

International Journal of Population Data Science

Journal Website: www.ijpds.org



Using social media metrics and linked survey data to understand survey behaviors

Tarek Al Baghal¹, Paulo Serodio¹, Shujun Liu², Luke Sloan², and Curtis Jessop³

¹University of Essex

²Cardiff University

³NatCen Social Research

Introduction & Background

Linking social media and survey data at the individual level has the potential to add evidence to a variety of research questions. To make this data openly available to others, social media data need to be converted into useful metrics that minimise issues of disclosure while maximising utility. This research explores linkages of Twitter data and survey data in the Understanding Society Innovation panel, focusing on the usage of non-disclosive metrics created from Twitter data alongside the similarly anonymised survey data.

Objectives & Approach

The Innovation Panel asked for consent to link Twitter data to survey responses and data has been collected from the Twitter API. However, Twitter's unstructured nature necessitates creating measures that can be used jointly with linked survey data. We have developed a framework to create social media metrics that can be combined with survey data that also remove any disclosive data, so these data can be widely shared for maximum utility. The current research analyses these data to understand what the metrics look like through presentation of descriptive statistics. We also begin to show these data may be used in combination with survey data through inclusion of a set of metrics in logistic regression models predicting attrition and measurement of mental health.

Relevance to Digital Footprints

Social media is a prevalent aspect of social life and leaves a substantial digital footprint. However, there are a number of limitations to these data, including a lack of understanding of who is producing the data, and having the ability to relate these to a variety of specific (and possibly higher quality) measures for a representative sample of the population. Linkage to surveys address these problems and can lead to new research opportunities using digital footprint data.

Results

While small sample sizes impact the power of some analyses, the methods developed are illustrative of ways to use this novel data source. Results show that there is high variation in the created metrics, and initial analysis shows that the inclusion of a set of user-level Twitter data is not significantly related to attrition. However, more accounts followed on Twitter and the number of user retweets are significantly related to higher levels of mental distress on the GHQ scale.

Conclusions & Implications

Overall, there is some evidence that social media helps to understand survey outcomes, perhaps more so on measurement outcomes. This study provides an initial start on how to use these curated linked social media and survey data, and we note there are other social media networks that we can apply this strategy to; for example, LinkedIn, particularly with changes made to Twitter (X).

