Editorial: Recent Advances in Survey Methods for Collecting Food Data

Survey Methods: Insights from the Field, Volume 1, issue 3 (2023): Special issue on Food Acquisition Research and Methods

Elina T. Page | Christopher Antoun | Jeffrey Gonzalez | Linda Kantor | Florian Keusch | Lauren Miller | Alexander Wenz

How to cite this article : Page, E.T., Antoun C., Gonzalez J., Kantor L., Keusch F., Miller L. & Wenz A. (2023). Recent Advances in Survey Methods for Collecting Food Data. Survey Methods: Insights from the Field, Special issue: 'Food Acquisition Research and Methods'. Retrieved from https://surveyinsights.org/?p=19311

DOI : 10.13094/SMIF-2023-00017

Copyright : © the authors 2023. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Abstract : High-quality food data collected from consumption and expenditure surveys are essential for researchers and policy makers to better understand trends in social, economic, and human development. Over the last decade, food surveys have faced various challenges to sustain high data quality, such as declining response rates and increasing respondent burden. At the same time, new opportunities around the use of new technologies and alternative data sources have emerged that can potentially address some of these challenges. This special issue, inspired by the International Food Acquisition Research and Methods (iFARM) Workshop held at the University of Maryland, College Park, in October 2022, presents recent methodological advancements in the collection of food survey data. The four papers contained in this issue contribute to research about increasing and sustaining data quality in such surveys. They compare traditional paper- and web-based approaches vis-à-vis novel smartphone-based approaches for collecting food data, explore the measurement of food preparation and consumption through time-use diaries, and examine the feasibility of innovative technologies to reduce respondent burden in food surveys.

Introduction

Evidence-based policies that effectively address public health concerns surrounding nutrition and wellbeing require complete and reliable food data. These data are foundational to any analysis of food security, nutrition, health, and poverty and are used by governments and non-governmental organizations to monitor and understand trends in social, economic, and human development (The Inter-Agency and Expert Group on Food Security, & Agricultural and Rural Statistics, 2019). The main sources of such data are individual and household food consumption and expenditure surveys. These surveys are diverse and serve many different policy objectives. The terms used to describe these surveys – consumption, expenditure, acquisition, purchase, intake – are often used interchangeably, which can sometimes lead to confusion. Generally, research that focuses on the role of dietary patterns in health will use dietary intake surveys, and analyses that focus on the economic drivers of food choice will rely on consumer expenditure surveys. Some consumer expenditure surveys focus solely on purchases, while others also

collect data on free foods such as charitable food acquisitions and meals from social networks. For the purpose of this paper, we will use the umbrella term “food surveys”.

Some examples of surveys that collect food data

Researchers use a variety of methods to collect food data but such surveys typically involve a set of common elements: They employ food diaries – either paper or electronic – and recall-based interviews, and are administered at either the household or the individual level. Three examples are the U.S. National Household Food Acquisition and Purchase Survey (FoodAPS), the U.S. National Health and Nutrition Examination Survey (NHANES), and the International Dietary Data Expansion Project (INDDEX).

The National Household Food Acquisition and Purchase Survey (FoodAPS) was conducted in the United States from 2012 to 2013 (Kirlin & Denbaly, 2017). The survey collected nationally representative data on household food purchases and acquisitions from households participating in the Supplemental Nutrition Assistance Program (SNAP), low-income households not participating in SNAP, and higher income households. Comprehensive and detailed information was collected about foods purchased or otherwise acquired for free for consumption at home and away from home. The study was administered as a week-long paper diary survey of all household members and included two in-person interviews to obtain household characteristics such as income, composition of the household, food security, health status, and diet and nutrition knowledge.

The biannual National Health and Nutrition Examination Survey (NHANES) collects nationally-representative indicators on the health and nutrition status of adults and children in the United States and has been continuously conducted since 1999 (National Center for Health Statistics, 2023). The unique survey format includes in-person medical and dental examinations and laboratory tests and a computerized 24-hour in-person and telephone dietary interview component, known as the What We Eat in America (WWEIA) Survey (O’Brien, 2018).

The International Dietary Assessment Platform, part of the International Dietary Data Expansion Project (INDDEX), is a pioneering effort that provides food composition data repositories and standardized data collection instruments to facilitate the collection, use, and interpretation of high-quality individual food intake data from individual respondents in low and middle income countries (LMICs). The INDDEX24 mobile app uses a dietary recall instrument modeled on the WWEIA survey (Coates et al., 2017).

Regardless of the context in which they work, researchers face various challenges when fielding food surveys. These include determining the appropriate mode(s) of data collection, managing declining response rates, collecting information that is not too detailed for respondents to report accurately yet detailed enough to meet researchers’ needs, and processing survey answers in an efficient manner after data collection. In 2017, a special issue in Food Policy was devoted to describing these challenges (Zezza et al., 2017). In this editorial we delve into the particular challenge of respondent burden and explore advances in data collection methods that might help address this and other issues related to food surveys.

Respondent burden

Respondent burden represents one of the biggest challenges facing survey researchers in general (Yan et al., 2020) but those who collect food data in particular. Surveys that collect these data tend to be lengthy and require respondents to report detailed information about food items, sometimes over multiple days,

making participation and survey completion less likely. Researchers can benefit from careful consideration of this issue when designing and evaluating food surveys.

Several survey design factors present advantages and disadvantages with respect to their impacts on respondent burden. First, there is the choice between recall and diary surveys (The Inter-Agency and Expert Group on Food Security, & Agricultural and Rural Statistics, 2019). A recall survey offers a clear advantage in terms of respondent ease and convenience because it can typically be completed at one time or in short interviews over one or two days. While this approach can make participation more likely (decreasing the risk of nonresponse error), such surveys also require respondents to retrieve detailed information from memory (e.g., item description, price, ingredients, portion size, and place type) which can produce large measurement errors (Beegle et al., 2012; Brzozowski et al., 2017; U.S. Department of Labor/Bureau of Labor Statistics/Division of Consumer Expenditure Survey, 2016). On the other hand, diaries – which require less participant recall – may increase cumulative burden and the likelihood of drop-off because they require respondents to report information over an extended study period. These trade-offs remain largely unexplored for food surveys and likely vary depending on the nature of the survey and target population.

The mode of data collection may also shape respondents' experiences and willingness to participate. Self-administration via paper or electronic diaries can increase respondents' comfort with disclosing sensitive information (e.g., about stigmatized food choices). Yet interviewer-administration makes it easier for respondents with low literacy or other limitations who need to seek clarification (e.g., what gets counted as food away from home?). Similarly, web or mobile applications may reduce post-processing burden, but can be problematic for some target populations, like the elderly and respondents living in areas with poor broadband access. In addition, food surveys that ask one respondent to proxy report for other household members can yield important information (e.g., who purchases food and how it is shared). But proxy reporting can also introduce challenges, especially for reporting of foods acquired away from home if the respondent lacks adequate knowledge about how other household members obtain food (Fiedler & Yadav, 2017).

Additional research in these areas can help improve our understanding of how to minimize respondent burden while meeting other survey objectives (level of detail about food items, available time, cost, etc.). To this end, research is needed that explores the survey design factors that influence the burden of responding – including the roles of survey frequency, length, and mode – along with innovative methods and tools to reduce it. If researchers heed this call, an important question is how they can take advantage of new opportunities for gathering food data, which is the topic we turn to next.

New opportunities for food surveys

In a changing technological landscape, several new opportunities have emerged with respect to collecting social and behavioral data (Couper, 2013) that are relevant for food surveys. First, the growing interest in user-centered design (Nielsen, 1994) is prompting researchers to not only focus on their research objectives but also on the users (i.e., respondents) and their needs. This perspective aims to create surveys – including smartphone-based designs – that facilitate and support respondents in the survey response task (Antoun et al., 2018). Given the concern about survey burden, a user-centered design approach should be a top priority in the development of food diaries (Chung et al., 2019).

In addition, survey designers are increasingly mixing modes of data collection by integrating self-administered modes into interviewer-administered surveys. While this approach can accommodate the

varying preferences of respondents and reduce costs, it may also introduce mode differences related to measurement or nonresponse (Olson et al., 2021). The contribution by Yan and Machado (2023) in this issue discusses some of the tradeoffs involved in using different modes of data collection for food diaries. For example, using web and mobile based instruments greatly decreases backend processing; however, this approach is only viable for respondents with Internet access and familiarity with digital technology.

A third opportunity relates to the integration of apps and sensors into data collection (Keusch & Conrad, 2022). Using smartphone technology in food surveys, for example, passive tracking of geolocation data or taking pictures with the camera, allows for the direct measurement of where people obtain food and what food they consume in much more detail than traditional self-reports (Christian, 2012). For example, the paper by Kaderabek (2023) in this special issue describes the use of optical character recognition (OCR) for capturing expenditure data from food purchase receipts. Smartphone receipt scanning is becoming a viable alternative to equipping people with handheld scanners to record all purchases via barcodes (Dubois et al., 2022; Wenz et al., 2023). DiGrande et al.'s (2023) augmented reality technology also described in this special issue can aid respondents with the cognitively burdensome task of accurately estimating portion sizes. As these methods continue to develop, it is imperative to consider respondents' privacy concerns, which may correspond to their willingness to participate and share such data (Keusch et al., 2019).

Finally, data from different sources (including administrative records and commercial scanner data) can augment or replace survey data. For example, food items reported in FoodAPS were linked to USDA nutrient and food composition databases to estimate their nutritional quality (Page et al., 2019). These opportunities offer ways to reshape how food data is gathered and processed, though further research examining their overall effects on data quality is needed.

The contribution of this Special Issue

This special issue is an outcome of the International Food Acquisition Research and Methods (iFARM) workshop held at the University of Maryland, College Park on October 20-21, 2022. The goal of the workshop, jointly organized by the U.S. Department of Agriculture's (USDA) Economic Research Service (ERS) and the Social Data Science Center (SoDa) at the University of Maryland, was to provide a venue for researchers from different disciplines and countries to discuss advancements in methods for collecting and processing food acquisition data. The workshop featured 15 talks, three keynote addresses, and two panel sessions, and brought together survey methodologists who are thinking about the processes that generate food data and substantive experts who are thinking about how to utilize that information to answer important research questions.

Three of the papers in this special issue were presented at the iFARM workshop. They show the breadth of methodological discussions around human-centered design to reduce respondent burden in food surveys through optical character recognition, traditional paper and web approaches in comparison novel app-based approaches, and measuring food preparation and intake through time-use diary studies. One additional paper on augmented reality technology in dietary intake surveys that was submitted through the call for papers rounds out this special issue. All four papers provide unique and timely contributions addressing current methodological issues of collecting food data.

Kaderabek (2023) demonstrates the usefulness of OCR technology for capturing expenditure data from food purchase receipts in a FoodAPS pilot study. Asking respondents to provide sales receipts rather than self-reporting purchases has the potential to reduce respondent burden and allows the collection of more

detailed data in food acquisition and purchase surveys. As the study shows, open-source OCR technology in concert with regular expressions enables researchers to efficiently extract information from receipt images and store them in a format suitable for statistical analysis. The article also demonstrates some of the challenges with OCR-based receipt processing, in particular with regard to image quality and non-standardized receipt formats. Blurry or otherwise illegible images can make accurate extraction of information difficult, potentially leading to results that deviate from the true values to a large extent. In addition, receipt formats vary considerably across establishments and the type of food purchased. This variation poses difficulties for accurately identifying and extracting different receipt elements, such as item descriptions and prices. Overall, the author finds that the OCR technology performs quite well in capturing the receipt information, with the OCR-based results aligning relatively closely to manual receipt coding. Future research is warranted to test the feasibility and quality of OCR for processing sales receipts in large-scale studies.

Yan & Machado (2023) provide a comprehensive review of various food diary administration approaches, including paper, web, and app-based methods. The strengths and weaknesses of each approach are assessed using the Total Survey Error (TSE) framework, considering their potential impacts on coverage, nonresponse, measurement, and post-survey processing errors. Although newer methods that take advantage of recent technology have great potential for improving data quality and are sometimes presented as a one-size-fits-all solution, the authors make it clear that the choice of method is not straightforward. Consider for instance app-based diaries which offer several features that eliminate the need for manual data entry – such as using the phone camera to scan barcodes, capture images of food items, or photograph receipts – but have the disadvantage of being costly to program and may exclude potential respondents who are less comfortable with the technology. Or consider paper diaries, which improve coverage of the target population but increase post-survey processing (e.g., scanning data into a database, checking for completeness, editing for accuracy). In summary, the authors believe that navigating this complex choice requires thoughtful consideration about what is best for a particular study given its target population, measurement requirements, budget, and so forth. The authors also point out the gap in knowledge and understanding around how best to design food diaries – irrespective of mode – and call for future research on this topic.

Rinderknecht et al. (2023) explore challenges associated with using time-diary data to estimate the time spent preparing and consuming meals. These challenges are (1) “colloquial double barreling” which the authors define as reporting two separate sequential activities in a single episode and (2) the fact that some activities are perceived to be too inconsequential, so they are underreported. This research has important implications not only for survey data collection of topics related to food acquisition, preparation, and consumption, but can also impact associated research on key correlates such as health and social development outcomes. The authors compared the sequence of activities in original time diary data collected via Amazon’s Mechanical Turk (MTurk) and Prolific to the American Time Use Survey (ATUS). The findings showed that when respondents report the time spent in primary meal preparation, approximately half of them report primary meal consumption, while a non-negligible percentage also report secondary eating. The study also showed that demographic factors can influence how activities are reported. This study highlights the complexities of collecting data on meal preparation and meal consumption via time diaries and emphasizes the importance of collecting detailed auxiliary information, such as secondary activity information, as it can significantly alter the understanding of how people spend their time.

DiGrande et al. (2023) assess the feasibility of adopting augmented reality (AR) technology in dietary intake surveys. Estimating portion sizes and quantities are a critical aspect of dietary recalls, but these concepts can be difficult for respondents to estimate without the use of aides, such as two-dimensional

images in a booklet or on a computer screen or three-dimensional food models. However, two-dimensional models are challenging to translate to scale and three-dimensional models are bulky and can be difficult to transport in face-to-face settings. AR technology can potentially overcome some of these obstacles by rendering three-dimensional digital images directly within a web- or mobile-based survey instrument. The authors found promise in these technologies: respondents were able to use the AR tool to estimate portions and reported relatively high levels of satisfaction.

Future directions for food surveys

The goal of this special issue is to raise awareness of the methodological challenges involved in collecting food data. In the process we have sought to apply findings from survey methodology at a general level to the specific domain of food surveys. However, this process is imperfect because of the distinctive requirements of studies on food acquisition and consumption (e.g., detailed information over multiple days). Thus, there is a need for further research into food surveys specifically, akin to what is presented in Zezza et al. (2017) and this special issue. These methodological investigations can help researchers understand the error properties of food surveys and how these surveys might be improved. A consistent concern in this context is respondent burden. Researchers have made some progress in identifying the survey design factors that impact the burden of responding (survey mode, recall versus diary format, individual versus household reporting), but these factors can simultaneously impact different sources of error. More work is needed to figure out these tradeoffs across food survey designs and populations of interest.

This editorial also described recent advancements in data collection methods that have been applied to food surveys. Efforts to use apps and sensors and integrate data from other sources show promise for reducing respondent burden. Indeed, these technologies are being applied in pilot studies for future rounds of FoodAPS. Other opportunities include expanding data products and producing statistics at smaller levels of granularity through modeling or small domain estimation methods. These methods "borrow strength" from larger domains or integrated data sources to produce estimates at the desired level of granularity, e.g., for population or geographic subgroups of interest (Rao & Molina, 2015). Survey designers might also take inspiration from the "quantified self" domain where multimodal tools have emerged for tracking food intake (Luo et al., 2021), though some of these tools may not extend to population surveys (e.g., because of low willingness to download research apps).

In summary, the papers presented in this special issue discuss opportunities around the use of new and emerging technologies and alternative data sources that can potentially address some of the challenges of collecting food data. Given rapid technological and societal change, we need more methodological work assessing these new methods to ensure high-quality food data for policy making. To achieve this goal, we will need to foster collaborations between survey methodologists, statisticians, and domain experts.

References

1. Antoun, C., Katz, J., Argueta, J., & Wang, L. (2018). Design Heuristics for Effective Smartphone Questionnaires. Social Science Computer Review, 36(5), 557–574. https://doi.org/10.1177/0894439317727072
2. Beegle, K., De Weerdt, J., Friedman, J., & Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. Journal of Development

Economics, 98(1), 3–18. https://doi.org/10.1016/j.jdeveco.2011.11.001
3. Brzozowski, M., Crossley, T. F., & Winter, J. K. (2017). A comparison of recall and diary food expenditure data. Food Policy, 72, 53–61. https://doi.org/10.1016/j.foodpol.2017.08.012
4. Christian, W. J. (2012). Using geospatial technologies to explore activity-based retail food environments. Spatial and Spatio-Temporal Epidemiology, 3(4), 287–295. https://doi.org/10.1016/j.sste.2012.09.001
5. Chung, C.-F., Wang, Q., Schroeder, J., Cole, A., Zia, J., Fogarty, J., & Munson, S. A. (2019). Identifying and Planning for Individualized Change: Patient-Provider Collaboration Using Lightweight Food Diaries in Healthy Eating and Irritable Bowel Syndrome. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(1), 7:1-7:27. https://doi.org/10.1145/3314394
6. Coates, J. C., Colaiezzi, B. A., Bell, W., Charrondiere, U. R., & Leclercq, C. (2017). Overcoming Dietary Assessment Challenges in Low-Income Countries: Technological Solutions Proposed by the International Dietary Data Expansion (INDDEX) Project. Nutrients, 9(3), Article 3. https://doi.org/10.3390/nu9030289
7. Couper, M. P. (2013). Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys. Survey Research Methods, 7(3), Article 3. https://doi.org/10.18148/srm/2013.v7i3.5751
8. DiGrande, L., Karns, S., Kinyara, E., & Pedrazzani, S. (2023). Usability of Augmented-Reality Portion Estimators During 24-hour Dietary Recall Interviews. Survey Methods: Insights from the Field. https://doi.org/10.13094/SMIF-2023-00016
9. Dubois, P., Griffith, R., & O'Connell, M. (2022). The Use of Scanner Data for Economics Research. Annual Review of Economics, 14(1), 723–745. https://doi.org/10.1146/annurev-economics-051520-024949
10. Fiedler, J. L., & Yadav, S. (2017). How can we better capture food away from Home? Lessons from India's linking person-level meal and household-level food data. Food Policy, 72, 81–93. https://doi.org/10.1016/j.foodpol.2017.08.015
11. Kaderabek, A. (2023). Exploring Optical Character Recognition (OCR) as a Method of Capturing Data from Food-Purchase Receipts. Survey Methods: Insights from the Field. https://doi.org/10.13094/SMIF-2023-00015
12. Keusch, F., & Conrad, F. G. (2022). Using Smartphones to Capture and Combine Self-Reports and Passively Measured Behavior in Social Research. Journal of Survey Statistics and Methodology, 10(4), 863–885. https://doi.org/10.1093/jssam/smab035
13. Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to Participate in Passive Mobile Data Collection. Public Opinion Quarterly, 83(S1), 210–235. https://doi.org/10.1093/poq/nfz007
14. Kirlin, J. A., & Denbaly, M. (2017). Lessons learned from the national household food acquisition and purchase survey in the United States. Food Policy, 72, 62–71. https://doi.org/10.1016/j.foodpol.2017.08.013
15. Luo, Y., Kim, Y.-H., Lee, B., Hassan, N., & Choe, E. K. (2021). FoodScrap: Promoting Rich Data Capture and Reflective Food Journaling Through Speech Input. Proceedings of the 2021 ACM Designing Interactive Systems Conference, 606–618. https://doi.org/10.1145/3461778.3462074
16. National Center for Health Statistics. (2023, May 31). NHANES – About the National Health and Nutrition Examination Survey. https://www.cdc.gov/nchs/nhanes/about_nhanes.htm
17. Nielsen, J. (1994). Usability Engineering. Morgan Kaufmann.
18. O'Brien, D. (2018). What Do We Eat in America? AgResearch, 66(33). https://agresearchmag.ars.usda.gov/2018/mar/survey/
19. Olson, K., Smyth, J. D., Horwitz, R., Keeter, S., Lesser, V., Marken, S., Mathiowetz, N. A., McCarthy, J. S., O'Brien, E., Opsomer, J. D., Steiger, D., Sterrett, D., Su, J., Suzer-Gurtekin, Z. T., Turakhia, C., & Wagner, J. (2021). Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys: AAPOR Task Force Report. Journal of Survey Statistics and Methodology, 9(3), 381–411. https://doi.org/10.1093/jssam/smz062

20. Page, E. T., Larimore, E., Kirlin, J. A., & Denbaly, M. (2019). The National Household Food Acquisition and Purchase Survey: Innovations and Research Insights. Applied Economic Perspectives and Policy, 41(2), 215–234. https://doi.org/10.1093/aepp/ppy034
21. Rao, & Molina. (2015). Small Area Estimation. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118735855.fmatter
22. Rinderknecht, R. G., Doan, L., & Sayer, L. C. (2023). Secondary Activities: Their Proximity to Primary Activities and Their Importance for Understanding Reports of Preparing and Consuming Meals. Survey Methods: Insights from the Field. https://doi.org/10.13094/SMIF-2023-00014
23. The Inter-Agency and Expert Group on Food Security, & Agricultural and Rural Statistics. (2019). Food Data Collection in Household Consumption and Expenditure Surveys: Guidelines for Low- and Middle-Income Countries (LSMS Guidebook 147689). World Bank Group. https://www.worldbank.org/en/programs/lsms/publication/Food-data-collection-in-household-consumption-and-expenditure-surveys-guidelines-for-low-and-middle-income-countries
24. U.S. Department of Labor/Bureau of Labor Statistics/Division of Consumer Expenditure Survey. (2016). 2015 Response Rates: Interview Survey and Diary Survey, Consumer Expenditure, Public Use Microdata. U.S. Department of Labor/Bureau of Labor Statistics/Division of Consumer Expenditure Survey. https://www.bls.gov/cex/pumd/2015/csxresponserates.pdf
25. Wenz, A., Jäckle, A., Burton, J., Couper, M. P., & Read, B. (2023). Quality of expenditure data collected with a mobile receipt scanning app in a probability household panel. Understanding Society (2023–02; Understanding Society Working Paper Series). University of Essex. https://www.understandingsociety.ac.uk/research/publications/547667
26. Yan, T., Fricker, S., & Tsai, S. (2020). Response Burden: What Is It and What Predicts It? In Advances in Questionnaire Design, Development, Evaluation and Testing (pp. 193–212). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119263685.ch8
27. Yan, T., & Machado, J. (2023). Review of Food Diaries Used to Collect Food Acquisition Data. Survey Methods: Insights from the Field. https://doi.org/10.13094/SMIF-2023-00013
28. Zezza, A., Carletto, C., Fiedler, J. L., Gennari, P., & Jolliffe, D. (2017). Food counts. Measuring food consumption and expenditures in household consumption and expenditure surveys (HCES). Introduction to the special issue. Food Policy, 72, 1–6. https://doi.org/10.1016/j.foodpol.2017.08.007