

**NatCen**

**Social Research** that works for society

# Linking Twitter and Survey Data

**Improving measurement of both data sources**

Curtis Jessop (NatCen), Tarek Al Baghal (University of Essex) & Luke Sloan (Cardiff University)

---

14 September 2022

# Acknowledgments

---

ESRC grant: “Understanding [Offline/Online] Society: Linking Surveys with Twitter Data”

- Luke Sloan – Cardiff University
- Tarek Al Baghal – University of Essex
- Matthew Williams – Cardiff University

# Contents

---

- Background: Why link social media & survey data?
- Existing work: Establishing feasibility
- Current work: Developing the approach
- Future plans: Scaling up

# Why link social media & survey data?



# What are we trying to do, and why?

---

- Link survey participants' answers to publically available information from their Twitter accounts
- Allows survey data to benefit from real-time, 'natural' behavioural and attitudinal data – including 'between waves'
- Adds the 'who' to Twitter data – creates a sample frame, and allows for the analysis of different groups

# Establishing feasibility

---

2

# Ethics

---

- Social media data have disrupted many elements of social research, including how we conduct it ethically
  - Distance between researcher and participant
  - Large scale, publicly available, searchable data
- Re-evaluating how we ensure:
  - Voluntary participation
  - Minimising harm
  - Maximising value

# Collecting informed consent

---

- As we are in contact with participants, have the opportunity to ask people for consent to access their Twitter data (and link it to their survey answers)
  
- Initial feasibility tests in 2017:
  - NatCen Panel: 27% consent rate
  - Understanding Society IP10: 31% consent rate

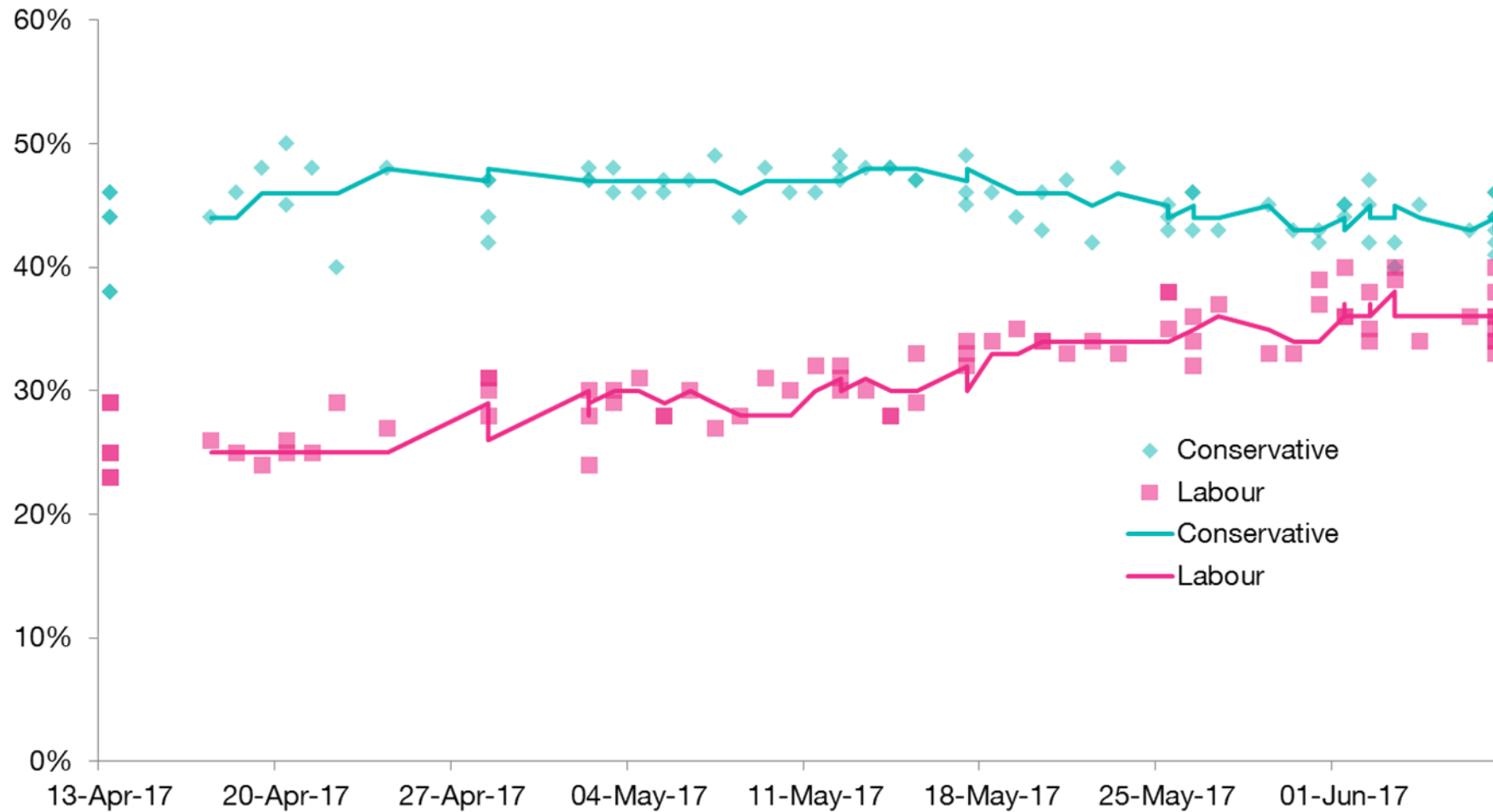


# Maintaining security

---

- Linking inherently identifiable Twitter data risks de-anonymising the survey data as well
- But anonymising raw Twitter data would undermine its value
  
- But there are other approaches to help minimise risks:
  - Systematic processing
  - Data reduction
  - Controlled access
  - Data deletion

# Case study: Understanding support for Labour



# Social media provided more 'texture'

## Parties or politicians?

"@jeremycorbyn doesn't avoid voters or hire a crowd, he reads out questions from real people #PMQs"

## Policy focus?

"Labour will build affordable homes and crack down on rogue private landlords"

## Pro-Labour...?

"Labour: minimum wage, school cuts reversed, universal child care. Tories: kill the foxes!"

## ...or anti-Conservative?

"Trying to humanise May just shows us she's personally awful as well as politically vile"

# Other studies

---

- Twitter use during the Covid-19 pandemic (Wenz *et al*, 2021)
- Linking Twitter and survey data: asymmetry in quantity and its impact (Al Baghal *et al*, 2021)

# Developing the approach



3

# Improving informed consent rates

- Low consent rates introduce bias and reduce statistical power
- Ensuring *informed* consent
- Qualitative study to understand public attitudes
  - No choices *fully* informed
  - Heuristic decision-making *but* don't change minds following discussions
  - Consider: risks, benefits, trust & control
- Experiment with new questionnaire design included in USoc IP 15

# Data sharing

---

- Successful secure sharing of data between team members/ organisations
- Next step is wider sharing with research community
  - Reproduction & expansion
- Expecting to publish Understanding Society IP data soon
  - Open access (summary variables) & secure access (detailed information)

# Methods work

---

- Enhancing Twitter data using survey data
  - Validation & enhancement of categorisation tools
- Enhancing survey data using social media data
  - Validation of survey measures
  - Additional data
  - Non-response adjustments



# Demonstrator study

---

- Public attitudes to ethnic minority groups in the UK
- Survey of 2000 adults + 4,000 Twitter users
  - Responses to events in past & future
  - Analysis of Twitter networks
  - Self-reported attitudes vs observed behaviours

# Plans for the future



# 4

# Development & expansion

- Engaging with methodological challenges
- Larger sample sizes
- Linking to other types of data
  - LinkedIn + ...?

# References

---

Al Baghal, T. *et al.* (2020) 'Linking Twitter and Survey Data: The Impact of Survey Mode and Demographics on Consent Rates Across Three UK Studies', *Social Science Computer Review*, 38(5), pp. 517–532.

Al Baghal, T. *et al.* (2021) 'Linking Twitter and survey data: asymmetry in quantity and its impact', *EPJ Data Sci.* 10, 32.

Jessop, C. (2017) 'Understanding political behaviour by linking survey & social media data', [https://natcen.ac.uk/media/1499364/Curtis\\_Jessop.pdf](https://natcen.ac.uk/media/1499364/Curtis_Jessop.pdf)

Jessop, C. (2021) 'Public attitudes to linking survey and Twitter data', <https://www.youtube.com/watch?v=t2do3l06Cfs>

Sloan, L. *et al.* (2020) 'Linking Survey and Twitter Data: Informed Consent, Disclosure, Security, and Archiving', *Journal of Empirical Research on Human Research Ethics*, 15(1–2), pp. 63–76.

Wenz, A. *et al.* (2021) 'Twitter use during the Covid-19 pandemic', <https://www.youtube.com/watch?v=eSUZjBTUi5c>

# Thanks!

---

**Curtis Jessop, Research Director**

**Email:** [curtis.jessop@natcen.ac.uk](mailto:curtis.jessop@natcen.ac.uk)

**Telephone:** 020 7549 7027

**Twitter:** @CurtisJessop

**NatCen**

**Social Research** that works for society