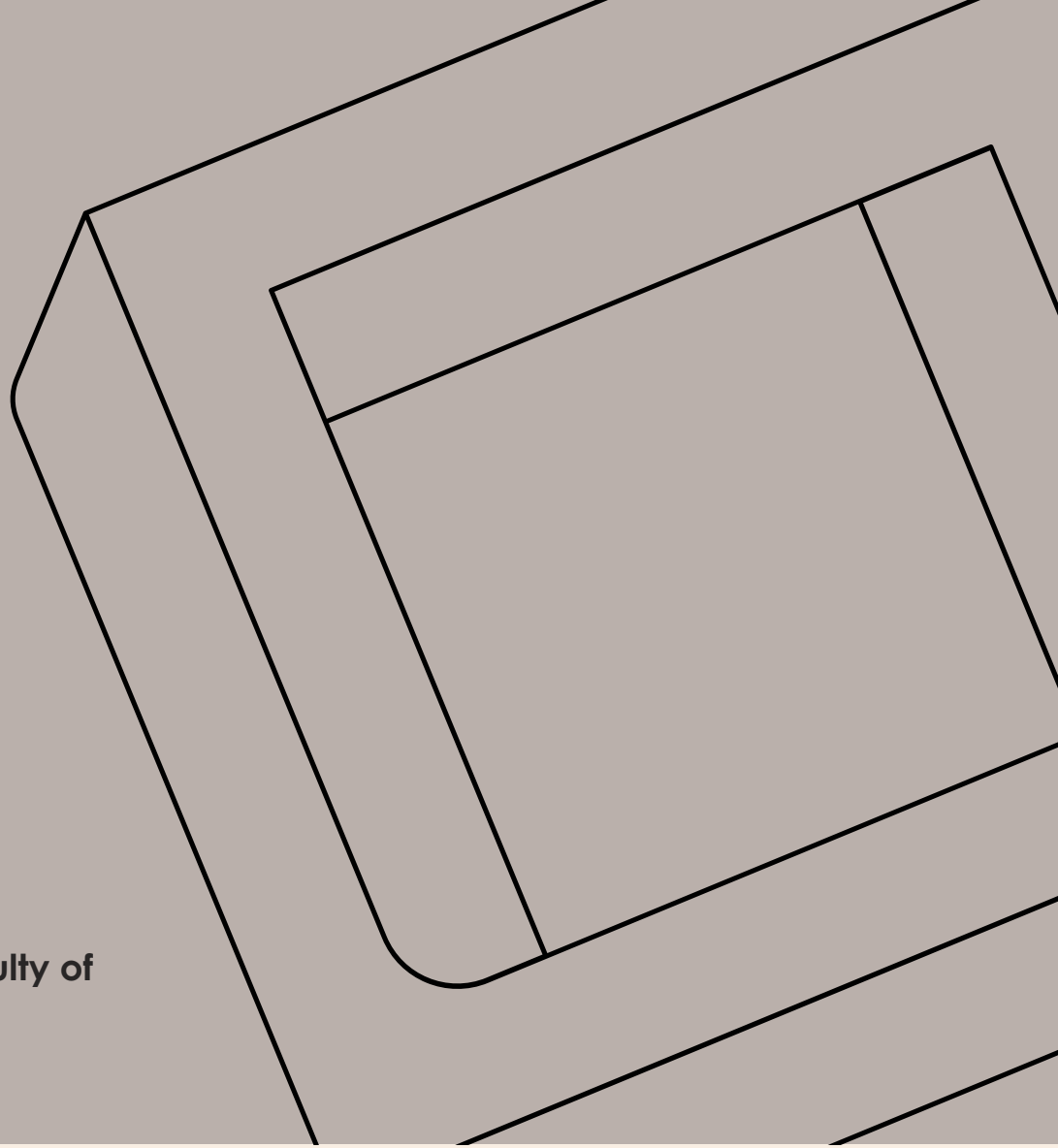
A new approach to mapping a minority population

2024 International Workshop on Comparative Survey Design and Implementation Program

Jamie Burnett
Head of Methods, International surveys
verian

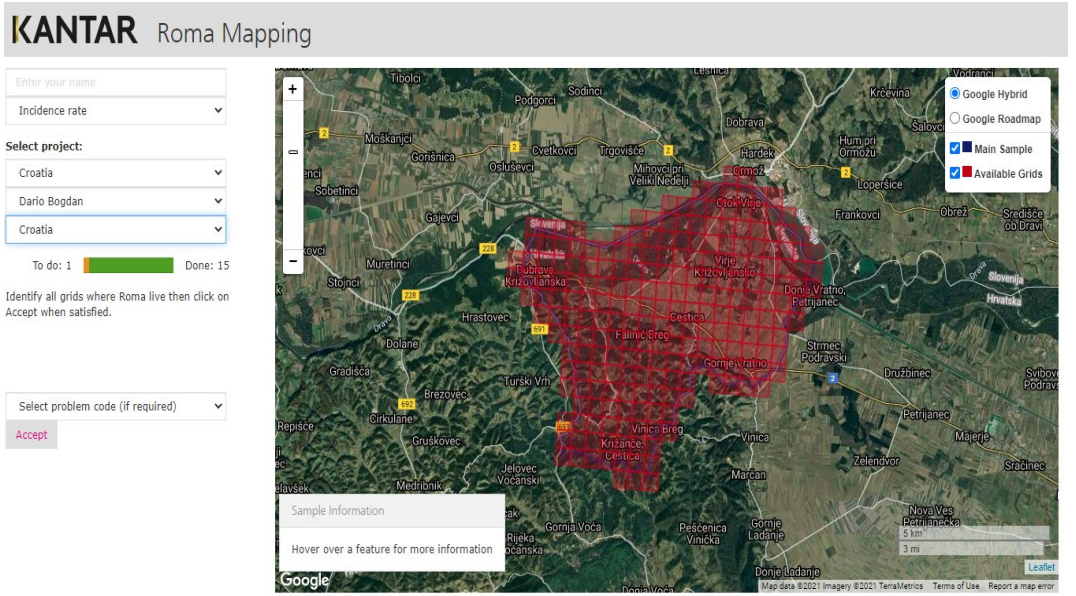
Dr David Simon
assistant professor
Department of Statistics, Faculty of Social Science, ELTE

20 March 2024



verian

Contents



Background

- Survey of the Roma population conducted by the European Union Agency for Fundamental Rights (FRA).
- Objective of the survey is to provide comparable data on the impact on the ground of EU and national anti-discrimination, anti-racism and equality legislation and policies.
- Roma 2021 was the fourth survey to be conducted across multiple European countries.
  - Face to face survey
  - Covering 10 countries in Europe
  - 8400 interviews
  - Four stage sample design
    1. PSUs (typically municipalities) with probability proportional to Roma population
    2. Within sampled PSUs Roma mapped and then smaller geographical areas (SSUs) selected with probability proportional to Roma population – variable number of addresses issued in each SSU based on estimated Roma incidence.
    3. Households selected using a random walk (Bauer) with a random start address
    4. Full enumeration of households and random selection of one adult to participate

Surveying the Roma population

- No registers on the Roma population available
- Official sources on the Roma population typically from the census but not published at a hyper local level.
- Estimates based on relatively large geographies (municipalities) make face to face interviewing unfeasible. Low incidence rates would lead to prohibitive costs to screen.
- Census estimates problematic as typically 10 years old and for a more transient population less reliable.
- Census estimates also known to underreport the Roma population – Roma less likely to participate and more likely to under report.

Mapping the Roma population necessary to make face to face interviewing feasible.

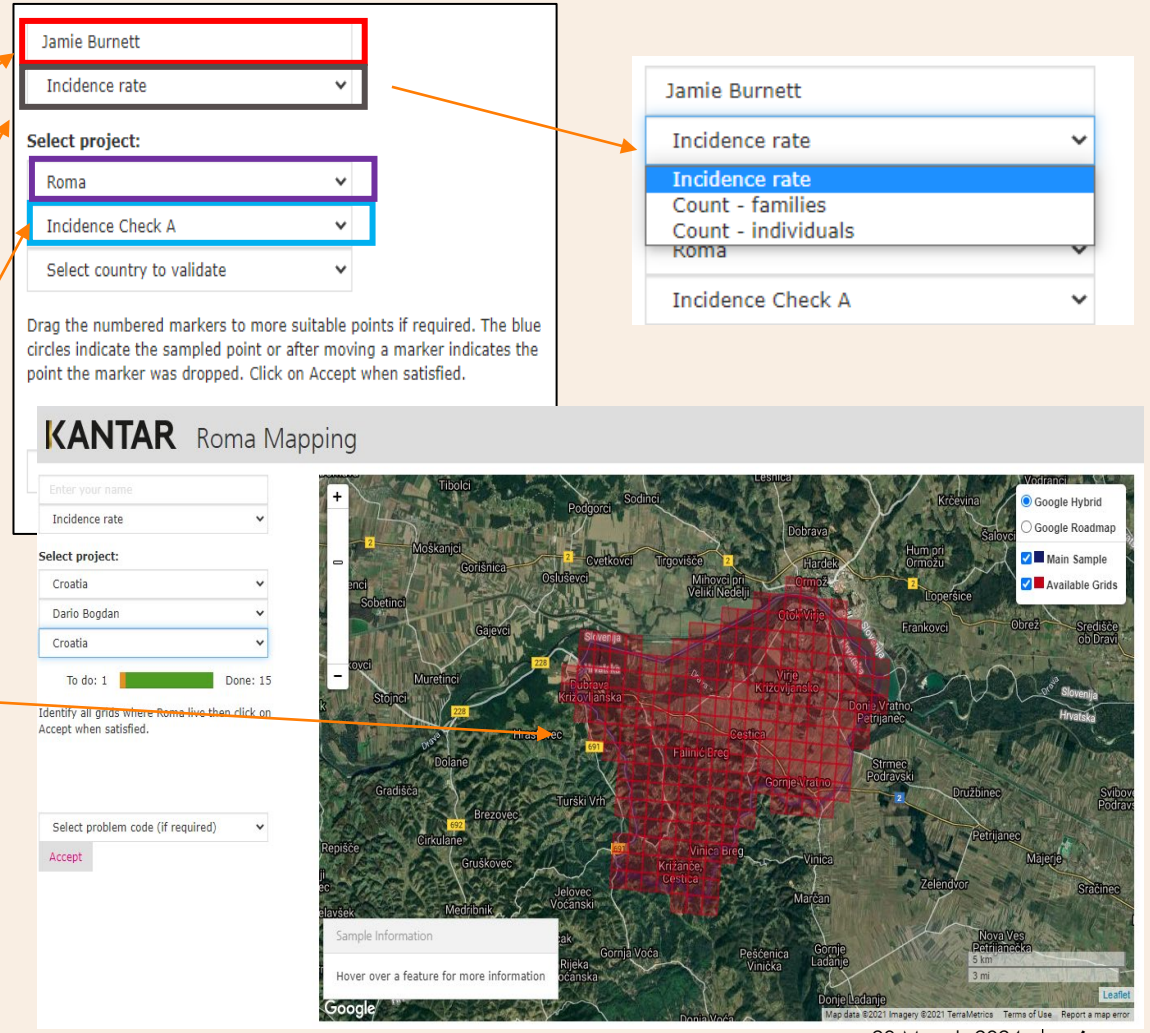
Mapped Roma population used to identify **where interviewers work** and to estimate the **total addresses to work in each location** (based on an incidence estimate).

Mapping the Roma population

- Roma experts employed to map the population in each country
- NGOs and Roma Organizations helped identify the Roma experts
- Multiple experts used to map the population in each country. Experts picked based on their local knowledge
- Before 2021, Roma mapping conducted **offline** – Roma experts supplied with maps of the selected primary sampling units (PSU) and instructed to provide details on where the Roma live within the sampled PSU
- Potential drawbacks of this approach:
  - Approach to mapping left up to the Roma experts – little oversight on the approach used.
  - Potential for processing errors due to the manual process
  - Potential to under-report if the process is time consuming/complex
- For these reasons, a new approach to mapping was developed by verian for Roma 2021 survey

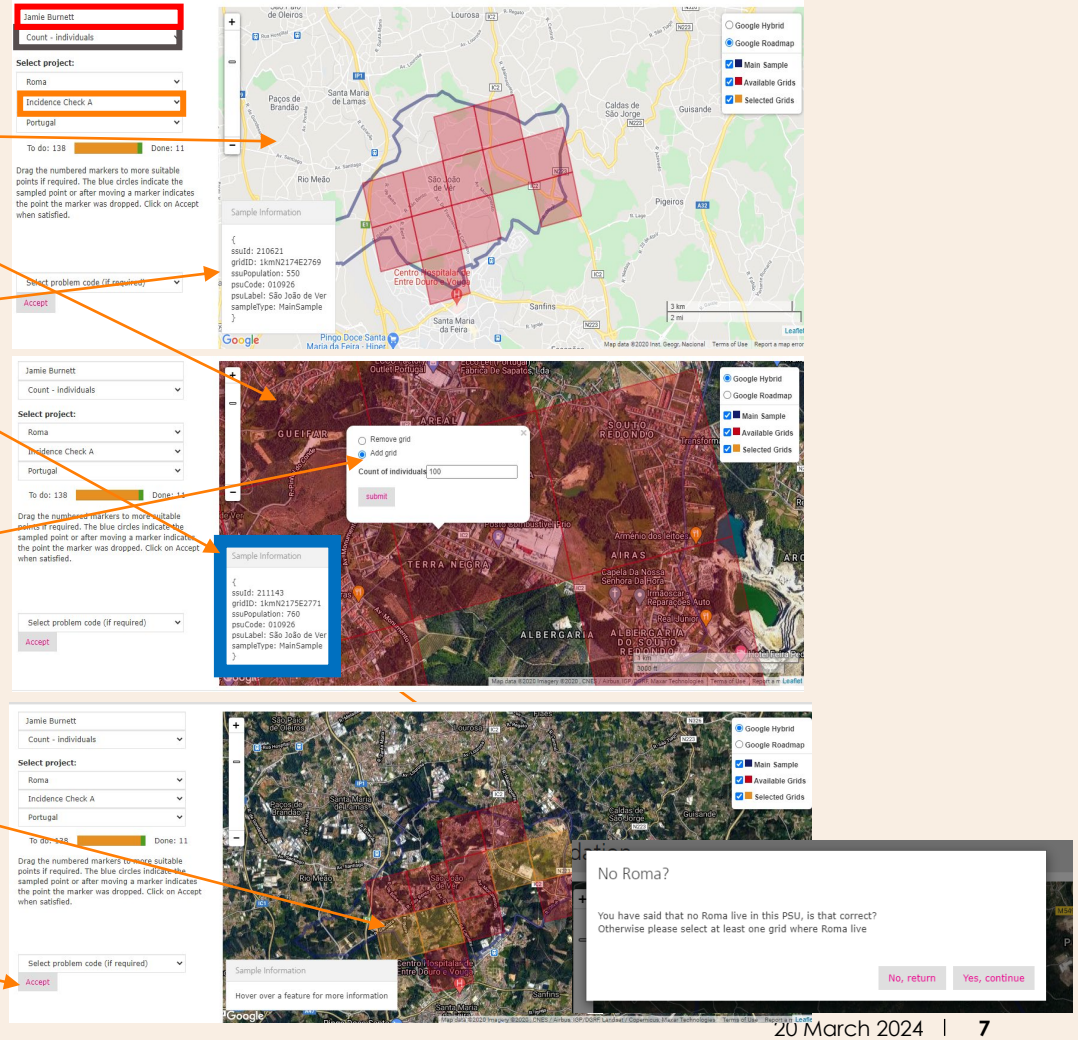
Mapping the Roma population – Functionality of the tool (1)

- Online portal accessible by the mappers
- Experts allocated to the PSU(s) they will map
- Experts need to enter their name before they start mapping.
- For each grid they select they are asked to supply the Roma population size. They can chose between 0+, families or % of total population.
- Selected PSUs are shown one at a time, experts can select the order in which they map them.
- Grid based frame is overlaid onto the PSU in red and experts are asked to select the grids where Roma live and provide population data. Three grid size options available: 250m^2 500m^2 and 1km^2
- Roma pop data is then captured in a database allowing the central team to monitor progress and quality



Mapping the Roma population – Functionality of the tool (2)

- Google maps used as the underlying layer
- Can toggle between hybrid and roadmap views
- Can zoom in and out of the map
- Roma and 0+ population data shown for the selected PSU and 0+ population data for each grid – grid based estimates supplied by Eurostat (Gisco)
- When the expert selects a grid a pop up forms appears where they can enter the Roma population information. If they make a mistake they can change the figure or deselect the grid
- All mapped grids are highlighted yellow
- Once complete the expert must confirm they have finished and select their next PSU
- If the selected PSU has no Roma in it they can also feed this information back



Mapping the Roma population – Functionality of the tool (3)

- The mapped Roma population could be downloaded from the tool at any time
- In the download you could extract the information listed opposite
- Quality control checks were performed comparing the:
  1. Mapped Roma population across all grids to the PSU Roma population from the PSU frame
  2. Mapped Roma population in each grid to the 0+ population from Gisco
- Issues flagged during QC process were feedback to the expert via the national fieldwork supplier

- Country
- Roma Expert ID
- PSU ID/Grid ID
- PSU population (Roma, 0+)
- Grid Population (Roma, 0+)
- Roma population metric (0+, families, incidence)
- Date and time mapped

Accuracy of the mapping

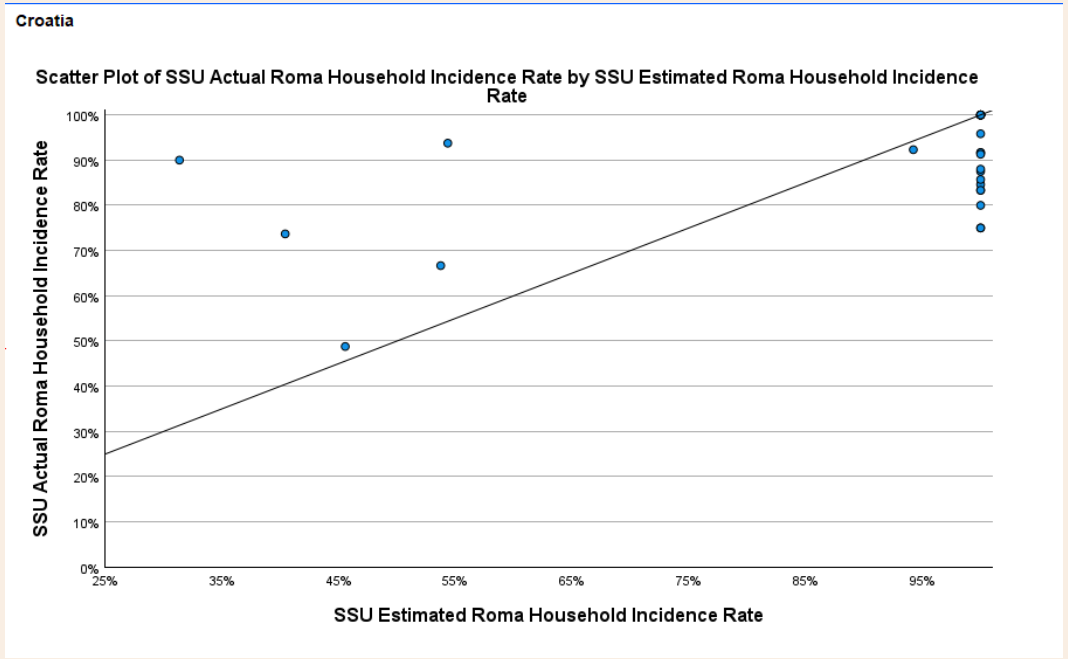
- One way to assess the accuracy of the mapping is to look at the number of replacement sample points. Replacement sample points were used when no Roma were identified in the sample point during fieldwork (dropping rule).
- High % of replacements observed in 3 countries
- When compared to Midis II we see mixed results – some higher and some lower with biggest difference observed in Croatia
- Grid size can play a part in the likelihood to replace a sample point (due to the selection of the start address)

	Roma 2021			MIDIS II		
Country	Total SPs	SPs replaced as no Roma	%	Total SPs	SPs replaced as no Roma	%
Serbia	71	11	15%			
Croatia	50	7	14%	50	1	2%
North Macedonia	51	7	14%			
Portugal	50	4	8%	100	5	5%
Romania	145	6	4%	145	1	1%
Greece	59	2	3%	41	0	0%
Italy	50	1	2%			
Czechia	88	1	1%	85	5	6%
Spain	92	0	0%	155	9	6%

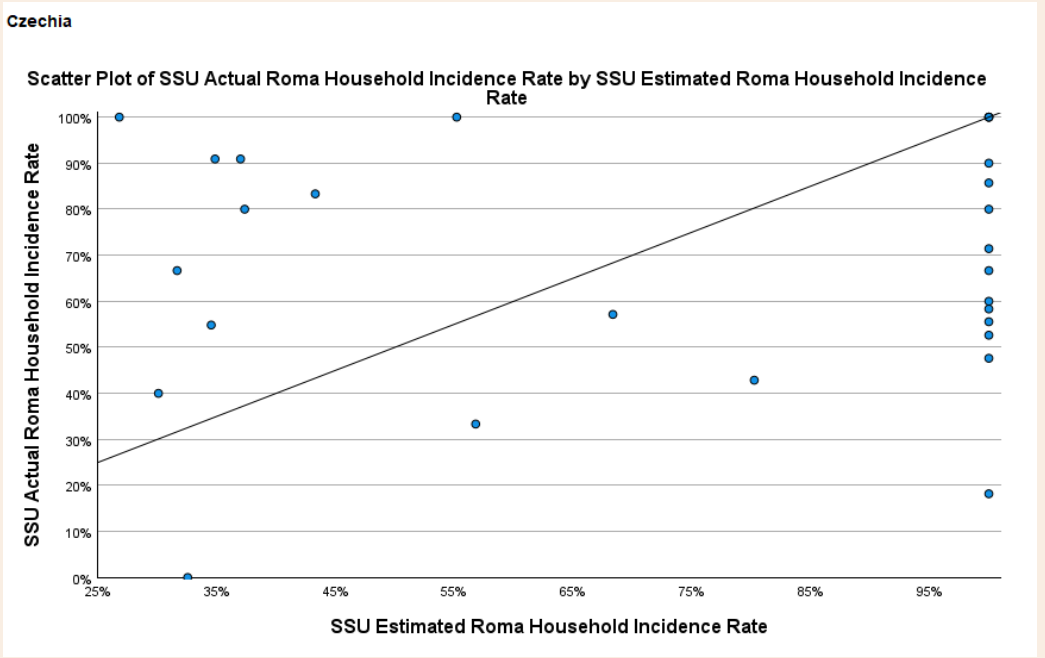
Accuracy of the mapping (1)

Another way to assess the accuracy of the mapping is to look at the incidence of the Roma population in the sample points – comparing expected from the mapping to actual based on fieldwork

Croatia

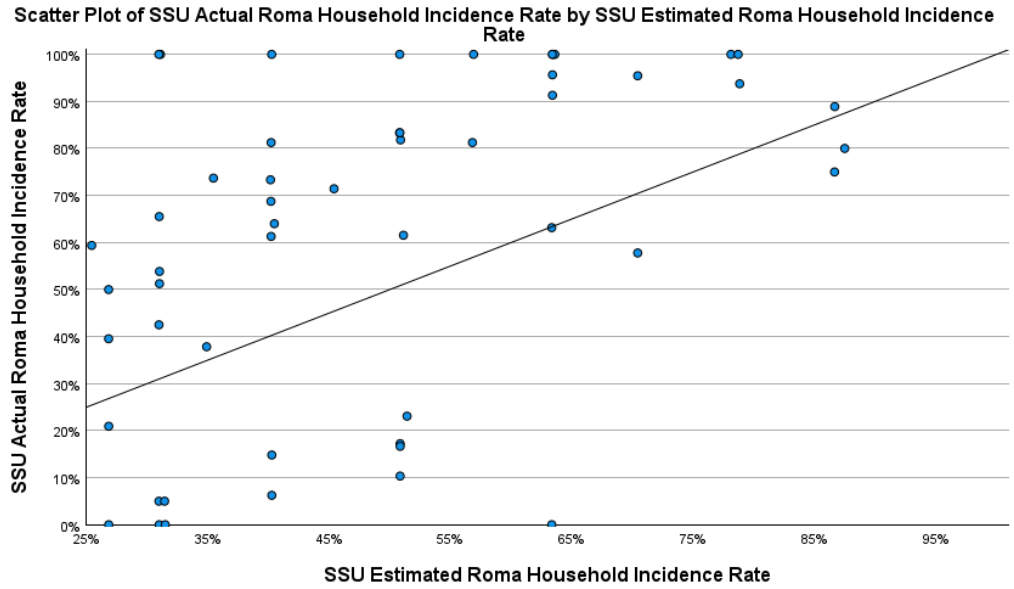


Czechia

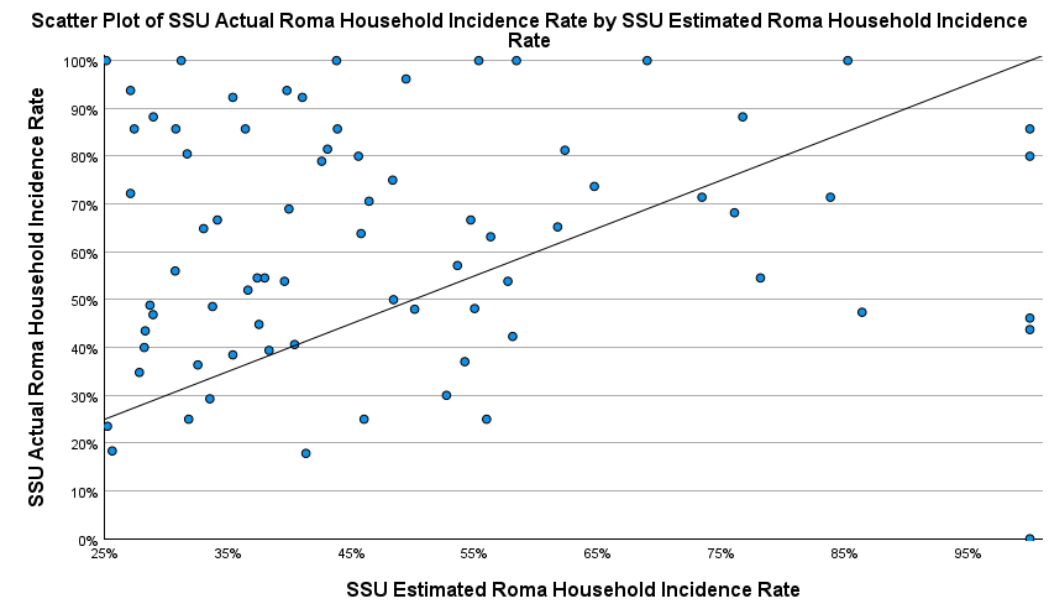


Accuracy of the mapping (2)

Greece



Romania



Accuracy of the mapping (3)

Key Observations

1) The expected incidence of Roma based on the mapping and actual eligibility based on the fieldwork do not correlate very strongly. Clear evidence of large disparities in these two estimates.
2) Evidence of a systematic bias – the eligibility during FW typically higher on average than the incidence based on the mapping.
3) These issues also present in MIDIS II – so unlikely to be a symptom of the mapping tool. More likely a combination of factors:
    - Experts find it easier to locate where Roma live but much more difficult to estimate population size
    - Large errors in mapped Roma estimates and in the model-based population estimates at grid level
    - Known issues with random walk designs – non-documentation or undocumented substitution of unproductive addresses by interviewers.

Lessons learnt

- Feedback from experts suggested that the tool was user friendly and the guidance clear
- Real time visibility of the mapped data allowed for QC checks to be done during the mapping process not after
- In the PSUs where the Roma population were dispersed the grid-based approach enabled the use of geo-spatial clustering techniques. This helped increase coverage.
- Mapping process was good at identifying where Roma live, but not good at estimating how many Roma live there.
- Disparities in the two rates (frame and fieldwork) has implications for the survey - managing interviewer workload and high variance in the design weights.

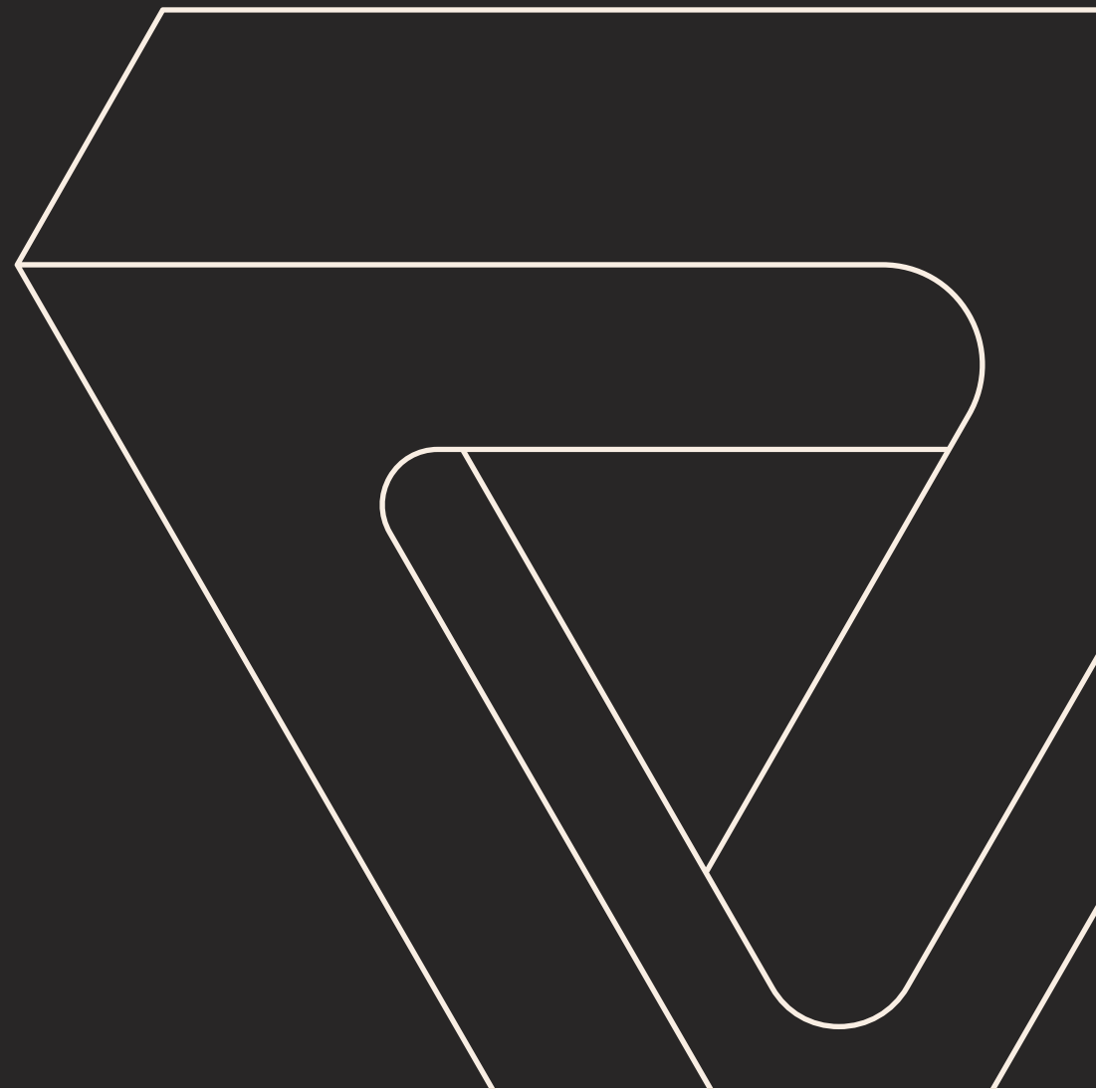
Potential improvements

- Use small grid sizes to maximize the chance that the random start address falls over or close to where Roma live, thereby minimizing the need to replace.
- If feasible have multiple experts map the same PSU
- Improve the functionality of the tool
  - Allow experts to select multiple neighboring grids before providing one population estimate
  - Include pop-up prompts that compare their Roma estimate to grid level population estimates and PSU level Roma estimates
  - To make the selection of a large number of small grids easier – first select larger grid, then progressively smaller.

Thanks

Jamie Burnett
Jamie.burnett@veriangroup.com

Dr David Simon
dr.david.simon@gmail.com



verian

Powering decisions
that shape the world.