Learning About the Poor Using Ethnography, Household Surveys, and Administrative Data

Bruce D. Meyer

Despite decades of work, researchers have not accurately accounted for who is poor, what is the nature of the poor’s predicament, what led to their situation, and what programs or policies improve their circumstances. Many researchers have suggested that ethnography is the key to answering some of these questions. Indeed, many excellent ethnographers have provided valuable insights, such as Kathryn Edin and Laura Lein who delineate the strategies single mothers used to support themselves and their children before welfare reform. In addition, Jennifer Romich and Thomas Wiesner describe how families use the earned income tax credit.

In this chapter, I suggest that ethnography can play an important role in this research agenda but also argue that many questions seemingly best addressed by ethnographic methods may be more easily and accurately answered by combining survey and administrative data in innovative ways. I begin by describing ethnography's strengths and weaknesses and the problem of measurement error. I then describe how linked administrative and survey data can answer many tough questions on the homeless and extreme poverty. I conclude with suggestions for capitalizing on the best aspects of ethnography and other research approaches.

Strengths and Weaknesses of Ethnography

Ethnography has many strengths. It can obtain hard-to-elicit information. In particular, ethnographers have succeeded in asking intrusive questions that require respondents’ trust to obtain good answers. These questions could involve illegal or socially sanctioned behavior, a topic that Edin and Lein cover. The hard-to-elicit information might also reflect individuals’ motivations—a topic Eberstadt emphasizes.

In this case, a researcher may learn how people explain their decisions to not work or to have a child outside of marriage. Ethnography can help survey researchers improve their questions, for example, by asking new questions and allowing people to respond in a way that traditional surveys do not allow. Another advantage of ethnography is the ability to suggest hypotheses to be further tested, as Robert Moffitt emphasizes. One example includes the mechanisms people use to manage their incomes and expenses when they are highly variable.

Ethnography also has notable weaknesses. Almost universally practiced, ethnography delivers a small, highly nonrepresentative sample from which generalization may be difficult. The unstructured interviews that are typically used may not ultimately elicit answers to some of the questions motivating the interviews in the first place.

And importantly, ethnography suffers from a lack of replicability. Replicability has long been recognized as a fundamental feature of good research in many disciplines. With some survey data, researchers may not be able to verify details of a given sample, but they may be able to replicate the approach with a different sample to check the main results. But
with ethnographic studies, a more limited version of replication is often harder because sampling relies on specific idiosyncratic relationships, such as those between neighbors or referrals by acquaintances, producing samples that another researcher cannot easily reproduce.

**Measurement Error and Bias in Surveys and Ethnographies**

An important though less emphasized drawback of surveys and ethnographies is potential bias due to measurement errors. Research results from either source should not be taken at face value. Evidence clearly shows that the problem of biased responses is particularly important for studies of the disadvantaged. We should expect some of the problems that plague conventional surveys to be potentially even more severe in ethnographies of the poor.

In flagship US household surveys, such as the Census Bureau’s Current Population Survey (CPS) or Survey of Income and Program Participation (SIPP), systematic errors leading to large biases are present. This problem is evident despite a long list of advantages these structured and codified studies have over ethnographic methods. The CPS methodology has been refined over 70 years and the SIPP’s over 35 years.

The Census Bureau takes many steps to address data quality. For example, Joanne Pascale, in describing the comprehensive research program developed to improve the CPS health insurance questions, reports:

Numerous small-scale studies—both qualitative and quantitative—were fielded and analyzed, and results were fed into subsequent small-scale tests in an iterative fashion. The research approach was to identify features of the questionnaire that were potentially contributing to measurement error, explore and modify those features, and test against the status quo to assess empirical evidence for improvements due to the changes. After a decade of this testing, a fundamentally redesigned questionnaire was crafted.9

The typical CPS field representative has had several years of experience and has received “extensive training on interviewing skills, such as how to handle non-interview situations, how to probe for information, ask questions as worded, and implement face-to-face and telephone techniques.”10 Both the CPS and SIPP surveys have had reinterview programs in which a share of respondent households were interviewed a second time, usually by a supervisor, to check the accuracy of the original interview responses. Phone interviews and laptop data entry from in-person interviews are recorded and can be later examined.

Despite the US Census Bureau’s professionalism and rigor, the CPS and SIPP have substantial errors in some key questions. Focusing on transfers important to those with low income, Bruce Meyer, Wallace Mok, and James Sullivan find that the shares of paid benefits captured in the CPS and SIPP are low and have declined sharply over time.11 In recent years, they find that the CPS is missing 50 percent of Temporary Assistance for Needy Families (TANF) dollars, 42 percent of Supplemental Nutrition Assistance Program (SNAP) dollars, 32 percent of unemployment insurance dollars, and 54 percent of workers’ compensation payments. The missing dollars in the SIPP appear to be less accurate than true reports.

These patterns are also evident in studies that link individual program records to surveys, such as Gathright Gathright and Tyler Crabbe who examine Social Security and Supplemental Security Income (SSI) reporting in the SIPP. Additionally, Bruce Meyer, Nikolas Mittag, and Robert George examine SNAP reporting in the American Community Survey (ACS), the CPS, and the SIPP.12 Comparing survey reports to administrative data from New York State, Meyer and Mittag find that 63 percent of public assistance recipients in the CPS and 44 percent of recipients in the SIPP do not report receipt of benefits.13 The corresponding numbers for SNAP recipients are 43 and 19 percent in the CPS and SIPP, respectively. Crucially,
measurement error in the poor’s income is not just noise. It has a nonzero mean and thus leads to large biases in many analyses.

These errors affect substantive conclusions about poverty and any measured poverty reduction due to transfer programs. Meyer and Mittag link CPS micro-data to New York administrative micro-data for the beneficiaries of four government transfer programs and find that unreported transfers exceed the recorded income of those with pretax income below half the poverty line. Adam Bee and Joshua Mitchell find that approximately half of pension payments are not reported in the CPS and that accounting for this and other underreporting reduces the elderly poverty rate by over 2 percentage points (about 24 percent). Bruce Meyer and Derek Wu examine the poverty reduction effects of several transfer programs in the SIPP, which has the least underreporting of the major surveys, and find that the survey data miss more than half the poverty reduction of several programs among single parents.

Ethnographies are typically more informal enterprises with few, if any, of the quality checks present in Census Bureau household surveys. If we obtain such skewed responses in a more rigorous and scrutinized situation, we should realize that ethnographic evidence should be viewed with at least the same skepticism applied to many formal survey statistics. This is not to say that ethnographers do not have high standards and impeccable motivations. They undoubtedly do. But even well-intentioned formal interviewers are unable to accurately elicit many basic features of the poor’s economic conditions.

Interviewing those with the greatest disadvantages presents additional difficulties. Motivations and attempts to build a relationship with respondents can distort responses in subtle ways. While a trusting relationship may encourage a respondent to reveal things he or she would not otherwise say, it can also compromise the interviewer. Empathy may make it difficult to question responses that are not entirely accurate.

The most disadvantaged frequently face situations in which they need to alter the truth to get by, whether asking a friend for money, talking with a welfare case-worker, or explaining difficulties to a stranger. The hard objective and scientific approach of formal surveys may do a better job eliciting accurate information. With surveys, we can check individual survey responses through data linkage; we frequently learn that they are not accurate. That way of checking answers is not possible with ethnographies.

**Skepticism and $2-a-Day Poverty**

We should view all data skeptically. We should not presume that any data are without bias. Ethnographic data are likely to have some of the same flaws as household survey data, plus some additional handicaps. If the focus is unusual situations, such as those with very low income, then such “outliers” are crucial to research. Unfortunately, outliers in many situations tend to be observations that are most likely erroneous.

Consider the literature on $2-a-day poverty, sometimes called “extreme poverty.” In a bestselling, award-winning book and in follow-up work, Kathryn Edin and Luke Shaefer argue that the number of families with children living on $2 a day is very high and has risen dramatically in recent decades.

The research is a mixture of ethnography and analysis of nationally representative household surveys—a commendable combination. However, when one further examines the data, using the linked data described below, one finds that claims of high rates and sharp growth do not hold up. Specifically, Bruce D. Meyer et al. link administrative data on earnings and program receipt with the survey data used in the $2-a-day poverty literature and find that the share of those living on extremely low incomes is a tiny fraction of that claimed; in fact, it is hard to find any families with children in the data that fit the descriptions given.

While there are multiple interpretations of this disagreement between the ethnographic and linked data—a failure to generalize, an overdramatization of ethnographic evidence, and a willingness to accept unbelievable stories—the fact remains that key claims of the $2-a-day literature are not true. The ethnography provides a vivid picture of the lives of families getting by on little income, how they ended up so
poor, their aspirations, and how they survive. But the example suggests that ethnographies can also have substantial biases that lead to erroneous conclusions. Because the ethnography’s details cannot be directly verified, the clearest conflict between the $2-a-day claims and what one finds in more accurate data is between the survey data taken at face value (with its reporting errors) and what the linked survey and administrative data provide.

The Promise of Linked Survey and Administrative Data as a Research Tool

At the Census Bureau, we are working on an unprecedented new project that assembles and links survey and administrative data on income, program receipt, expenditures, and closely related information. We call the project the Comprehensive Income Dataset (CID). The project links several Census Bureau surveys including the CPS, SIPP, ACS, and Consumer Expenditure Survey (CE) to tax records and means-tested social insurance program records, such as public and subsidized housing data from the Department of Housing and Urban Development, Social Security retirement and disability payments, SSI from the Social Security Administration, Veterans Affairs means-tested pensions and disability benefits, TANF, and SNAP (formerly food stamps) benefits from about half of the states.

Our goal is to assemble a nearly complete set of administrative income sources and related information to assess material disadvantage. Although the Census Bureau has not obtained all the necessary interagency data use agreements, the process is ongoing, and the remaining missing sources are likely small. We can obtain a comprehensive picture of the population’s income sources and programs received. We believe the CID will allow us to answer fundamental questions about poverty, including who is poor, under a variety of measures differing on what income or consumption sources are included and which poverty cutoffs are considered.

In assessing who is poor and validating our definitions, we will examine many markers of disadvantage. In doing so, we hope to identify the population that is most materially deprived. We will have information on:

- Household expenditures and consumption to assess the quantities and types of housing, food, transportation, clothing, and other goods and services consumed (in the CE Survey);
- Housing conditions, such as holes in walls and floors, broken windows, electrical problems, leaky plumbing or roofs, pests, etc. (in the SIPP) and other housing characteristics (in the CE Survey and potentially the American Housing Survey);
- The ownership of basic and other appliances (in the SIPP and CE Survey);
- Assets owned such as a home that can be lived in, a car that can be driven, and financial resources that can be drawn on (in the SIPP and the CE Survey); and
- Responses to questions (in the SIPP) about hardships faced, such as difficulty paying utility bills or rent and not having enough food.

We have already found that a majority of those in the bottom few percentage of the reported income distribution have badly measured income, and a large share have incomes several multiples higher than their reported incomes, with housing characteristics and freedom from economic hardship similar to the middle class. The demographic makeup of those who truly are at the bottom is quite different from those who reported being at the bottom in the SIPP and CPS.

Implicit in the advocacy of assembling and using linked data is the presumption that, in many domains, administrative data will be more accurate than traditional survey responses or ethnography will be. While administrative data are not without error—most commonly incompleteness—for questions of income and program receipt, traditional survey data are so frequently erroneous that administrative data sources are likely much more accurate. In other areas, such
as health information and criminal histories that are potentially obtained through administrative data, again administrative data are likely to be more accurate.

**Learning Who Is Poor and How the Safety Net Supports Them**

Armed with better measures of poverty, we will be able to better answer questions such as: What are the characteristics of those who are most materially deprived? How do they look in terms of family structure, employment, education, age, and other characteristics? With the administrative data, we will be able to identify who receives what government programs and whether recipients are among the families most in need. We will be able to answer: Who does the safety net miss? Are those who do not receive assistance less disadvantaged, and are there identifiable families that are especially deprived and missed by the safety net?

A further question is: Who participates in multiple welfare or social insurance programs? Multiple program receipt appears to be the norm among low-income populations. Eberstadt also raised the need to know the patterns of multiple program receipt. For example, the overlaps among recipients of TANF, SNAP, Medicaid, and housing benefits all appear to be high.

Given the large number of overlapping programs and high frequency of reporting errors, the chances of getting the inventory right for a given household in survey data (or likely ethnography) are low. With the linked administrative and survey data, we could see if those who receive multiple benefits have a high level of disadvantage according to many other indicators, including inability to pay bills and low consumption. We can examine if non-recipients can maintain a moderate or high level of consumption and if they need to rely on savings or credit that allow them to get by in bad times.

Alternatively, we may learn that families that appear to be eligible non-recipients are not as deprived and thus may not participate. It is likely that a large share of non-receipt characterized by those eligible for transfer programs is really non-reported receipt in the error-ridden survey responses.

**The Economic Situation of the Disadvantaged**

More generally, the CID-linked data are a valuable source to assess the bottom’s standard of living by examining income, consumption, assets, liabilities, and the hardship measures and indicators of material well-being discussed above. Individual histories of employment and earnings can also be studied using tax data on employment and earnings. Linked data can also improve household surveys.

**Determining Barriers to Success**

Understanding why certain individuals are poor requires knowing what barriers inhibit their success. Usually this idea is conceived as barriers to employment, but the idea is more general and could be applied to barriers to success as an employee, homemaker, caregiver, or other roles. The SIPP covers important domains, including lack of education, disability, housing problems, some information on job displacement, and limited information on drug and alcohol abuse.

In his discussion of the decline in male labor force participation, Eberstadt provides a helpful list of potential barriers to employment that could be investigated through ethnography. These topics include “disability and dependence” and “criminality and mass felonization.” While not currently part of the CID data set, in the future, it may be possible to link survey data and administrative income and program receipt data to medical claims data to learn about mental health and substance abuse. Linking to crime data to understand criminal histories that may be a barrier to employment or an alternative income source is also something that could be done in the near future. I suggest other topics below that could be examined by expanded survey questions.

**Changing the Role of Surveys**

This volume is about imagining better data to better understand the nature of poverty and related policy choices, so let me think big for a minute and
discuss data that require changes in major surveys’ content, which takes time and money. Some of our major household surveys could shift away from asking detailed income questions if such information is better obtained from government records. Reducing the burden of existing surveys would leave room to ask some of the questions Eberstadt and Moffitt proposed that cannot be answered by currently available linked or potentially linkable data in the near future.

For example, Eberstadt suggests a class of questions about people’s hopes, aspirations, and frustrations. He poses other questions about the stigma of nonwork and people’s work ethic. These questions are less traditional but might illuminate barriers to employment and financial success that people face. We might also add questions about family commitments such as the need to be a caregiver or family dysfunction and abuse that might make success difficult. Eberstadt also notes that information on reasons for nonwork is slim in census surveys, thus suggesting that a more detailed probing of reason for nonwork would be informative.

Using Linked Data to Learn About the Homeless

Linked data (such as the CID) may offer welcome insights into the homeless, likely the most disadvantaged group in our country and certainly an understudied one. This is an ideal group for ethnographic research. Most household surveys omit (or largely omit) the homeless. Some specialized surveys have given us valuable insights into the homeless. For example, Martha Burt et al. provide a wealth of information, but the study is over 20 years old.

However, recent and ongoing research shows that the homeless can be studied using linked administrative data. For example, Stephen Metraux et al. examine the earnings of the homeless over more than two decades by successfully linking New York City Department of Homeless Services data to earnings records from the Social Security Administration.

National information on major components of the homeless population can be obtained from two Census Bureau sources, allowing researchers to generalize their results to the entire US. A large share of those interviewed can be linked over time to administrative data and other sources.

The ACS includes the homeless sleeping in transitional shelters. It asks extensive questions about income and demographics and some limited questions about mental and physical disabilities. These shelter homeless are about two-thirds of the overall homeless at any point in time (and presumably a larger share of those ever homeless). About 70 percent of those sampled are linkable to administrative records. One finds that a significant share of those in a shelter have had formal earnings recorded in a tax return in that year, and the vast majority receive government benefits, most often SNAP or Medicaid. One can trace their earnings and benefit receipt over 10 years with these data and look at their movement across the country and their longevity.

The 2010 Decennial Census includes both those sleeping in shelters and those who are unsheltered (the street homeless). The demographic characteristics are limited, including only age, gender, race, and Hispanic origin. While the shelter homeless can be linked at a high rate to administrative data, the...
unsheltered are more problematic, as most cannot be linked to other data sources. Nevertheless, a substantial share—though not a majority—of unsheltered homeless can be linked to other data sources on income and program receipt, providing a window into this hard to study population.

In addition, linking administrative data to local shelter homeless data that are longitudinal is also possible. The Census Bureau has long histories of Homeless Management Information System data for several large cities including Los Angeles, California; Chicago, Illinois; and Houston, Texas. Researchers have begun to use such information to learn about the homeless. For example, Metraux et al. examine the employment patterns of the sheltered homeless in New York City. It would be valuable to learn about other aspects of this population, such as drug and alcohol use, mental illness, difficulties in finding and keeping work, connections to other individuals, and histories of living arrangements. Ethnography will play a vital role in this research. Linked survey and administrative data likely will too.

Conclusions

In conclusion, ethnography has a valuable role in asking intrusive questions and suggesting hypotheses to be tested. However, impediments may mean that the goals of the ethnographies Moffitt and Eberstadt proposed are hard to realize.

This chapter suggests that not all statements from ethnographies should be taken at face value given the potential for biased measurement in all types of surveys, both quantitative and qualitative. Many of the topics that surveys have been incapable of addressing can be examined with linked survey and administrative data. The content of household surveys should shift toward the topics often left to ethnography, while many quantitative questions on topics such as income and program receipt can be answered by linked administrative and survey data.

Acknowledgments

Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the US Census Bureau. I would like to thank Alexa Grunwaldt and Paula Worthington for helpful comments and the Russell Sage Foundation, Alfred P. Sloan Foundation, and Charles Koch Foundation for financial support.
Notes


5. Eberstadt, “Understanding America’s ‘Men Without Work’ Problem.”


20. We will miss income that is informal and not reported to tax authorities or the Census Bureau. We may be able to infer such missing income from information on expenditures and assets.


22. I have in mind incomplete coverage of a complex set of programs (such as public and subsidized housing) or missing data that come from one contributor of administrative records (defined by program or geography, for example) being omitted over time.


25. Takeup will always be hard to determine especially without linked data—survey data rarely have the information used to determine eligibility and administrative data have it only for recipients, if at all.


27. Eberstadt, “Understanding America’s ‘Men Without Work’ Problem.”


29. Eberstadt, “Understanding America’s ‘Men Without Work’ Problem.”


32. While many poor individuals do not appear in tax records, the low-income working population is highly represented due to the advantages of claiming the earned income tax credit, which requires filing a tax return.


34. The CPS and SIPP include small numbers of individuals in homeless shelters, but they are not close to being fully covered in the sample frame.


38. While non-linking is not random, the survey data include survey measures of many of the outcomes of interest, such as income and program receipt. Thus, adjustment for non-linking, such as inverse probability weighting that uses those variables, is likely to eliminate a large share of any bias when examining administrative income or program receipt as outcomes.