International Journal of Social Research Methodology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tsrm20

The demise of the survey? A research note on trends in the use of survey data in the social sciences, 1939 to 2015

Patrick Sturgis & Rebekah Luff

To cite this article: Patrick Sturgis & Rebekah Luff (2020): The demise of the survey? A research note on trends in the use of survey data in the social sciences, 1939 to 2015, International Journal of Social Research Methodology, DOI: 10.1080/13645579.2020.1844896

To link to this article: https://doi.org/10.1080/13645579.2020.1844896

Published online: 30 Nov 2020.

INTERNATIONAL JOURNAL OF SOCIAL RESEARCH METHODOLOGY
https://doi.org/10.1080/13645579.2020.1844896

Routledge
Taylor & Francis Group

The demise of the survey? A research note on trends in the use of survey data in the social sciences, 1939 to 2015

Patrick Sturgis[a] and Rebekah Luff[b]

[a]Department of Methodology, London School of Economics and Political Science, London, UK; [b]Social Care Institute for Excellence, London, UK

ABSTRACT

We assess the case for a decline in the use of survey data in the social sciences during a period in which conventional survey research has faced existential challenges to its ongoing feasibility and growing competition from new forms of 'Big Data'. Presser (1983) and Converse (1987) undertook content analysis of articles published in a set of leading social science journals, finding a trend of increasing the use of survey data between 1939 and 1980. In an extension of Presser's analysis to the mid-1990s, Saris and Gallhofer (2007) found a further small increase in the rate of survey use, though with notable variability across disciplines. We update these studies to include the period 2014 to 2015. While our analysis reveals the emergence of a small proportion of articles using Big Data, we find no evidence of a concomitant decline in the use of survey data. On the contrary, the use of surveys increased, being used in nearly half of all published articles in this set of journals in 2014/15 and, where articles reported using Big Data, many of them also used survey data. Additionally, we find a substantial increase in the use of secondary survey data over the reference period.

ARTICLE HISTORY

Received 18 August 2020
Accepted 28 October 2020

Introduction

There can be little doubt that the sample survey constituted the pre-eminent social science research method of the twentieth century (Presser, 1984; Savage, 2010). Emerging in the 1930s out of the confluence of social reform movements, innovations in random sampling, and the nascent market research industry, the survey provided a powerful new tool for describing and understanding individual behavior and population dynamics (Ayrton, 2017; Converse, 1987). Professionalization and expansion flowed from technical and methodological developments that came in response to the information demands of the second world war. This period also produced a cadre of talented survey practitioners who emerged from military research centres to establish pioneering survey institutes across the United States (Converse, 1987). By the latter decades of the twentieth century, these developments in the US had spread internationally and the survey attained near-hegemonic status as the methodological vehicle of choice in quantitative social science (Groves, 2011; Savage, 2010).

Despite, or perhaps because of its dominant status, the survey has faced sustained criticism relating to, *inter alia*, its positivist epistemological orientation (Blumer, 1955), the separation of respondents from their social contexts (Cicourel, 1964), and the over-reliance on error-prone subjective self-reports (Nisbett & Wilson, 1977). However, by the 1990s and 2000s more existential

CONTACT Patrick Sturgis p.sturgis@lse.ac.uk Department of Methodology, London School of Economics and Political Science, London WC2A2AE, UK

challenges were evident, threatening as some saw it, the long-term viability of the method itself (Groves, 2011; Savage & Burrows, 2009). In particular, sharply rising costs and declining response rates meant that surveys were perceived to be offering population inference of uncertain accuracy at a snail's pace and at eye-watering prices (Miller, 2017).

As survey commissioners and consumers questioned the cost-effectiveness of the conventional survey, excitement was growing about the potential research applications of the new forms of data that were emerging from the growth of the Internet and the use of online digital devices (Cukier & Mayer-Schoenberger, 2013). Most notable in this regard was the increasing availability of passively 'given off' transactional and sensor data, digital traces from social media users, and vast online archives of textual and visual data, the so-called 'Big Data revolution' (Kolb & Kolb, 2013). When this kind of 'organic' data covering entire populations was becoming ever more readily available, often in real time and for free, what need was there for out-moded and expensive sample surveys with single-digit response rates? For some, these new forms of data portended the demise of the ailing survey method (Savage & Burrows, 2007, 2009).

Others, though, were more sanguine about the implications of these developments for the future of survey research, seeing the Big Data revolution as an opportunity rather than a threat. Pointing to many of the inherent limitations of Big Data for addressing social scientific research questions, Couper (2013) for example, argued that these new forms of data were likely to complement rather than replace surveys. Or, as Groves put it, 'the biggest payoff will lie in new combinations of designed data and organic data, not in one type alone' (Groves, 2011, p. 896).

So, are we witnessing the beginning of the end of the survey, or a period of revitalised growth? In this research note, we assess the evidence for a decline in the use of survey data by extending previous analyses of its prevalence in the social sciences over time. Presser (1984) analysed the frequency of use of survey data in articles published in a set of top-ranking social science journals between 1949 and 1980. This time-series was subsequently extended backward to 1939 by Converse (1987) and forward to 1995 by Saris and Gallhofer (2007). The headline trend from these analyses showed surveys accounting for an increasing proportion of published articles, growing from 14% in 1939 to 36% in 1995, though following an initially rapid increase, growth had largely stabilised by 1965. This overall average, however, masked considerable variability between disciplines, with particularly marked growth in the use of surveys in sociology and social psychology over the period. In addition to updating the time series of survey data use in these journals to 2015, we also assess the extent to which analysis of Big Data is evident and examine whether new forms of digital data are used in isolation or, as Groves (2011) and Couper (2013) suggest, in combination with surveys.

Method

We content analysed all 1451 articles published in the same set of journals considered by Presser (1984), Converse (1987) and Saris and Gallhofer (2007) in the years 2014 and 2015. The journals, the number of articles published in each and their Google Scholar H5 Index rank are shown in Table 1. Presser selected these particular journals on the basis that they were, by his assessment, the highest-ranking general journals within each major social science discipline at the time. While some of the journals are now outside the top 10, they remain amongst the most important in each disciplinary area. It is clear that these journals do not provide anything like comprehensive coverage of all academic social science, nor of any of the individual disciplines. The results of these analyses then should be treated as indicative of the content of top-ranking general journals, rather than as representative of all academic research in the social sciences.

A team of seven coders coded the articles, first according to whether or not empirical data were used and, for the subset of empirical articles, whether the data and analysis were quantitative, qualitative, or used mixed methods. Quantitative articles[1] were further coded according to whether

Table 1. Journals used in the analysis, their discipline, ranking and number of articles.

Discipline	Journal	Disciplinary ranking*	Number of Articles
Economics	American Economic Review	1	263
	Journal of Political Economy	7	61
	The Review of Economics and Statistics	10	149
Sociology	American Sociological Review	1	94
	American Journal of Sociology	8	67
	Social Forces	11	124
Political Science	American Journal of Political Science	1	119
	American Political Science Review	2	99
	Journal of Politics	4	156
Social Psychology	Journal of Personality and Social Psychology	1	236
Public Opinion	Public Opinion Quarterly	15	83

*H5 Google Index Ranking.

the article used one of the four broad data types: survey; administrative; census; or Big Data. For articles using survey data, a further code was applied to denote whether the survey was designed and administered by the authors of the paper (primary data) or was an analysis of an existing survey collected by other researchers (secondary data). Following Presser (1984), surveys were defined as any data collection operation that gathers information from a sample of humans by means of a questionnaire, irrespective of the sampling method or mode of data collection. Articles were coded as using Big Data when they reported analysis of large-scale transactional data sets, survey paradata, social media data, or large corpuses of textual data (Couper, 2013; Groves, 2011).

Each coder was allocated approximately 250 articles by stratified, systematic random allocation. If a coder was unsure how to code an article, it was referred to the authors (of this note) for resolution. A random sample of 49 articles was coded by all seven coders in order to assess reliability. This showed the average pairwise agreement to be 87% with a Cohen's Kappa of 0.70, indicating moderate to strong agreement (Krippendorff, 2004).

Results

A total of 88% of articles published in the selected journals in 2014/15 were empirical in nature, with the highest rate of non-empirical articles found in economics (20%) and political science (18%) with much smaller minorities in sociology (3%) and public opinion (2%) and none at all in social psychology. Nearly all empirical articles in economics (98%) and public opinion (97%) used quantitative data, followed by political science (87%), sociology (80%) and social psychology (73%). Table 2 shows the proportion of articles in each discipline that used survey data in 2014/15 alongside the corresponding figures from Presser, Converse, and Saris and Galhofer.[2]

The percentage of articles that reported using survey data in 2014/15 was 43%, an increase of 7 percentage points compared to 15 years previously. This ranged from a high of 84% in public opinion to a low of 25% in economics, with political science (34%), sociology (50%), and social psychology (69%) in between these upper and lower bounds. In all disciplinary areas except public

Table 2. Proportion of published articles using survey data by discipline and year.

Discipline	1939–40	1949–50	1964–65	1979–80	1994–95	2014–15
Sociology	18% (277)	24% (282)	55% (259)	56% (285)	47% (287)	50% (285)
Political Science	3% (113)	3% (114)	19% (160)	35% (203)	27% (303)	34% (374)
Economics	10% (175)	6% (141)	33% (155)	29% (317)	20% (461)	25% (473)
Social Psychology	21% (34)	22% (59)	15% (233)	21% (377)	49% (347)	69% (236)
Public Opinion	28% (51)	43% (86)	56% (61)	91% (53)	90% (46)	84% (83)
TOTAL %	**14% (91/650)**	**19% (129/684)**	**34% (292/868)**	**36% (449/1235)**	**36% (520/1444)**	**43% (619/1451)**

Data for 1939–40 compiled by Converse (1987), 1949–80 by Presser (1983) and 1994–5 by Saris and Gallhofer (2007).

Table 3. Different types of quantitative data by discipline, 2014–2015.

Discipline	Survey	Admin	Census	Big data	n
Sociology	51%	42%	16%	4%	277
Political Sciences	41%	58%	9%	4%	308
Economics	32%	74%	19%	3%	374
Social Psychology	69%	5%	0%	2%	235
Public Opinion	86%	16%	3%	5%	81
TOTAL	**49**%	**47**%	**11**%	**3**%	**1275**

opinion, there was an increase in the rate of survey data use compared to the mid-1990s. We find no evidence here, then, of a decline in the use of survey data across these social science disciplines. Indeed, this evidence points to a *growing* reliance on surveys 20 years after the introduction of the World Wide Web, with surveys now used in nearly half of published articles in these leading journals.

That surveys were highly prevalent in these journals in 2014/15 does not preclude the possibility that social scientists were also starting to use Big Data. Table 3 shows the percentage of articles in each discipline using different kinds of quantitative data (the base for Table 3 is now all articles using quantitative data). Here we see evidence of the entrance of Big Data into mainstream social science, with articles using Big Data published in all disciplinary areas, representing 3% of all articles. The use of administrative data, at 47%, is at approximately the same level as surveys, with political science and economics showing notably high rates of administrative data use, at 74% and 58%, respectively. There is some disagreement, given its longstanding history of use in the social sciences, about whether administrative data should be included under the heading of Big Data (Japec et al., 2015) and we have treated it as a separate category here. Nonetheless, it seems likely that at least part of this body of research constitutes a departure from more conventional kinds of administrative data and, therefore, constitutes further evidence of the penetration of Big Data in the social sciences. Turning to the question of whether new forms of data are used on their own or in combination with surveys, we find that, of the 597 papers that used administrative data, a quarter (24%) also used survey data, while the corresponding figure for articles using Big Data was a third (34%).

Lastly, we consider whether published articles used primary or secondary survey data. Figure 1 shows the proportion using secondary survey data in each year, using Presser's (1984) estimates for the period 1949 to 1980. A clear growth in the use of secondary survey data is evident over this time frame, with a marked increase in the most recent period, amounting to 62% of all survey-based articles using secondary data in 2014–15 compared to 33% in 1949–50.

Discussion

In their provocative 2007 article, Savage and Burrows proposed that the dominance of the survey as the pre-eminent form of data in the social sciences was in imminent danger of being usurped by new forms of digital transactional data. Indeed, they even warned that those relying on surveys to conduct research 'might want to reflect on whether this might leave them exposed to marginalisation or even redundancy.' (pp.892). The analyses we have presented here provide little support for this rather apocalyptic vision. Building on existing content analyses of articles published in a set of leading social science journals (Converse, 1987; Presser, 1984; Saris & Gallhofer, 2007), we find evidence of the emergence in 2014/15 of studies using Big Data, such as social media, transactional, sensor, and textual data. While such articles constituted only a small minority of the total output, it seems reasonable to expect the rate to increase in the future. Be that as it may, we currently find no support for the idea that these new kinds of data pose a threat to the longstanding dominance of surveys. This is because the frequency of articles using survey data has continued to grow, even as publications using Big Data have started to emerge. An additional off-setting factor is that, where

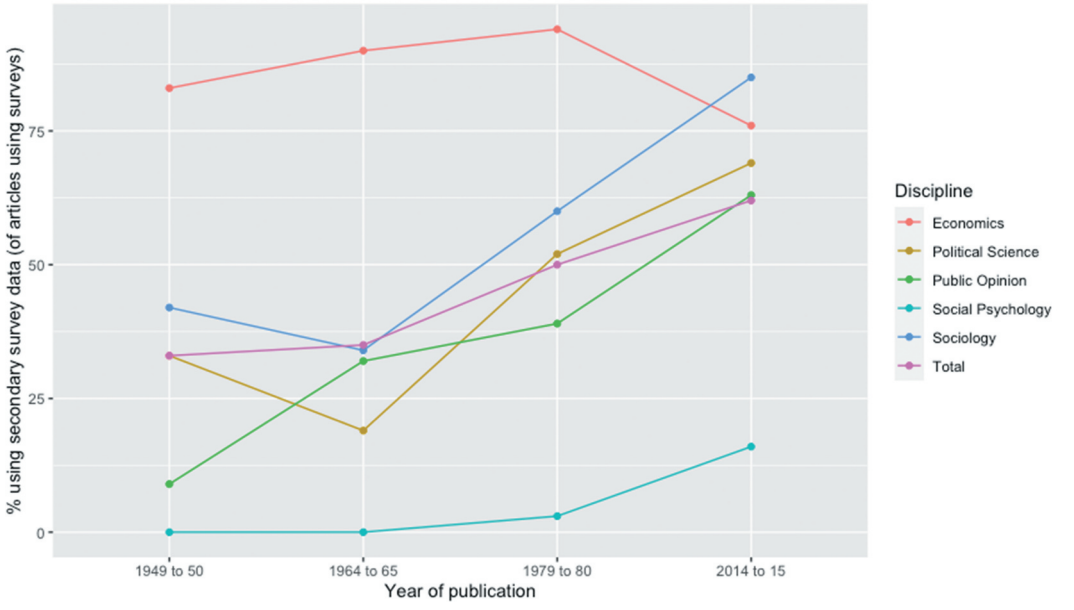

Figure 1. Rates of secondary analysis of survey data in published articles 1949 to 2015.

scholars have made use of Big Data, it is common for this to be done in conjunction with survey data, rather than in isolation. This, then, serves as partial confirmation of the contentions of Groves (2011) and Couper (2013) that the analytical power of big data is enhanced by combining it with representative survey data (see also Japec et al., 2015).

An additional trend of note is an increasing reliance on secondary survey data in published academic articles. It is difficult to say exactly why this trend has emerged, although it may have resulted, at least in part, from increasing costs of fieldwork rendering bespoke data collection exercises more difficult to support and justify. In this sense, while increasing costs may be limiting the number of high-quality random surveys that are carried out, it is not reducing the frequency with which such surveys are analysed.

It is necessary to acknowledge the limitations of our empirical strategy, particularly with regard to our ability to generalise these estimates of data use from this set of journals to the wider universe of social science research output (Platt, 2016). The journals we have focused on were, in effect, proscribed by the choices of the scholars who preceded us in this endeavour and there can be little doubt that an approach based on the selection of articles via stratified random sampling would yield a more representative sample of outputs. In particular, it seems likely that the journals originally selected by Stanley Presser in 1984 over-represent quantitative articles relative to qualitative and theoretical pieces. They are also likely to under-represent articles using Big Data which we would expect to be concentrated, initially at least, in more specialist journals. Nonetheless, despite these limitations, it seems clear that the survey retained its position as the pre-eminent form of data in leading social science journals as recently as 2015 and, insofar as social scientists are turning to Big Data to address substantive questions, this is often done as a complement rather than as an alternative to surveys.

Notes

1. Sub-codes were also applied to qualitative and mixed-methods articles but we do not describe them here.
2. Saris and Gallhofer (2007) provide two sets of figures for 1994/5, one set which codes all articles citing statistical bureaus as using survey data and a second set which bases the coding on whether the article reported using survey data. We use the latter set of figures in Table 2.

Acknowledgments

We gratefully acknowledge the support of the Economic and Social Research Council (National Centre for Research Methods grant RES-576-47-5001).

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Patrick Sturgis is a Professor of Quantitative Social Science in the Department of Methodology at the London School of Economics. His research focuses on survey methodology, public opinion and political behaviour, and social mobility.

Rebekah Luff is a Research Analyst at the Social Care Institute for Excellence. Prior to that, she was a Senior Research Fellow at the National Centre for Research Methods at the University of Southampton. Her research interests lie predominantly within adult health and social care, particularly long-term care of older adults.

References

Ayrton, R. (2017) Time for a revival? A historical review of the social survey in Great Britain and the United States, *National Centre for Research Methods Methodological Review paper*, University of Southampton, Southampton.

Blumer, H. (1955). Sociological analysis and the "variable". *American Sociological Review*, *21*(6), 683–690. https://doi.org/10.2307/2088418

Cicourel, A. V. (1964). *Method and measurement in sociology*. Free Press.

Converse, J. (2009). *Survey research in the United States: Roots and emergence 1890–1960* (2nd ed.). University of California Press.

Couper, M. P. (2013). Is the sky falling? New technology, changing media, and the future of surveys. *Survey Research Methods*, *7*(3), 145–156. https://doi.org/10.18148/srm/2013.v7i3.5751

Cukier, K., & Mayer-Schoenberger, V. (2013, May–June). Rise of big data: How it's changing the way we think about the world. *Foreign Affairs*. The Council on Foreign Relations. http://fam.ag/1bKTv6t

Groves, R. M. (2011). Three eras of survey research. *Public Opinion Quarterly*, *75*(5), 861–871. https://doi.org/10.1093/poq/nfr057

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., O'Neil, C., & Usher, A. (2015). Big data in survey research: AAPOR task force report. *Public Opinion Quarterly*, *79*(4), 839–880. https://doi.org/10.1093/poq/nfv039

Kolb, J., & Kolb, J. (2013). *The big data revolution*. Applied Data Labs.

Krippendorff, K. (2004). Content Analysis, an Introduction to Its Methodology, 2nd Edition. Thousand Oaks, CA: Sage Publications.

Miller, P. V. (2017). Is there a future for surveys? *Public Opinion Quarterly*, *81*(S1), 205–212. https://doi.org/10.1093/poq/nfx008

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Platt, J. (2016). Using journal articles to measure the level of quantification in national sociologies. *International Journal of Social Research Methodology*, *19*(1), 31–49. https://doi.org/10.1080/13645579.2014.947644

Presser, S. (1984). The use of survey data in basic research in the social sciences. In C. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. 2, pp. 93–114). Russell Sage Foundation.

Saris, W. E., & Gallhofer, I. N. (2007). *Wiley series in survey methodology. Design, evaluation, and analysis of questionnaires for survey research*. Wiley-Interscience.

Savage, M. (2010). *Identities and social change in Britain since 1940*. Oxford University Press.

Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, *41*(5), 885–899. doi:10.1177/0038038507080443

Savage, M., & Burrows, R. (2008). Wither the survey? *International Journal of Market Research*, *50*(3), 305–307. https://doi.org/10.1177/147078530805000302

Savage, M., & Burrows, R. (2009). Some further reflections on the coming crisis of empirical sociology. *Sociology*, *43*(4), 762–772. doi:10.1177/0038038509105420.