the realizations of the spatial sampling null model. The dashed line indicates the gender ratio computed on the spatially perturbed null model.

To assess the significance of these gender imbalances, it is crucial to compare the observed values of  $r_k$  against the values obtained by generating POI locations according to null models, or against values obtained by spatially perturbing the original POI locations. Therefore, we compare the observed value with those obtained by using two different spatial null models that randomly generate synthetic POI locations (see Materials and Methods for details). The first null model is based on randomly sampling phone towers according to the probability of visit by any user and then perturb locations so that they no longer lie on the regular grid. The second null model is aimed at verifying the sensitivity of our results with respect to the specific set of locations of each POI type. It is built by generating a spatially-perturbed version of the same set of POIs, by randomly adding or subtracting an offset of 0.01 from the latitude and longitude of each POI.

The results of this analysis are reported in Fig. 4, panels B, C, and D, for the 3 POI types showing the largest gender imbalance. The shaded area indicates, for each value of the bandwidth  $d$ , the 95% confidence interval of the values  $r_k$  computed on the realizations of the first null model with the same number of points  $N_k$  as the original POI type. For these POI types, the observed gender imbalances are strongly significant against the chosen null

model for a broad interval of intermediate bandwidth values around  $-2.5$ . The dashed lines describe the behavior of  $r_k$  using the second null model, that is the spatially-perturbed version of the same set of POIs. We notice how perturbing the POI locations pushes the ratio  $r_k$  towards unity and into the 95% confidence interval, making the deviation from unity insignificant for a broad range of bandwidth values  $d$  (and in particular for  $\log_{10} d \approx -2.5$ ).

To further check the robustness of our results we compute the ratio  $r_k$  considering different assumptions in the POI distributions and users' movements (see Text E in the Supplementary Materials). In all cases, the gender ratio  $r_k$  appears to be strongly imbalanced for the same set of POI types, with females visiting more places near hospitals, malls and taxi stops. In particular, we check for spatial correlations among the POI layers showing the largest imbalances in  $r_k$  and find them to be minimally correlated (see Table S3 in the Supplementary Materials). We also check that the observed gender imbalance in  $r_k$  can not be ascribed to a higher female calling activity at those POI locations (see Text E in the Supplementary Materials) and that higher values of  $r_k$  are still observed when restricting the set of visited locations to those that are more than 5 km distant from users' home, and are not their work locations (see Fig. S7 in the Supplementary Materials).

## Discussion

The main contribution of our study can be summarized as follows: women's travel patterns in the metropolitan area of Santiago are different than men's in various aspects, and such differences can be exposed by the analysis of large-scale anonymized mobile phone data disaggregated by sex.

In their daily movements, women visit fewer locations than men and they are more localized, that is, they tend to distribute their time within a few preferred locations. Such reduced mobility for women might result from the interplay of cultural, infrastructure, resource, and safety constraints (Kwan, 1999). Although we mainly focused on capturing behavioral differences, we were able to relate mobility inequalities to a number of socio-demographic factors that may potentially explain the gender effect. First, the gender inequality in mobility is widened in presence of income inequality. Indeed, a smaller gender gap in mobility characterizes the affluent municipalities of the Santiago Metropolitan Region. More in general, the analysis of spatial patterns showed that income, employment and gender mobility equality are all positively correlated.

Our results are in line with previous studies that showed how mobile phone derived mobility metrics can be used as proxy for human development (Pappalardo et al., 2016; Eagle et al., 2010).

At the same time, we proved how the complex relations between gender, mobility and poverty can be elucidated by combining high-resolution telecommunication data with demographic statistics.

Transport and gender are tightly intertwined. By linking the gender gap in mobility with open data about public and private transport, we found that access to transport reduces mobility differences across socioeconomic segments for men but significantly less so for women. Lower income remains a relevant factor that constrains women's mobility even when public transport is available, calling for more gender inclusive design policies for the transportation system of Santiago. Finally, we found that not only women's mobility patterns in Santiago are different in terms of spatial and temporal features, but also in the *type* of locations most frequently visited. Indeed, we showed that it is possible to use Point-Of-Interest (POI) geographic databases to expose gender differences in the type of visited locations, as described by the presence or spatial proximity of specific points of interest. This allows to relate gender differences to categorized spatial features, suggesting hypotheses for further research as well as informing potential interventions. Specifically, we found that visits to cell towers close to *hospitals*, *malls* and *taxi stands* are significantly gendered, and that this gender imbalance is strongly significant when contrasted to null models for POI spatial distribution and when tested against different assumptions on users mobility and POIs filtering (see Text E in the Supplementary Materials for a sensitivity analysis). In summary, we found that visits to non-home, distant, non-frequently-visited (hence, likely, non-work) locations display significant gender differences for specific POI types, with females visiting more places near hospitals, malls and taxi stops. This might indicate, in particular, that women carry a larger burden of caring for family members or related individuals in hospitals. Our results thus demonstrate that mobile phone data are sensitive enough to capture the different mobility needs of men and women, related to their trips' purpose, thus representing a relevant source of information for urban planners to design gender responsive solutions. Providing specific policy recommendations goes beyond the scope of our study, however, we think our results can be valuable for policymakers. In 2018, the Chilean Ministry of Transport and Telecommunications issued an agenda report in which a number of key activities to address the gender gap in transportation and mobility are outlined (Gobierno de Chile, 2018). Most of the

activities planned by the Chilean government in the period 2018–2022 relate to identifying and measuring gender differences in mobility in three Chilean cities. We believe that our study can help answer some of the questions defined in the Ministry agenda, by providing a cost-effective approach to define a quantitative baseline and keep track of the progress towards gender equality.

Further research is needed to assess whether the findings of our study generalize beyond the case of Santiago, Chile. However, we expect our results to hold for different countries, as demonstrated by a recent comparative analysis on travel behavior by gender that found similar gender effects (e.g., shorter trips made by women) in eight cities, across three different continents (Ng and Acker, 2018). Besides it would worth investigate the gender gap in a rural setting.

The use of mobile phone data to study gender mobility disparities—while promising—suffers from some limitations. One of them is the bias inherent to the data which can be related to two main issues: users' representativity and differences in calling activity. First, the users' sample might not be representative of the population under study: the sample size and composition will depend on the operator market share and in general users' demographics will not be fully representative of the population demographics. Second, mobility patterns are inferred through user's calling activity, which is known to be affected by age and gender, among other individual features (Frias-Martinez et al., 2010). In our work we have controlled at our best for these biases, but we could not address some potential confounding effects, for instance those related to users' age, a variable not available to us. To overcome the issue of a biased activity sampling for some users, the analysis of high-frequency x-Detail Records (XDR) could provide a valuable alternative. Also, when considering to extend our study to developing countries, we must mention biases in phone ownership, with women being more likely to use shared phones than men (Blumenstock and Eagle, 2010), although mobility estimates have been shown to be surprisingly robust to such biases at population level (Wesolowski et al., 2013). While limitations arising from a possible mismatch between the line owner and the actual phone user may be relevant in some developing countries, the current study should not suffer from such bias as mobile phones are largely adopted in Chile, with the highest penetration level in Latina America (GSMA, 2020).

Our work falls within the growing research efforts to leverage large-scale data from digital traces to tackle crucial humanitarian and development questions. Previous studies have explored the use of big data such as transactions records, call detail records or social network penetration to map and investigate gender disparities (Reed et al., 2016; Lenormand et al., 2015; Fatehikia et al., 2018; Garcia et al., 2018). While these novel data sources have proved their potential for social good, they also raise some important privacy concerns (Jacques, 2018; De Montjoye et al., 2013). We are aware of such concerns and we took them seriously into account by adopting a range of strategies to preserve users' privacy (see Materials and Methods). Nevertheless, while recognizing the need of systematic and established approaches to the privacy-conscious use of mobile phone data (de Montjoye et al., 2018), we believe that their benefit can outweigh the risks, especially if analyzed in aggregated form, as done here.

Our study is the result of a joint collaboration between research centers, international organizations and private companies within the framework of a *data collaborative* (Susha et al., 2017). We hope that our results will foster the creation of new data partnerships to further investigate the urban mobility experiences of women and girls within cities, so as to inform urban planners' decision-making process.

**Data availability**

Upon request, the authors can provide the data of the average mobility scores by cell tower, disaggregated by gender, along with appropriate documentation for replication of the related analysis. Upon request, authors can also allow other researchers to run queries on datasets stored in the Telefónica servers, to reproduce the results of the study in full. Raw counts or individual information of mobile phone users cannot be made publicly available, or leave Telefónica monitored servers, due to contractual and privacy reasons.

Received: 14 November 2019; Accepted: 12 May 2020;

Published online: 17 June 2020

**Note**

1 [https://wiki.openstreetmap.org/wiki/Downloading\\_data](https://wiki.openstreetmap.org/wiki/Downloading_data)

**References**

- Alexander L, Jiang S, Murga M, González MC (2015) Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Res Part C* 58:240–250
- Bank W (2011) World development report 2012: gender equality and development. World Bank Publications
- Beiró MG, Bravo L, Caro D, Cattuto C, Ferres L, Graells-Garrido E (2018) Shopping mall attraction and social mixing at a city scale. *EPJ Data Science* 7(1):28
- Blondel VD, Decuyper A, Krings G (2015) A survey of results on mobile phone datasets analysis. *EPJ Data Sci* 4(1):10
- Blumenstock J, Eagle N (2010) Mobile divides: gender, socioeconomic status, and mobile phone use in Rwanda. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*. ACM, p 6
- Brown D, McGranahan G, Dodman D (2014) Urban informality and building a more inclusive, resilient and green economy. IIED, London
- Buvinic M, Furst-Nichols R, Koolwal G (2014) Mapping gender data gaps. Data2X Report. United Nations Foundation, Washington, DC
- Buvinic M, Levine R (2016) Closing the gender data gap. *Significance* 13(2):34–37
- Calabrese F, Diao M, Lorenzo GD, Ferreira J, Ratti C (2013) Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Res Part C* 26:301–313
- Chant S (2013) Cities through a “gender lens”: a golden “urban age” for women in the global South? *Environ Urban* 25(1):9–29
- Clarke M, Dix M, Jones P (1981) Error and uncertainty in travel surveys. *Transportation* 10(2):105–126
- Çolak S, Lima A, González MC (2016) Understanding congested travel in urban areas. *Nat Commun* 7(1):1–8
- Couper MP (2000) Web surveys: A review of issues and approaches. *Public Opin Q* 64(4):464–494
- Cresswell T, Uteng TP (2016) Gendered mobilities: towards an holistic understanding. In *Gendered mobilities*. Routledge, p 15–26
- de Montjoye Y-A, Gams S, Blondel V, Canright G, de Cordes N, Deletaille S, Engø-Monsen K, Garcia-Herranz M, Kendall J, Kerry C et al. (2018) On the privacy-conscious use of mobile phone data. *Sci Data* 5:1–6
- De Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the Crowd: the privacy bounds of human mobility. *Sci Rep* 3(1):1376
- Eagle N, Macy M, Claxton R (2010) Network diversity and economic development. *Science* 328(5981):1029–1031
- Fatehikia M, Kashyap R, Weber I (2018) Using Facebook ad data to track the global digital gender gap. *World Dev* 107:189–209
- Fisher RA (1924) The distribution of the partial correlation coefficient. *Metron* 3:329–332
- Frias-Martinez V, Frias-Martinez E, Oliver N (2010) A gender-centric analysis of calling behavior in a developing economy using call detail records. *AAAI Symposium on Intelligence for Development*, p 37–42
- García D, Mitike Kassa Y, Cuevas A, Cebrian M, Moro E, Rahwan I, Cuevas R (2018) Analyzing gender inequality through large-scale Facebook advertising data. *Proc Natl Acad Sci USA* 115(27):6958–6963
- Gobierno de Chile MdTyT (2018) Agenda para la politíca de equidad de género en transportes. Technical Report
- González MC, Hidalgo CA, Barabási A-L (2009) Understanding individual human mobility patterns. *Nature* 458(7235):238–238

- Graells-Garrido E, Caro D, Parra D (2018) Inferring modes of transportation using mobile phone data. *EPJ Data Sci* 7(1):49
- Graells-Garrido E, Ferres L, Caro D, Bravo L (2017) The effect of Pokémon Go on the pulse of the city: a natural experiment. *EPJ Data Sci* 6(1):1–19
- Groves RM (2006) Nonresponse rates and nonresponse bias in household surveys. *Public Opin Q* 70(5):646–675
- GSMA (2020) The mobile economy. Latin America 2019. Technical Report, GSMA
- Hanson S (2010) Gender and mobility: new approaches for informing sustainability. *Gen Place Cult* 17(1):5–23
- Ho J, Tumkaya T, Aryal S, Choi H, Claridge-Chang A (2019) Moving beyond P values: data analysis with estimation graphics. *Nat Methods* p 1
- Hwang J-N, Lay S-R, Lippman A (1994) Nonparametric multivariate density estimation: a comparative study. *Trans Sig Proc* 42(10):2795–2810
- Jacques DC (2018) Mobile phone metadata for development. Technical Report. arXiv: 1806.03086
- Kwan M-P (1999) Gender, the home-work link, and space-time patterns of nonemployment activities. *Economic Geogr* 75(4):370–394
- Law R (1999) Beyond ‘women and transport’: towards new geographies of gender and daily mobility. *Prog Hum Geogr* 23(4):567–588
- Lenormand M, Louail T, Cantú-Ros OG, Picornell M, Herranz R, Arias JM, Barthelemy M, Miguel MS, Ramasco JJ (2015) Influence of sociodemographics on human mobility. *Sci Rep* 5(1):10075
- Loukaitou-Sideris A (2014) Fear and safety in transit environments from the women’s perspective. *Security J* 27(2):242–256
- Lovera CAS (2015) Urban sprawl and infrastructural lands: revamping internal spaces in Santiago de Chile. *Geoforum* 67:36–40
- Naboulsi D, Fiore M, Ribot S, Stanica R (2016) Large-scale mobile traffic analysis. *IEEE Commun Surveys Tutorial* 18(1):124–161
- Ng W-S, Acker A (2018) Understanding urban travel behaviour by gender for efficient and equitable transport policies. International Transport Forum Discussion Paper
- Pappalardo L, Vanhoof M, Gabrielli L, Smoreda Z, Pedreschi D, Giannotti F (2016) An analytical framework to nowcast well-being using mobile phone data. *Int J Data Sci Anal* 2(1):75–92
- Psylla I, Sapiezynski P, Mones E, Lehmann S (2017) The role of gender in social network organization. *PLoS ONE* 12(12):e0189873
- Puertas OL, Henríquez C, Meza FJ (2014) Assessing spatial dynamics of urban growth using an integrated land use model. Application in Santiago Metropolitan Area, 2010–2045. *Land use policy* 38:415–425
- Reed PJ, Khan MR, Blumenstock J (2016) Observing gender dynamics and disparities with mobile phone metadata. *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development-ICTD ’16*. ACM Press, New York, USA, p 1–4
- Rosenbloom S (2004) Understanding women’s and men’s travel patterns. In *Research on Women’s Issues in Transportation: Report of a Conference*
- Schneider CM, Belik V, Couronné T, Smoreda Z, González MC (2013) Unravelling daily human mobility motifs. *J R Soc Interface* 10(84):20130246
- Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. *Science* 327(5968):1018–1021
- Susha I, Janssen M, Verhulst S (2017) Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: taxonomy development. In: *Proceedings of the 50th Hawaii International Conference on System Sciences*
- Tacoli C (2012) Urbanization, gender and urban poverty: paid work and unpaid carework in the city. Technical Report. International Institute for Environment and Development
- United Nations (2018) Department of economic and social affairs, population division. Technical report, United Nations
- Uteng TP, Cresswell T (2016) Gendered mobilities: towards an holistic understanding. In: *Gendered mobilities*. Routledge, p 15–26
- Vaitla B, Bosco C, Alegana V, Bird T, Pezzulo C, Hornby G, Sorichetta A, Steele J, Ruktanonchai C, Ruktanonchai N, Wetter E, Bengtsson L, Tatem AJ, Di Clemente R, Luengo-Oroz M, González MC, Nielsen R, Baar T, Vacarelu F, de Choudhury M, Sharma S, Logar T, Eekhout W (2017) Big Data and the Well-being of Women and Girls: Applications on the Social Scientific Frontier. Technical Report. Data2X
- Vanhoof M, Reis F, Ploetz T, Smoreda Z (2018) Assessing the quality of home detection from mobile phone data for official statistics. *J Off Stat* 34(4): 935–960
- Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO (2013) The impact of biases in mobile phone ownership on estimates of human mobility. *J R Soc Interface* 10(81):20120986

**Acknowledgements**

We thank Telefonica I+D for providing access to the data and the infrastructure to analyze it, in particular Pablo Garcia Briosso. We thank Eduardo Graells-Garrido for help with data visualization. This project was funded in part by the Data2X Initiative through the Big Data for Gender Challenge Awards. The authors acknowledge financial support from Movistar-Telefonica Chile, the Chilean government initiative CORFO

13CEE2-21592 (2013-21592-1-INNOVA\_PRODUCION2013-21592-1), Conicyt PAI Networks (REDES170151) "Geo-Temporal factors in disease spreading and prevention in Chile". The support of the CRT Foundation through the Lagrange Project is gratefully acknowledged. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Author contributions

LG, MT, AY, NA,SV, LF, and CC designed research; LG, MT, SP, AY, NA, SV, LF, and CC performed research; LG, MT, SP, LF, and CC analyzed data; and LG, MT, NA, SV, LF, and CC wrote the paper.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1057/s41599-020-0500-x>.

**Correspondence** and requests for materials should be addressed to L.G.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020