Survey Research Methods (2013)
Vol. 7, No. 3, pp. 145-156
ISSN 1864-3361
http://www.surveymethods.org

Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys*

Mick P. Couper
Survey Research Center
University of Michigan

In this paper I review three key technology-related trends: 1) big data, 2) non-probability samples, and 3) mobile data collection. I focus on the implications of these trends for survey research and the research profession. With regard to big data, I review a number of concerns that need to be addressed, and argue for a balanced and careful evaluation of the role that big data can play in the future. I argue that these developments are unlikely to replace transitional survey data collection, but will supplement surveys and expand the range of research methods. I also argue for the need for the survey research profession to adapt to changing circumstances.
Keywords: big data; organic data; social media; mobile surveys; non-probability surveys

> "To everything there is a season, and a time to every purpose under the heaven … a time to be born, a time to die, a time to plant, and a time to pluck up that which is planted … " (Ecclesiastes 3:1)

1 Introduction

Has survey research's time come to an end? There are many who suggest that the glory days of surveys are behind us, and we face a future of marginalization if not redundancy (see, e.g., Savage and Burrows, 2007). There are three elements to this. First, with the rise of Big Data[1], when one can collect data on everything that people do, who needs surveys of small subsets of a population? Second, with the rise of opt-in panels, Google Consumer Surveys, Amazon's Mechanical Turk, etc., and other ways to get responses from large numbers of people in relatively little time and at very low cost, who needs probability sample surveys? And third, with the rise of do-it-yourself (DIY) survey tools (e.g., SurveyMonkey), who needs survey professionals? Anyone can do a survey, and – it seems these days – almost everyone does.

Are we redundant? I believe not. In this paper, I review some of the massive changes currently underway in the use of technology – especially social media use and mobile computing – and the implications of these trends on the survey profession. Some take the view that "big data" represents a "brave new world" that will soon replace surveys as the major (or only) source of data on people's attitudes, behaviors, intentions, and the like. This perspective, together with the challenges to traditional surveys in terms of coverage and nonresponse, along with rising costs, may suggest that the survey method has outlived its usefulness. I take a different view, and argue for the important role of surveys – and especially high quality surveys – in our understanding of people and the societies in which we live. I believe that surveys still play a vital role in society, and will continue to make important contributions in the future. However, this does not mean we can be complacent – we do need to adapt as the world around us changes.

It is not my plan to review the technology developments in detail here. This is a well-worn path. There are many who extol the virtues of big data. Similarly, almost every recent presentation on mobile Web surveys reviews all the wonderful things one can do with mobile devices and talks about the rapid penetration of the technology. This was the same kind of excitement that greeted the advent of the Internet, and the development of computer-assisted telephone interviewing (CATI) before that – we are not immune from the hype around new technology. The growth in social media has been similarly well-documented. My goal is to focus not on the technology trends themselves, but on the implications of these trends for the survey profession.

I focus on three key technology-related trends: 1) big data, 2) non-probability samples, and 3) mobile data collec-

Contact information: Mick P. Couper, Survey Research Center, University of Michigan, USA (mcouper@umich.edu)

* Editors' note: This article is not the kind of paper usually published in SRM. However, SRM is the journal of the European Survey Research Association (ESRA), and this text is a reviewed and revised version of the keynote address of the 2013 ESRA conference in Ljubljana. By publishing this article, SRM documents the keynote speech of this important event of its funding institution.

[1] Several others are writing about this topic. For example, Prewitt's (2013) paper appeared as this paper was being completed. In it he talks about the "digital data tsunami" and raises many of the issues regarding big data that are addressed here.

tion. While these are seemingly unrelated, I attempt to show how they raise similar questions for the future of survey research. I discuss each of these in turn before offering some observations on what we can do as survey researchers to respond to the challenge posed by these developments.

2 Big Data

Groves (2011) coined the term "organic data" to describe digital data automatically generated by systems. There are characteristics other than size that describe such data, and "Big Data" (often capitalized) may make one think of "Big Brother," with all the negative connotations[2]. However, "big data" is now part of the modern lexicon, so I will use the two terms interchangeably.

There are three attributes that are generally agreed to describe organic data (see, e.g., Daas, Roos, van de Ven, & Neroni, 2012):

1. volume (exceeds capacity of traditional computing methods to store and process),
2. velocity (streaming data or complex event processing), and
3. variety or variability (raw, messy, unstructured, not ready for processing, does not fit into a relational structure).

In addition to these characteristics of big data, we can identify a number of broad types of organic data, with different implications for access and analysis. These include[3]:

1. Administrative data – data provided by persons or organizations for regulatory or other government activities. Users may assume that the data are confidential and used only for the intended purpose by the agency collecting the data.
2. Transaction data (credit cards, highway/public transport passes, loyalty cards, phone records, browsing behavior, etc.) – data generated as an automatic byproduct of transactions and activities. Users may recognize that the data are being captured and used for the primary purpose of processing the transaction or to facilitate user activities, but may not be aware of secondary uses of the data (e.g., marketing).
3. Social media or social networking data – created by people with the express purpose of sharing with (at least some) others. User expectations about who has access to the data and for what purpose may vary.

Most of my focus is on the second and third types. There are those who argue that with so much data being generated, surveys are no longer of any value. In one provocative view, the title of a 2008 article in *Wired Magazine* posited "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete" (Anderson, 2008). In similar vein, Savage and Burrows (2007, 891) argued that, "...where data on whole populations are routinely gathered as a by-product of institutional transactions, the sample survey seems a very poor instrument." In my view, this confuses quantity with quality. My goal here is not to argue for the benefits of big data – I think there are many useful and interesting things that can be done with these data, and they offer exciting opportunities for researchers. My goal is to argue for a balanced view on big data – like all other sources of data (including surveys), organic data have strengths and weaknesses, and understanding these is important to making appropriate use of them.

Some possible limitations of big data – and reasons why I think big data will complement survey data rather than replace surveys – include the following:

Single variable, few covariates

If all we were interested in was temporal trends or geographical variation in a single measure (e.g., the price of fuel, or the spread of influenza), social media analysis or web scraping tools might well give us what we want. But surveys are about much more than estimating a single variable. Social media and transaction data do not have much demographic data. For example, Keeter and Christian (2012) note that demographic information is not available for about 30-40% of Google Consumer Survey (GCS) respondents. For the rest, GCS either uses demographic data provided through Google+ or "assumes" or "imputes" characteristics based on browsing behavior. In their comparison of inferred characteristics from Google Consumer surveys to reported characteristics, Keeter and Christian (2012) found that the inferred gender matches reported gender in about 75% of cases. For age categories, the two match about 44% of the time, but this goes up to 75% of the time when adjacent categories are considered. Similarly, about one-third of Facebook users have no demographic information available (Link, 2013). This limits the kinds of multivariate analyses or subgroup comparisons that are the staple of survey research. Or, as Prewitt (2013) recently put it, big data are case rich but variable poor, while survey data are variable rich but case poor.

Further, the type of data is often limited. Transaction data reveals behaviors – what people are doing, but not why they do it, or what they intend to do in the future. Similarly, social media data might reveal people's thoughts, feelings, preferences, etc., but not necessarily the behaviors that go with those reported views. If one only wanted to measure prices of consumer goods, the Billion Prices Project (see bpp.mit.edu) may give us timely and detailed information. But the Consumer Expenditure Survey (for example; see http://www.bls.gov/cex/) allows us to understand how increases in costs in one type of commodity may result in changes in household expenditures in other areas. For example, do households spend less on medications when food costs go up (or vice versa)? Similarly, we might know that transportation costs are going up, but we don't know whether and how people are changing their travel and other behavior

[2] The recent disclosures about the U.S. National Security Administration's (NSA) use of big data remind one of some of the risks of such data to those who generate it.

[3] There are other types of big data of relevance to national statistics (e.g., passive traffic monitors, movement of goods, environmental monitors). I focus on those involving provision of information by humans.

as a result. We might correlate fuel prices with ridership of public transport or purchase of fuel-efficient cars at an aggregate level using big data, but this may be harder to do at the individual level. For that we need surveys.

Bias

Two types of bias are of concern with regard to organic data. The first is selection bias. Big data tends to focus more on the "haves" and less on the "have-nots". This may also be true of much market research, but social research has traditionally been more interested in the "have-nots". And while almost all of us are users of the new media, we must remind ourselves not to generalize from our own experiences, and remember that while the number of active Facebook users (for example) is enormous, not everyone is on Facebook. Similarly, while almost everyone has heard of Twitter, the number of people who actively tweet is still relatively small (about 13% of the US online population, according to Link, 2013), and highly selective. That is, we should make a distinction between the producers of social media and the consumers of such media. The former may not be representative of the latter, and neither may be representative of the general population. Studying Twitter posts (for example) may be closer to studying elites than the general population. Similarly, there are still sizable minorities of the population who do not use the Internet – thus, for example, those most affected by influenza (the poor, the elderly, the marginalized) may be least likely to search the Internet for help. To the extent that these characteristics are geographically clustered, we may miss key areas affected. Does this invalidate Google Flu Trends as a method of studying the spread of the virus? No, but we must be clear about the selection biases inherent in these kinds of analyses (as with surveys), and understand how they could affect the conclusions we draw.

We also need to understand the limits of transaction data – not everyone uses loyalty cards (for example) or credit or debit cards. Mobile phone (and especially smart phone) penetration is not at 100%. Not everyone communicates by e-mail, and those who do may use different accounts for different purposes. Selection bias can occur at the individual level (e.g., those still using cash) and at the transaction level (e.g., some types of purchases – such as alcohol, cigarettes, condoms, or fast food – may be more likely to be paid for in cash). There are still many ways in which transactions can be conducted without leaving a trace, and many tips and techniques for avoiding being traced (e.g., Singer, 2013). Selection bias is a key feature of organic data (especially of transaction data and social media data, but also of administrative data) and understanding the extent and impact of these biases is a key challenge – and one where we can make a contribution. As with survey data, these biases may be negligible or ignorable in some cases but large and misleading in others.

The second type of bias is measurement bias. Again, this is something that is well known to survey researchers, but has tended to be ignored in the heady rush to exploit the volume of organic data becoming available. Despite the stories one reads about the things people post on Facebook or other social media sites, social media is primarily about impression management (see Boyd & Ellison, 2008). To what extent do people's posts represent their "true" values, beliefs, behaviors, etc.? Similarly, if we counted the number of Facebook friends one has as an indicator of true social network size, we may be seriously wrong. The average Facebook user is estimated to have 229 "friends" (Hampton, Goulet, Rainie, & Purcell, 2011). Again, I'm not saying that Facebook is useless for research purposes, I'm just saying that we need to understand who is using the medium and why they are doing so, in order to understand what biases may exist with social media data.

Volatility or lack of stability

Social media may come and go (remember MySpace? Second Life?), but surveys are relatively constant. This is especially important for trends over long time frames (decades or generations). The rapid rise of Facebook (which was founded in 2004) gets our attention now, but what will Facebook look like five or ten years from now? Will it even exist – what is the half-life of Facebook? Social media may be useful for short-term trends (days or weeks), but may not be stable enough for longer time trends (years, decades). For example, Twitter (which only began in 2006) grew 5000% in the last five years. This means that Twitter today is very different from Twitter five (or even two) years ago. Who knows what Twitter will look like five years from now, or whether it will even still exist? Google itself is just a teenager, with the domain being registered in 1997. There is also rapid evolution in what people share on these sites, and the limits they place on access to their information, especially in response to external events (such as the recent leaks about the US National Security Administration's PRISM surveillance program[4]). As soon as we (research institutions) become interested in a social network or media site, it is probably already past its prime.

Privacy

Related to the issue of volatility is the changing behavior of people using social media and other websites based on concerns about privacy, along with legislation (particularly in Europe) aimed at giving users control over what is collected when they go online. The more the collection and use of big data become broadly known, the more concerned people may become about sharing their information freely – e.g., Wilson, Gosling, and Graham (2012) document some of the changes in Facebook privacy settings over time. This will likely result in an increase in opting out of tracking, rejection of cookies, changes in the amount and type of information shared, use of alternatives to "hide" activities (e.g., paying cash for alcohol and tobacco; using fake e-mail addresses and multiple browsers to confound cookies, etc.), and the development of tools to give users control over what is shared with whom. For example, advertisers have reacted negatively to

[4] http://www.washingtonpost.com/blogs/wonkblog/wp/2013/06/12/heres-everything-we-know-about-prism-to-date/

Microsoft's decision to make the Do Not Track option the default in its new browser (Internet Explorer 10)[5]. Max Frankel (*New York Times*, June 23rd, 2013) noted that: "Privacy is a currency that we all now routinely spend to purchase convenience." But that may not always be the case, and it may not be true of all activities.

Access

Much of the big data being generated is proprietary. It is being used for commercial purposes and has a value (i.e., a price) to those who collect it. Access to data is also restricted for confidentiality purposes, either to protect the identity of participating individuals or to protect the business interests of the entities collecting the data. This means that it may not be freely available – or available at all – to the broader research community. For example, Facebook is not likely to make their database of members available to researchers for sampling or analysis, even at a fee. In addition, the availability of such data may change over time, further adding to concerns about stability. One of the key strengths of surveys, by way of contrast, is public access to the data – conditional on confidentiality restrictions and disclosure limitations. This facilitates reanalysis and replication, which strengthens the underlying value of the data and our faith in the conclusions drawn from the data.

Opportunity for mischief

It is harder to find evidence of this, but I believe that the more people realize that analysis of organic data can influence decision-making, the more likely we are to see attempts to manipulate the system – e.g., to generate interest in a topic or produce the desired results by directly manipulating social media. This is the social media equivalent of ballot-stuffing, which required time and money for call-in polls, but is virtually effortless in the online world, given the ability to write code to generate such content automatically, to create multiple accounts, to generate buzz by re-tweeting, and so on. A story in *The Guardian*[6] in 2011 revealed a US spy operation that manipulated social media, claiming "Military's 'sock puppet' software creates fake online identities to spread pro-American propaganda." Similarly, a recent online story[7] claimed that nearly half of Justin Bieber's 37 million Twitter followers were either fake or inactive. It was recently estimated that about 83 million Facebook accounts (or 8.7% of all accounts) were fake, with 4.8% being "duplicate accounts," 2.4% being "misclassified accounts" (that represent an entity other than the user), and 1.5% being "undesirable accounts" (that purposefully violate Facebook's terms of service, such as spamming)[8]. With increased visibility and importance of big data may come increased attempts to manipulate the data for financial or political gain, or merely to make mischief.

Size is not everything

The characteristic of big data most often mentioned is size. I believe this is the biggest mistake people make with regard to big data. Before we get too excited about the large numbers of people who are using social media, we need to remember that bigger is not necessarily better. Let's take one old example: a sample of 10 million records yielded a response rate of over 23 percent. That's over 2.36 million records – sizeable by any standard. The study was conducted by an organization that had correctly predicted the outcome of 5 previous elections. But the result was a spectacular failure – this is the infamous *Literary Digest* poll of 1936 (see Squire, 1988; Lusinchi, 2012), which called the US election for Landon over Roosevelt. This debacle led to the demise of the *Digest*. Big, but wrong!

On the other hand, it is remarkable that we have to go as far back as 1936 to find such a spectacular failure in election polling. This is an example of cherry-picking that I'll address later – selectively presenting evidence to support arguments against big data. Actually, the 1948 election (Dewey defeats Truman) has been used for decades as an argument for the failure of quota sampling (used by Gallup and all other leading pollsters at the time), and led to the rise of probability sampling. This brings us to the US election of 2012, where Gallup (using probability sampling methods) was one of the furthest from the final outcome[9]. This suggests that all methods need constant evaluation. Election polling (with a few exceptions that would be expected by chance) has had a remarkable run. In several countries, pre-election polls have been used to contest the outcome of elections (i.e., asserting evidence of fraud), suggesting that such polls can at times be even more accurate than a (flawed) count. But it's not just about the size of the sample. And, being accurate once (or even several times) is no guarantee of continued accuracy. This brings me to the final concern about big data.

The file drawer effect

This issue goes well beyond the big data debate, and is worth further attention. The term is attributed to Rosenthal (1979), who wrote: "For any given research area, one cannot tell how many studies have been conducted but never reported. The extreme view of the 'file drawer problem' is that journals are filled with the 5% of the studies that show Type I errors, while the file drawers are filled with the 95% of the studies that show nonsignificant results" (Rosenthal, 1979).

The concern is that much of what we've seen so far is based on selective reporting of findings that support the hypothesis in favor of big data. Aside from the well-known Google Flu Trends (e.g., Dugas et al., 2013), there are many other published papers using Internet searches or Twitter

[5] http://adage.com/article/digital/advertising-week-microsoft-blasted-track/237532/

[6] http://www.guardian.co.uk/technology/2011/mar/17/us-spy-operation-social-networks

[7] http://www.digitalspy.com/music/news/a471915/justin-bieber-twitter-followers-50-percent-are-fake-says-report.html

[8] http://usatoday30.usatoday.com/tech/news/story/2012-08-03/cnbc-facebook-fake-accounts/56759964/1

[9] http://fivethirtyeight.blogs.nytimes.com/2012/11/10/which-polls-fared-best-and-worst-in-the-2012-presidential-race/

analyses to "predict" a variety of things, including voting behavior, problem drinking, mental health, consumer behavior, economic conditions, and the like (see, e.g., Choi & Varian, 2012; Frijters, Johnston, Lordan, & Shields, 2013; Ghosh & Guha, 2013; Lansdall-Welfare, Lampos, & Cristianini, 2012; Paul & Dredze, 2011). While these papers trumpet the success of the method (by showing high correlations between the organic data and benchmark measures), we do not know how many efforts to find such relationships have failed. In one exception, Murphy and colleagues (2011; see also Kim, Hansen, and Murphy, 2012; Kim et al., in press) compared trend analyses regarding the drug *salvia divinorum*, using Twitter feeds and Google search, to data from the National Survey of Drug Use and Health (NSDUH). They find that the trends are quite dissimilar. Specifically, a huge spike in tweets about the drug was associated with a YouTube video of Miley Cyrus smoking *salvia*, without a corresponding change in actual drug use at the time. Similar recent results have been found for Google Flu Trends, with significant errors in both 2009 and 2013 (see Cook, Conrad, Fowlkes, & Mohebbi, 2011; Butler, 2013).

In a humorous example, Leinweber (2007), in a paper originally written in 1995, showed how one can "predict" the S&P 500 index of the US stock market[10] with an R^2 of 0.99 using just three variables: 1) butter production in Bangladesh and the US, 2) cheese production in the US, and 3) sheep production in Bangladesh and the US. The same three variables were useless outside the fitted time period.

The file drawer problem is not limited to new technologies and trends. For example, Hirschhorn and colleagues (2002) conducted a review of 600 positive associations between gene variants and common diseases. Out of 166 reported associations studied 3 or more times, only 6 were replicated consistently. Similarly, Ioannidis (2005) argues that "in modern research, false findings may be the majority or even the vast majority of published research claims" (see also Moonesinghe, Khoury, and Janssens, 2007). In a comparison of publications in 18 empirical areas, Fanelli (2011) found ratios of confirmed hypotheses ranging from 70% (space science) to 92% (psychology and psychiatry). This rate of 92% is far above what should be expected, given typical effect sizes and statistical power of psychological studies (see also Asendorpf et al., 2013; Yong, 2012). This has led some fields to develop ways to encourage the reporting of nonsignificant effects, replications, and the like (see, e.g., http://www.psychfiledrawer.org/).

We as survey researchers are facing a similar dilemma. The papers that "demonstrate" the utility of an exciting new method are more likely to get published than later papers doing the careful but less sexy evaluation of those methods. One simple solution is for journals like *Survey Research Methods* to have a special section for short research notes where such reports are encouraged. Another thing we need is independent evaluations of the trends produced from organic data by those who don't have a vested interest in the outcome – e.g., contrast McDonald, Mohebbi, and Slatkin (2012) with Keeter and Christian (2012). As Carl Sagan and many others have noted, "absence of evidence is not evidence of absence" (Sagan, 1995, 213).

While this is a problem facing all fields of study, because of the large amount of data involved, analysis of organic data may be more likely to yield Type I errors (i.e., finding significant effects or associations where no substantively meaningful effect exists).

My review of concerns about big data may sound like I'm arguing against big data or organic data. This is not the case. I'm convinced that social media research (and big data more generally) has much to offer. The analysis of transaction data is likely to yield many important insights into human behavior that could not be garnered in other ways. Similarly, administrative data have a huge potential. However, I'm equally convinced that these approaches are unlikely to replace survey research. Our role as researchers is to figure out how best to make use of these new opportunities, to expand the range of data we use to understand the societies in which we live. There's a wealth of interesting research opportunities out there for quantitatively-minded researchers. We need to figure out when big data is useful, what biases and flaws may exist, and how we can overcome them. To do this, we need to strip away the hype and examine the evidence in detail – that is, we need to do the research. The same methods and criteria we use for surveys should be useful. As Groves (2011, 869) noted, "The challenge to the survey profession is to discover how to combine designed data with organic data, to produce resources with the most efficient information-to-data ratio." This is where we have important contributions to make.

3 Non-Probability Samples

I will devote less space to the second trend. This is not a new trend, but it is still instructive to review. Non-probability surveys have been around for a long time (see AAPOR, 2013; Baker et al., 2013), but the recent attention that has been paid to such methods can be attributed to the rise of Internet surveys and, more specifically, the development of volunteer opt-in or access panels. Understanding the short history of Web surveys will help us prepare for future technology shifts.

The rise of online opt-in or access panels in the early part of the 21st century was meteoric. Promoters of such panels were claiming that that they make other methods of survey data collection obsolete. One of my favorite quotes from that time is from Gordon Black, then chairman and CEO of Harris Interactive, who stated that "Internet research is a 'replacement technology'—by this I mean any breakthrough invention where the advantages of the new technology are so dramatic as to all but eliminate the traditional technologies it replaces: like the automobile did to the horse and buggy. Market research began with door-to-door household surveys which gave way to telephone polling in the mid-1960s and is now making a quantum leap forward with new Internet research techniques" (Harris Interactive press release, August 1, 1999; see also Couper, 2000).

In the heady early days of Internet panels, the belief was that there was an infinite number of potential survey respon-

[10] http://us.spindices.com/indices/equity/sp-500

dents. It was unthinkable then that the demand for surveys would exceed the supply of respondents. But this is indeed what seems to have happened over the last decade or so. There is increasing evidence that a relatively large number of surveys are completed by a relatively small number of active panelists, many of whom belong to several panels (e.g., Vonk, Willems, & van Ossenbruggen, 2006; Tourangeau, Conrad, & Couper, 2013). The number of surveys requests sent to panelists has sky-rocketed over time. This has led to a rise of concerns about fraudulent or inattentive behavior on the part of panelists, leading some to question the quality of data from such panels (e.g., AAPOR, 2010; Baker et al., 2010). This led the AAPOR Task Force on Online Panels (2010) to conclude that while such panels have a number of uses, "Researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values."

But it's not just the online panels that contributed to this problem. Almost any online transaction these days results in a follow-up satisfaction survey. For those who travel a lot (for example), this can mean several surveys for one trip, including one (or more) for each flight, hotel, rental car, and other activity. Sometimes these surveys take longer to complete than the actual transaction being asked about.

In a way, the very success of Internet surveys has contributed to their possible downfall. There is a parallel to the way the rise of telemarketing affected the telephone survey industry. When something is almost costless and treated like a commodity, it tends to lose value. Beniger was remarkably prescient when he wrote in 1998 about the rise in Web surveys, "Good luck to any serious survey firms which pin much of their futures on the hope of being heard for long above the mounting background noise and confusion of this swelling tide of amateur and slapdash pseudopolls" (Beniger, 1998, 446). Replace "pesudopolls" with "big data analytics" (or any other popular trend) and we see the situation we face today.

Efforts to fix the problems faced by volunteer online panels include a variety of alternative recruitment and selection methods such as river sampling, respondent-driven sampling (RDS), and sample matching (see AAPOR, 2013, for a description of these methods), and the use of Google Consumer Surveys (see, e.g., McDonald, Mohebbi, & Slatkin, 2012; Keeter & Christian, 2012) and Amazon's Mechanical Turk (see, e.g., Berinsky, Huber, & Lenz, 2012). Attention has also focused on improving the design or content of the surveys, with terms like "gamification" and "surveytainment" gaining popularity. In my view, none of these approaches fix the fundamental problem – of demand exceeding supply, of our appetite for data overwhelming the capacity of participants to provide it. Ironically the very success of these panels points to the value of survey data, while at the same time making it harder for everyone to do good surveys because of the saturation problem. For this reason, if the rise in big data means fewer surveys, then maybe this is a good thing. Fewer surveys might mean that those that are done will be of better quality. Scarcity of surveys may also raise their value among potential respondents. It is the ubiquity of surveys and the corresponding commoditization of surveys (Tourangeau, 2010) that have led (in part) to some of the problems we face.

The use of the terms "gamification" and "surveytainment" (see, e.g., Downes-Le Guin, Baker, Mechling, & Ruyle, 2012; Findlay & Alberts, 2011; Puleston, 2011; Tress, Winkler, & Schmidt, 2012) is unfortunate. Trying to turn an otherwise bad survey into a game or a form of entertainment is like putting lipstick on a pig. To be fair, this is not what the proponents of gamification are arguing. Survey engagement (in my view) is a better concept. The idea is not to trivialize the survey enterprise. We want people to take what we do (and what we ask them to do) seriously. Gamifying surveys undermines this and sends a different message. While gamification has been shown to improve a number of metrics such as idea generation, length of open responses, and the like (see Puleston, 2011), this may not be the domain of much standardized survey measurement. However, I believe we should design surveys (both content and presentation) with the goal of fostering user or respondent engagement. This is the basis of user-centered design. We need to see the survey from the respondents' perspective, not our own. In my view, we have become arrogant in our design of surveys, placing increasing demands on respondents, with little thought to their motivation, interest, ability, etc. When concerns are raised, we throw trivial amounts of money at them, in the form of token incentives. I believe we need to meet respondents halfway.

While volunteer online surveys remain enormously valuable and serve many useful purposes[11], they are undergoing a transformation, in part because of the challenges presented by over-saturation, but also in part due to the opportunities presented by big data alternatives. It's going to be interesting to see how this plays out over the next few years.

4 Mobile Data Collection

This brings me to the final technology trend, that of the "mobile revolution." A distinction can be made between three types of mobile use:

1. data collectors (interviewers) using mobile devices (tablets, smartphones, mobile Web) to conduct surveys and collect data,
2. respondents using mobile devices to complete regular Web surveys, and
3. respondents using mobile devices for enhanced data collection (e.g., GPS, photos, ecological momentary assessment (EMA), diary studies, food consumption measures, health monitoring, etc.).

Of most relevance here is the last of these types, but I will indulge in a short detour on the first and a brief comment on the second. The move to tablet-based or hand-held computers finally appears to be here. It has been a long time coming. Based on ergonomic studies conducted in the early 1980s, Statistics Sweden determined the ideal weight of a

[11] To be clear, I have used such panels in much of my own recent work, and think they are an important tool in the survey toolkit.

handheld CAPI computer to be less than 1 Kg (see Lyberg, 1985). In our testing in the US (Couper & Groves, 1992), we came up with a number around 1.6 Kg, which was significantly lighter (by a factor of 5) than all of the available machines at the time. We've had to wait almost 20 years for suitable products to come on the market. The iPad weighs about 0.6 Kg, while the Microsoft Surface is about 0.68 Kg. My point is simply that there are many who criticize the survey profession as being slow to adapt. I believe that, in several instances, the need for the technology is recognized well before such technology is ready for widespread use. Another example is audio-CASI, where the first implementation required interviewers to carry a separate device to generate the sound files because the DOS-based laptop used at the time could not generate sound (see O'Reilly, Hubbard, Lessler, Biemer, & Turner, 1994). It all sounds so quaint looking back, but these were important advances at the time.

Regarding respondents' use of mobile Web, there is a belief (or hope) that mobile Web would bring in different types of people – especially the young, who are currently disproportionately missing from other types of surveys – that is, that technology would compensate for nonresponse bias. So far, the results seem to suggest that we may just be getting more of the same. Those using new technologies to complete our surveys generally seem to be those who would do them anyway using more traditional methods. In this sense, mobile Web may offer more complications to an existing mode rather than solutions to problems we face. But this is an area where further research is needed, and again opportunities abound.

To return to the third type, there are many exciting opportunities for using mobile devices to capture data with greater frequency and fidelity and reduce the need for self-reporting, and there is no shortage of researchers pointing out all the marvelous things that could be done using these devices (e.g., Palmer, Espenshade, Bartumeus, & Chung, 2013). However, to date, almost all of the studies that have demonstrated the use of these devices and apps have been based on volunteers. These volunteers usually have to download and install an app, activate a peripheral device, or otherwise take an active part in collecting the data. These studies have often been restricted to users of particular devices, or to small groups of highly motivated users.

Work on the Dutch LISS panel[12] is one promising exception, and the French ELIPSS panel[13], which is equipping panelists with tablet computers, offers exciting opportunities. But until we can successfully move from small-scale studies of volunteers to implementation among probability-based samples of the general population, these will remain niche technologies (from a general population survey perspective).

Two recent papers from the NTTS conference in Brussels[14] illustrate the challenge. One paper (Biler, Šenk, & Winklerová, 2013) surveyed people in the Czech Republic about their willingness to participate in a travel survey using a GPS device. Only 8% said that they would be willing, while 67% said no (the remainder being uncertain). Another (Armoogum, Roux, & Pham, 2013) asked participants in the 2007-2008 French National Travel Survey about their willingness to accept a GPS device to monitor their travel: 29.8% said yes without condition, 5.1% said yes as long as they could turn it off, and 64.3% said no. Even trained professionals (i.e., interviewers) are not fully compliant – Olson and Wagner (2013) report, for example, that equipping interviewers with GPS-enabled smart phones and having them activate an app to track their work-related travel each day, yielded GPS files for 59.4% of the interviewer-days.

We are all excited about the cool things we as researchers could do with mobile devices, but the question remains, what are people willing and able to do? If we can't answer these questions, we won't be able to defend probability-based surveys against the threat of large data or volunteer surveys. This is one thread that binds these three trends – a point I'll return to later.

Another challenge remains that of coverage. Despite the apparent ubiquity of mobile devices – a recent headline[15] claimed that the number of active mobile phones will exceed the world population by 2014, with more than 100 countries where active cell phones already exceed the countries' population[16] – not everyone has a mobile phone, and not everyone has (or uses) a smartphone. The latest US numbers (June 2013) from the Pew Internet Project[17] suggest that about 91% of telephone-answering adults[18] have a mobile phone, and about 56% have a smartphone. Again, understanding the differences between the "haves" and "have-nots," and what this means for inference to the broader population, is a critical element of good survey research.

Having briefly examined three selected trends driven by technology changes, let me turn to offer a few thoughts on what this all means for the future of surveys, and the future of the survey profession.

5. The Future of Surveys ... and the Surveys of the Future

What ties these three trends (big data, online panels, and mobile data collection together)? While all are at different points in their trajectory, they are all technology trends that have had, or will have, a potentially large impact on the survey profession and the methods we use. In each case, the early proponents of the new methods are (or were) claiming that they will replace "traditional" methods of survey data collection, making current approaches obsolete. At the other end of the spectrum, there are those who bemoan the threat

[12] http://www.lissdata.nl/lissdata/Home

[13] http://www.elipss.fr/elipss/recruitment/

[14] http://www.cros-portal.eu/content/ntts-2013-programme

[15] http://www.digitaltrends.com/mobile/mobile-phone-world-population-2014/

[16] According to the International Telecommunications Union (http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx), there are 126.5 active mobile cellular subscriptions per 100 inhabitants in Europe in 2013 and 109.4 in the Americas.

[17] http://www.pewinternet.org/~/media//Files/Reports/2013/PIP_Smartphone_adoption_2013.pdf

[18] This is a survey conducted by telephone with response rates of 10% for the landline sample and 13% for the cell sample.

that these trends pose for our tried-and-true approaches. My sense is that there are many similarities between these trends – and any future technology trends – that are instructive for the profession.

In 1999, Gordon Black provocatively proclaimed "It's a funny thing about scientific revolutions. People who are defenders of the old paradigm generally don't change. They are just replaced by people who embrace new ideas" (*Wall Street Journal*, April 13, 1999). Are we facing a revolution, a Kuhnian paradigm shift? I'm not convinced that we are. While the changes facing the survey profession are many and large, I don't see surveys going away any time soon. Although surveys will survive as a method of scientific inquiry, we will have to adapt. I see several key areas of adaptation, many of which have been raised by others before. The first two are more specific and practical, while the latter ones are more a call-to-arms or challenge for the profession.

Reducing survey length or burden

We need to match our survey requests to the lifestyles of our potential participants. There has been an increasing disconnect between what we are asking for, and what people may think is reasonable to provide. This is being driven by the rising demand for high-quality survey data, but this is increasingly disconnected from reality. The model where we ask 1-2 hours' worth of questions in a single sitting is no longer sustainable. Who has that kind of time anymore? Or, more critically, those who do are likely to be very different from those who don't.

This model of overly-lengthy surveys is driven by the high cost of asking the first question in interviewer-administered surveys. Given the enormous investment required to find the sample person and get them to agree to do the survey, researchers want to maximize the return on that investment. But this may in fact be counter-productive. It may be – and this is an untested assertion – that a significantly shorter questionnaire may reduce the costs of contacting and persuading sample persons to participate. In part this is the assumption underlying the development of probability-based online panels. We should strive to make the barriers to initial participation low, and build loyalty and commitment over time.

To be provocative, I would go further, and assert that many of the researchers who design the questionnaire would themselves not be willing to do the interview. We have become too removed from the people we are studying. Here's one radical proposal – no one gets to asks questions on a survey unless they themselves will sit down and be interviewed (or complete the questionnaire) as part of the survey pretest. Given the increasing sophistication and complexity of surveys, researchers are increasingly removed from the data collection process, and this has to change. In the team-based approach to survey research, researchers may only feel responsible for a small part of the questionnaire, and may never experience the gestalt of the instrument. Respondents are a precious commodity, a scarce resource, and we should treat them as such. We have to match our survey requests to the lifestyles and expectations of our potential respondents.

How do we find ways to reduce the number of questions we ask? There are several possible approaches, and I think we should be looking into all of them:

1. Work on improving the validity and reliability of single-item measures or short scales, or using item-response theory (IRT) and computerized adaptive testing (CAT) to minimize the number of items asked.
2. Increase the use of data from other sources (whether administrative data or transaction data). We should not ask people to provide answers to questions we can get in other ways – or to questions to which they may not know the answer (see later).
3. Ask less detail, measure with less precision. Our analytic models demand data of increasing fidelity and detail, often exceeding respondents' ability (not to mention willingness) to provide the information. We need more modeling to make estimates based on what we have, rather than increasing our insatiable demand for more data, more variables, and more precision.
4. Make much more use of planned missingness or matrix sampling approaches.

If increased incentives and/or increased effort are the only tools in our toolkit, we are doomed to failure. Until we can also give on the content or length of the survey, we are unlikely to get out of the dilemma we are in. Here I'm a believer in the "less is more" precept of minimalist design popularized by architect Mies van der Rohe.

Using technology

Turning to the second area of adaptation, how can technology help us? We need to think about technology use both by respondents and interviewers. Again, we need to meet respondents halfway, and use the technologies they're already using, and the things they're already sharing, and have them help us. This may mean shorter, repeated measurements rather than single long surveys. Making contact, recruitment, and persuasion are still the key – but we're using old style methods to achieve this at great expense.

Mixed-mode data collection – despite the initial setbacks – is (I believe) still the future of survey research. Responsive or adaptive designs (see, e.g., Groves & Heeringa, 2006; Couper & Wagner, 2011; Schouten, Calinescu, & Luiten, 2011) are gaining ground, but I believe much more could be done. First, we could focus more of our attention on nonresponse bias rather than response rates. But second, we could be thinking about tailored or adaptive designs on a larger scale, including not only mode, incentive, and timing of effort, but also survey content. We need to be more nimble. The era of one-size-fits all approaches may be behind us.

Taken to the extreme, this suggests customized or individualized surveys. We're already doing this with complex computer-assisted interviewing (CAI) instruments with fills, skips, etc., and the increasing use of computerized adaptive testing (CAT), but I'm talking about doing this on a much larger scale. What does this mean for our conception of surveys as standardized measurements on a representative sample of persons? If different subsets of the sample are getting different sets of measures, either based on randomization or on their willingness to participate and provide this information, how do we create rectangular datasets for analysis? In some sense we're already doing this with questions on income (for example). A large number often don't respond, and get followed up with unfolding brackets (which are sometimes themselves not the same for all respondents). With imputation, a single income measure is constructed. This approach also has big technology and process implications – not only for instrument design (CAI programming and testing), but also for documentation and dataset production.

We have to understand how best to do this, and understand what new errors we may be introducing. This is where survey research may be at conflict with itself. One of the fundamental tenets of the survey method is standardization of methods and measurement – everyone is treated the same. In the early days this meant equal probability samples, identical measurement instruments, and standardized interviewing protocols. We have already moved far away from this in terms of sampling – unequal probability samples are now the norm rather than the exception. With the introduction of computer-assisted interviewing (CAI), measurement instruments also became increasingly more customized. Now we use multiple modes of data collection, differential incentives, and a variety of other adaptive approaches. How can we balance the notion of standardization with adaptive and responsive design? This will need good theory, good statistical methods, and good technologies to support.

Understanding the nonresponse problem

This issue has been around since the beginning of surveys, but is increasingly becoming the most pressing issue for probability-based samples. The fundamental problem facing surveys remains that of nonresponse – making contact with people and getting them to respond to surveys. What distinguishes probability-based sample surveys from many other quantitative methods of scientific study (experiments, observational studies, case-control studies, etc.) is that we do not rely primarily on volunteers. But increasingly this is changing, both explicitly (e.g., opt-in or access panels, river samples, etc.) or implicitly (low response rate surveys). What we need is to understand how volunteers differ from non-volunteers on the variables we are interested in measuring and the populations we are interested in studying. This won't be easy, as the very nature of non-volunteers or nonrespondents makes them elusive research subjects. But this is one of the big challenges for survey research in the next decade. There are two related questions we need to try to answer:

1. For probability samples, in what ways are respondents different from nonrespondents, and how this may differ across surveys? This is not just in terms of sociodemographic characteristics (the things we have frame data for, or could correct for), but attitudes, values, behaviors, intentions, etc. More important, we need to answer the question of why they may be different.
2. For non-probability samples and big data analytics, how do volunteers (those who choose to do surveys, sign up for panels, or agree to share their data) differ from non-volunteers?

Tackling these research questions will take new and innovative research methods. Developing theories to explain such differences is the single biggest challenge for surveys. Unlocking this key will help define the role of probability-based surveys for future decades – or lead to the conclusion that probability samples may not be that special after all.

Developing better quality metrics

Next, we need to develop quality metrics to help users differentiate between different types of surveys, or different types of estimates. Unfortunately, the recent work by Groves (2006) and Groves and Peytcheva (2008) makes it clear that this is a hard task. Error – whether sampling or measurement error (as has long been understood), or coverage or nonresponse error (as is only more recently being acknowledged), is a property of a statistic, not of a survey. Replacing response rates with other estimate-level metrics of nonresponse error (for example; see Wagner, 2012) will be a tough sell. But without this, how do we respond to the claims that organic data (and non-probability online surveys) are big, fast, and cheap, and that these factors alone may compensate for lower quality? We can't simply argue that more money means better quality.

The total survey error (TSE) paradigm is a useful framework and a good starting point. But it is rooted in the principles and procedures of probability sampling. We need other ways to quantify the risks of selection bias or non-coverage in big data or non-probability surveys. We need to focus more on costs, not just on errors. TSE remains relevant as an organizing framework but needs to be expanded.

The notion of fitness for purpose has also been around a long time. Quality is not an absolute. It must be evaluated relative to the stated aims of the survey and the purpose to which is put, and the investment (time and money) in obtaining the data. Non-probability surveys and organic data have their place, but so do probability surveys. And we need to develop methods to guide our decisions about which to use when. This is an issue that affects both the producers of data and the consumers of such data, whether analysts or the general public.

Like good wine, the provenance of the data we analyze is important, as is quality. We need to educate users on how to consume data. Sometimes I fear this may be a lost cause. Analytic software makes it too easy for people to conduct analyses without concern for where the data come from or how they are produced. The analytic software we use is agnostic as to the source of the data. Also, the sheer volume of data, and the number of people who directly consume data without regard for source, makes this an almost impossible task. But we must try, at least among ourselves – in the papers we present, in the journal articles we submit and review, in the reports we write. We should take care to point out what we did, and alert readers to the risks of using the data.

Using (and developing) different statistical tools

The kinds of design and analytic problems we are facing require different analytic tools. The methods that many of us learned, which assumed probability-samples with little or no error (other than sampling error) producing rectangular and complete datasets, are increasingly inadequate to handle the complex and messy datasets we now encounter. There's a lot of development already going on in this area, for example, in dealing with missing data, complex hierarchical designs, small area estimation, estimation in the presence of coverage and nonresponse bias, and mixed-mode designs with measurement error (to name but a few). But we also need (for example) new statistical tools to make sense of the masses of messy paradata being generated (see Kreuter, 2013).

On a broader level, we need to be open to other statistical frameworks and approaches to inference, especially for dealing with inference from non-probability based surveys or organic data (see AAPOR, 2013). The probability-based sample survey and frequentist statistical framework are not the only paths to inference. I'm not arguing we should all abandon the frequentist view and become Bayesians (c.f., Little, 2012). But I do agree with Silver (2012, 15) who says "We must become more comfortable with probability and uncertainty. We must think more carefully about the assumptions and beliefs that we bring to a problem." We need tools that match the data we have.

To summarize, I believe surveys will still be around, but they will need to change. We can't cling to the old ways and oppose any new method or approach. Nor can we throw the baby out with the bathwater, and rush to adopt every new method that arises. Big data are here to stay, as are non-probability samples. We have to figure out what method makes sense for which problem.

I find it interesting that those who argue for the superiority of non-probability surveys often use probability-based surveys to demonstrate the quality of their estimates. Similarly, big data estimates are often correlated with survey estimates to evaluate their utility. What would happen if the probability-based surveys were to disappear? We need well-designed and well-executed surveys to serve as benchmarks by which we can evaluate alternative approaches. While high-quality surveys serve this important role of providing a foundation for a vast array of other research, it seems likely that the number and scope of such high-quality benchmark surveys will decline. So far, the demand for all types of surveys – including large-scale, high quality studies like the European Social Survey (ESS), the Survey of Health, Ageing and Retirement in Europe (SHARE), and European Union Statistics on Income and Living Conditions (EU-SILC) – does not seem to have abated, even though there are pressures to do more with less. But I can imagine an effort to consolidate and focus on a few key benchmark surveys while reducing or eliminating overlap or redundancy.

There are lots of interesting opportunities and challenges. Many different skills are needed. We need to set a research agenda that will get us there in the next few years. We're already embarked on this journey, and much good work is already being done in this area. This gives me confidence in the future of our profession.

6 Conclusions

To return to the title of this paper, I don't believe the sky will fall anytime soon. Let me end with two related thoughts. First, a gentle reminder that surveys are tools, and we should not lose focus on the ultimate goals of what we do. Second, I end with some advice for young researchers or those considering getting into this field.

Surveys are a set of tools. More specifically, surveys are a set of tools. There are many different types of surveys and many ways to conduct surveys. So, surveys are like screwdrivers. There are many different types and sizes of screwdrivers, for a range of different purposes. They also vary in quality and cost. But there are also many other tools in a toolbox. Screwdrivers and hammers (for example) serve different functions. Surveys are one of a number of tools we have available for understanding the world around us. They are certainly not the only method, nor are they necessarily always the best. Surveys are particularly good for some things, but not at all good for others.

Sometimes we as survey methodologists fall into the trap of thinking that surveys are the only possible tool. We also get caught up in building the perfect tool, and forget that the tools are not a goal in themselves, but are used for a purpose. Our job is to make better tools, to give the users a range of tools to use in their work, and to guide them in which tool is best for which job. The ultimate goal is to use the tools to make sense of the world around us and, in doing so, help to make a better world.

My view is that we should welcome – rather than fear or oppose – these new developments. They expand the range of tools available to us to understand society. They force is to rethink our assumptions and take a closer look at the methods we're currently using. To continue with the toolkit metaphor, they represent shiny new tools that we can add to our toolkit, enabling us to do things that we couldn't do as well before. But we shouldn't throw away our old tools – and our knowledge of which tools to use for what purpose, and how best to use the tools, remains fundamental. Powerful tools need trained professionals.

Finally, at the risk of sounding arrogant, let me offer some advice for those relatively new to the field. The talk of the obsolescence of surveys may make you wonder what you're getting into. I believe this continues to be a vibrant, rewarding, and fascinating field to work in. There are lots of opportunities to innovate, to develop new methods, and to contribute to our understanding of societies that are rapidly changing. I believe that the training that you have (or are getting) will remain valuable, no matter which direction we take. This will be true even if there are dramatic changes to the way we conduct surveys or measure society. Here are some specific thoughts:

1. Be open to new ideas, but don't be too quick to reject "old" methods. A lot of the theories and methods that have evolved over the decades still apply. One example is the reinvigoration of mail surveys, thought to be near death after the growth of Internet surveys. But there is still clearly a place for mail – at least until the postal service disappears.
2. Look towards the future, but don't ignore the past. It's helpful to remember that the "total survey error paradigm" dates back to the 1940s (Deming, 1944). It's instructive to look back as well as looking forward. Read the old literature – a lot of it is still surprisingly relevant today.
3. Get as much technical and statistical knowledge as you can. Modeling and data analytic skills will always be valuable, I believe. These skills will never be wasted.
4. But don't underestimate the value of good theory. A lot of the issues we face today are crying out for theoretical development – both social science and statistical theories.

For those who are not quite so new, survey research is a dynamic field. Our skills and experience are still relevant today, but are not static. We constantly need to hone our skills, update our knowledge, and expose ourselves to new developments in other disciplines and fields of research and application. This is what makes survey research exciting. While based on strong foundations and a long history of success, survey research is a vibrant, dynamic, and forward-looking field. Long live surveys!

Acknowledgments

This paper was presented as a keynote address at the ESRA conference in Ljubljana in July 2013. I thank Patrick Sturgis and the organizers for giving me the opportunity to present these ideas and encouraging submission of the paper. I am also grateful to Eleanor Singer, Jaak Billiet, and the reviewers for their helpful comments on an earlier draft.

References

AAPOR. (2010). *Report of the AAPOR task force on online panels*. Deerfield, IL: American Association for Public Opinion Research.

AAPOR. (2013). *Report of the AAPOR task force on non-probability sampling*. Deerfield, IL: American Association for Public Opinion Research.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine 16:07*. Retrieved 05/11/2013, from http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Armoogum, J., Roux, S., & Pham, T. (2013). *Total nonresponse of a GPS-based travel surveys.* (Paper presented at the conference on New Techniques and Technologies for Statistics, Brussels, March)

Asendorpf, J., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J., Fiedler, K., … Wicherts, J. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*(2), 108-119.

Baker, R., Blumberg, S., Brick, J., Couper, M., Courtright, M., Dennis, J., … Zahs, D. (2010). Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly*, *74*(4), 711-781.

Baker, R., Brick, J., Bates, N., Battaglia, M., Couper, M., Dever, J., … Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling (with discussion and rejoinder). *Journal of Survey Statistics and Methodology*. doi: 10.1093/jssam/smt008

Beniger, J. (1998). Survey and market research confront their futures on the World Wide Web. *Public Opinion Quarterly*, *62*, 442-452.

Berinsky, A., Huber, G., & Lenz, G. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351-368.

Biler, S., Šenk, P., & Winklerová, L. (2013). *Willingness of individuals to participate in a travel behavior survey using GPS devices.* (Paper presented at the conference on New Techniques and Technologies for Statistics, Brussels, March)

Boyd, D., & Ellison, M. (2008). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, *13*, 210-230.

Butler, D. (2013, February 13). When Google got flu wrong: US outbreak foxes a leading Web-based method for tracking seasonal flu. *Nature*. Retrieved 05/11/2013, from http://www.nature.com/news/when-google-got-flu-wrong-1.12413

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, *88*, 2-9.

Cook, S., Conrad, C., Fowlkes, A., & Mohebbi, M. (2011). Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS ONE*, *6*(8), e23610. doi: 10.1371/journal.pone.0023610

Couper, M. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, *64*(4), 464-494.

Couper, M., & Groves, R. (1992). Interviewer reactions to alternative hardware for computer-assisted personal interviewing. *Journal of Official Statistics*, *8*(2), 201-210.

Couper, M., & Wagner, J. (2011). *Using paradata and responsive design to manage survey nonresponse.* (Invited paper presented at the 58th Congress of the International Statistical Institute, Dublin, Ireland, August)

Daas, P., Roos, M., van de Ven, M., & Neroni, J. (2012). *Twitter as a potential data source for statistics. (Discussion paper 201221)*. Den Haag/Heerlen, The Netherlands: Statistics Netherlands.

Deming, W. (1944). On errors in surveys. *American Sociological Review*, *9*(4), 359-369.

Downes-Le Guin, T., Baker, R., Mechling, J., & Ruyle, E. (2012). Myths and realities of respondent engagement in online surveys. *International Journal of Market Research*, *54*(4), 613-633.

Dugas, A., Jalalpour, M., Gel, Y., Levin, S., Torcaso, F., Igusa, T., & Rothman, R. (2013). Influenza forecasting with Google flu trends. *PLoS ONE*, *8*(2), e56176.

Findlay, K., & Alberts, K. (2011). *Gamification: What it is and what it's not.* (Paper presented at the ESOMAR Congress, Amsterdam, September)

Frijters, P., Johnston, D., Lordan, G., & Shields, M. (2013). Exploring the relationship between macroeconomic conditions and problem drinking as captured by Google searches in the US. *Social Science and Medicine*, *84*, 61-68.

Ghosh, D., & Guha, R. (2013). What are we 'tweeting' about obesity? mapping tweets with topic modeling and geographic information system. *Cartography and Geographic Information Science*, *40*(2), 90-102.

Groves, R. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646-675.
Groves, R. (2011). Three eras of survey research. *Public Opinion Quarterly*, *75*(5), 861-871.
Groves, R., & Heeringa, S. (2006). Responsive design for household surveys: Tools for actively controlling survey nonresponse and costs. *Journal of the Royal Statistical Society: Series A*, *169*(3), 439-457.
Groves, R., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, *72*(2), 167-189.
Hampton, K., Goulet, L., Rainie, L., & Purcell, K. (2011). *Social networking sites and our lives: How people's trust, personal relationships, and civic and political involvement are connected to their use of social networking sites and other technologies.* Washington, DC: Pew Internet and American Life Project.
Hirschhorn, J., Lohmueller, K., Byrne, E., & Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genetics in Medicine*, *4*(2), 45-61.
Ioannidis, J. (2005). Why most published research findings are false. *PLOS Medicine*, *2*(8), 696-701.
Keeter, S., & Christian, L. (2012). *A comparison of results from surveys by the Pew Research Center and Google Consumer Surveys.* Washington, DC: Pew Research Center for the People & The Press.
Kim, A., Hansen, H., & Murphy, J. (2012). *Methodological considerations in analyzing Twitter data.* (Paper presented at the annual meeting of the American Association for Public Opinion Research, Orlando, FL, May)
Kim, A., Murphy, J., Richards, A., Hansen, H., Powell, R., & Haney, C. (in press). Can tweets replace polls? A U.S. health care reform case study. In C. Hill & E. Dean (Eds.), *Social media, sociality, and survey research* (p. Chapter 3). New York: Wiley.
Kreuter, F. (Ed.). (2013). *Improving surveys with paradata: Analytic uses of process information*. New York: Wiley.
Lansdall-Welfare, T., Lampos, V., & Cristianini, N. (2012). Nowcasting the mood of the nation. *Significance*, *9*, 26-28.
Leinweber, D. (2007). Stupid data miner tricks: Overfitting the S&P 500. *The Journal of Investing*, *16*(1), 15-22.
Link, M. (2013). *Emerging technologies: New opportunities, old challenges.* (Keynote presentation at FedCASIC, Washington, DC, March.)
Little, R. (2012). Calibrated Bayes: An alternative inferential paradigm for official statistics (with discussion and rejoinder). *Journal of Official Statistics*, *28*, 309-372.
Lusinchi, D. (2012). 'President' Landon and the 1936 Literary Digest poll: Were automobile and telephone owners to blame? *Social Science History*, *36*(1), 23-54.
Lyberg, L. (1985). Plans for computer assisted data collection at Statistics Sweden. In International Statistical Institute (Ed.), *Proceedings of the 45th session, book iii* (p. Topic 18.2). The Hague/Leidschenveen: P. Marin.
McDonald, P., Mohebbi, M., & Slatkin, B. (2012). *Comparing Google Consumer Surveys to existing probability and non-probability based Internet surveys. (White paper).* Mountain View, CA: Google Inc.
Murphy, J., Dean, E., Hill, C., & Richards, A. (2011). *Social media, new technologies and the future of health survey research.* (Paper presented at the Conference on Health Care Research Methods, Peachtree City, GA, April)
Olson, K., & Wagner, J. (2013). *A field experiment using GPS devices to measure interviewer travel behavior.* (Paper presented at the annual conference of the American Association for Public Opinion Research, Boston, May)
O'Reilly, J., Hubbard, M., Lessler, J., Biemer, P., & Turner, C. (1994). Audio and video computer assisted self-interviewing: Preliminary tests of new technologies for data collection. *Journal of Official Statistics*, *10*(2), 197-214.
Palmer, J., Espenshade, T., Bartumeus, F., & Chung, C. (2013). New approaches to human mobility: Using mobile phones for demographic research. *Demography*, *50*(3), 1105-1128.
Paul, M., & Dredze, M. (2011). You are what you tweet: Analyzing Twitter for public health. In *Fifth international aaai conference on weblogs and social media barcelona, july 17-21.* (p. 265-272). Palo Alto, CA: AAAI Publications.
Prewitt, K. (2013). The 2012 Morris Hansen lecture: Thank you Morris, et al., for Westat, et al. *Journal of Official Statistics*, *29*(2), 223-231.
Puleston, J. (2011). Improving online surveys. *International Journal of Market Research*, *53*(4), 557-560.
Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641.
Sagan, C. (1995). *The demon-haunted world*. New York: Ballantine Books.
Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, *41*(5), 885-899.
Schouten, B., Calinescu, M., & Luiten, A. (2011). *Optimizing quality of response through adaptive survey designs. (Discussion paper 201118)*. The Hague/Heerlen: Statistics Netherlands.
Silver, N. (2012). *The signal and the noise: Why most predictions fail – but some don't*. New York: Penguin.
Singer, N. (2013, June 20). Ways to make your online tracks harder to follow. *New York Times, Special Supplement on Big Data*, F4.
Squire, P. (1988). Why the 1936 Literary Digest poll failed. *Public Opinion Quarterly*, *52*(1), 125-133.
Tourangeau, R. (2010). Incentives, falling response rates, and the respondent-researcher relationship. In L. Aday & M. Cynamon (Eds.), *Proceedings of the ninth conference on health survey research methods* (p. 244-253). Hyattsville, MD: National Center for Health Statistics.
Tourangeau, R., Conrad, F., & Couper, M. (2013). *The science of Web surveys*. New York: Oxford University Press.
Tress, F., Winkler, T., & Schmidt, S. (2012). *Surveytainment 2.0: Why investing 10 more minutes in constructing your questionnaire is worth considering.* (Paper presented at the General Online Research conference, Mannheim, March)
Vonk, T., Willems, P., & van Ossenbruggen, R. (2006). *Uitkomsten Nederlands Onlinepanel Vergelijkingsonderzoek (NOPVO).* Amsterdam: Markt Onderzoek Associatie. Retrieved 05/11/2013, from http://www.moaweb.nl/nopvo
Wagner, J. (2012). A comparison of alternative indicators for the risk of nonresponse bias. *Public Opinion Quarterly*, *76*(3), 555-575.
Wilson, R., Gosling, S., & Graham, L. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science*, *7*(3), 203-220.
Yong, E. (2012). Bad copy. *Nature*, *485*(7398), 298-300.