



**General Online Research Conference  
GOR 19  
6 March to 8 March 2019, TH Köln – University of  
Applied Sciences, Cologne, Germany**

Johannes Breuer, GESIS – Leibniz Institute for the Social Sciences  
Sebastian Stier, GESIS – Leibniz Institute for the Social Sciences  
Pascal Siegers, GESIS – Leibniz Institute for the Social Sciences  
Tobias Gummer, GESIS – Leibniz Institute for the Social Sciences  
Arnim Bleier, GESIS – Leibniz Institute for the Social Sciences

**Linking survey data with social media data and the importance of  
informed consent**

Contact: [johannes.breuer@gesis.org](mailto:johannes.breuer@gesis.org)

Suggested citation: Breuer, Johannes, Stier, Sebastian, Siegers, Pascal, Gummer, Tobias, Bleier, Arnim. 2019. "Linking survey data with social media data and the importance of informed consent". General Online Research (GOR) Conference, Cologne.



# Background & motivation

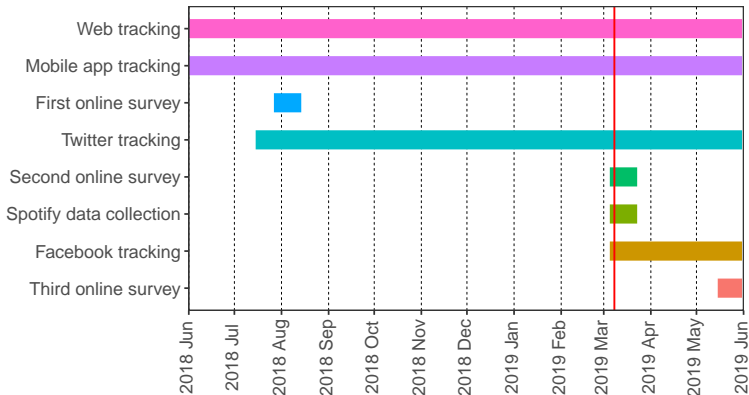
- surveys and social media data both have limitations
- combining them holds great promise for research on online behavior
- linking at the individual level requires informed consent
- with GDPR in Europe, researchers have to provide detailed information about
  - what data they collect
  - how they collect it
  - and for what purpose

## Research interest

- **Research question:** *Which factors influence consent to the tracking and linking of social media data?*
- replication & extension of Al Baghal, T., Sloan, L., Jessop, C., Williams, M. L., & Burnap, P. (2019). Linking Twitter and survey data: The impact of survey mode and demographics on consent rates across three UK studies. *Social Science Computer Review*. Advance online publication. <https://doi.org/10.1177/0894439319828011>

# Project timeline

GESIS Data Linking Project



# Methods

- *respondi* web tracking panel ( $N \sim 2000$  for each month)
- online survey among panelists ( $N = 1347$ )
  - 48.11% female
  - age: 16 to 70 ( $M = 44.66$ ,  $SD = 13.2$ )
- survey included questions on
  - media use
  - (political attitudes)
  - sociodemographics
  - consent to tracking & linking of Twitter data

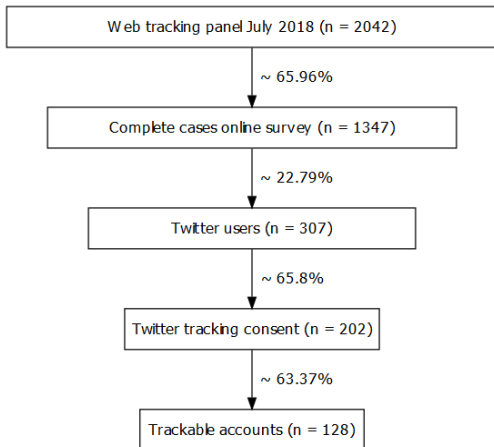
## Twitter consent

- short informed consent in questionnaire
- extended data privacy information on GESIS website (link in questionnaire)
  - translated & adapted with permission from Al Baghal et al. (2019)
  - only max.  $n = 6$  read the extended privacy information (respondents who visited gesis.org during the survey field phase, avg. duration per visit = 57.09 sec) of which  $n = 4$  consented to Twitter tracking
- (additional) incentive of 5 Euro for Twitter tracking
- 2 incentive conditions
  - paid regardless of consent
  - paid only if consented

## Twitter use

- 22.79% of respondents reported having a Twitter account
- 14.48% of respondents reported that they had used Twitter in the last 7 days
  - for comparison: ARD-ZDF Online Study 2018: 4% use Twitter weekly
- **men (29.27%) were more likely to have a Twitter account** than women (15.59%),  $\chi^2(1, 1344) = 35.68, p < 0.001, r = 0.16$
- **Twitter users were significantly younger** ( $M = 42.04, SD = 13.35$ ) than nonusers ( $M = 45.44, SD = 13.07$ ),  $t(493.27) = 3.94, p < 0.001, \text{Cohen's } d = 0.26$

# Participation

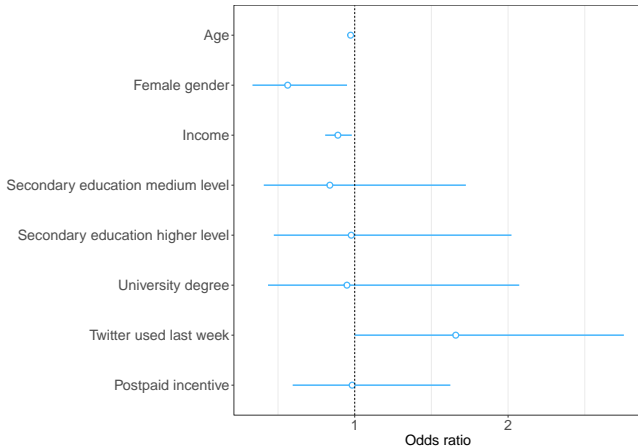




# Twitter tracking consent - Bivariate

- among the  $n = 307$  Twitter users:
  - **consenters were significantly younger** ( $M = 40.52$ ,  $SD = 13.2$ ) than nonconsenters ( $M = 44.86$ ,  $SD = 13.27$ ),  $t(204.45) = 2.71$ ,  $p < 0.01$ , *Cohen's d* = 0.33
  - **men (69.61%) were more likely to consent** than women (58.42%),  $\chi^2(1, 303) = 4.26$ ,  $p = 0.039$ ,  $r = 0.12$
  - **incentive condition (pre- vs. postpaid) did not have an effect** on consent rates,  $\chi^2(1, 305) = 0.03$ ,  $p = 0.87$ ,  $r = 0.01$

# Twitter tracking consent - Multivariate



## Reasons for opting out

Reason	Percentage
Tweets are private	45.63
Privacy concerns	38.83
Usefulness unclear	24.27
Risk of non-scientific use	16.50
Other	12.62

*Note:*

Multiple selection was possible.

- all of the  $n = 13$  respondents who provided another reason in the open-ended item indicated that they use Twitter only rarely and/or passively

## Conclusion & Outlook

- results for NatCen Panel from Al Baghal et al. (2019) replicated for gender & age
- detailed privacy information ignored by most participants
  - still: detailed informed consent important to have
- consent rate in our study possibly = upper bound due to nature of the sample & study design
- next steps:
  - similar analyses for Facebook & Spotify data
  - investigate additional predictors of consent
  - further explore biases in the sample (e.g., with regard to privacy attitudes)