

Combining Active and Passive Mobile Data Collection: A Survey of Concerns

Florian Keusch¹, Frauke Kreuter^{1,2,3}, Bella Struminskaya⁴, & Martin Weichbold⁵

¹University of Mannheim, ²University of Maryland, ³Institute for Employment Research, ⁴Utrecht University, ⁵University of Salzburg

 f.keusch@uni-mannheim.de

 @floriankeusch



OCTOBER 25-27, 2018 ■ BARCELONA, SPAIN

Acknowledgement

- This work was supported by the German Research Foundation (DFG) through the Collaborative Research Center SFB 884 “Political Economy of Reforms” (Project A9) and the Austrian Federal Ministry of Science, Research and Economy through the project “Platform for survey research, methods and empirical analyses” (PUMA).

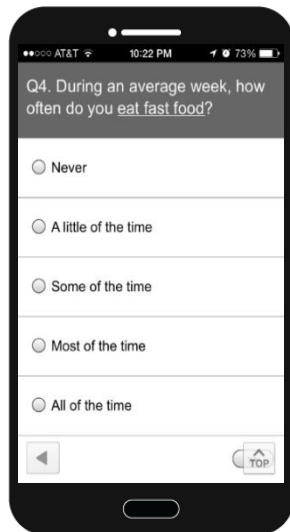
New Data Collection Opportunities

- Popularity of smartphones creates new opportunities for researchers to answer social science research questions
- Examples
 - Emotional well-being in different geographic locations (MacKerron and Mourato 2013; York Cornwell and Cagney 2017)
 - Social networks and ties (Eagle and Pentland 2006; Kassarnig et al. 2017, Stopczynski et al. 2014)
 - Job search behavior (Kreuter et al. forthcoming; Sugie 2018)
 - Consumer expenditure (Jäckle et al. 2017)
 - Mobility and travel behavior (Geurs et al. 2015, Goodspeed et al. 2018, Green et al. 2016, Scherpenzeel 2017)

Role of Participant (Wenz et al. 2017)

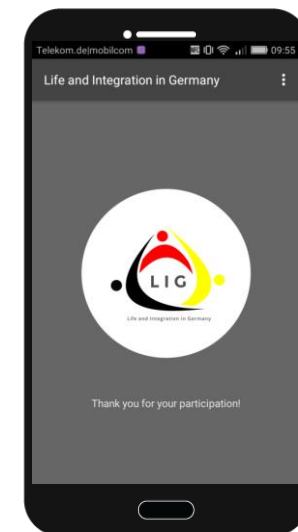
Active

- e.g., mobile survey, taking pictures
- Control over what data are shared



Passive

- e.g., location tracking, call logs, app use, activity data
- Little or no control during data collection other than turning off measurement



Passive Smartphone Data as Big Data

- Smartphone sensor data have many characteristics of “Big Data”
 - Large volume, high velocity/frequency of measurement, variety of data formats
- Compared to surveys, passive smartphone measurement has potential to..
 - ...provide richer data
 - ...decrease respondent burden
 - ...reduce measurement error
- At same time, providing these data poses risks to participants
 - Data streams could be intercepted by unauthorized party
 - Connecting multiple streams of data could re-identify previously anonymous users
 - Information could be used to impact credit, employment, or insurability

Previous Research

- Lower hypothetical willingness for passive tracking than actively completing tasks (Revilla et al. 2016, 2018; Wenz et al. 2017)
 - Control over when data are shared with researchers increases willingness (Keusch et al., under review)
- Smartphone users differ in their behavior and engagement with device
 - Smartphone skills and use of smartphones for different activities correlate with reported willingness (Couper et al. 2017; Keusch et al. 2017; Struminskaya et al. 2018; Wenz et al. 2017)

Concern with Smartphone Data Collection

- Does concern about participating in smartphone data collection differ by type of data collected on a smartphone?
 - Lower concern assumed for tasks allowing users to curate data before sending them to researcher
- Does concern vary across subgroups of users with different levels of smartphone skills and smartphone use habits?
 - More prolific smartphone users assumed to have lower concern

Samples

Sample 1

- 2,647 smartphone users
- Germany
- Non-probability online panel 1
- December 2016

Sample 2

- 1,398 smartphone users
- Germany
- Non-probability online panel 2
- December 2017

Sample 3

- 2,186 smartphone users
- Germany
- Probability online panel (GIP)
- November 2017 (Wave 32)

Sample 4

- 632 smartphone users
- Austria
- Probability web survey (PUMA)
- Spring 2018 (Wave 5)

Outcome Measure

Smartphones can collect a variety of data that provide researchers with information of the everyday life of the users.

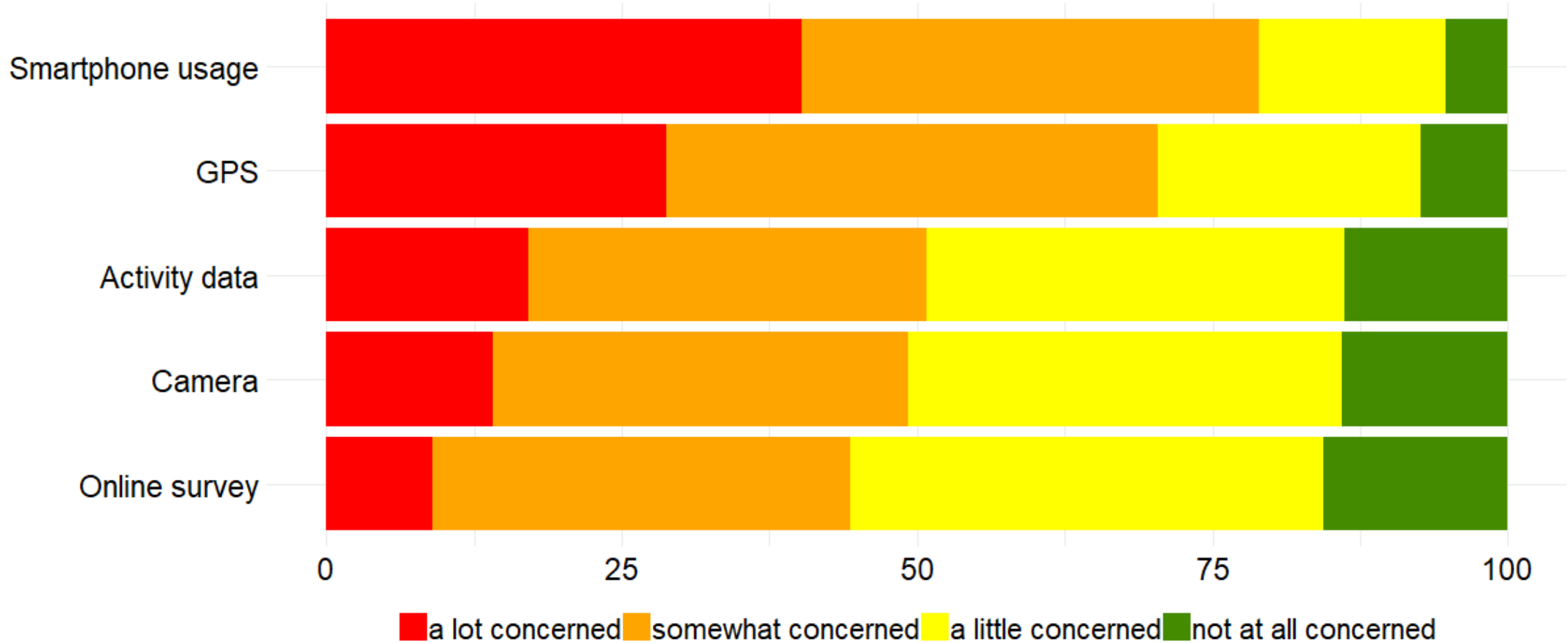
Below you will see a number of activities that you could do with your smartphone. How concerned would you be about the security of providing information in the following ways for research?

(Scale: not at all concerned, a little concerned, somewhat concerned, very concerned)

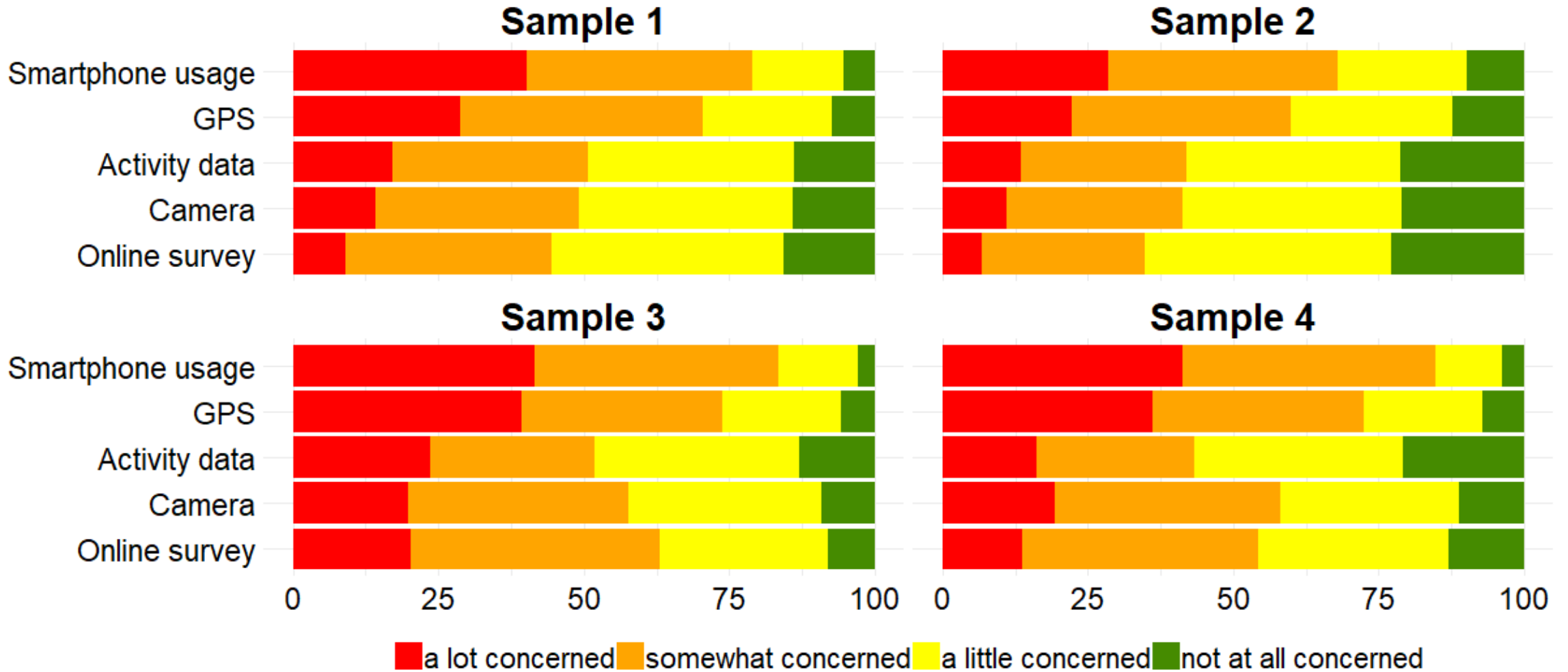
- Complete an online questionnaire on your smartphone*
- Download an app which collects data about how you use your smartphone*
- Use the camera of your smartphone to take photos or scan barcodes (for example photos of receipts, barcodes of purchased products)*
- Allow built-in features of your smartphone to measure the frequency and speed at which you walk, run or cycle*
- Share the GPS position of your smartphone (for example to measure time spent in urban vs. green spaces)*

Does Concern Differ by Type of Data Collected?

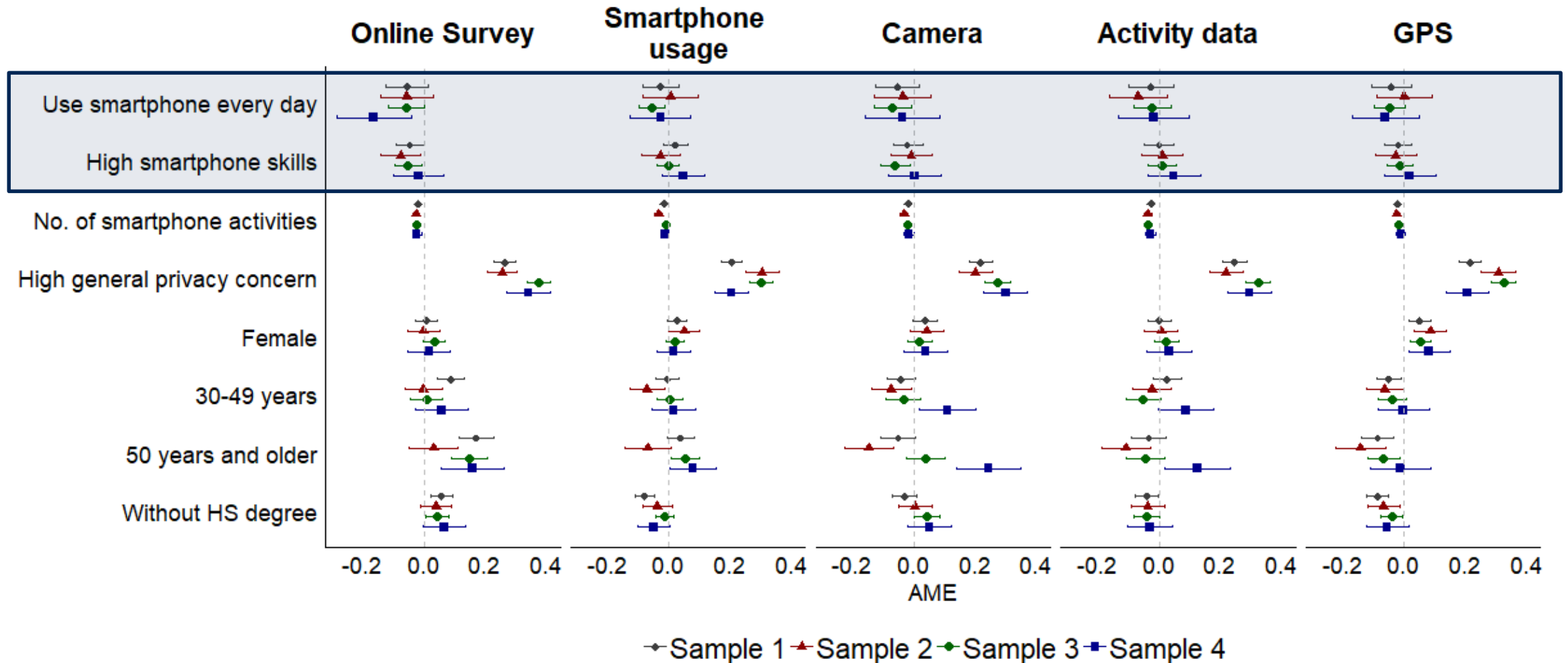
Sample 1



Does Concern Differ by Type of Data Collected?

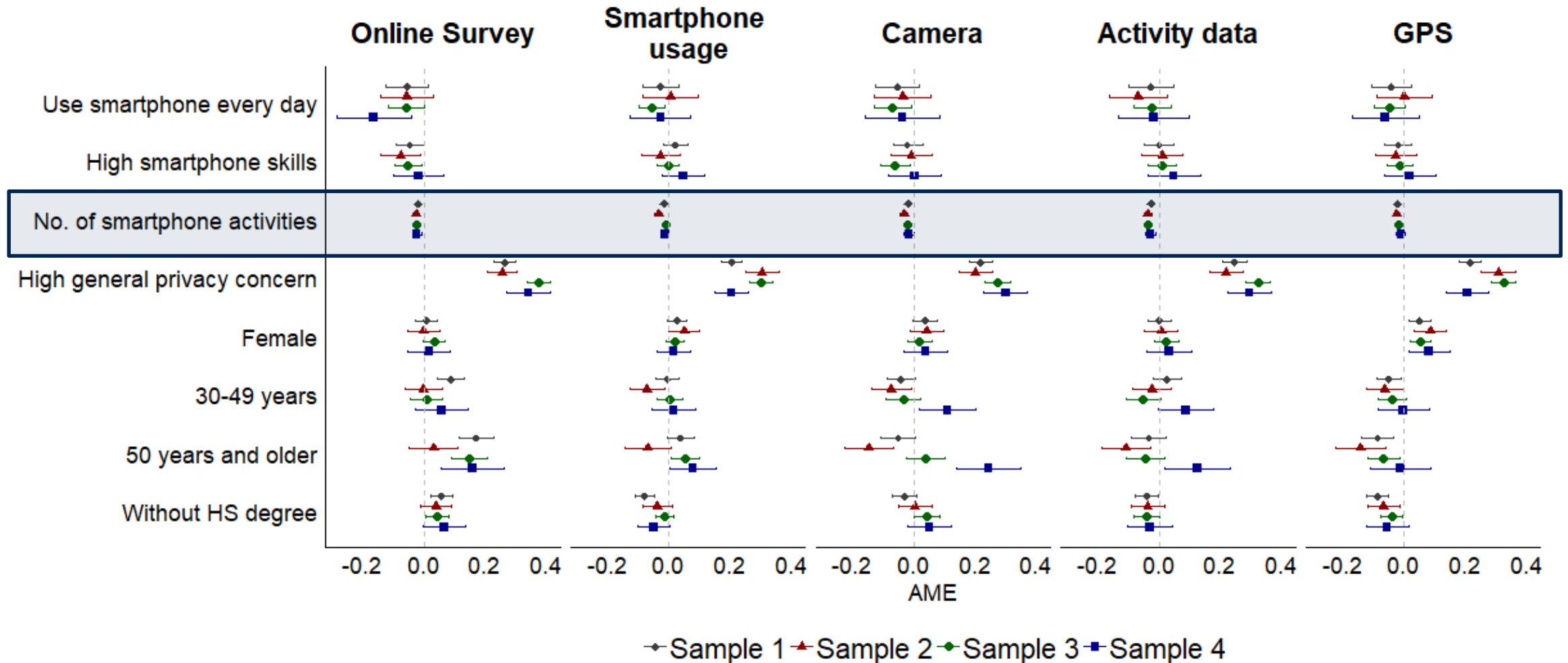


Does Concern* Vary Across Subgroups?



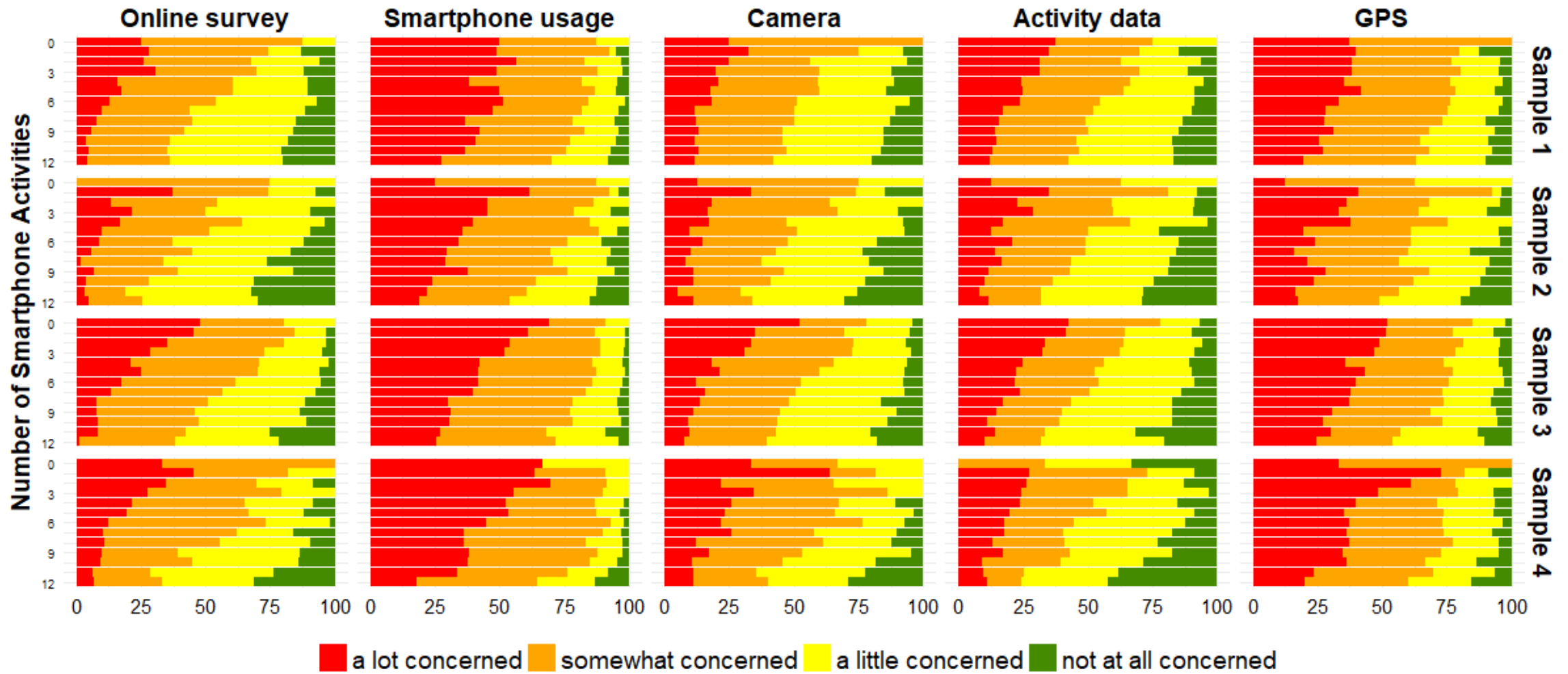
*Outcome variable dichotomized: ('a lot concerned' and 'somewhat concerned') vs. (combining 'a little concerned' and 'not at all concerned')

Does Concern* Vary Across Subgroups?

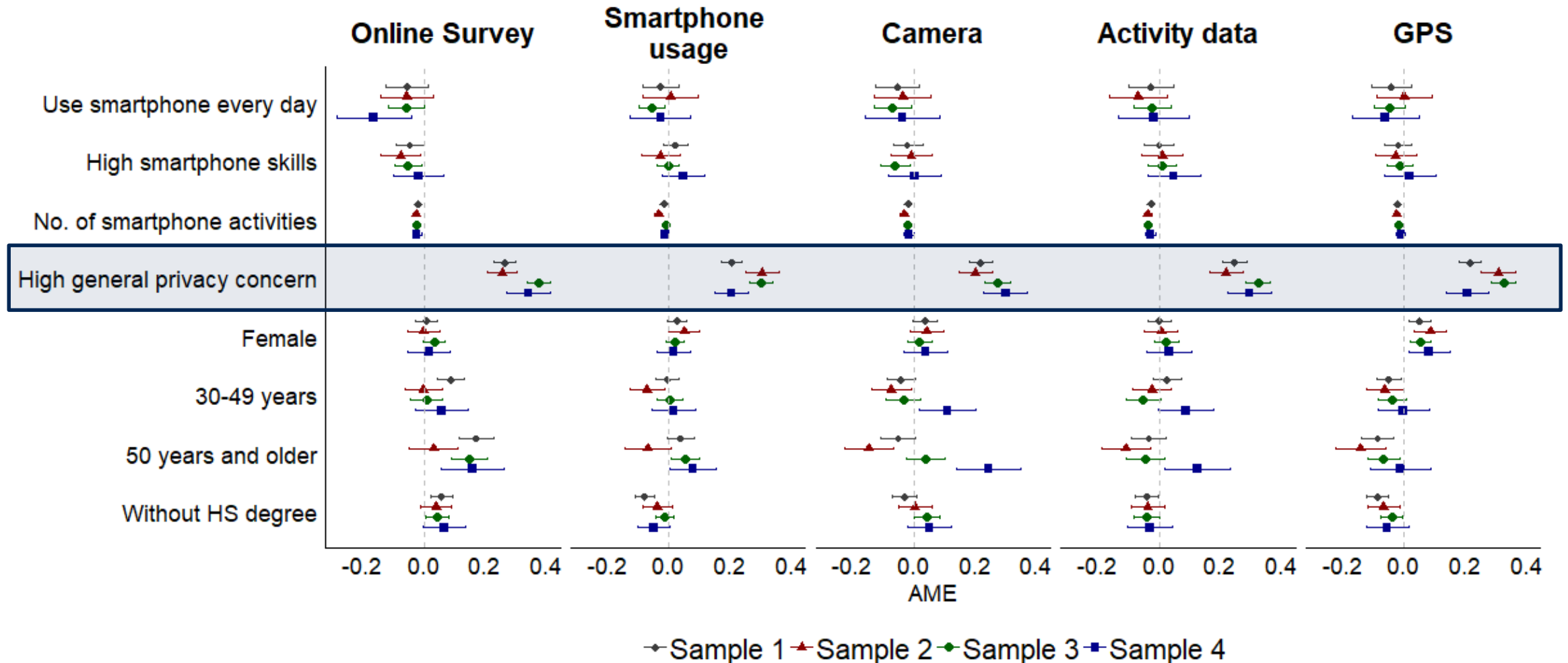


*Outcome variable dichotomized: ('a lot concerned' and 'somewhat concerned') vs. (combining 'a little concerned' and 'not at all concerned')

Concern Differs by No. of Smartphone Activities

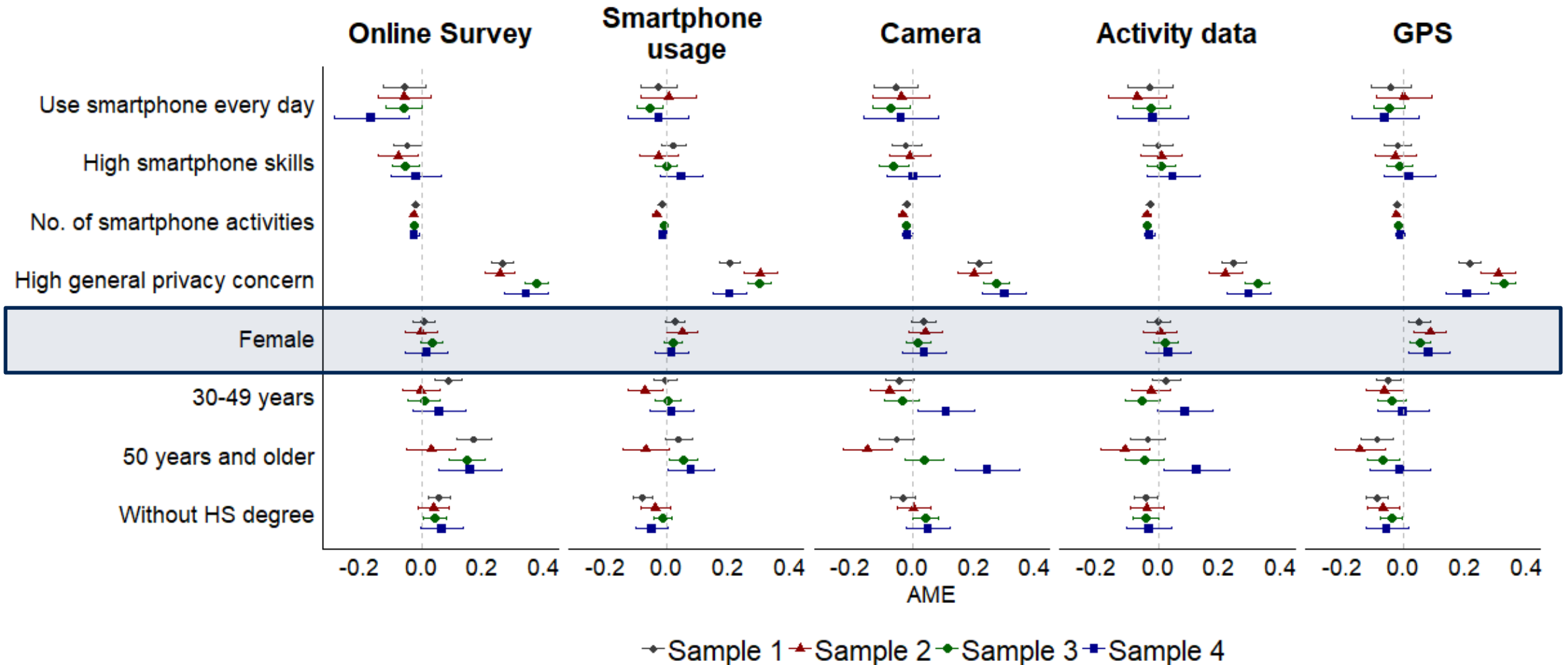


Does Concern Vary Across Subgroups?



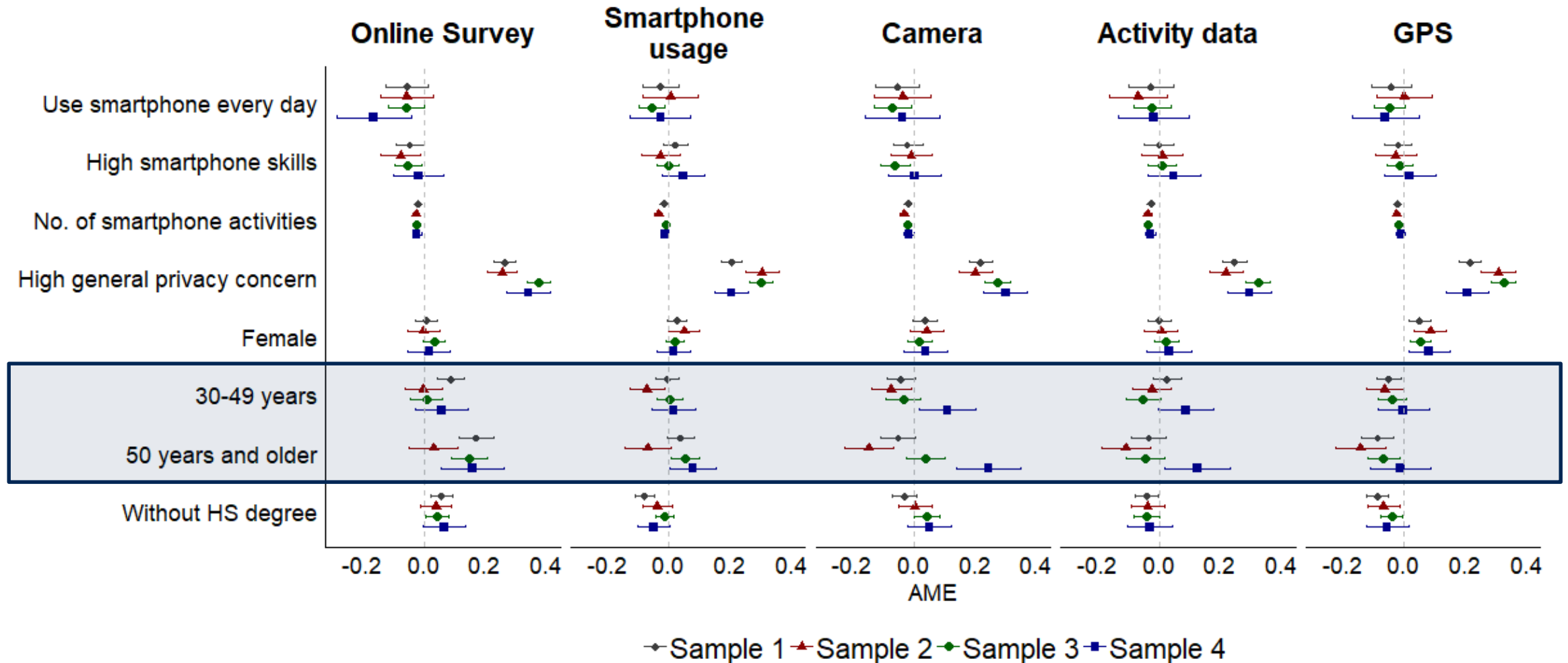
Average marginal effects (AME) with 95%-CIs from logistic regression models predicting dichotomized concern ('a lot concerned' and 'somewhat concerned') vs. (combining 'a little concerned' and 'not at all concerned')

Does Concern Vary Across Subgroups?



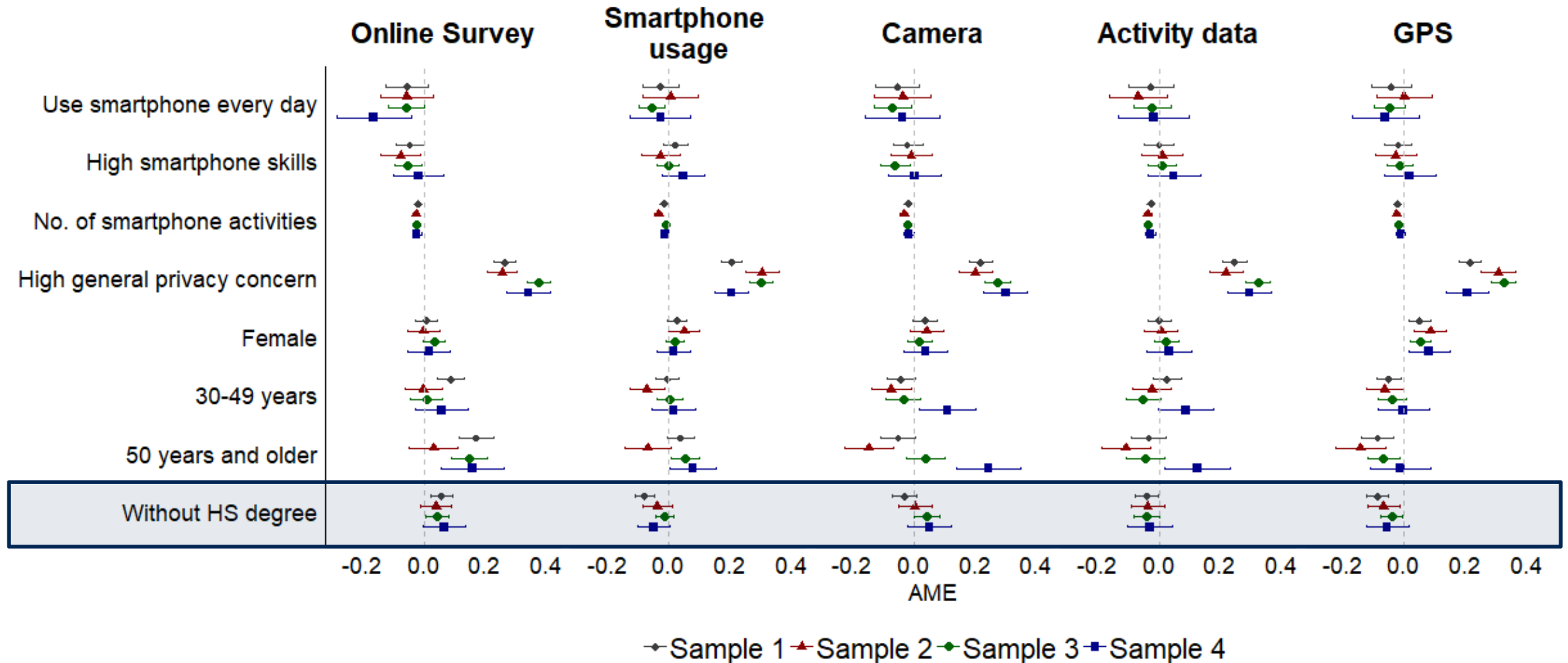
Average marginal effects (AME) with 95%-CIs from logistic regression models predicting dichotomized concern ('a lot concerned' and 'somewhat concerned') vs. (combining 'a little concerned' and 'not at all concerned')

Does Concern Vary Across Subgroups?



Average marginal effects (AME) with 95%-CIs from logistic regression models predicting dichotomized concern ('a lot concerned' and 'somewhat concerned') vs. (combining 'a little concerned' and 'not at all concerned')

Does Concern Vary Across Subgroups?



Average marginal effects (AME) with 95%-CIs from logistic regression models predicting dichotomized concern ('a lot concerned' and 'somewhat concerned') vs. (combining 'a little concerned' and 'not at all concerned')

Conclusion

- Level of concern with smartphone data collection differs by type of data collected

Data	Concern	Role of participant	Control	Sensitivity of data
Smartphone usage GPS	High	Passive	Low	High
Activity data	Medium	Passive	Low	Low
Online survey Camera	Low	Active	High	High/Low

Conclusion (2)

- Number of different activities for which people use smartphones correlates negatively with concern about both passive and active mobile data collection
- Very similar patterns of concern across all four samples
 - Concern in general lower in non-probability than probability online panels
 - Ranking of types of data by concerns highly comparable across samples

Implications

- Easy access to data and control when and what data passively collected could reduce concern and increase willingness
- Since concern highly correlates with number of different smartphone activities, bias in passive measurement of these activities expected due to differential nonparticipation
- Concern might have different reasons (e.g., data too sensitive, fear of data breach, lack of technical knowledge) that need to be identified and addressed when inviting smartphone users
 - Future research should investigate how concern can be reduced in communication with users in experimental setting, ideally with behavioral outcome measure

Thank You!



Florian Keusch

University of Mannheim

School of Social Sciences

Statistics and Methodology

 f.keusch@uni-mannheim.de

 <http://floriankeusch.weebly.com/>

 [@floriankeusch](https://twitter.com/floriankeusch)