Combining smart and traditional survey methods: Mode effects and other data integration considerations

# Introduction

Over the past two decade, sensor arrays and machine intelligence have moved from the exclusive domain of technophiles to become so mundane that we often take them for granted. Most people have a smartphone, and more people than ever report that they feel comfortable interacting with them (F. Keusch, Wenz, and Conrad 2022; Couper et al. 2018). The sensors contained in a typical smartphone, such as cameras, accelerometers, GPS receivers, ambient light sensors, or gyroscopes, have become embedded in users’ everyday life tasks with the goal of making things easier, faster, and more accurate (Khan et al. First 2013). Users have become accustomed to the ways in which these devices can improve their experience.

In the same two decades, response rates to a broad range of long-running surveys have declined, requiring institutions such as National Statistical Institutes (NSIs) to expend more resources to achieve comparable sample sizes (Luiten, Hox, and de Leeuw 2020; Stedman et al. 2019). The causes behind the falling response rates aren’t clear, although the sheer number of requests for participation and increase in surveyors from the commercial space has been proposed as a factor (Dillman 2015). In the past, surveys following up with non-responders have suggested issues of salience [“wasn’t relevant”], burden [“no time”], and interest [“too boring”] (Couper et al. 2007; Tait, Reynolds, and Gutstein 1995; Singer and Couper 2017). Unsurprisingly, these same aspects are also well-represented in surveys of what respondents find bothersome in surveys (Mayer 2019; Husebø et al. 2018; Johnston 2014). The places where surveys fail to perform, such as in asking repetitive questions, requiring heavy time investment, and precise and accurate measurement of things like time or space, are exactly the places in which sensors and algorithms shine. This fortuitous overlap has not gone unnoticed by survey researchers, and the last decade has been marked by an increase in “smart surveys” seeking to augment existing methodology by using the tools readily at hand within smartphones (Couper et al. 2018; Struminskaya et al. 2020; Link et al. 2014).

Although these smart surveys can be deployed in isolation, researchers whose current surveys might make use of some of the theoretical benefits are interested in integrating results from smart surveys with historical data sources and ongoing, established surveys. This presents an unfortunate conundrum, as the format of data gathered by sensors are often very different from data acquired via survey questions, and can require considerable cleaning and processing before it can even be directly compared (McCool, Schouten, and Lugtig 2021; Kaplan, Kopp, and Phipps 2020; Harding et al. 2021; Florian Keusch et al. 2023). Incomplete coverage of smartphones, combined with the potential for a differential self-selection bias between smart surveys and traditional surveys complicate the matter further (Stone et al. 2023; Wenz and Keusch 2023). At the moment, there exists no comprehensive methodology proposing steps for the integration of smart survey data, despite both researcher interest impending necessity.

While the usage of smartphone-acquired sensor data is certainly a new challenge, the field of survey methodology has contended with similar issues in the past. Mixed-mode design, in which a survey is delivered across multiple platforms (e.g. via telephone and face-to-face), has been used for decades to improve low response rates, and adjust for issues of selection and coverage (Schouten et al. 2021b; Klausch 2014; E. de Leeuw and Hox 2008). Lessons learned on mode effect estimation and data integration of other disparate modes can provide a framework for smart surveys.

Finally, The Total Survey Error (TSE) framework offers an attractive entry point for discriminating between different sources of error that may arise between the differing modes of data collection (Biemer and Lyberg 2003). TSE provides a taxonomy of errors that may be introduced at each step of the survey process, which can then be translated into existing methodology to both estimate and account for systematic differences. Usage of this framework carries with it the additional benefit that researchers have expanded the framework over time to consider various situations that may be good approximations of smart data, including big data (Amaya, Biemer, and Kinyon 2020), found data (Biemer and Amaya 2020), and web-tracking data (Bosch and Revilla 2022). Drawing from these varied approaches provides theoretical backing for this new area of research.

This literature review aims to accomplish the following:

1. Identify, classify, and quantify sources of error that may pose risks for the integration of smart surveys with traditional survey methods
2. Establish patterns of similarity between smart/traditional survey integration and previous research on mixed mode surveys

In Section 2, we describe and outline different examples of smart surveys that provide concrete examples for the sections that follow. In Section 3, we briefly describe and review the literature on Total Survey Error to provide the necessary vocabulary for following sections. In Section 4, we present relevant literature on mixed mode survey methodology and its relationship to the question at hand. In Section 5, we present the literature describing initial findings on mode effects in smart surveys. In Section 6, we present results on estimation methodology and data integration. Finally, in Section 7, we synthesize the findings from the literature, provide recommendations, and suggest experimental methods for closing the gaps in existing literature.

# Smart surveys

## Smartphones and apps

Well before apps gained their current level of prevalence, researchers were investigating the usage of smartphones and other mobile devices independently of their capacity to provide complementary external data to the survey. Couper et. al (2017) offers a comprehensive review of the literature on web surveys completed on mobile devices. Important considerations included differences in coverage, non-response, break-offs, and how best to design web surveys to accommodate the new device (Toepoel and Lugtig 2015; Pearce and Rice 2013; Peterson et al. 2017).

The primary difference between completing a web survey on a smartphone and using an app on a smartphone is the length of time a person will need to interact with the device. Aspects such as coverage and differential non-response remain pertinent and to smart surveys. Although smartphone penetration has increased in the United States and Europe, the differences between who has them and who does not have remained.

### Coverage

The original concern of smartphone coverage was that smartphone coverage may have been limited by key demographics, but it may now be the case that access to the web on non-smartphone devices may be limited by key demographics.

Online surveys are done on many different devices. Mixed-device surveys. Online surveys not yet adapted to small screens. Variability in display of questions. No long matrix questions. Measurement differences can be minimal. (Toepoel and Lugtig 2015)

## Levels of smartness

The difference between a web survey conducted on a smartphone and a smart survey, while conceptually easy to distinguish, is sometimes fuzzy. In this way, it may be useful to describe different levels of ‘smartness’ that a survey may have. A survey that a user accesses by typing in a website to the browser in their phone, then fills in questions in the same way as they would if they were answering in any other mode, would not be considered a smart survey. On the other hand, the addition of search bar for an input field for, say, the food you last ate could be considered a smart feature as it makes use of a smartphone’s capacity to serve up search results to reduce a user’s total effort in typing out a full phrase. This is a low level of smartness, but if it is compared to an entry blank on a pen and paper survey, it is easy to identify that it is present.

A high level of smartness for a similar question might involve scanning the barcode or taking a picture of the food you ate in order to answer a similar question. Although the gradation is not clear-cut, a high level of smartness tends to involve device sensors as its smart features because these offer an extended set of tools for meeting the goals of smart surveys: reducing burden and measuring concepts that respondents are unlikely to know or cannot measure. Schouten et. al (2021a) list a number of smart features that smart surveys may have: device intelligence, internal sensors, external sensors, access to public online data, access to personal online data, or linkage consent. Often, fully-developed smart surveys will employ combinations of many of these at once.

Following are three examples of smart survey types that are currently in use among NSIs. These surveys share a common history as complicated pen-and-paper diaries that often required interviewer assistance. Their high burden and the presence of questions that are difficult to measure or recall have made them ideal targets for novel methodologies over the years, which allows investigation into the impact of using differing combinations of smart features.

## Mobility

Surveys looking to measure people’s travel behavior identified shortcomings from the beginning (Clarke, Dix, and Jones 1981). The goal of these studies is to reliably measure travel behavior for a sample within a given geographic area, including aspects of the travel such as mode of transportation, precise start and stop times for each trip, and addresses for visited places, which is accomplished by asking respondents to record this information in diaries spanning varied lengths of time depending on the study (Axhausen 1995). At the Second International Conference for New Survey Methods in Transport, differences in reporting were identified between days incorporating interviewer assistance and not, and between recorded behavior and road sensors (Ampt, Richardson, and Brög 1985). Ashley et. al. (2009) discuss trip underreporting across multiple studies that rely on varied modes for comparison against the self-administered travel diary. Although the particulars differ, they share the commonality that the data from self-administered travel diaries consistently shows fewer trips than other sources.

The mobility survey represents the first of the included surveys to incorporate smart features. Early in the 90s, researchers began to make use of standalone GPS receivers for the purposes of recording all trips (Bricka et al. 2009; Yalamanchili et al. 1999; Sarasua and Meyer 1996). This worked quite well at capturing user locations, but because it was an external sensor, linkage with data from the respondent was not simple, and methods varied across studies. Although some proponents posed the GPS logger as a complete solution that would eliminate the need for respondent involvement altogether (Wolf, Guensler, and Bachman 2001), the capacity for accurately determining trip purpose, transportation mode, and the identification of missing data has yet to prove itself as accurate as user input (Nguyen et al. 2020; Sadeghian, Håkansson, and Zhao 2021; Bähr et al. 2020; Gong et al. 2014).

At the same point in time, the increase in Internet penetration and web familiarity led other researchers to experiment with bringing travel diaries online (Adler, Rimmer, and Carpenter 2002; Arentze et al. 2001). This allowed for the introduction of different smart features, including machine intelligence that added checks to the data entry stages that prevented impossible or unlikely entries, and linkage to personal data to decrease respondent burden (Hoogendoorn-Lanser, Schaap, and OldeKalter 2015). Much like with the GPS loggers, the data collected from the enhanced web-based travel diaries was often quite different.

In the early 2010s, smartphones began to come with embedded GPS technology and other sensors that made it feasible for them to record user locations, and researchers began to develop smart surveys for mobility behavior that made use of these features (Cottrill et al. 2013; Nitsche et al. 2014; Greaves et al. 2015; Berger and Platzer 2015). Here, too, the specific smart features differed per app: some made use of additional device sensors, fusing the GPS records with accelerometer data (Prelipcean, Gidófalvi, and Susilo 2018), and some integrated the machine-based check mechanisms with user feedback (Greaves et al. 2015). Soon, recommendations began to emerge for how best to make use of all possible smart features in order to improve data quality and reduce user burden (Harding et al. 2021).

As the travel diary became increasingly smart, it introduced new avenues that could account for previous sources of error, as well as new avenues for error to occur. While GPS coordinates could help to reduce recall error for respondents, the sensor could also fail in a number of ways that pen-and-paper studies were unlikely to fail. Determining the reasons for different outcomes between surveys with and without smart features requires considering each of these levels independently.

## Expenditure

Expenditure data, often gathered in the form of recall or diary studies, is important to researchers and policy-makers alike (Browning, Crossley, and Weber 2003). Declining response rates and data that don’t align well with aggregate measures have become a problem for researchers (Crossley and Winter 2014). There have been multiple efforts to improve the quality of the data generated, but recent work suggests that the intensive burden of having to report all expenditures by writing down amounts and details is a hindrance to both nonresponse and measurement quality (Wenz 2023). Issues of diary fatigue, where reported expenditure declines over the measurement period, are quite common (Brzozowski, Crossley, and Winter 2017).

Respondents are generally quite good at being aware of when and how they have spent money, and the task is rarely complex. Thus, of the three benefits that smart surveys offer, expenditure research can be said only to benefit from a reduction in burden. This can be done either by offloading laborious tasks onto the available sensors, by taking pictures of the receipts to automatically fill in line items (Wenz 2023; Jäckle et al. 2019), or by using geolocation to offer reminders when people are in areas where they are likely to make purchases . As with the mobility case, previous efforts to decrease the burden have involved moving the data collection online, incorporating decision rules to attempt to prevent motivated misreporting (Eckman 2022).

* Riegler 2015: cawi, cati, papi mode effect
* French et al. 2008: annotated receipts)

Talk about scanning only vs scanning plus diary vs smartphone diary only

## Time use

Text

# Total Survey Error

Total Survey Error (TSE) is a paradigm in which the varied ways that error can permeate through a survey can be described, and provides a basis for their joint and independent evaluation for contribution to the overall quality of the survey estimates (Groves and Lyberg 2010; Biemer 2010). Total survey error, conceptually, describes the difference between a parameter as it might be measured within a population, and the estimate of the same parameter as it might be measured by a survey (Biemer and Lyberg 2003, 36). If the objective is to compare a smart survey against its non-smart counterpart, we are ultimately interested in the comparison of each of these against the population. While adaptations of the TSE framework have been proposed for big data, found data, and metered data (Amaya, Biemer, and Kinyon 2020; Biemer and Amaya 2020; Bosch and Revilla 2022), none have been proposed for smart surveys. This article therefore relates the scheme as presented by Biemer and Lyberg (2003) to the case studies at hand. This version offers sufficient flexibility to categorize and demonstrate the differences in potential error sources between smart surveys and traditional surveys. Figure 1 is a graphical overview of the categorization levels: within total survey error, we distinguish between sampling error, caused by the process by which the sample is drawn from the population, and nonsampling error, of which we distinguish five categories.

For the purposes of this article, briefly describing each of these sources of nonsampling error is sufficient to provide context for the following sections. Specification error arises when there is a mismatch between the parameters of interest for the researchers and the information that the survey will capture. Frame error, also referred to as coverage, results from the failure of the sampling frame to adequately represent the population. Nonresponse error comes from a sampled person’s failure to respond to the survey instrument, either completely, or in part. Measurement error arises when a respondent answers in a way that differs from the truth, whether intentionally or not. Finally, processing error comes from processing, coding, editing, or working with the data.

The categories of nonresponse error benefit from an additional structural layer. Nonresponse can be called unit nonresponse if the sampled person does not respond to any part of the survey, or item nonresponse for when a sampled person has some response, but it is incomplete. Importantly, for diary studies, which are intensive and longitudinal, item nonresponse may be more complex (Lynn and Lugtig 2017). While traditional item nonresponse is often conceptualized as questions left unanswered, the existence of patterns occurring over time, such as response that decreases over time or ends prematurely suggest the need for a third category of nonresponse or a classification of differential item response patterns.

In addition to the longitudinal aspects of diary survey methods, there are additional considerations specific to smart surveys making use of passive data collection. Bosch and Revilla (2022) note two important deviations for passively-collected data from actively-collected data: it is difficult to distinguish missing data from absence of behavior, and similarly difficult to categorize missing data as either item nonresponse or measurement error. In their adaptation of the TSE framework to Big Data, Amaya et al (2020) address this by assessing the concept of missing data error in place of nonresponse error, noting that the confounding can be abated when the generation mechanism of the missingness is identifiable.

The following chapter uses the TSE framework as a basis for discussion of the differences that are known to arise within each of these components between different modes of survey administration. Chapter 5 presents what is currently known specifically on the comparison between smart surveys and their non-smart counterparts.

# Mixed-mode surveys and multi-source statistics

A survey may be administered through one or more methods, including face-to-face, paper-based, telephone, or via a smartphone app. The choice of mode by which a survey is administered is known to influence the accuracy of the data collected (DeLeeuw 2018). When the same survey content is assessed by researchers by differing modes of response, the survey design is considered to be mixed-mode, as distinct from single-mode (E. D. de Leeuw, Hox, and Dillman 2015a). Each mode of administration in a mixed-mode survey will accumulate error within the non-sampling error components discussed in Chapter 3: specification error, frame error, nonresponse error, measurement error, and processing error. When differences in error exist between different modes, we speak of mode effects.

To some degree, mode effects in fact represent the desirable element of conducting mixed-mode surveys. A telephone-based survey is limited in its coverage by default to persons who possess a telephone and web-based surveys will encounter coverage errors related to Internet access, but the development of a survey design that incorporates both modes will have greater coverage of the total population, assuming that the two modes differ in their coverage error. Most researchers who employ mixed-mode designs make use of this fact in order to improve coverage and response (DeLeeuw 2018). On the other hand, within the field of mixed-mode design, mode effects not contributing to an increase total survey coverage of these are frequently seen as nuisance elements to be avoided or corrected for against some gold standard measurement (Burton and Jäckle 2020; Klausch, Hox, and Schouten 2013). This view is at least partially at odds with the goals of smart surveys, which often seek to combine the benefits of both active and passive measurements precisely because of the lack of a gold standard.

In their book Mixed-Mode Official Surveys, Schouten et al. devote a chapter to the discussion of smart devices as an emerging new mode, noting that the new types of data “challenge the comparability of response with and without” the data (2021a, 223). The task of combining data generated by smart and non-smart surveys may ultimately bear greater resemblance to combining data from different sources if the variables arising from traditional surveys and smart surveys differ in their level of aggregation or frequency, corresponding to situations 7 or 8 respectively as discussed by Waal, Delden, and Scholtus (2020). We can therefore contrast the mixed-mode paradigm with the multi-source paradigm in which the existence of differential error between data sources can provide a method by which to compensate for the disadvantages of each (De Broe et al. 2021). Although the perspectives between mixed-mode and multi-source statistics differ, the methodology for the estimation of differences between the two is very similar, and so this paper condenses literature out of both disciplines. We will assess the relevant literature on mode/source differences at each level of nonsampling error within the TSE framework.

## Specification error

There has been relatively little attention paid explicitly to the concept of specification error as it relates to mixed-mode survey design, although the importance of proper concept specification as the “backbone” of survey quality has been repeatedly emphasized (Salant and Dillman 2008; E. D. de Leeuw, Hox, and Dillman 2015b). Specification is the process by which the concepts of interest are translated into a variable that can be measured by the survey instrument, and specification error the mismatch between the two. Careful alignment of theory and questions by involving everyone in the process, along with a pretesting stage, can identify specification error (E. D. de Leeuw, Hox, and Dillman 2015b). Regardless of whether the operationalization has been sound, survey modes that don’t differ in their presentation of the question are unlikely to elicit differences here -- except perhaps longitudinally (Lynn and Lugtig 2017). In this way, the unified mode approach, in which all modes have questions phrased as similarly as possible, limits the introduction of mode specification effects (Dillman, Smyth, and Christian 2014; Dillman and Edwards 2016). The line between mode specification effect and mode measurement effect is not always clear in the data. In their chapter on Mixed-Mode Research, Hox et al. (2017) note the potential for instruments to “reflect different constructs across modes,” in the worst case scenario of mode measurement effects.

Unlike in the mixed-mode domain, the difficulties arising from mode specification effect come up regularly in multi-source literature, both because the data sources under consideration may be created independently of each other, and because the collecting instrument may limit the ways that the concept can be operationalized (Zhang 2012). Here, too, there is confounding with measurement effect, but often the presence of clear differences in the operationalization of a concept lend itself to seeing differences at the level of specification rather than measurement.

## Frame Error

## Nonresponse error

### Unit nonresponse

* Bla

### Item nonresponse

* Bla

## Measurement Error

Measurement error – can differ both in the amount of random error and systematic effect

* Dillman et al, 2014 resp in interview may interpret question differently from web survey
* Two types – mode effects that shift the response distribution (differences in estimates between modes, but not between correlations), vs change in question-answer process which can produce answers not equivalent between modes (Hox, de Leeuw, and Klausch 2017)

## Processing error

* Bla

Notes:

* Variants
  + Mode choice bias vs real measurement differences (2012
  + Sequential vs concurrent
  + Single vs multiple per resp.
* Mode effects are evaluated relative to some benchmark mode
* Diary vs recall (“on a typical day…”) as comparison?

<Small paragraph on coverage, but probably this is the other WP2 deliverable>

# Mode effects in smart surveys

TSS1 – WP2 deliverable 2.1 consumption interview vs none

* Diary vs app comparisons
  + Combined mode effect
    - (Wenz 2023): Compares data from scanned receipts, scanned receipts plus direct entry with national budget survey, uses inverse probability weighting to match the sample composition of the app to the diary (could be considered reducing mode selection effect)
      * Distribution from SR+DE aligns with LCF for median total expenditure, but receipts alone underestimate expenditure (101.30 vs 122.80 vs 70.10)
      * Percentage zero expenditures is higher in the two-week period from scanned-only
      * SR for food and groceries is not significantly different from the benchmark, but SR + DE overestimates
  + Mode measurement effect
  + Mode selection effects
    - Coverage and participation bias in budget app (Jäckle et al. 2019)

# Estimation Methodology and data integration

* Comparing mixed-mode to single mode to separate selection and measurement bias (observed variables insensitive to survey mode)
  + Vannieuwenhuyze, J., G. Loosveldt, and G. Molenberghs. 2010. A method for evaluating mode effects in mixed-mode surveys. Public Opinion Quarterly 74:1027–1045.
  + Vannieuwenhuyze, J., and G. Loosveldt (2013), Evaluating Relative Mode Effects in Mixed-Mode Surveys: Three Methods to Disentangle Selection and Measurement Effects, Sociological Methods & Research, 42, 82–104.
* Probability samples with experiments for relative differences (probability sample, assigned a data collection mode, test hypotheses about differences)
  + Van den Brakel, J.A. (2013). Design based analysis of factorial designs embedded in probability samples. Survey Methodology, 39, pp. 323-349.
  + Van den Brakel, J.A. (2008). Design-based analysis of embedded experiments with applications in the Dutch Labour Force Survey. Journal of the Royal Statistical Society, Series A, 171, pp. 581-613.
  + Van den Brakel, J.A., and R. Renssen (2005). Analysis of Experiments Embedded in Complex Sample Designs. Survey Methodology, 31, pp. 23-40.
* Interview/re-interview to disentangle selection and measurement bias
  + Biemer, P. (2001). Nonresponse bias and measurement bias in a comparison of face to face and telephone interviewing. Journal of Official Statistics 17:295-320.
* Weighting/regression-based inference for generic selection effects in non-experimental settings
  + Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the effect of data collection mode on measurement. International Statistical Review, 78(1), 3-20.
* Estimate classification error rates Measurement error categorical variables with overlapping datasets - Latent class analysis
  + Pavlopoulos and Vermunt (2015)
  + Boeschoten et al. (2016), and Oberski (2017).
* Propensity score matching resp. using covariates, difference in survey estimates is measurement effect, assuming covs. explain selection
  + Lugtig, P., Lensvelt-Mulders, G. J., Frerichs, R., & Greven, A. (2011). Estimating nonresponse bias and mode effects in a mixed-mode survey. International Journal of Market Research, 53(5), 669-686.
  + Capacci, S., Mazzocchi, M., & Brasini, S. (2018). Estimation of unobservable selection effects in on-line surveys through propensity score matching: An application to public acceptance of healthy eating policies. PloS one, 13(4), e0196020.
* Split sample – random assignment to a mode, then follow-up repeated measurement with one mode
  + Schouten, B., van den Brakel, J., Buelens, B., van der Laan, J., & Klausch, T. (2013). Disentangling mode-specific selection and measurement bias in social surveys. Social Science Research, 42(6), 1555-1570.
  + Klausch, T., J. Hox, and B. Schouten (2015), “Selection Error in Single- and Mixed Mode Surveys of the Dutch General Population,” Journal of the Royal Statistical Society: Series A (Statistics in Society), 178(4), 945–961.
* Mode preference as covariate for explaining selection effects
  + Vandenplas, C., Loosveldt, G., & Vannieuwenhuyze, J. T. (2016). Assessing the use of mode preference as a covariate for the estimation of measurement effects between modes: a sequential mixed mode experiment. Methods, data, analyses: a journal for quantitative methods and survey methodology, 10(2), 119-142.
* Longitudinal random allocation in a panel – quasi simplex model
  + Cernat, A. (2015). The impact of mixing modes on reliability in longitudinal studies. Sociological Methods & Research, 44(3), 427-457.
  + Cernat, A., Couper, M. P., & Ofstedal, M. B. (2016). Estimation of mode effects in the health and retirement study using measurement models. Journal of survey statistics and methodology, 4(4), 501-524.
* Classification error (generic, maybe other error types) with audit data
  + Delden, A. van, S. Scholtus, and J. Burger (2016), Accuracy of Mixed-Source Statistics as Affected by Classification Errors. Journal of Official Statistics
* Estimate measurement error Overlapping units and variables – SEM
  + S. Scholtus, B.F.M. Bakker and A. van Delden (2015), Modelling Measurement Error to Estimate Bias in Administrative and Survey Variables. Discussion Paper 201517, Statistics Netherlands, The Hague. Available at: www.cbs.nl.
  + B.F.M. Bakker (2012), Estimating the Validity of Administrative Variables. Statistica Neerlandica 66, pp. 8–17.
  + (Boeschoten, Oberski and De Waal 2016).
* MTMM
  + A.C. Scherpenzeel and W.E. Saris (1997), The Validity and Reliability of Survey Questions: A Meta-Analysis of MTMM Studies. Sociological Methods and Research 25, pp. 341–383.
* Prediction of a “true” values of numeric variables of interest, given multiple data sources covering only subsets of the population – latent variable
  + Guarnere U., Variale R. Estimation from contaminated multi-source data based on latent class models. Statistical Journal of the IAOS, vol. Preprint, no. Preprint, pp. 1-8, 2015
* Linking two data sources at individual level – estimating quality of probabilistic linkage
  + Chipperfield, J.O. and Chambers, R. (2015). Using the bootstrap to account for linkage errors when analysing probabilistically linked categorical data. Journal of Official statistics 31, 397–414.
  + Chambers, R., J. Chipperfield, W. Davis and M. Kovacevic (2009). Inference based on estimating equations and probability-linked data. Centre for statistical and survey methodology, University of Wollongong, Working paper 18-09. Available at <http://ro.uow.edu.au/cssmwp/38>.
* Variance estimation in Combining several sources of aggregate data - Macro-integration by inequality restriction/balancing
  + Boonstra, H.J., C.J. de Blois and G.-J.M. Linders (2011), Macro-Integration with Inequality Constraints: An Application to the Integration of Transport and Trade Statistics. Statistica Neerlandica 64, pp. 407-431.
  + Knottnerus, P. (2016), On New Variance Approximations for Linear Models with Inequality Constraints. Statistica Neerlandica 70, pp. 26-46.
  + Chen: A Balanced System of Industry Accounts for the U.S. and Structural Distribution of Statistical Discrepancy (41 pp, URL https://tinyurl.com/mwuzbty, retrieved April 2, 2017).
* Reducing variance estimates with small area estimation, which can be used not only for geographic areas, but distinct small categories that are related
  + Rao, J.N.K. and I. Molina (2015), Small Area Estimation. John Wiley.
  + Boonstra, H.J., van den Brakel, J., Buelens, B., Krieg, S. and M. Smeets (2008). Towards small area estimation at Statistics Netherlands. Metron 66 (1), pp. 21-49.
* Reconciling high frequency and low frequency data – quadratic minimization
  + Bikker R., Daalmans J., Mushudiani N. (2011) Macro Integration. Data reconciliation. Statistical methods (201104). The Hague/Heerlen, Statistics Netherlands.
* Measurement error assuming “Overlapping Numerical Variables with a Benchmark”
  + (Fasulo and Solari 2017)
* \*\*Mixed-mode measurements as treatments in causal modelling framework counterfactual.. potential outcomes through regression, then overall estimates are combos of real answers and predicted answers
  + Suzer-Gurtekin, Z. T., Heeringa, S., & Vaillant, R. (2012). Investigating the Bias of Alternative Statistical Inference Methods in Sequential Mixed-Mode Surveys. Proceedings of the JSM, Section on Survey Research Methods, 4711-2.
* \*\*Multiple imputation of non-observed answers
  + Kolenikov, S., & Kennedy, C. (2014). Evaluating three approaches to statistically adjust for mode effects. Journal of survey statistics and methodology, 2(2), 126-158.
  + Park, S., Kim, J. K., & Park, S. (2016). An imputation approach for handling mixed-mode surveys. The Annals of Applied Statistics, 10(2), 1063-1085.
* \*\*Reweighting of survey response – calibration correction of survey response to fixed-mode distributions
  + Buelens, B., & van den Brakel, J. A. (2015). Measurement error calibration in mixed-mode sample surveys. Sociological Methods & Research, 44(3), 391-426.
  + Buelens, B., and J. van den Brakel (2017). “Comparing two inferential approaches to handling measurement error in mixed-mode surveys,” Journal of Official Statistics, 33(2), 513-531.
* \*\* Covariate adjustments to correct for measurement effects
  + Vannieuwenhuyze, J.T., Loosveldt, G., Molenberghs, G. (2014) Evaluating mode effects in mixed-mode survey data using covariate adjustment models. J. Off. Stat. 30(1), 1–21
* \*\*IRT-based Bayesian hierarchical model/latent traits
  + Mariano, L. T., & Elliott, M. N. (2017). An Item Response Theory Approach to Estimating Survey Mode Effects: Analysis of Data from a Randomized Mode Experiment. Journal of Survey Statistics and Methodology, 5(2), 233-253.
* \*\*Re-interview after observational study and adjustment to benchmark mode
  + Klausch, T., Schouten, B., Buelens, B., & Van Den Brakel, J. (2017). Adjusting measurement bias in sequential mixed-mode surveys using re-interview data. Journal of Survey Statistics and Methodology, 5(4), 409-432.

# Synthesis

TOPICS

Specific areas

Maybe it’s like differences between other modes

Because the data are so different, maybe it’s more like integrating other big data sources

Missing data

Adler, Thomas, Leslie Rimmer, and David Carpenter. 2002. “Use of Internet-Based Household Travel Diary Survey Instrument.” *Transportation Research Record* 1804 (1): 134–43.

Amaya, A., P. P. Biemer, and D. Kinyon. 2020. “Total Error in a Big Data World: Adapting the TSE Framework to Big Data.” *Journal of Survey Statistics And*. https://academic.oup.com/jssam/article-abstract/8/1/89/5728725.

Ampt, Elizabeth S., Anthony Joseph Richardson, and Werner Brög. 1985. *New Survey Methods in Transport: Proceedings of 2nd International Conference, Hungerford Hill, Australia, 12-16 September 1983*. VSP.

Arentze, Theo, Martin Dijst, Elenna Dugundji, Chang-Hyeon Joh, Loek Kapoen, Stephan Krygsman, Kees Maat, and Harry Timmermans. 2001. “New Activity Diary Format: Design and Limited Empirical Evidence.” *Transportation Research Record* 1768 (1): 79–88.

Ashley, D., T. Richardson, and D. Young. 2009. “Recent Information on the Under-Reporting of Trips in Household Travel Surveys.” In *Australasian Transport Research Forum (ATRF), 32nd, 2009, Auckland, New Zealand*. Vol. 32. trid.trb.org. https://trid.trb.org/view/1149551.

Axhausen, Kay W. 1995. “Travel Diaries: An Annotated Catalog.” Institut für Strassenbau und Verkehrsplanung. https://rosap.ntl.bts.gov/view/dot/13806.

Bähr, Sebastian, Georg-Christoph Haas, Florian Keusch, Frauke Kreuter, and Mark Trappmann. 2020. “Missing Data and Other Measurement Quality Issues in Mobile Geolocation Sensor Data.” *Social Science Computer Review*, August, 0894439320944118.

Berger, Martin, and Mario Platzer. 2015. “Field Evaluation of the Smartphone-Based Travel Behaviour Data Collection App ‘SmartMo.’” *Transportation Research Procedia* 11 (January): 263–79.

Biemer, Paul P. 2010. “Total Survey Error: Design, Implementation, and Evaluation.” *Public Opinion Quarterly* 74 (5): 817–48.

Biemer, Paul P., and Ashley Amaya. 2020. “Total Error Frameworks for Found Data.” *Big Data Meets Survey Science*. Wiley. https://doi.org/10.1002/9781118976357.ch4.

Biemer, Paul P., and Lars E. Lyberg. 2003. *Introduction to Survey Quality*. John Wiley & Sons.

Bosch, Oriol J., and Melanie Revilla. 2022. “When Survey Science Met Web Tracking: Presenting an Error Framework for Metered Data.” *Journal of the Royal Statistical Society. Series A,*  185 (Suppl 2): S408–36.

Bricka, Stacey, Johanna Zmud, Jean Wolf, and Joel Freedman. 2009. “Household Travel Surveys with GPS: An Experiment.” *Transportation Research Record* 2105 (1): 51–56.

Browning, Martin, Thomas F. Crossley, and Guglielmo Weber. 2003. “Asking Consumption Questions in General Purpose Surveys.” *Economic Journal (London, England)* 113 (491): F540–67.

Brzozowski, Matthew, Thomas F. Crossley, and Joachim K. Winter. 2017. “A Comparison of Recall and Diary Food Expenditure Data.” *Food Policy* 72 (October): 53–61.

Burton, Jonathan, and Annette Jäckle. 2020. “Mode Effects.” *ISER: Understanding Society Working Paper Series*, no. 2020–05. https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2020-05.pdf.

Clarke, Mike, Martin Dix, and Peter Jones. 1981. “Error and Uncertainty in Travel Surveys.” *Transportation* 10 (2): 105–26.

Cottrill, Caitlin D., Francisco Câmara Pereira, Fang Zhao, Inês Ferreira Dias, Hock Beng Lim, Moshe E. Ben-Akiva, and P. Christopher Zegras. 2013. “Future Mobility Survey: Experience in Developing a Smartphone-Based Travel Survey in Singapore.” *Transportation Research Record* 2354 (1): 59–67.

Couper, Mick P., Christopher Antoun, and Aigul Mavletova. 2017. “Mobile Web Surveys.” In *Total Survey Error in Practice*, 133–54. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Couper, Mick P., Garret Gremel, William Axinn, Heidi Guyer, James Wagner, and Brady T. West. 2018. “New Options for National Population Surveys: The Implications of Internet and Smartphone Coverage.” *Social Science Research* 73 (July): 221–35.

Couper, Mick P., Andy Peytchev, Victor J. Strecher, Kendra Rothert, and Julia Anderson. 2007. “Following up Nonrespondents to an Online Weight Management Intervention: Randomized Trial Comparing Mail versus Telephone.” *Journal of Medical Internet Research* 9 (2): e16.

Crossley, Thomas F., and Joachim K. Winter. 2014. “Asking Households about Expenditures: What Have We Learned?” In *Improving the Measurement of Consumer Expenditures*, 23–50. University of Chicago Press.

De Broe, Sofie, Peter Struijs, Piet Daas, Arnout van Delden, Joep Burger, Jan van den Brakel, Olav ten Bosch, Kees Zeelenberg, and Winfried Ypma. 2021. “Updating the Paradigm of Official Statistics: New Quality Criteria for Integrating New Data and Methods in Official Statistics.” *Statistical Journal of the IAOS* 37 (1): 343–60.

DeLeeuw, E. D. 2018. “Mixed-Mode: Past, Present, and Future.” *Survey Research Methods*. https://ojs.ub.uni-konstanz.de/srm/article/view/7402.

Dillman, Don A. 2015. “On Climbing Stairs Many Steps at a Time: The New Normal in Survey Methodology.” presented at the Faculty and Graduate Student Seminar Series, The School of Economics Washington State University, September 18. https://web.archive.org/web/20181221174204if\_/http://ses.wsu.edu:80/wp-content/uploads/2015/09/DILLMAN-talk-Sept-18-2015.pdf.

Dillman, Don A., and Michelle L. Edwards. 2016. “Designing a Mixed-Mode Survey.” In *The SAGE Handbook of Survey Methodology*, 255–68. 1 Oliver’s Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons.

Eckman, Stephanie. 2022. “Underreporting of Purchases in the US Consumer Expenditure Survey.” *Journal of Survey Statistics and Methodology* 10 (5): 1148–71.

French, Simone A., Scott T. Shimotsu, Melanie Wall, and Anne Faricy Gerlach. 2008. “Capturing the Spectrum of Household Food and Beverage Purchasing Behavior: A Review.” *Journal of the American Dietetic Association* 108 (12): 2051–58.

Gong, Lei, Takayuki Morikawa, Toshiyuki Yamamoto, and Hitomi Sato. 2014. “Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies.” *Procedia - Social and Behavioral Sciences* 138 (July): 557–65.

Greaves, Stephen, Adrian Ellison, Richard Ellison, Dean Rance, Chris Standen, Chris Rissel, and Melanie Crane. 2015. “A Web-Based Diary and Companion Smartphone App for Travel/Activity Surveys.” *Transportation Research Procedia* 11 (January): 297–310.

Groves, R. M., and L. Lyberg. 2010. “Total Survey Error: Past, Present, and Future.” *Public Opinion Quarterly* 74 (5): 849–79.

Harding, C., A. Faghih Imani, S. Srikukenthiran, and E. J. Miller. 2021. “Are We There yet? Assessing Smartphone Apps as Full-Fledged Tools for Activity-Travel Surveys.” *Transportation*. https://link.springer.com/article/10.1007/s11116-020-10135-7.

Hoogendoorn-Lanser, Sascha, Nina T. W. Schaap, and Marie-José OldeKalter. 2015. “The Netherlands Mobility Panel: An Innovative Design Approach for Web-Based Longitudinal Travel Data Collection.” *Transportation Research Procedia* 11 (January): 311–29.

Hox, Joop, Edith de Leeuw, and Thomas Klausch. 2017. “Mixed-Mode Research.” In *Total Survey Error in Practice*, 511–30. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Husebø, Anne Marie Lunde, Ingvild Margreta Morken, Kristina Sundt Eriksen, and Oda Karin Nordfonn. 2018. “The Patient Experience with Treatment and Self-Management (PETS) Questionnaire: Translation and Cultural Adaption of the Norwegian Version.” *BMC Medical Research Methodology* 18 (1): 147.

Jäckle, Annette, Jonathan Burton, Mick P. Couper, and Carli Lessof. 2019. “Participation in a Mobile App Survey to Collect Expenditure Data as Part of a Large-Scale Probability Household Panel: Coverage and Participation Rates and Biases.” *Survey Research Methods* 13 (1): 22.

Johnston, Karenpreview Author Details; Ryan. 2014. “Researching the Respondents.” *Market & Social Research* 22 (1): 39–46.

Kaplan, Robin L., Brandon Kopp, and Polly Phipps. 2020. “Contrasting Stylized Questions of Sleep with Diary Measures from the American Time Use Survey.” *Advances in Questionnaire Design, Development, Evaluation and Testing*. Wiley. https://doi.org/10.1002/9781119263685.ch27.

Keusch, F., A. Wenz, and F. Conrad. 2022. “Do You Have Your Smartphone with You? Behavioral Barriers for Measuring Everyday Activities with Smartphone Sensors.” *Computers in Human Behavior* 127. https://doi.org/10.1016/j.chb.2021.107054.

Keusch, Florian, Sebastian Bähr, Georg-Christoph Haas, Frauke Kreuter, and Mark Trappmann. 2023. “Coverage Error in Data Collection Combining Mobile Surveys With Passive Measurement Using Apps: Data From a German National Survey.” *Sociological Methods & Research* 52 (2): 841–78.

Khan, Wazir Zada, Yang Xiang, Mohammed Y. Aalsalem, and Quratulain Arshad. First 2013. “Mobile Phone Sensing Systems: A Survey.” *IEEE Communications Surveys & Tutorials* 15 (1): 402–27.

Klausch, Thomas. 2014. *Informed Design of Mixed-Mode Surveys: Evaluating Mode Effects on Measurement and Selection Error*. Utrecht University.

Klausch, Thomas, Joop J. Hox, and Barry Schouten. 2013. “Measurement Effects of Survey Mode on the Equivalence of Attitudinal Rating Scale Questions.” *Sociological Methods & Research* 42 (3): 227–63.

Leeuw, Edith D. de, Joop J. Hox, and Don A. Dillman. 2015a. “Mixed-Mode Surveys.” In *International Handbook of Survey Methodology*. Routledge.

———. 2015b. “The Cornerstones of Survey Research.” In *International Handbook of Survey Methodology*. Routledge.

Leeuw, Edith de, and Joop Hox. 2008. “Mixing Data Collection Methods: Lessons from Social Survey Research1, 2.” *Advances in Mixed Methods Research: Theories and Applications*, 138.

Link, Michael W., Joe Murphy, Michael F. Schober, Trent D. Buskirk, Jennifer Hunter Childs, and Casey Langer Tesfaye. 2014. “Mobile Technologies for Conducting, Augmenting and Potentially Replacing SurveysExecutive Summary of the AAPOR Task Force on Emerging Technologies in Public Opinion Research.” *Public Opinion Quarterly* 78 (4): 779–87.

Luiten, Annemieke, Joop Hox, and Edith de Leeuw. 2020. “Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study across Countries and Surveys.” *Journal of Official Statistics* 36 (3): 469–87.

Lynn, Peter, and Peter J. Lugtig. 2017. “Total Survey Error for Longitudinal Surveys.” In *Total Survey Error in Practice*, 279–98. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Mayer, Adam. 2019. “‘Your Survey Is Biased’: A Preliminary Investigation into Respondent Perceptions of Survey Bias.” *Survey Practice* 12 (1): 1–8.

McCool, Danielle, J. G. Schouten, and P. Lugtig. 2021. “An App-Assisted Travel Survey in Official Statistics. Possibilities and Challenges.” *Journal of Official Statistics* 37 (1): 149–70.

Nguyen, Minh Hieu, Jimmy Armoogum, Jean-Loup Madre, and Cédric Garcia. 2020. “Reviewing Trip Purpose Imputation in GPS-Based Travel Surveys.” *Journal of Traffic and Transportation Engineering (English Edition)* 7 (4): 395–412.

Nitsche, Philippe, Peter Widhalm, Simon Breuss, Norbert Brändle, and Peter Maurer. 2014. “Supporting Large-Scale Travel Surveys with Smartphones – A Practical Approach.” *Transportation Research Part C: Emerging Technologies* 43 (June): 212–21.

Pearce, Katy E., and Ronald E. Rice. 2013. “Digital Divides from Access to Activities: Comparing Mobile and Personal Computer Internet Users.” *The Journal of Communication* 63 (4): 721–44.

Peterson, Gregg, Jamie Griffin, John LaFrance, and Jiaojiao Li. 2017. “Smartphone Participation in Web Surveys.” In *Total Survey Error in Practice*, 203–33. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Prelipcean, Adrian C., Győző Gidófalvi, and Yusak O. Susilo. 2018. “MEILI: A Travel Diary Collection, Annotation and Automation System.” *Computers, Environment and Urban Systems* 70 (July): 24–34.

Riegler, Marc Plate Romana. 2015. “Integrating the Web Mode In.” cros-legacy.ec.europa.eu. 2015. https://cros-legacy.ec.europa.eu/system/files/Presentation%20S17CP2.pdf.

Sadeghian, Paria, Johan Håkansson, and Xiaoyun Zhao. 2021. “Review and Evaluation of Methods in Transport Mode Detection Based on GPS Tracking Data.” *Journal of Traffic and Transportation Engineering (English Edition)* 8 (4): 467–82.

Salant, Priscilla, and Don A. Dillman. 2008. *How to Conduct Your Own Survey*. Nashville, TN: John Wiley & Sons.

Sarasua, W., and M. Meyer. 1996. “New Technologies for Household Travel Surveys.” In *Conference Proceedings 10: Conference on Household Travel Surveys: New Concepts and Research Needs*, 170–82. books.google.com.

Schouten, Barry, Jan van den Brakel, Bart Buelens, Deirdre Giesen, Annemieke Luiten, and Vivian Meertens. 2021a. “Multi-Device Surveys.” In *Mixed-Mode Official Surveys*, 223–49. Boca Raton: Chapman and Hall/CRC.

———. 2021b. *Mixed-Mode Official Surveys: Design and Analysis*. CRC Press.

Sekula, W., M. Nelson, K. Figurska, M. Oltarzewski, R. Weisell, and L. Szponar. 2005. “Comparison between Household Budget Survey and 24-Hour Recall Data in a Nationally Representative Sample of Polish Households.” *Public Health Nutrition* 8 (4): 430–39.

Singer, Eleanor, and Mick P. Couper. 2017. “Some Methodological Uses of Responses to Open Questions and Other Verbatim Comments in Quantitative Surveys.” *Methods, Data, Analyses* 11 (2): 20.

Stedman, Richard C., Nancy A. Connelly, Thomas A. Heberlein, Daniel J. Decker, and Shorna B. Allred. 2019. “The End of the (Research) World as We Know It? Understanding and Coping with Declining Response Rates to Mail Surveys.” *Society & Natural Resources* 32 (10): 1139–54.

Stone, Arthur A., Stefan Schneider, Joshua M. Smyth, Doerte U. Junghaenel, Mick P. Couper, Cheng Wen, Marilyn Mendez, Sarah Velasco, and Sarah Goldstein. 2023. “A Population-Based Investigation of Participation Rate and Self-Selection Bias in Momentary Data Capture and Survey Studies.” *Current Psychology* , March. https://doi.org/10.1007/s12144-023-04426-2.

Struminskaya, Bella, Peter Lugtig, Florian Keusch, and Jan Karem Höhne. 2020. “Augmenting Surveys With Data From Sensors and Apps: Opportunities and Challenges.” *Social Science Computer Review*, December, 0894439320979951.

Tait, A. R., P. I. Reynolds, and H. B. Gutstein. 1995. “Factors That Influence an Anesthesiologist’s Decision to Cancel Elective Surgery for the Child with an Upper Respiratory Tract Infection.” *Journal of Clinical Anesthesia* 7 (6): 491–99.

Toepoel, Vera, and Peter Lugtig. 2015. “Online Surveys Are Mixed-Device Surveys. Issues Associated with the Use of Different (Mobile) Devices in Web Surveys.” *Methods, Data, Analyses*. https://doi.org/10.12758/MDA.2015.009.

Waal, Ton, Arnout Delden, and Sander Scholtus. 2020. “Multi‐source Statistics: Basic Situations and Methods.” *International Statistical Review = Revue Internationale de Statistique* 88 (1): 203–28.

Wenz, Alexander. 2023. “Quality of Expenditure Data Collected with a Mobile Receipt Scanning App in a Probability Household Panel.” understandingsociety.ac.uk. 2023. https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2023-02.pdf.

Wenz, Alexander, and Florian Keusch. 2023. “Increasing the Acceptance of Smartphone-Based Data Collection.” *Public Opinion Quarterly* 87 (2): 357–88.

Wolf, Jean, Randall Guensler, and William Bachman. 2001. “Elimination of the Travel Diary: Experiment to Derive Trip Purpose from Global Positioning System Travel Data.” *Transportation Research Record* 1768 (1): 125–34.

Yalamanchili, Lalit, Ram M. Pendyala, N. Prabaharan, and Pramodh Chakravarthy. 1999. “Analysis of Global Positioning System-Based Data Collection Methods for Capturing Multistop Trip-Chaining Behavior.” *Transportation Research Record* 1660 (1): 58–65.

Zhang, Li-Chun. 2012. “Topics of Statistical Theory for Register-Based Statistics and Data Integration.” *Statistica Neerlandica* 66 (1): 41–63.