Data 1 - innovative data and methods for population science

The Impact of Religious Intervention on Fertility Rates: A Synthetic Control Analysis of Patriarch Ilia II's Influence in Georgia
Alina Hordiienko - Minerva University

This study evaluates the impact of Patriarch Ilia II's baptism initiative on fertility rates in Georgia using the Synthetic Control Method. The analysis reveals a substantial and sustained increase in fertility rates attributable to this culturally and religiously grounded intervention, estimating approximately 159,759 additional births from 2008 to 2022. Representing a 1.24% increase relative to the cumulative reproductive-age female population, these findings highlight the effectiveness of non-financial, culturally embedded pronatalist policies. The research underscores the importance of institutional trust and socio-cultural contexts in demographic interventions and provides valuable insights for policymakers aiming to mitigate demographic declines through culturally sensitive approaches.

Email: alinahordiienko@gmail.com

Assessing Bias in Migration Stock Estimates from Facebook and Instagram Using a Social Media-Based Survey
Maciej Danko - Max Planck Institute for Demographic Reserach, Francesco Rampazzo - University of Manchester, Jessica Donzowa - Max Planck Institute for Demographic Research, Jisu Kim - University of Utrecht, Emilio Zagheni - Max Planck Institute for Demographic Research

This study evaluates the reliability of Facebook and Instagram digital trace data for estimating migration and mobility patterns. While traditional migration statistics are delayed and lack granularity, social media platforms offer real-time insights—albeit with concerns about representativeness and algorithmic bias. We combine estimates from Meta's advertising platform with a novel online survey—the International Mobility Survey—targeting Polish migrants in the UK and Germany. The survey, promoted via Facebook and Instagram, collects detailed data on respondents' migration history, social media usage, and future mobility intentions, allowing direct comparison with platform-derived metrics. To improve representativeness, we apply post-stratification using three methods: age-sex aggregation, Generalized Additive Models, and imputation for sparse strata. Results highlight systematic biases in Meta data, including underrepresentation of older men and individuals with lower education levels, as well as discrepancies in reported versus inferred migration status. The survey sample also showed high dropout among certain demographic groups, affecting data completeness. Our findings indicate that while Meta's platform offers valuable mobility signals, its data require careful correction to be useful for demographic estimation. This study contributes to the development of improved hybrid models that integrate traditional and digital sources for more timely and accurate migration monitoring.

Email: francesco.rampazzo@sociology.ox.ac.uk

Correcting biases in human mobility data from mobile phone applications
Carmen Cabrera - University of Liverpool, Francisco Rowe - University of Liverpool

Access to human mobility data is key for a wider variety of social challenges, including urban planning, sustainability, public health and economic development. Location trace data collected through digital technology, such as mobile applications have become widely available to study human mobility, and overcome key limitations of traditional data streams, such as surveys and censuses. Yet, digital trace data are not representative of the general population, and consequently require statistical adjustment to mitigate existing biases and make robust statistical inferences. While efforts have been made to correct human mobility data from digital sources, no generalisable solution has been proposed to satisfactorily adjust human mobility flows by attributes at the origin and destination areas. I will present the work of ESRC-funded project DEBIAS, which aims to develop a generalisable framework to measure, assess and correct existing biases in human mobility data extracted from digital trace data. First, I will present a measure to quantify biases in spatial population counts derived from digital trace data, and identify key demographic,

socioeconomic and geographic features underlying these biases. Second, I will introduce the proposed framework to mitigate biases in spatial population count data, and generate bias-adjusted DF human mobility counts. Third, I will present plans for future work on this area.

Email: c.cabrera@liverpool.ac.uk

Who Is at Risk—and When? Integrating Static and Dynamic Data for Typhoon Risk Assessment in the Philippines
Anna Amalia B. Vibar - University of Namur, Elma Laguna - University of the Philippines Population Institute (UPPI), Jose Andres Ignacio - University of the Philippines Population Institute (UPPI), Sabine Henry - University of Namur, Sebastien Dujardin - University of Namur

Disaster risk management has become increasingly critical due to climate change. There are more frequent and severe typhoons, and resources for recovery is limited in some countries. Identifying the most vulnerable population to typhoon risk is essential to improving preparedness and enabling targeted response efforts.

This study presents a novel approach to capturing spatio-temporal variations in exposure and vulnerability to typhoons by combining traditional and dynamic data sources. Using the Philippine context, it integrates static but reliable census-based social vulnerability indicators with near real-time geolocated population data from Meta-Facebook during Typhoon Goni (2020) and Typhoon Rai (2021).

Through demographic analysis and advanced spatial methods, the study examines how variations in population densities throughout a disaster event interact with factors of vulnerability such as housing conditions, age, education, and gender. Preliminary analysis of Typhoon Rai indicates that changes in population density at specific times and locations can influence the impact of a disaster on specific segments of the population. These patterns will then be compared with the experience of Typhoon Goni and a deeper understanding of the overall risk will be provided.

By addressing the limitations of static data, this study underscores the importance of integrating dynamic datasets into disaster risk assessments. The approach offers a more nuanced and relevant understanding of who is most at risk and when. It hopes to contribute to disaster risk management by enabling more responsive climate-resilient strategies for the people who need it most.

Email: anna-amalia.vibar@unamur.be

Data 2 - innovative data and methods for population science

Mapping Life Trajectories: A Machine Learning Approach to Clustering Multidimensional Life Courses
Zerui Tian - University of Oxford, Zhuang Hao - Huazhong Agricultural University

This study introduces a novel approach to clustering multidimensional life course trajectories using transformer-based embeddings (with deep learning architecture) and unsupervised machine learning. Drawing on the 1970 British Cohort Study, it analyzes two key life domains—partnership and activity (employment, education, caregiving) histories—spanning ages 16 to 46. The study aims to (1) identify typologies of life trajectories, (2) examine their predictive power for later-life well-being outcomes, and (3) assess whether early-life characteristics predict cluster membership.

Transformer models capture complex temporal dependencies across life events and reduce high-dimensional sequence data into lower-dimensional latent embeddings. Clustering is then performed using K-Means and Gaussian Mixture Models (GMM), with the optimal model selected via evaluation metrics. Results reveal four life trajectory clusters for women (e.g., "late bloomers," "career-family jugglers") and five for men (e.g., "cohabiting careerists," "young settlers"), reflecting gendered patterns in life-course structuring.

Cluster membership is significantly associated with life satisfaction, health, and income at age 46 and 51, especially for women. Including trajectory clusters improves model fit for predicting later-life outcomes beyond conventional covariates. Moreover, early-life characteristics (e.g., father's occupational class, ethnicity) predict cluster membership, suggesting structural determinants in shaping life paths.

Methodologically, this research advances life course analysis by combining transformer-based sequence modeling with clustering. Substantively, it uncovers how multidimensional trajectories create distinct "risk environments" that moderate the impact of social inequalities. The findings offer practical tools for policymakers to identify vulnerable populations using administrative data, guiding targeted interventions and informing life-course-sensitive policy design.

Email: zerui.tian@gtc.ox.ac.uk

Understanding the impacts and aftermath of COVID-19 on reshaping global migration patterns
Ruth Neville - University College London, University of Liverpool, Francisco Rowe - University of Liverpool

Understanding how international migration patterns shifted during and after the COVID-19 pandemic is critical for informing global mobility governance and resilience planning. While early reports documented widespread disruption to migration flows, systematic, global-scale evidence remains limited, particularly regarding the longer-term aftermath and differentiated patterns of recovery across corridors. This study addresses this gap by examining the impact of COVID-19 on international migration flows using high-frequency digital trace data from Facebook's International Migration dataset (2018–2022). We use two complementary methodological strategies: interrupted time series analysis (ITSA) to quantify the pandemic's immediate and sustained impact on migration volumes, and cluster and network analysis to assess evolving global migration structures and corridor-specific dynamics.

We expect that migration flows declined sharply across nearly all regions during 2020–2021, coinciding with the imposition of global travel restrictions. We also anticipate evidence of migration corridors linked to labour - particularly in the Gulf Cooperation Council (GCC) countries - showing signs of rebounding by 2022, contrasting with a slower recovery in leisure or tourism corridors. We also expect to identify potentially emerging patterns of return migration to less densely populated regions becoming more prominent post-pandemic, indicating potential long-term shifts in migration preferences.

Our findings seek to reveal the timing and geographical differences in recovery of global migration flows, highlighting vulnerabilities and adaptive responses across different migration systems. By leveraging digital trace data and advanced time series and network methods, our study contributes to the growing literature on pandemic-driven mobility disruptions and offers actionable evidence for policymakers and international organisations seeking to build more adaptive, equitable migration frameworks in a post-pandemic world.

Email: ruth.neville@ucl.ac.uk

Decoding Public Perceptions of Chronic Illnesses through Digital Trace Data: A Qualitative Topic Modelling Analysis of YouTube Comments

Ayushi Das, Preeti Dhillon - International Institute for Population Sciences (IIPS), Mumbai

This study aims to explore public perceptions and experiential narratives related to chronic conditions—dementia, hypertension, and diabetes—by leveraging user-generated content from YouTube. In doing so, it highlights the potential of digital trace data to offer nuanced insights into lived experiences, emotional responses, and socio-cultural framings of health.

Comments were collected using YouTube Application Programming Interface (version 3) from the top 100 most-viewed videos on each condition, selected through Google Trends-identified keywords to ensure relevance and reach. The final dataset comprised 129,613 English-language comments: 83,016 on dementia, 23,301 on hypertension, and 23,296 on diabetes. Following standard natural language pre-processing, we applied Latent Dirichlet Allocation (LDA) to identify coherent themes within each disease corpus.

The results revealed deeply emotional and socially embedded narratives. Dementia-related discussions focused heavily on caregiver burden (29%) and emotional grief (27%), indicating the isolating nature of the disease and the lack of structured support. Hypertension conversations prioritized alternative treatments (28.9%), lifestyle concerns (26.4%), and scepticism around medication. Diabetes dialogues were dominated by anxieties around diet (37.2%), especially within the context of Indian cuisine, and challenges balancing insulin, blood pressure, and food habits (20.7%). Across all conditions, users shared distrust in formal healthcare systems and a reliance on experiential, community-based advice.

These findings emphasize how digital platforms serve as informal health forums, capturing culturally embedded and emotionally driven accounts of illness. Recognizing these narratives can help public health professionals design culturally sensitive communication strategies and leverage AI-driven tools to better understand and respond to community health needs.

Email: ayushi.das1999@gmail.com

Tracing the Flow: A Statistical Analysis of Shared Mobility Services in Venice

Vanshika Keshwani - University of Padova, Stefano Mazzuco - University of Padova, Francesco Rampazzo - University of Manchester

Shared micro-mobility services have become an attractive alternative to private vehicles, changing urban traffic by providing sustainable and flexible transportation. User-dynamic studies help in making transportation sector more equitable by responding to the gap between service operators and user-specific patterns of different sociodemographic groups. The results provide findings for mobility companies to optimize usage, improve service accessibility, and strengthen urban transport sustainability through digital data and statistical modelling. Using high-resolution data on trips with electric, non-electric bikes and e-scooters obtained through a collaboration with the Municipality of Venice, this study tracks routes and usage patterns in 2024. We leverage sociodemographic characteristics of users and usage dynamics from spatial data and open street networks to optimize micro-mobility usage in the city. The temporal trends over seasons and weekly commuting behavior are analyzed to see how travel behavior and socio-demographic characteristics influence bike-sharing and e-scooter and usage. Additionally, a network-based method is used to detect travel mobility based on the actual trip data, to assess the most utilized routes in the city. Results show users of ages 18-24 and male users have higher trip frequencies. We also identify distinct behavioral patterns across different vehicle types. E-scooters are mostly used for short-duration trips lasting less than 3 minutes, unlike bike-sharing, which exhibits a more structured commuting pattern with moderate distance traveled (3-6 km). The findings underline would help Municipality of Venice to track the peripheral areas where the degree of sharing is low and optimize the service under changing population patterns.

Email: vanshika.keshwani@studenti.unipd.it

Producing population-level estimates of internal displacement in Ukraine using GPS mobile phone data

Francisco Rowe, Rodgers Iradukunda, Elisabetta Pietrostefani - University of Liverpool

Nearly 110 million people are forcibly displaced people worldwide. However, estimating the scale and patterns of internally displaced persons in real time, and developing appropriate policy responses, remain hindered by traditional data streams. They are infrequently updated, costly and slow. Mobile phone location data can overcome these limitations, but only represent a population segment. Drawing on an anonymised large-scale, high-frequency dataset of locations from 25 million mobile devices, we propose an approach to leverage mobile phone data and produce population-level estimates of internal displacement. We use this approach to quantify the extent, pace and geographic patterns of internal displacement in Ukraine during the early stages of the Russian invasion in 2022. Our results produce reliable population-level estimates, enabling real-time monitoring of internal displacement at detailed spatio-temporal resolutions. Accurate estimations are crucial to support timely and effective humanitarian and disaster management responses, prioritising resources where they are most needed.

Email: fcorowe@liverpool.ac.uk