Designed Big Data:
Blending Surveys and Digital Trace Data

Bella Struminskaya

Utrecht University

Short introduction

Assistant Professor Utrecht University

Affiliated Researcher Statistics Netherlands

Data Collection Committee ODISSEI Observatory (GGP, ESS, SHARE, EVS, NKO)

Board member German Society for Online Research (DGOF)

Board member IEDI – Integrated Data Collection Infrastructure

AAPOR Education Committee member

Senior researcher / junior researcher at GESIS (2010-2016)

PhD Methodology & Statistics, Utrecht University (2014)

M.A. Sociology University of Mannheim (2010)

Utrecht University

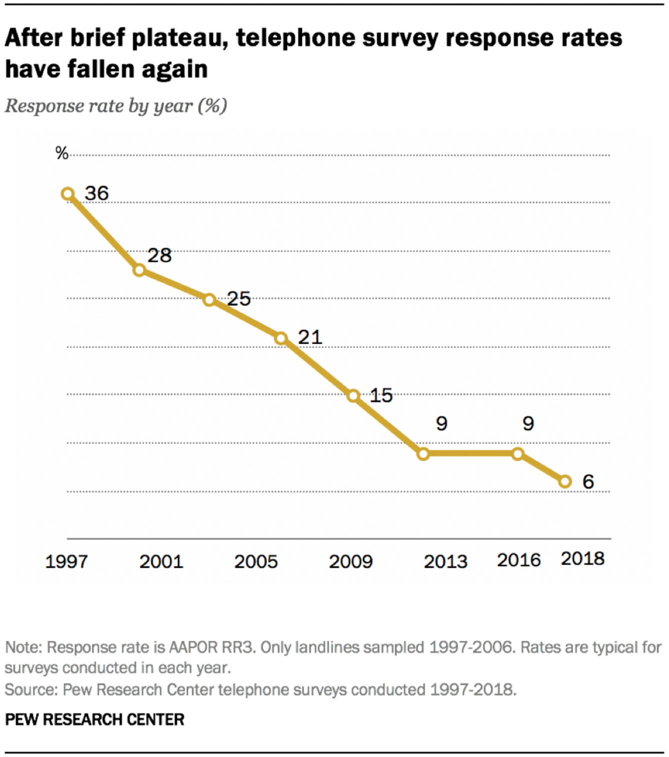
Statistics Netherlands

ODISSEI

DGOF

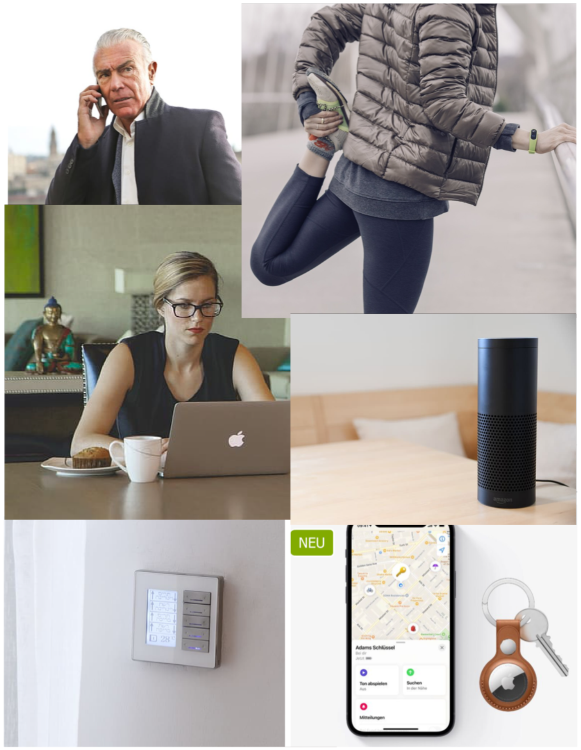
AAPOR

Problem



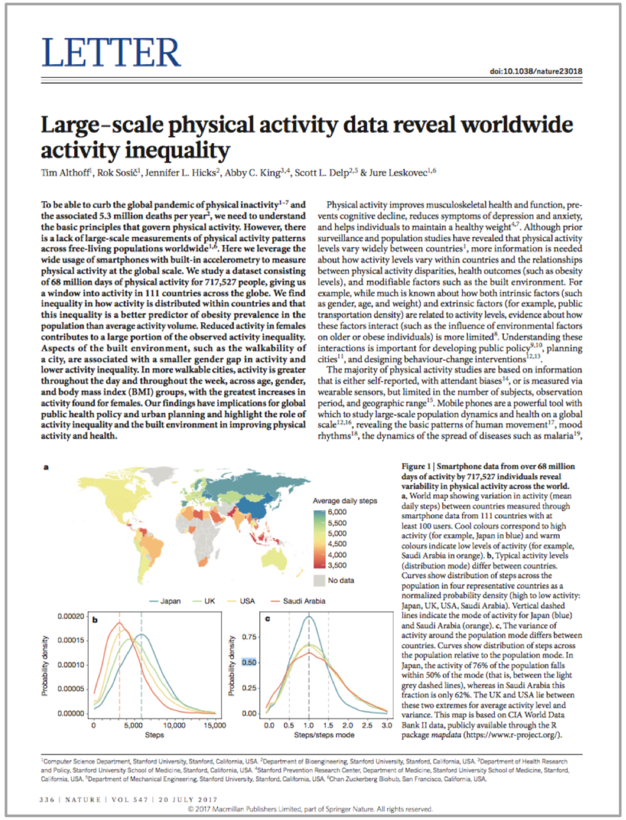
Source: Pew Research Center 2019

Opportunity



Solution (?)

Can we study physical activity using passive data?



LETTER

doi:10.1038/nature23018

Large-scale physical activity data reveal worldwide activity inequality

Tim Althoff[1], Rok Sosič[1], Jennifer L. Hicks[2], Abby C. King[3,4], Scott L. Delp[2,5] & Jure Leskovec[1,6]

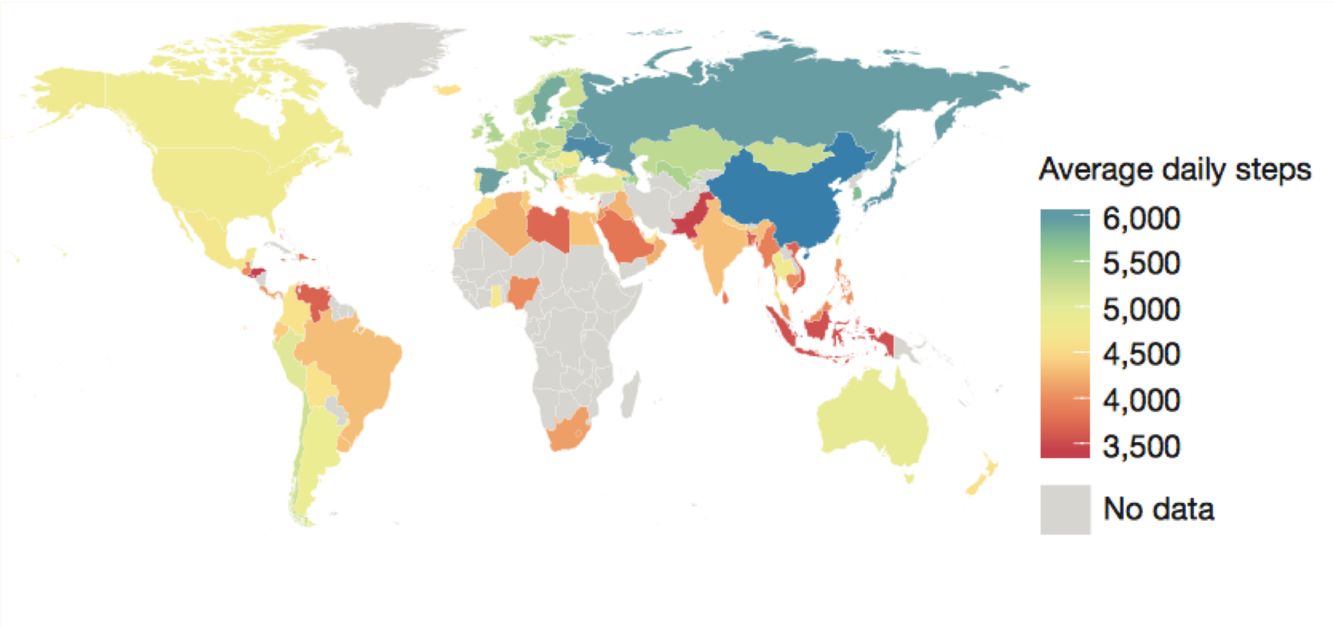
To be able to curb the global pandemic of physical inactivity[1–7] and the associated 5.3 million deaths per year[2], we need to understand the basic principles that govern physical activity. However, there is a lack of large-scale measurements of physical activity patterns across free-living populations worldwide[1,6]. Here we leverage the wide usage of smartphones with built-in accelerometry to measure physical activity at the global scale. We study a dataset consisting of 68 million days of physical activity for 717,527 people, giving us a window into activity in 111 countries across the globe. We find inequality in how activity is distributed within countries and that this inequality is a better predictor of obesity prevalence in the population than average activity volume. Reduced activity in females contributes to a large portion of the observed activity inequality. Aspects of the built environment, such as the walkability of a city, are associated with a smaller gender gap in activity and lower activity inequality. In more walkable cities, activity is greater throughout the day and throughout the week, across age, gender, and body mass index (BMI) groups, with the greatest increases in activity found for females. Our findings have implications for global public health policy and urban planning and highlight the role of activity inequality and the built environment in improving physical activity and health.

Physical activity improves musculoskeletal health and function, prevents cognitive decline, reduces symptoms of depression and anxiety, and helps individuals to maintain a healthy weight[4,7]. Although prior surveillance and population studies have revealed that physical activity levels vary widely between countries[1], more information is needed about how activity levels vary within countries and the relationships between physical activity disparities, health outcomes (such as obesity levels), and modifiable factors such as the built environment. For example, while much is known about how both intrinsic factors (such as gender, age, and weight) and extrinsic factors (for example, public transportation density) are related to activity levels, evidence about how these factors interact (such as the influence of environmental factors on older or obese individuals) is more limited[8]. Understanding these interactions is important for developing public policy[9,10], planning cities[11], and designing behaviour-change interventions[12,13].

The majority of physical activity studies are based on information that is either self-reported, with attendant biases[14], or is measured via wearable sensors, but limited in the number of subjects, observation period, and geographic range[15]. Mobile phones are a powerful tool with which to study large-scale population dynamics and health on a global scale[12,16], revealing the basic patterns of human movement[17], mood rhythms[18], the dynamics of the spread of diseases such as malaria[19],

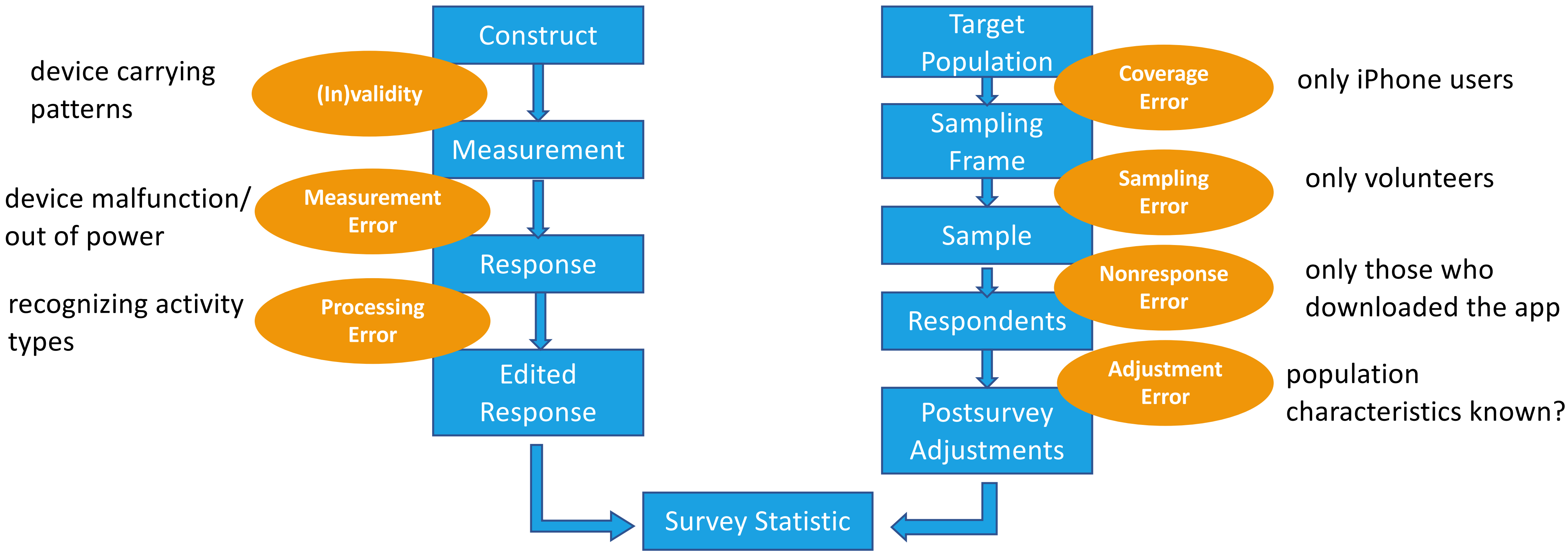
Figure 1 | Smartphone data from over 68 million days of activity by 717,527 individuals reveal variability in physical activity across the world. **a,** World map showing variation in activity (mean daily steps) between countries measured through smartphone data from 111 countries with at least 100 users. Cool colours correspond to high activity (for example, Japan in blue) and warm colours indicate low levels of activity (for example, Saudi Arabia in orange). **b,** Typical activity levels (distribution mode) differ between countries. Curves show distribution of steps across the population in four representative countries as a normalized probability density (high to low activity: Japan, UK, USA, Saudi Arabia). Vertical dashed lines indicate the mode of activity for Japan (blue) and Saudi Arabia (orange). **c,** The variance of activity around the population mode differs between countries. Curves show distribution of steps across the population relative to the population mode. In Japan, the activity of 76% of the population falls within 50% of the mode (that is, between the light grey dashed lines), whereas in Saudi Arabia this fraction is only 62%. The UK and USA lie between these two extremes for average activity level and variance. This map is based on CIA World Data Bank II data, publicly available through the R package *mapdata* (https://www.r-project.org/).

[1]Computer Science Department, Stanford University, Stanford, California, USA. [2]Department of Bioengineering, Stanford University, Stanford, California, USA. [3]Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California, USA. [4]Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, California, USA. [5]Department of Mechanical Engineering, Stanford University, Stanford, California, USA. [6]Chan Zuckerberg Biohub, San Francisco, California, USA.



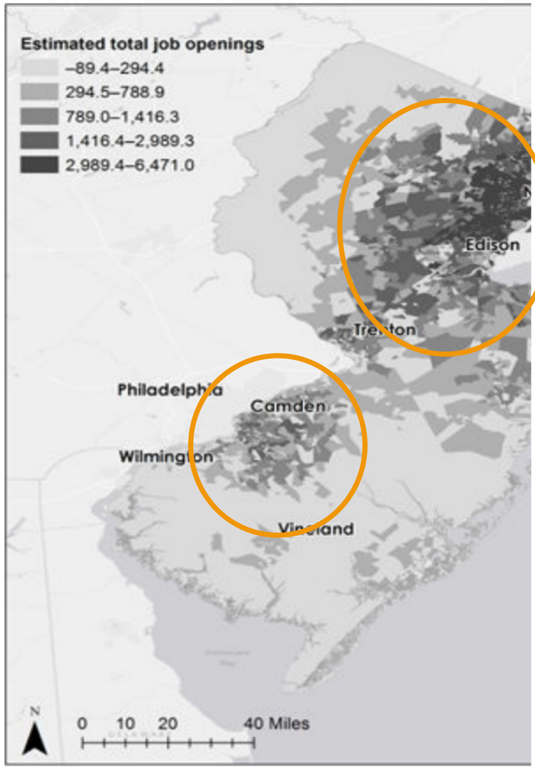
Althoff, T., Hicks, J. L., King, A. C., Delp, S. L., & Leskovec, J. (2017). Large-scale physical activity data reveal worldwide activity inequality. *Nature*, *547* (7663), *336-339*

Total (Survey) Error & Althoff et al.

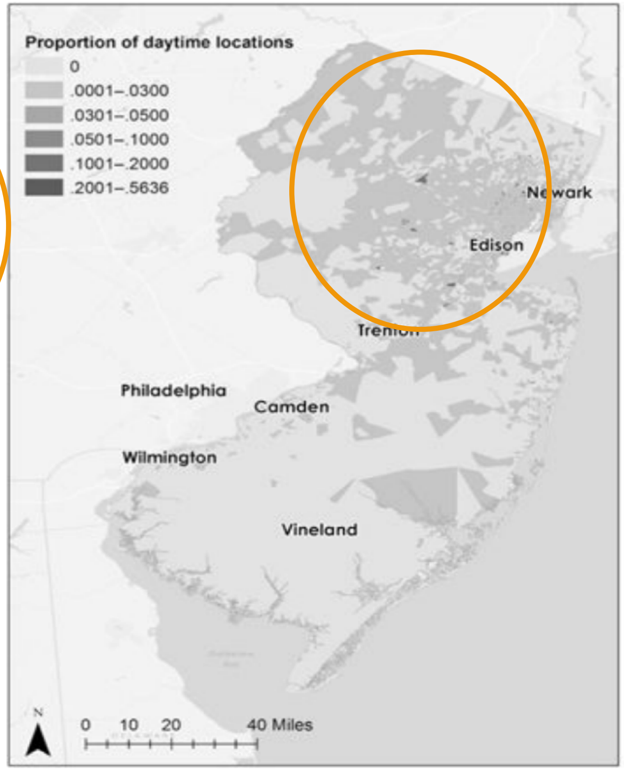


(Groves et al. 2004)

Can we study job search and work behavior of marginalized job seekers?



Job openings



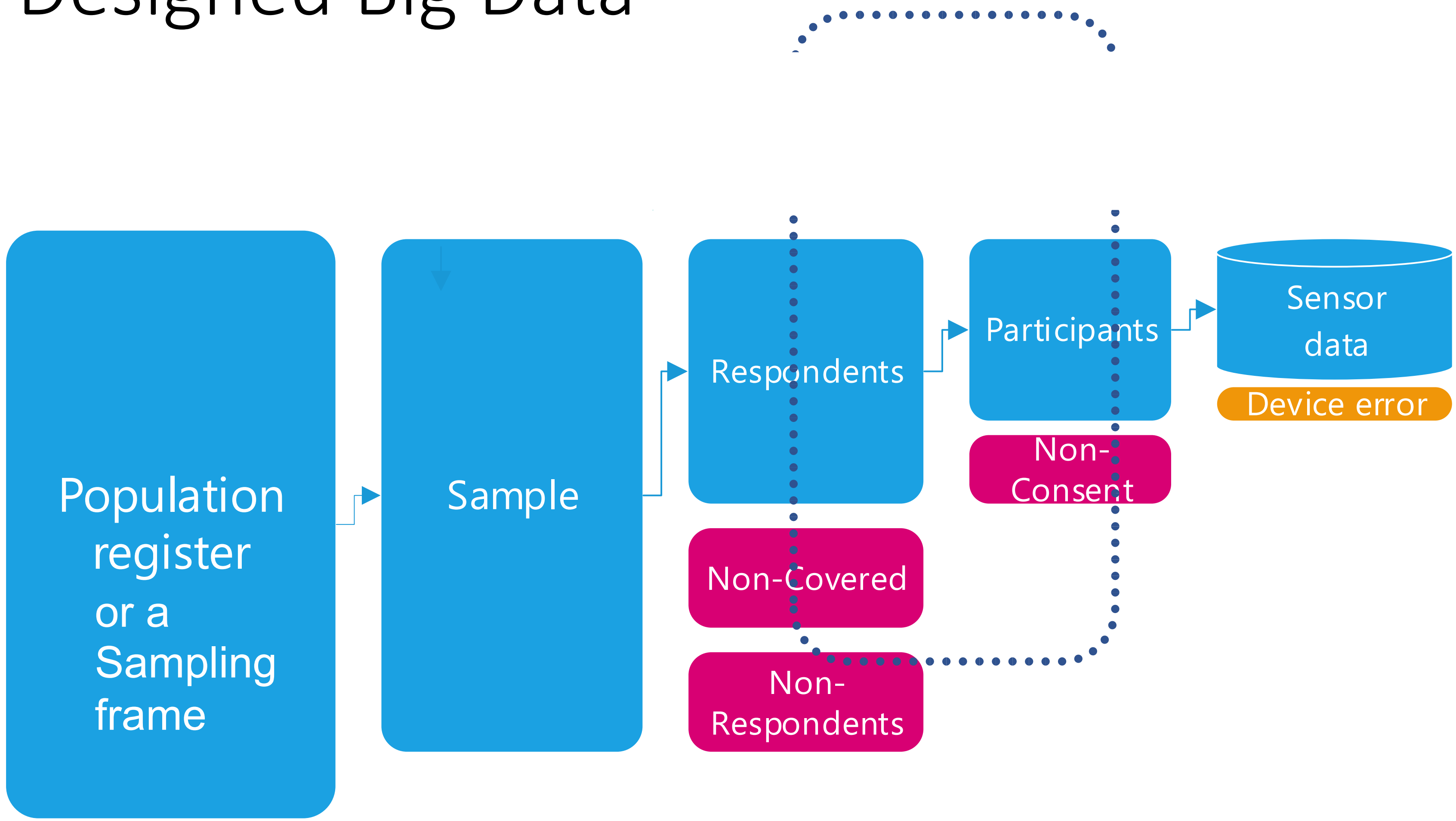
Daytime locations of parolees

- 133 parolees (RR 89%)
- Geolocation, calls & texts, EMA
- Spatial mismatch: low-skilled, nonwhite job seekers within central cities, job opportunities in outlying areas
- Residential mismatch lengthens time to employment
- Mobility can compensate for residential deficits

(Sugie 2018; Sugie and Lens 2017)

Bridging the gap between surveys and Big Data

“Designed Big Data”



(Struminskaya et al. 2020)

What are the mechanisms of WTS and sharing of app, sensor, and digital trace data?

Public Opinion Quarterly, Vol. 84, No. 3, 2020, pp. 725–759

UNDERSTANDING WILLINGNESS TO SHARE SMARTPHONE-SENSOR DATA

BELLA STRUMINSKAYA*
VERA TOEPOEL
PETER LUGTIG
MARIEKE HAAN
ANNEMIEKE LUITEN
BARRY SCHOUTEN

Abstract The growing smartphone penetration and the integration of smartphones into people's everyday practices offer researchers opportunities to augment survey measurement with smartphone-sensor measurement or to replace self-reports. Potential benefits include lower measurement error, a widening of research questions, collection of *in situ* data, and a lowered respondent burden. However, privacy considerations and other concerns may lead to nonparticipation. To date, little is known about the mechanisms of willingness to share sensor data by the general population, and no evidence is available concerning the stability of willingness. The present study focuses on survey respondents' willingness to share data collected using smartphone sensors (GPS, camera, and wearables) in a probability-based online panel of the general population of the Netherlands. A randomized experiment varied study sponsor, framing of the request, the emphasis on control over the data collection process, and assurance of privacy and confidentiality.

BELLA STRUMINSKAYA is an assistant professor in the Department of Methodology and Statistics at Utrecht University, Utrecht, the Netherlands. VERA TOEPOEL is an assistant professor in the Department of Methodology and Statistics at Utrecht University, Utrecht, the Netherlands. PETER LUGTIG is an associate professor in the Department of Methodology and Statistics at Utrecht University, Utrecht, the Netherlands. MARIEKE HAAN is an assistant professor in the Department of Sociology at the University of Groningen, Groningen, the Netherlands. ANNEMIEKE LUITEN is a data collection specialist at Statistics Netherlands, Heerlen, the Netherlands. BARRY SCHOUTEN is a senior methodologist at Statistics Netherlands, The Hague, the Netherlands, and professor in the Department of Methodology and Statistics at Utrecht University, Utrecht, the Netherlands. The authors thank Max van der Velde for research assistance and three anonymous reviewers for feedback on an earlier draft of the manuscript. This paper makes use of data of the LISS (Longitudinal Internet Studies for the Social Sciences) panel administered by CentERdata (Tilburg University, the Netherlands). Data collection was funded by Statistics Netherlands. *Address correspondence to Bella Struminskaya, Utrecht University, Padualaan 14, 3584CH Utrecht, the Netherlands; email: b.struminskaya@uu.nl.

doi:10.1093/poq/nfaa044 Advance Access publication 13 February 2021

Public Opinion Quarterly, Vol. 85, Special Issue, 2021, pp. 423–462

SHARING DATA COLLECTED WITH SMARTPHONE SENSORS
WILLINGNESS, PARTICIPATION, AND NONPARTICIPATION BIAS

BELLA STRUMINSKAYA*
PETER LUGTIG
VERA TOEPOEL
BARRY SCHOUTEN
DEIRDRE GIESEN
RALPH DOLMANS

Abstract Smartphone sensors allow measurement of phenomena that are difficult or impossible to capture via self-report (e.g., geographical movement, physical activity). Sensors can reduce respondent burden by eliminating survey questions and improve measurement accuracy by replacing/augmenting self-reports. However, if respondents who are not willing to collect sensor data differ on critical attributes from those who are, the results can be biased. Research on the mechanisms of willingness to collect sensor data mostly comes from (nonprobability) online panels and is hypothetical (i.e., asks participants about the likelihood of participation in a sensor-based study). In a cross-sectional general population randomized experiment, we investigate how

BELLA STRUMINSKAYA is an assistant professor in the Department of Methodology and Statistics at Utrecht University, Utrecht, The Netherlands. PETER LUGTIG is an associate professor in the Department of Methodology and Statistics at Utrecht University, Utrecht, The Netherlands. VERA TOEPOEL is an assistant professor in the Department of Methodology and Statistics at Utrecht University, Utrecht, The Netherlands. BARRY SCHOUTEN is a senior methodologist at Statistics Netherlands, The Hague, The Netherlands, and a professor in the Department of Methodology and Statistics at Utrecht University, Utrecht, The Netherlands. DEIRDRE GIESEN is a senior methodologist at Statistics Netherlands, The Hague, The Netherlands. RALPH DOLMANS is an information and communication technology developer at the Blaise team at Statistics Netherlands, Heerlen, The Netherlands. The authors thank Ole Mussmann for assistance in preparing the sensor plug-ins; Jelmer de Groot, Annemieke Luiten, and Vivian Meertens for assistance with data collection; three anonymous reviewers for feedback on an earlier draft of the manuscript; and Frederick Conrad for his invaluable feedback during the finalization of the paper. Data collection was funded by an internal grant from Statistics Netherlands. *Address correspondence to Bella Struminskaya, Utrecht University, Padualaan 14, 3584CH, Utrecht, The Netherlands; email: b.struminskaya@uu.nl.

doi:10.1093/poq/nfab025 Advance Access publication September 3, 2021

Public Opinion Quarterly, Vol. 83, Special Issue 2019, pp. 210–235

WILLINGNESS TO PARTICIPATE IN PASSIVE MOBILE DATA COLLECTION

FLORIAN KEUSCH*
BELLA STRUMINSKAYA
CHRISTOPHER ANTOUN
MICK P. COUPER
FRAUKE KREUTER

Abstract The rising penetration of smartphones now gives researchers the chance to collect data from smartphone users through passive mobile data collection via apps. Examples of passively collected data include geolocation, physical movements, online behavior and browser history, and app usage. However, to passively collect data from smartphones, participants need to agree to download a research app to their smartphone. This leads to concerns about nonconsent and nonparticipation. In the current study, we assess the circumstances under which smartphone users are willing to participate in passive mobile data collection. We surveyed 1,947 members of a German nonprobability online

FLORIAN KEUSCH is an assistant professor in the School of Social Sciences at the University of Mannheim, Mannheim, Germany, and an adjunct research assistant professor in the Joint Program in Survey Methodology at the University of Maryland, College Park, MD, USA. BELLA STRUMINSKAYA is an assistant professor in the Department of Methodology and Statistics at Utrecht University, Utrecht, the Netherlands. CHRISTOPHER ANTOUN is an assistant research professor in the College of Information Studies and the Joint Program in Survey Methodology at the University of Maryland, College Park, MD, USA. MICK P. COUPER is a research professor in the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA, and a research professor in the Joint Program in Survey Methodology, University of Maryland, College Park, MD, USA. FRAUKE KREUTER is the director of the Joint Program in Survey Methodology, University of Maryland, College Park, MD, USA, a professor of statistics and methodology at the University of Mannheim, Mannheim, Germany, and head of the Statistical Methods Research Department at IAB, Nuremberg, Germany. The authors wish to thank Theresa Ludwig and Ruben Layes for research assistance, Edgar Treischl for help with the factorial design, Christoph Kern for support with the data analysis, and the members of the FK[2]RG and three anonymous reviewers for feedback on an earlier draft of the manuscript. This work was supported by the German Research Foundation (DFG) through the Collaborative Research Center SFB 884 "Political Economy of Reforms" (Project A9) [139943784 to Markus Frölich, F.K., and F.K.].
*Address correspondence to Florian Keusch, University of Mannheim, A5, 6, 68159 Mannheim, Germany; email: f.keusch@uni-mannheim.de.

doi:10.1093/poq/nfz007 Advance Access publication June 6, 2019

Struminskaya et al. 2021

- WTS & actual sharing
- Cross-section* (NL)
  COOP2=54%
- GPS, photos, video; no app

Struminskaya et al. 2020

- Willingness to share (WTS)
- Prob. LISS Panel (NL)
  2 waves, RR1 = 89%, 84%
- Share GPS, photos, video

Keusch et al. 2019

- Willingness to share (WTS)
- Nonprob. panel (DE)
  2 waves
- Download tracking app

Implementation (Struminskaya et al. 2021)

General consent

Framing, autonomy, & privacy explanation

GPS measurement

Photos & Video

Struminskaya et al. 2021

- WTS & actual sharing
- Cross-section* (NL)
- GPS, photos, video; no app
- Requests with rand. assig.:
  Autonomy over data collection
  Benefit framing
  Confidentiality assurance
- Fixed order of measurements
- Privacy concern, tech skills, prev. exp., survey exp.

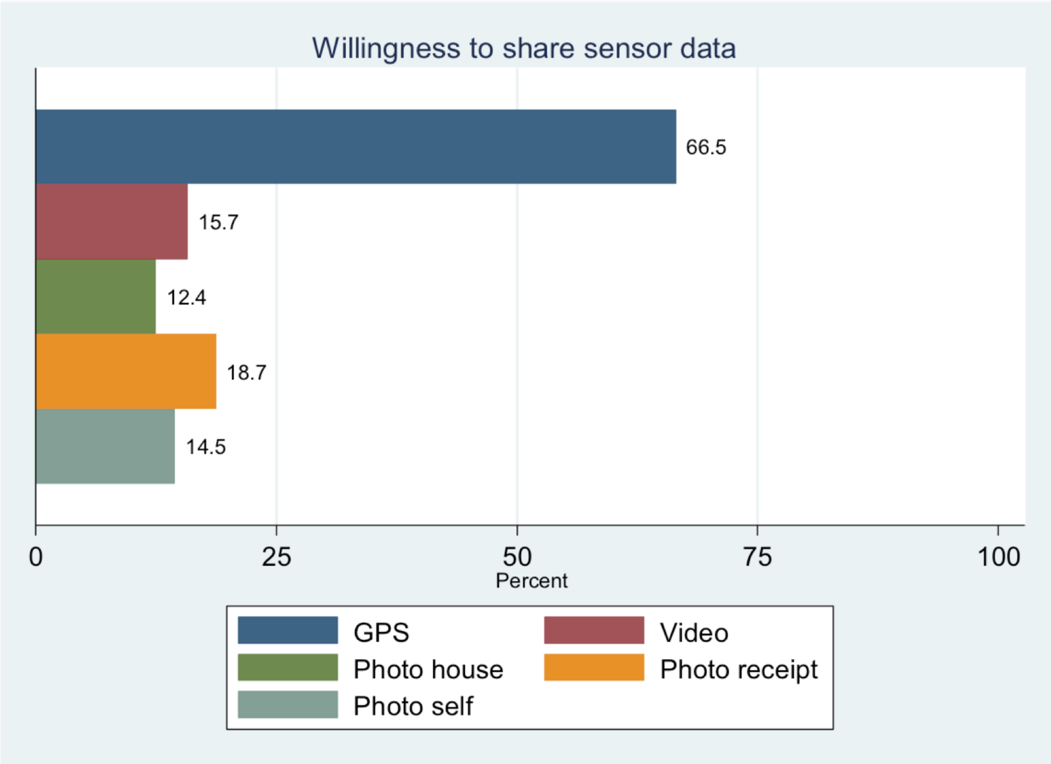
Struminskaya et al. 2020

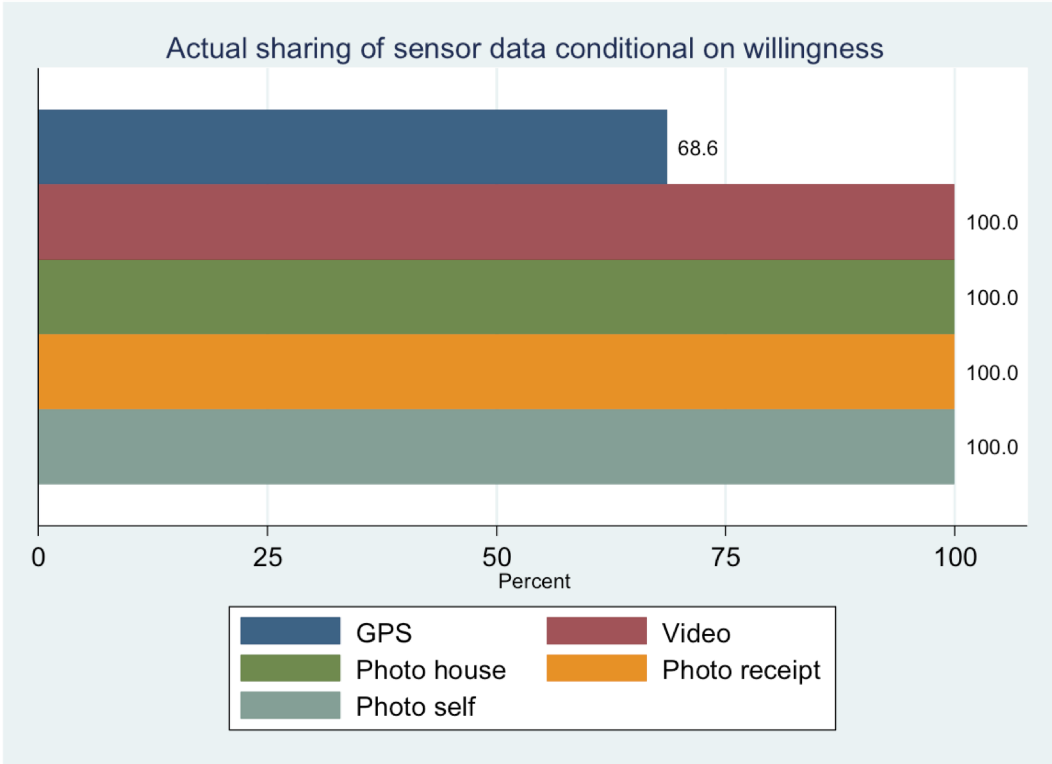
- Willingness to share (WTS)
- Prob. LISS Panel (NL)
  2 waves
- Share GPS, photos, video
- Vignettes w rand. assig.:
  Sponsor
  Autonomy over data collection
  Benefit framing
  Confidentiality assurance
- Randomized order of tasks
- Privacy concern, tech skills, prev. exp., survey exp.

Keusch et al. 2019

- Willingness to share (WTS)
- Nonprob. panel (DE)
  2 waves
- Download tracking app
- Vignettes w rand. assig.:
  Sponsor
  Autonomy over data collection
  Duration
  Topic
  Incentive
  Questions in-app
- Randomized order of vignettes
- Privacy concern, tech skills, prev. exp., survey exp.

Willingness and actual sharing (Dutch cross-section)



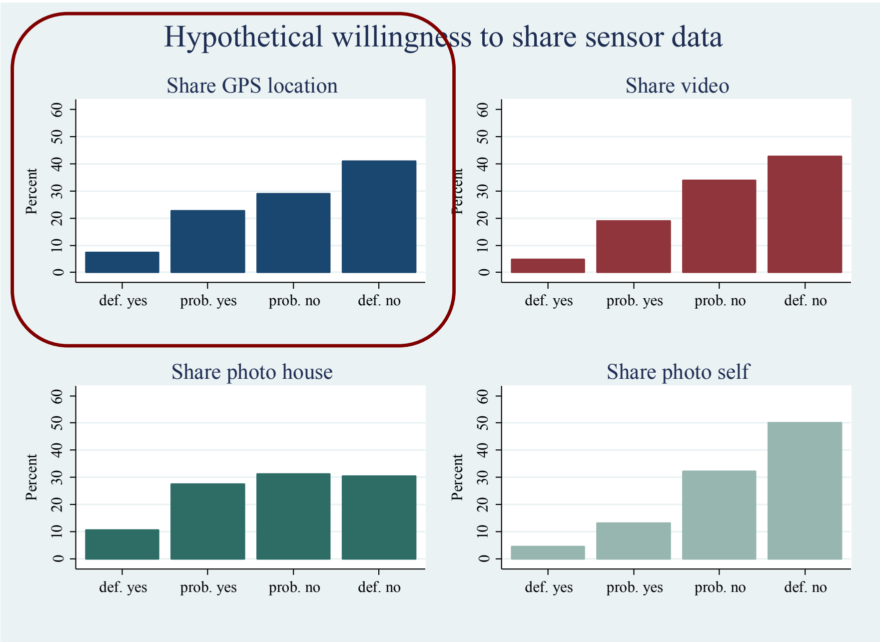


Participation rate GPS: 45.6%; n=1883 Dutch smartphone and tablet users

(Struminskaya et al. 2021)

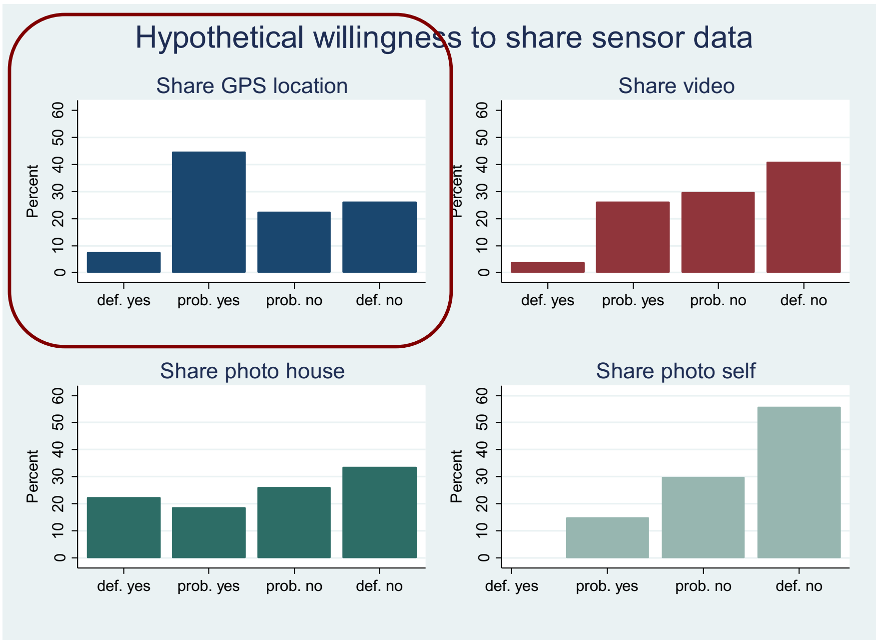
Hypothetical willingness & Order effects

Overall, randomized order



Order effect: Average marginal effect +5.6 p.p
(Struminskaya et al. 2020)

Order: GPS, Video, Photo house, Photo self



% Willing to share GPS:

- **If asked first: 41%**
- **If asked last: 26%**

Willingness mechanisms

Predictors	WTS GPS	Share GPS	Share video	Share photo house	Share photo receipt	Share photo self
Benefit framing	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Autonomy over data collection	.11***	-.06*	n.s.	n.s.	.04*	n.s.
Privacy	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

n=1,853; Average marginal effects; covariates not shown

(Struminskaya et al. 2021)

Willingness mechanisms

Predictors	WTS GPS	Share GPS	Share video	Share photo house	Share photo receipt	Share photo self
Benefit framing	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Autonomy over data collection	.11***	-.06*	n.s.	n.s.	.04*	n.s.
Privacy	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

n=1,853; Average marginal effects; covariates not shown

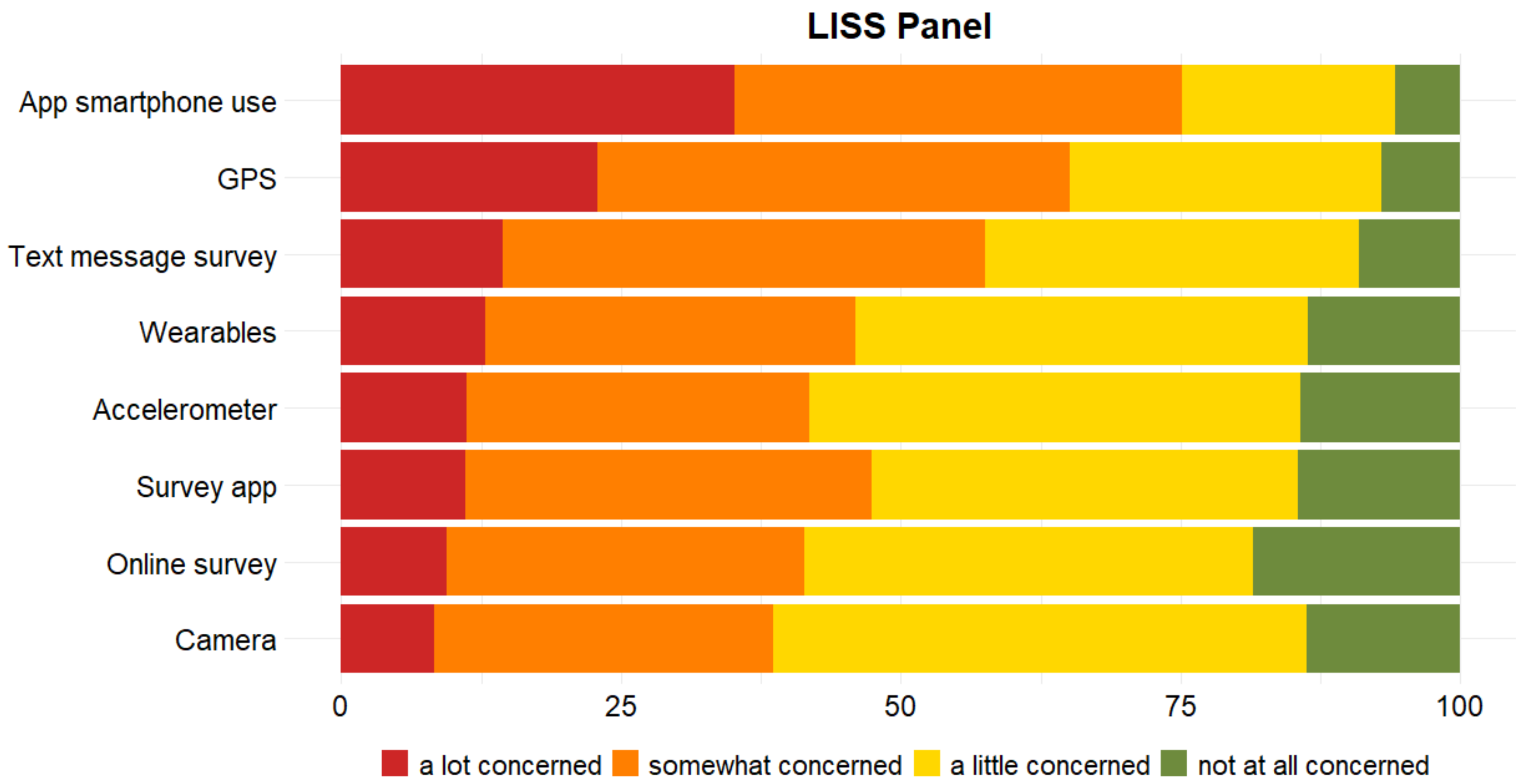
Predictors	Sharing
Order (asked first)	0.02 **
Sponsor University	0.09***
Sponsor Market Research	n.s.
Benefit framing	–0.02*
Autonomy over data collect.	n.s.
Privacy	n.s.

n=2,669; Average marginal effects; covariates not shown

In all 3 studies: sig. effects of smartphone use behaviors, mixed findings about the effect of privacy concerns, attitudes toward surveys, prior app download

(Struminskaya et al. 2020; 2021)

Concern by Type of Collected Data

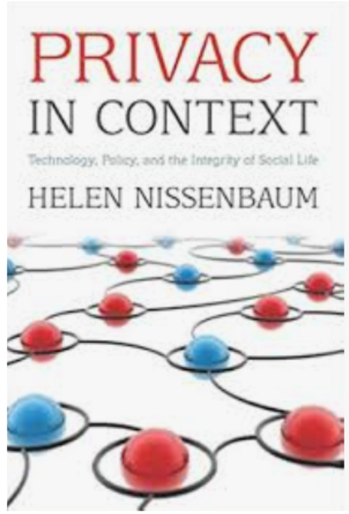


(Struminskaya & Keusch / ODISSEI study; see also Keusch, Struminskaya et al. 2020)

Summary (so far)

- Decisions about sharing are situation-specific, nuanced
- Hypothetical behavior differs from actual participation behavior
- The nature of the task more relevant than sensor
- Clear communication of who asks to share & for what purpose
- Balance between maximizing sharing and providing detailed information about the data (“backfire effects”)
- Ceiling effects possible due to loyalty, trust in sponsor

“Is this your current location? Yes/No”



How much does nonparticipation matter?

Nonresponse vs. nonparticipation bias

- Survey & sensor data linked to Dutch registries:
  general population register, education register, households register, register of motorized vehicles, dwelling register, employment register, tax register

$$Non - Response\ Bias\ (\bar{y}_{ADMIN}) = \bar{y}_{ADMIN,\ respondents} - \bar{y}_{ADMIN,\ gross\ sample}$$

$$Non - Participation\ Bias\ (\bar{y}_{ADMIN}) = \bar{y}_{ADMIN,\ consenters} - \bar{y}_{ADMIN,\ respondents}$$

Administrative data variables	Sample value (%)	Non-response bias (%)	Nonparticipation bias (%) GPS shared	Video surround.	Photo house	Photo receipt	Photo self
Age (25–34)	21.9	**-1.7***					
Gender (man)	42.5	0.9					
Education (high)	37.1	**3.4*****					
Ethnic background (non-Dutch)	16.3	**-1.8****					
Marital status (married)	45.8	**2.7****					
No. hh. members (2 people)	35.8	**2.9*****					
Owns a car	46.5	**2.5****					
Has a driver's license	82.9	**2.8*****					
Homeowner	74.4	**2.3****					
Urban (>=1500 addresses/km^2)	51.5	-0.9					
Size of township (>50,000)	54.2	-1.0					
In paid work	60.5	-0.4					
Income percentile (75th–100th)	40.0	**4.2*****					
Average abs. bias		2.1					

Administrative data variables	Sample value (%)	Non-response bias (%)	Nonparticipation bias (%) GPS shared	Video surround.	Photo house	Photo receipt	Photo self
Age (25–34)	21.9	**-1.7***	-1.3	**2.8****	-2.4	-0.5	**-2.0***
Gender (man)	42.5	0.9	1.1	1.4	**6.3*****	**-3.9*****	**3.8****
Education (high)	37.1	**3.4*****	-2.2	-0.9	**-6.8*****	-2.4	**-7.5*****
Ethnic background (non-Dutch)	16.3	**-1.8****	**-1.6***	0.8	**3.0****	**2.4***	**-2.6*****
Marital status (married)	45.8	**2.7****	0.5	**-2.7***	**3.5****	1.4	0.6
No. hh. members (2 people)	35.8	**2.9*****	-0.8	-0.7	**3.5****	-0.3	1.5
Owns a car	46.5	**2.5****	-1.4	1.0	**2.5***	-0.9	-0.7
Has a driver's license	82.9	**2.8*****	-0.3	0.6	**1.5***	-1.2	**1.6***
Homeowner	74.4	**2.3****	0.1	**-2.0***	**-4.6*****	**-2.3***	-1.2
Urban (>=1500 addresses/km^2)	51.5	-0.9	0.1	**5.9*****	**7.7*****	**7.0*****	0.7
Size of township (>50,000)	54.2	-1.0	-0.6	**5.2*****	**6.3*****	**3.8****	0.7
In paid work	60.5	-0.4	0.5	0.8	**-2.5***	-1.3	-2.0
Income percentile (75th–100th)	40.0	**4.2*****	-0.2	1.8	1.3	-1.9	-0.6
Average abs. bias		2.1	0.8	2.0	4.0	2.2	2.0

(Struminskaya et al. 2021)

- Small biases, but depends on research question

Administrative data variables	Sample value (%)	Non-response bias (%)	Nonparticipation bias (%) GPS shared	Video surround.	Photo house	Photo receipt	Photo self
Age (25–34)	21.9	**-1.7***	-1.3	**2.8****	-2.4	-0.5	**-2.0***
Gender (man)	42.5	0.9	1.1	1.4	**6.3*****	**-3.9*****	**3.8****
Education (high)	37.1	**3.4*****	-2.2	-0.9	**-6.8*****	-2.4	**-7.5*****
Ethnic background (non-Dutch)	16.3	**-1.8****	**-1.6***	0.8	**3.0****	**2.4***	**-2.6*****
Marital status (married)	45.8	**2.7****	0.5	**-2.7***	**3.5****	1.4	0.6
No. hh. members (2 people)	35.8	**2.9*****	-0.8	-0.7	**3.5****	-0.3	1.5
Owns a car	46.5	**2.5****	-1.4	1.0	**2.5***	-0.9	-0.7
Has a driver's license	82.9	**2.8*****	-0.3	0.6	**1.5***	-1.2	**1.6***
Homeowner	74.4	**2.3****	0.1	**-2.0***	**-4.6*****	**-2.3***	-1.2
Urban (>=1500 addresses/km^2)	51.5	-0.9	0.1	**5.9*****	**7.7*****	**7.0*****	0.7
Size of township (>50,000)	54.2	-1.0	-0.6	**5.2*****	**6.3*****	**3.8****	0.7
In paid work	60.5	-0.4	0.5	0.8	**-2.5***	-1.3	-2.0
Income percentile (75th–100th)	40.0	**4.2*****	-0.2	1.8	1.3	-1.9	-0.6
Average abs. bias		2.1	0.8	2.0	4.0	2.2	2.0

Mobility

Type of community

Living conditions

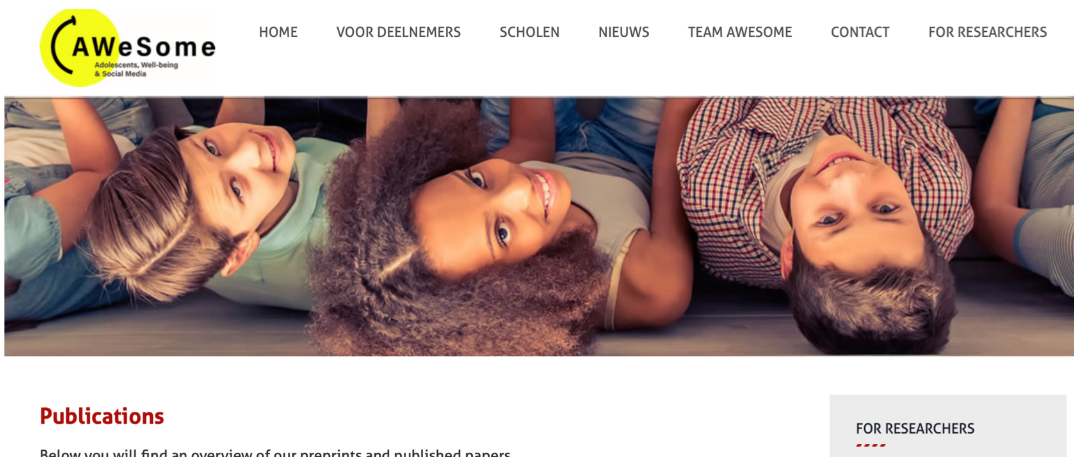
Financial situation

(Struminskaya et al. 2021)

Does this hold for digital traces?

Selectivity in donation of social media data

- Project AWeSome (Adolescents, Well-being, and Social Media) by University of Amsterdam
- Topics: social media use, well-being, social relationships, self-regulation
- Teenagers 13-15 yo in NL, recruited f2f at school, parental consent provided (N = 388)
- 80% have Instagram account(s)
- 32% donated Instagram data (raw)



Privacy (again) and loyalty (again) are key

Sociability

- Social comparison
- # good friends
- Friendship quality
- Parental phone rules
- Parental knowledge
- Adolescent dis-closure & secrecy*** (AME=.10)

Psychological chars

- Affective well-being
- Cognitive well-being
- Positive affect
- Negative affect
- Self-esteem* (AME=-.08)
- Loneliness
- Self-regulation*** (AME=1.23)

Social media, SP use

- # accounts
- Sphone self-monitor.
- Sphone type (iPhone)
- # followers
- Importance followers
- # likes on post
- Eval. of reactions
- Eval. # of reactions
- Importance positive reactions

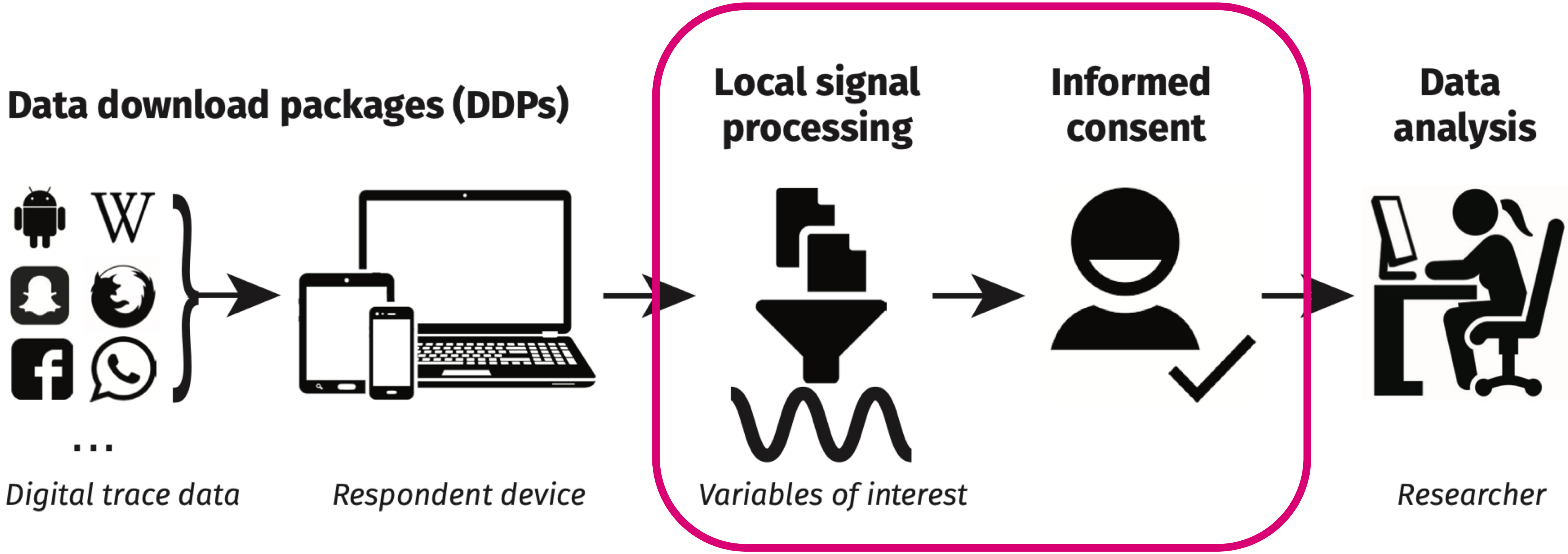
Study design

- # completed ESM1
- # completed ESM2*** (AME=.004)
- # completed surveys

Logistic regressions donated (1) vs. did not donate (0), models include sex, grade, education, N varies 146-316.

Giving agency to participants

Privacy-preserving Data Donation Workflow

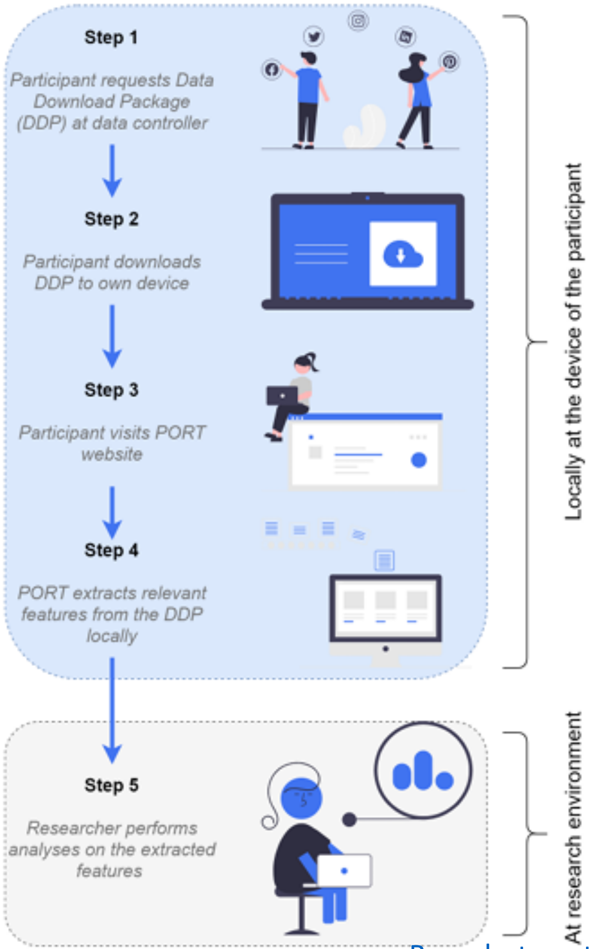


[Boeschoten et al. 2020]

Platform Digitale Infrastructuur
Social Sciences & Humanities

Digital Data Donation Infrastructure (D3I)

- With Laura Boeschoten, Daniel Oberski et al.
- 6 Dutch universities, funding for 3 years
- Data donation with local extraction (PORT)
- Agency (changing data)
- For: Google*, Meta*, Twitter, Netflix, Spotify
- Methodological questions:
  - Understanding of consent
  - Representativeness
  - UX
  - Measurement quality
  - Validity & Reliability

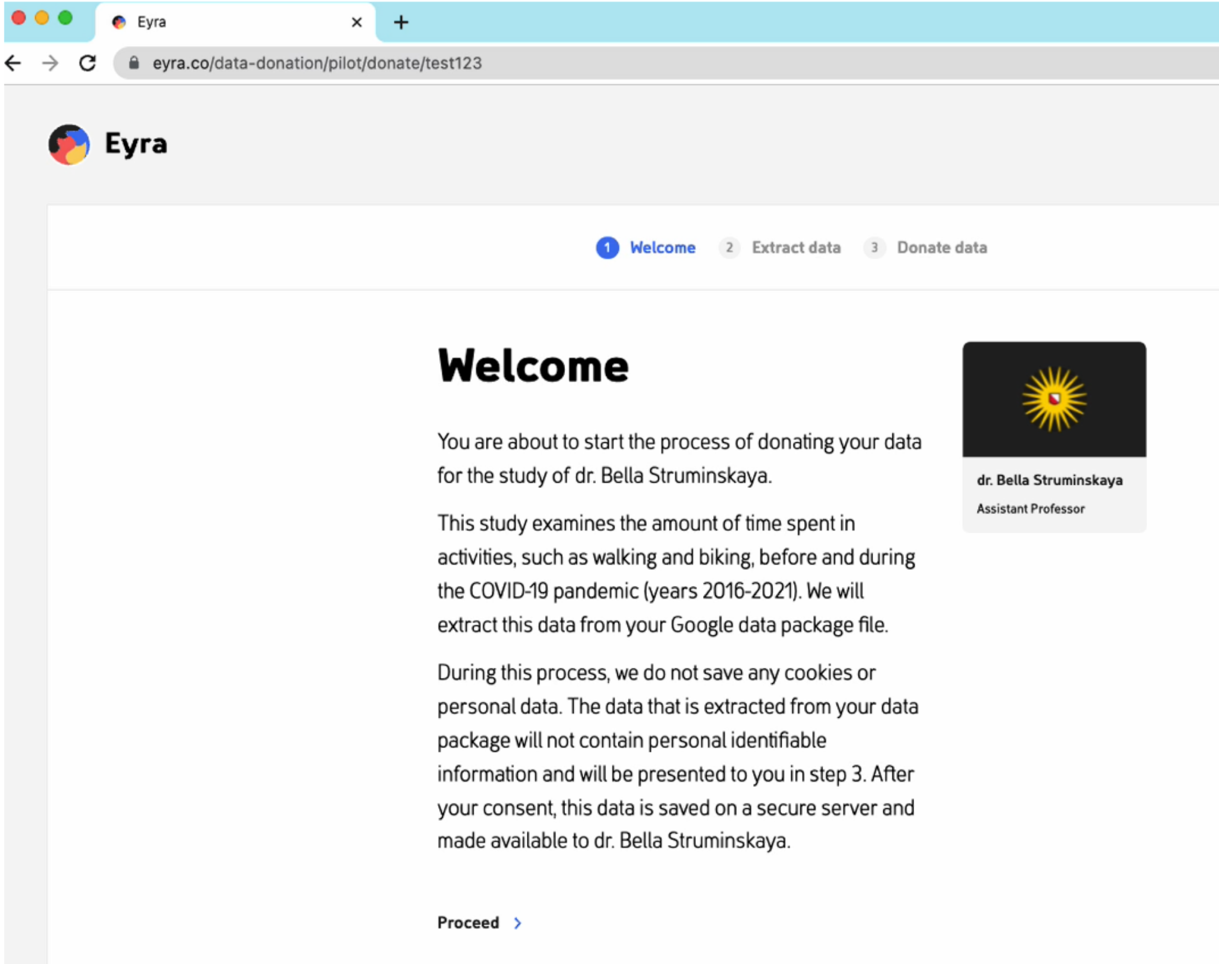


Boeschoten et al. 2022

Platform Digitale Infrastructuur
Social Sciences & Humanities

Google Location History Data Donation

- Seed project Focus Area Applied Data Science
- Together with geographers and computer scientists
- Testing the data donation workflow and ERB procedure
- Changes to Python script for large-scale implementation

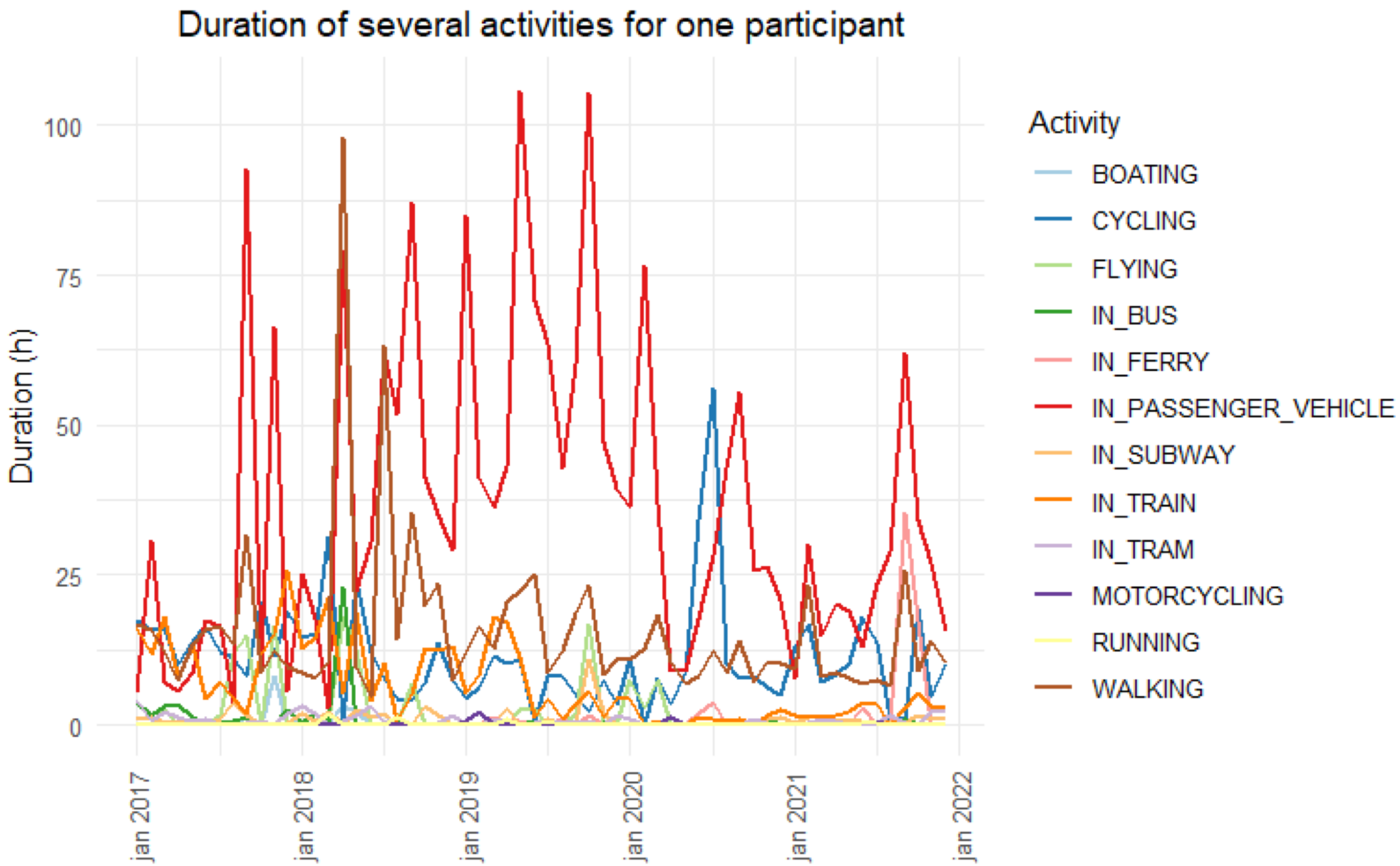


Platform Digitale Infrastructuur
Social Sciences & Humanities

Google Location History Data Donation

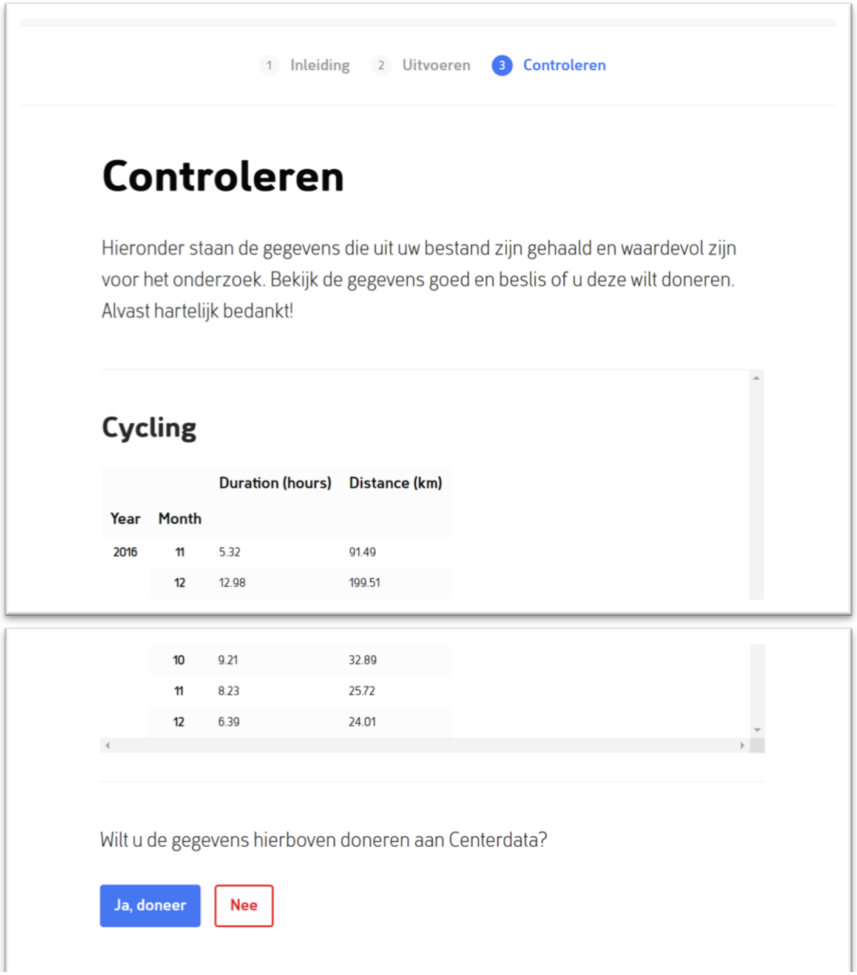
Video

Google Location History Data Donation



Google Location History Data Donation

- Study in CentERpanel August 2022
- N=1035 (75% AAPOR RR1)
- Integration of data donation (PORT)
- Willing = 30%, 144 donated (14%)
- Methodological questions:
  - Visualization prior to request
  - Understanding of consent request
  - Incentive amount (5€ vs. 10€)
  - Nonparticipation bias



Incentive & visualization

- No difference in incentives
  - 5€: willing to donate 32% (n=147)
  - 10€: willing to donate 34% (n=159)
  - Chi2(1) = 0.32, p=.574
  - Donated: 48% vs. 46% (Chi2(1) = 0.17, p=.676)

- No difference by showing how data looks like
  - Visualized: willing to donate 34% (n=159)
  - Not visualized: willing to donate 32% (n=147)
  - Chi2(1) = 0.56, p=.456
  - Donated: 46% vs. 48% (Chi2(1) = 0.17, p=.676)

De reisbewegingen kunt u in deze vragenlijst delen met Centerdata. Het is goed om te weten dat locaties die u hebt bezocht niet uit het pakketje worden gehaald en dus ook niet met Centerdata worden gedeeld. Er wordt **alleen** informatie gedeeld **hoe** u zich heeft verplaatst en **hoeveel tijd** u hieraan hebt besteed per maand en jaar.

[*if condition = 1* Een voorbeeld van hoe deze informatie eruitziet ziet u hieronder:

Cycling

		Duration (hours)	Distance (km)
Year	Month		
2021	8	1.14	6.32

In Bus

		Duration (hours)	Distance (km)
Year	Month		
2021	8	1.97	28.23

In Passenger Vehicle

		Duration (hours)	Distance (km)
Year	Month		
2021	8	23.31	375.84

Understanding the consent request

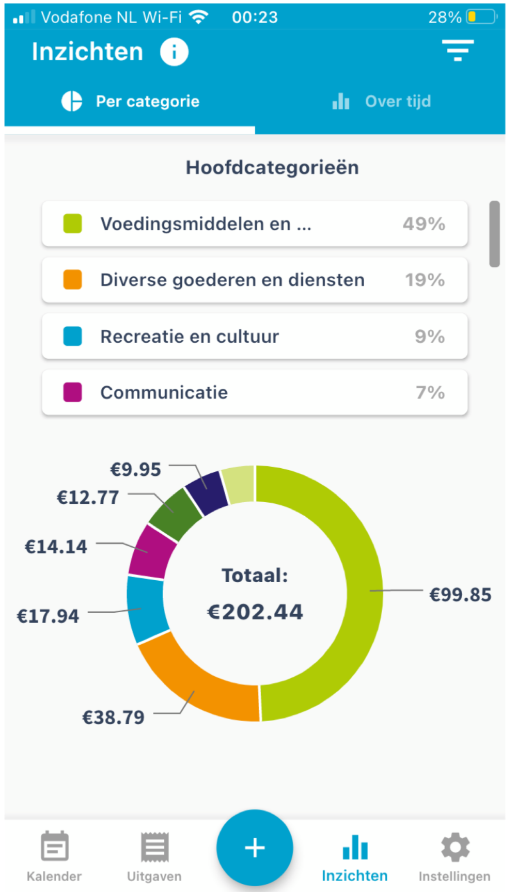
Statements asked to respondents	Correct %	Incorrect %	Don’t know %
You are asked to download information from Google. TRUE	48.8	19.8	31.4
The software implemented in the survey will extract the information on the number of hours you cycle, walk, take public transport, travel by car. TRUE	62.3	6.1	31.2
Information on all the locations you visited will be shared with Centerdata. FALSE	39.2	31.4	29.4
Google collects information on location about everyone. FALSE	24.8	46.6	28.5
From the data you will provide, the information can be traced back to you. FALSE	45.3	22.2	32.5
You will be able to inspect the data before sending it to Centerdata. TRUE	59.0	7.8	33.1
It is impossible to identify you as an individual from the data that you provide. TRUE	43.4	19.6	37.0

Understanding the consent request

- 5.5% had everything correct
- Mean correct: 3.23, median = 4
- People with more correct answers more likely to be willing & to donate:
  - 4.54 correct statements for willing
  - 2.56 correct statements for non-willing
  - OR = 1.572, $p < .001$

  - 5.33 correct statements for donated
  - 3.94 correct statements for not donated
  - OR = 1.795, $p < .001$

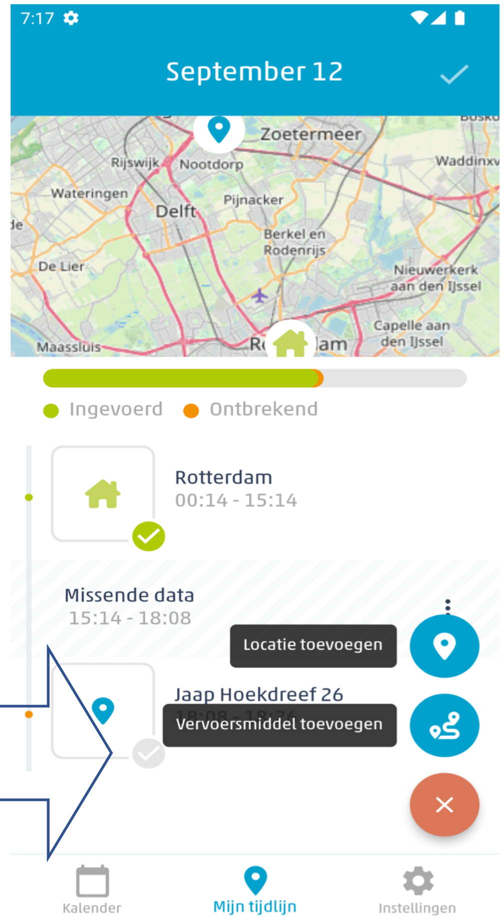
Respondents' agency in app-based surveys



Rodenburg, Schouten, Struminskaya

Household Budget Survey (HBS)
fall 2021,
NL, ES, LU
N=3916,
Completion = 16%
No influence of feedback on representativeness, data quality

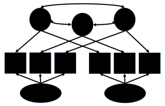
Travel app possibility to provide context to passive data, add data



Research agenda

Participant agency, informed consent, legal guidelines for data donation

- D3I pilots with various data sources (w/Araujo, Boeschoten, et al.)



DBD measurement quality, reliability & validity

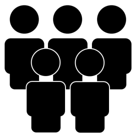
- Algorithmic bias (w/Ruben Bach, U Mannheim)
- Interventions based on digital trace data
- Book under contract CRC Press (w/Keusch, Eckman, Guyer)

Trusted Smart Surveys

- Increasing representation & adherence, UX and app design (w/*Danielle*, Peter, Barry, Eurostat)
- Data donation in official statistics (w/Laura, Thijs, *Danielle*, Barry, CBS)
- Feedback effects (w/*Evelien Rodenburg*, Barry Schouten, *Danielle Remmerswaal*)

Data Integration

- Integration of prob & nonprob surveys & sensors (w/*Camilla Salvatore* et al.)
- Integration of DBD and surveys (w/CBS)



Extending traditional data collection infrastructures

- Secure data linkage, informed consent to novel data types (w/Emery et al.)

Some final bits

- Most recent
  - Seed funding Institute 4 preventive health
  - Will appreciate brainstorm!

PROACT: PREVENTION OF ACUTE EVENTS OF OLDER PEOPLE

Preventing unplanned hospitalizations requires in-depth knowledge of patient-related, environmental and care-related factors. Using various digital tools, the research team tests and measures how older people can remain self-reliant for longer and what interventions are needed to prevent admission to the ED. *Researchers: Helianthe Kort (TU/e), Laura Genga (TU/e), Bella Struminskaya (UU), Nienke Bleijenberg (UMC Utrecht)*

- Some other ongoing
  - DFG-sponsored project “Panel conditioning in longitudinal studies”, UU external PhD candidate Fabienne Krämer, visiting M&S March-April 2023
  - More traditional survey methods: PhD students Isabella Minderop, Alexandra Asimov, Hannah Schwarz

PANEL CONDITIONING

Acknowledgements

Peter Lugtig, Barry Schouten,
Danielle McCool, Katie Roth,
Laurent Smeets, Ole Mussman,
Jelmer de Groot, Vera Toepoel,
Deirdre Giesen, Tom Oerlemans,
Ralph Dolmans,Annemieke Luiten,
Vivian Meertens, Jesper van Thor, Jeldrik
Bakker, Evelien Rodenburg, Anne Elevelt,
Tom Oerlemans, Laura Boeschoten,
Thijs Carriere, Niek de Schipper,
Daniel Oberski, Danielle Remmerswaal,
Florian Keusch, Christopher Antoun,
Mick Couper, Frauke Kreuter, and more

Thank you!

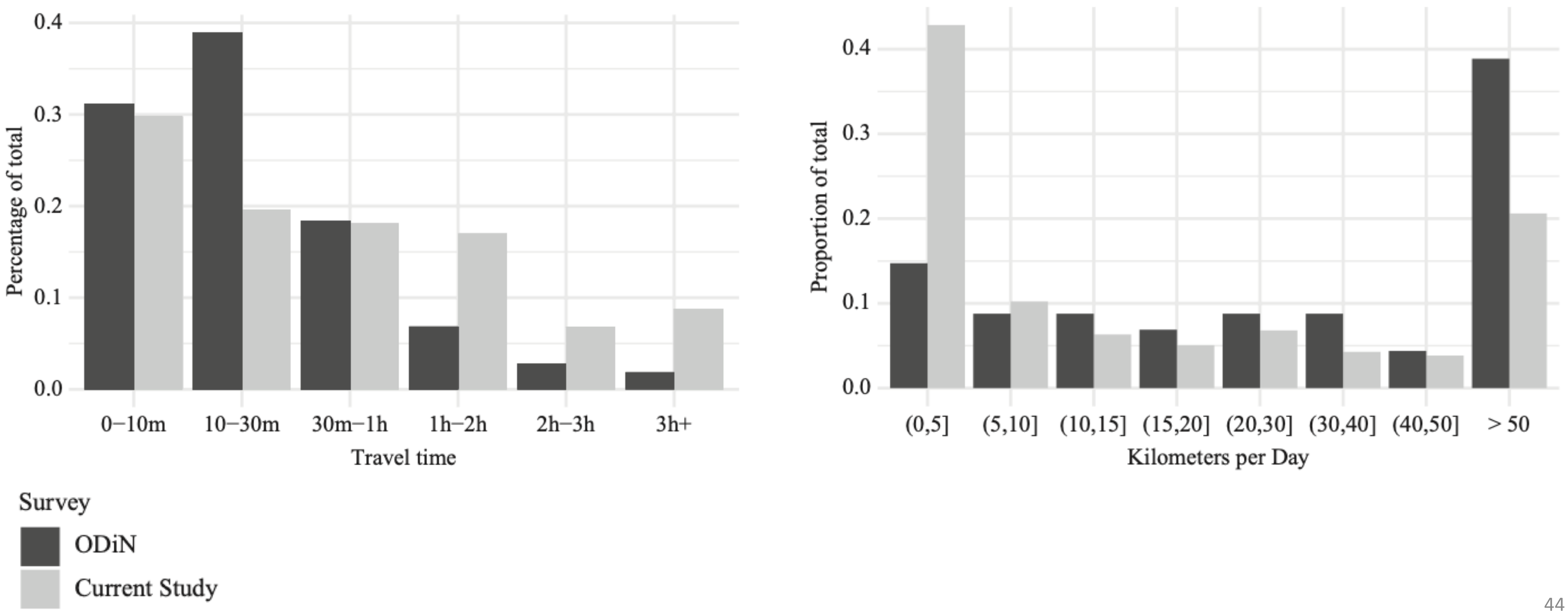
Contact:

b.struminskaya@uu.nl

https://bellastrum.com

@bellastrum

App data compared to web diary self-report (ODiN)



McCool et al. (2021)

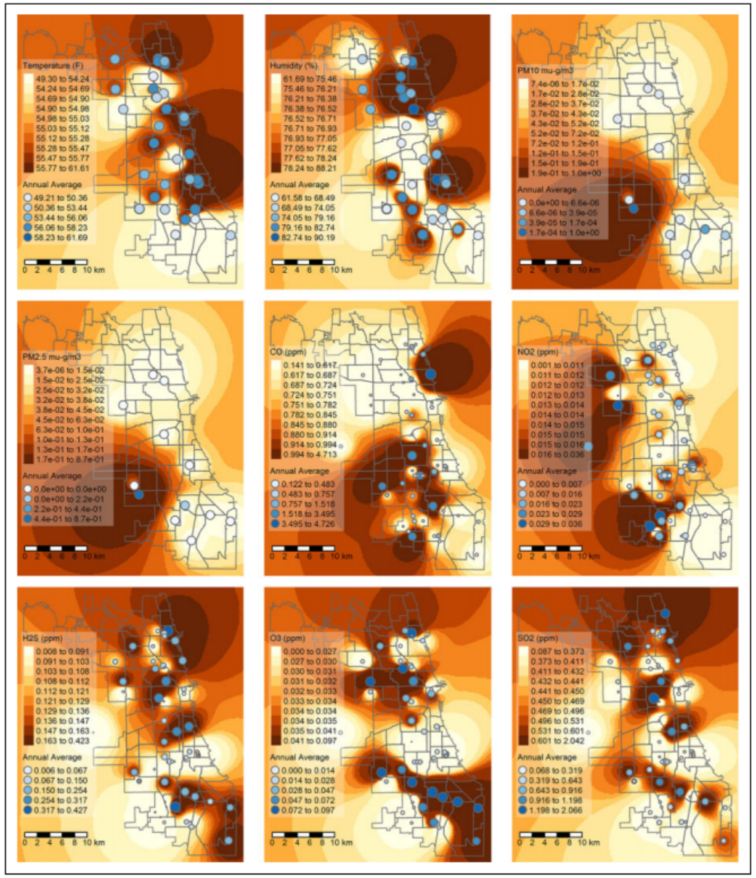
Article

Making Sense of Sensor Data: How Local Environmental Conditions Add Value to Social Science Research

Social Science Computer Review
1-16

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0894439320920601
journals.sagepub.com/home/ssc
SAGE

Ned English[1], Chang Zhao[1], Kevin L. Brown[1], Charlie Catlett[2], and Kathleen Cagney[3]



Linked the two data sources and created heat maps to understand the relationship between air pollution and health outcomes.

Exposure to pollutants together with having lived in the neighborhood for a longer period of time is significantly related to respiratory health issues.

DOI: 10.1177/0894439320920601

Predicting Homophily and Social Network Connectivity From Dyadic Behavioral Similarity Trajectory Clusters

Social Science Computer Review
1-17

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0894439320923123
journals.sagepub.com/home/ssc
SAGE

Brandon Sepulvado[1], Michael Lee Wood[2], Ethan Fridmanski[3], Cheng Wang[4], Matthew J. Chandler[5], Omar Lizardo[6], and David Hachen[3]

Table 2. Coefficient Estimates for Logistic Regression Models Predicting Dyadic Sociodemographic Similarity and Connectivity in the Communication Network as a Function of Activity Trajectory Cluster Membership.

(Intercept)	Gender	Race	Religion	Edge
	−0.02 (0.03)	0.12 (0.03)***	0.41 (0.03)***	−3.76 (0.11)***
Cluster				
1	0.10 (0.05)	−0.08 (0.05)	−0.01 (0.05)	−0.25 (0.18)
2	0.11 (0.10)	0.39 (0.10)***	−1.35 (0.11)***	−0.07 (0.35)
3	0.13 (0.05)**	−0.11 (0.05)*	−0.07 (0.05)	−0.09 (0.17)
4	−0.11 (0.09)	0.31 (0.09)***	0.09 (0.09)	0.67 (0.22)**
5	0.07 (0.05)	−0.05 (0.05)	−0.06 (0.05)	−0.29 (0.18)
6	−0.02 (0.07)	0.10 (0.07)	0.21 (0.07)**	−0.96 (0.34)**
7	−0.10 (0.07)	0.05 (0.07)	−0.57 (0.07)***	0.40 (0.20)*
8	0.02 (0.06)	−0.65 (0.06)***	0.05 (0.06)	−0.29 (0.23)
9	0.08 (0.05)	−0.17 (0.05)**	−0.15 (0.05)**	0.13 (0.17)
10	−0.04 (0.07)	−0.40 (0.07)***	0.19 (0.08)*	0.00 (0.25)
11	Ref.	Ref.	Ref.	Ref.

Equiped >600 students with fitbits to study development of social ties over time. Dyads that tend to adopt similar temporal trajectories in terms of physical activity are more likely to be similar in ethnoracial staus and religious affiliation and more likely to end up connected in the communication network. Emergence of social ties seems to be impacted by synchronization in dynamic behavioral space among individuals.

Being in the same behavioral trajectory cluster is less useful for predicting whether a given dyad is composed of individuals with the same or different gender: similarity in physical activity trajectories is less tied to this marker than it is to race and religion

DOI: 10.1177/0894439320923123