EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

From a smart travel-survey proof of concept towards an official statistic

Jonas Klingwort, Yvonne Gootzen, Mike Vollebregt - Statistics Netherlands

Daniëlle Remmerswaal – Utrecht University

Barry Schouten - Statistics Netherlands and Utrecht University

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

Smart travel surveys – opportunities and challenges

Travel surveys like to know how public infrastructure is being used, why, with whom and for what purpose.

Opportunities

Track travel trajectories

Stop-track segmentation

Travel mode prediction

Stop purpose prediction

Challenges

Willingness to go smart

Diversity in devices

UI-UX

Database management

AI-ML

Legal acceptance

Studies SN

Field test 2018

- PoC
- Randomized stop-track
- Randomized incentives

Field test 2022-2023

- Randomized push-to-smart
- Randomized respondent edits
- Randomized study duration
- Tailored battery management
- Analysis DPIA

Push/nudge-to-smart

Lessons learned

Field test 2018

- Relatively high response rate
- Selection on age and education
- Around 20% drop-out in a week

Field test 2022-23

- Much lower response rate
- Similar drop-out as 2018
- Offering online + app confusing
- 7 days RR > one day RR

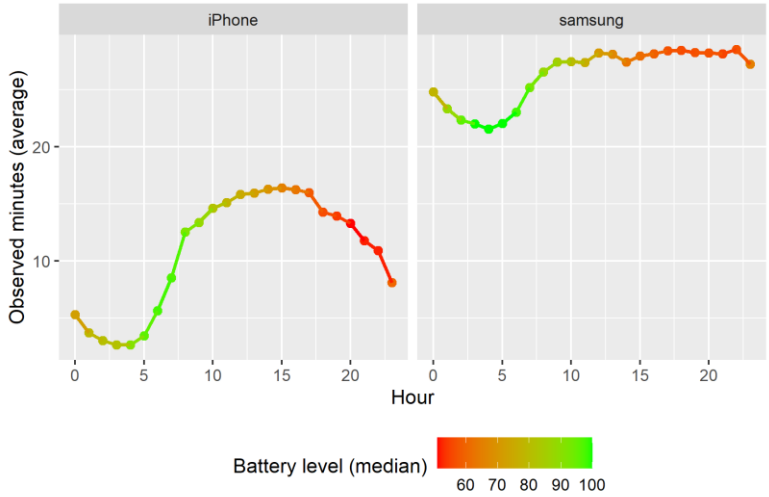
	Sample		Incentive experiment		
	Former respondents	Fresh sample	5 + 5 + 5	5 + 0 + 10	5 + 0 + 20
Number	422	252	191	231	252
Percentage	44%	27%	30%	36%	40%

MODE	Number of days		Timing questionnaire		
	1 days	7 days	Invitation	Reminder 1	Reminder 2
App	11%	13%	11%	12%	12%
Questionnaire	7%	8%	10%	7%	4%
Total registration	17%	20%	20%	19%	16%

Data quality validation and technical performance

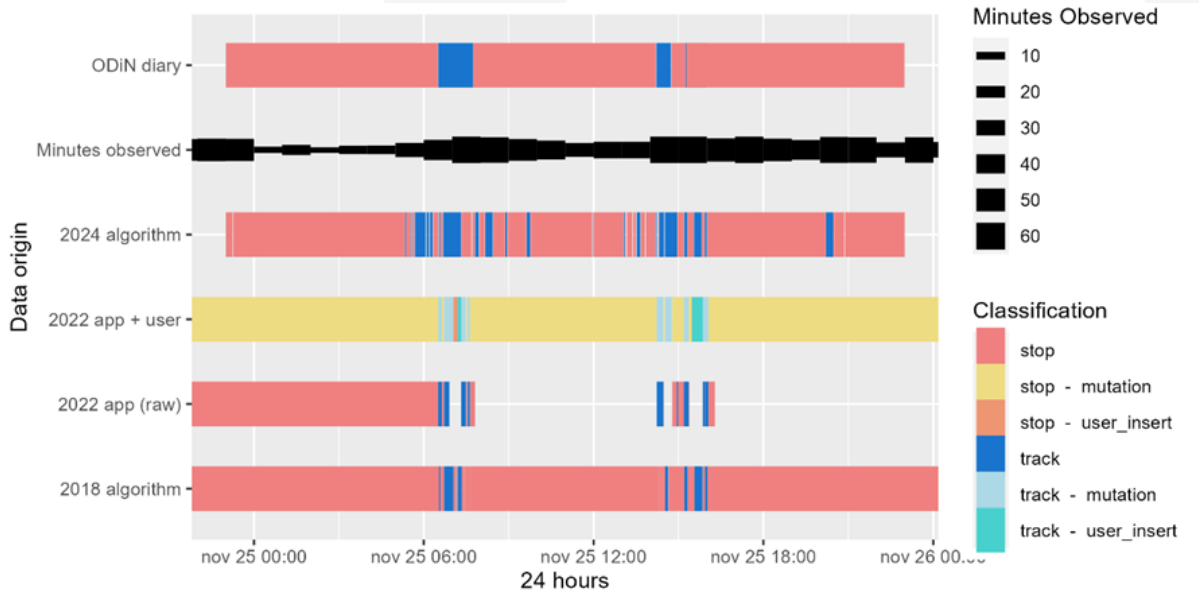
Technical performance:

- Depends on OS, brand and model
- Within a certain range performance is comparable.
- Chinese brands cause issues as native routines cannot be used



Comparison diary – app data:

- Respondents tend to adjust app data towards the diary
- Segmentation routines tend to be too sensitive
- Decision rules need more work



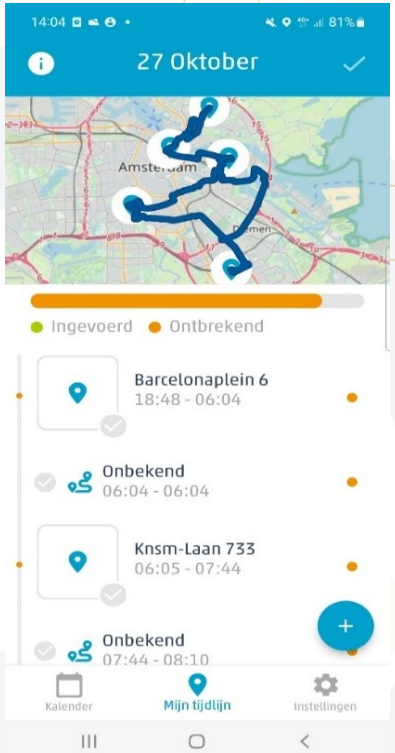
Respondent editing and database management

Respondents used all editing options (mutate start-end times, labelling, delete, add)

Not offering options leads to less enjoyment

Database management allowing for mutations and reconstructions (paradata) is key

Indicators (%)	Day 1	Day 2	Day 6	Day 7
Deleted >0 stop-tracks	80%	92%	96%	89%
Labeled stop-tracks	59%	77%	78%	72%
Modified time stop-tracks	31%	51%	56%	53%
Added >0 stop-tracks	26%	30%	35%	31%

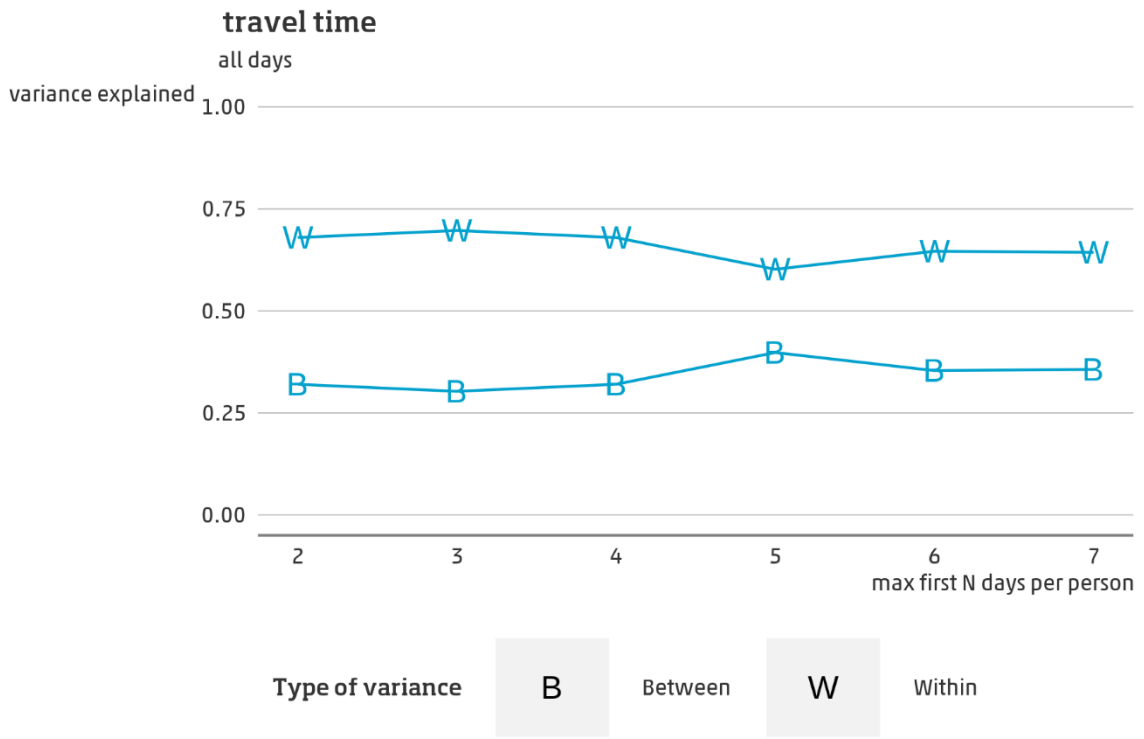


Study duration and legal acceptance

The length of the location tracking reporting period is an influential decision both methodologically and legally (privacy-by-design).

+ No reference day needed, app RR↑, sample size↓
– Data minimization, integration with non-smart diary

A multi-level model shows that within variance is larger than between variance. Seven days provides the same information as 2.5 different persons.



AI-ML predictions

AI-ML is employed for stop-track segmentation and may be employed for travel mode prediction and/or stop purpose prediction. Experiences:

- Predictions can be made near real-time, but including ‘memory’ is imperative; frequency of visits to a location is the strongest feature
- Respondent editing will remain to be needed
- Incomparability may result from varying granularity in POI databases

Model	ANN	RF	XGB	SVM	NB
Overall	72.3	77.5	77.7	70.8	42.7
Pick-up	75	78.2	75.3	70.7	50
Edu	83.3	80.3	83.1	80.4	50
Others	62.9	69	74.1	64.4	50
Transit	80.1	79.1	83.3	71.7	50
Sport	71.1	73.1	81.2	68.1	50
Home	91.3	92.8	93.2	90.7	52.6
Visit	67.5	62.3	74.6	67.3	50
Work	84.1	90.8	87.9	83.9	49.9
Shop	81.5	86	83.8	80.1	55.5

Discussion

Experiences:

- Creating a baseline app is relatively straightforward and fast, but the AI-ML and UI-UX are easily underestimated. Collaboration between statistical institutes would be efficient
- Travel surveys probably have the strongest business case in going smart from a respondent burden perspective, but are also the most complex due to the continuous nature of smart data collection

Next steps:

- Another go at the push/nudge-to-smart recruitment and motivation strategy
- Further elaboration of AI-ML procedures

Contact: Jonas Klingwort (j.klingwort@cbs.nl)or Barry Schouten (jg.schouten@cbs.nl)

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS

2024 ESTORIL - PORTUGAL

Instituto Nacional de Estatística
Statistics Portugal

eurostat

The conference is partly
financed by the European Union