

HOW TO BETTER MEASURE POVERTY

THE COMPREHENSIVE INCOME DATASET

KEVIN CORINTH AND BRUCE D. MEYER

INTRODUCTION

In 2015, Kathryn Edin and Luke Schaefer released their book, *\$2.00 a Day: Living on Almost Nothing in America*, making the extraordinary and widely disseminated claim that over 3 million children in the United States live on less than \$2 per day.¹ A report of the Special Rapporteur for the United Nations Human Rights Council in 2018 declared that “5.3 million [Americans] live in Third World conditions of absolute poverty.”²

If millions of Americans—many of them children—were truly living in a state of deprivation as bad as that faced by the poorest people in the poorest countries in the world, our entire approach to alleviating poverty in the United States would need to be reevaluated. Not only would policymakers need to quickly mend the scandalous holes in the safety net, but on-the-ground interventions by social service providers would warrant an overhaul. Social service providers would need to shift their focus to ensuring that families could simply subsist before they could even consider the possibility of helping them climb the ladder of opportunity.

Fortunately, it turns out that the shocking claims of extreme poverty in America portrayed by Edin and Schaefer and the Special Rapporteur to the

United Nations were wrong. Thanks to the Comprehensive Income Dataset (CID) Project—an unprecedented effort to link government surveys with dozens of sources of administrative data on taxes and government program receipt—we could, for the first time ever, accurately measure incomes at the very bottom of the distribution. When linking all the data sources together and including all sources of income, we could no longer find a single child in the linked dataset living on less than \$2 per day.³ While deprivation is very real in the United States, it does not, in fact, rival the severe levels of deprivation found in the poorest countries in the world.

This example illustrates that understanding the big picture of deprivation in the United States is imperative, not only for policymakers but also for social service providers. Unless we know who suffers from the highest levels of deprivation and the types of challenges they face, it will be difficult to ensure specific interventions are targeted to those most in need, and that they focus on the biggest problems. Nor will we be able to assess how the successes of practitioners and communities add up to overall societal progress.

Unfortunately, existing evidence on the big picture of disadvantage in the United States is inaccurate and incomplete. It relies on surveys that suffer from growing reporting errors and misses some of the most vulnerable segments of the population.

The CID Project seeks to improve on the existing evidence by creating the most accurate dataset on economic well-being ever created for the United States.⁴ The CID combines survey data with an unprecedented set of administrative data from tax filings and government programs. We conduct rigorous research and apply cutting-edge statistical techniques to combine these data sources in a way that maximizes the accuracy of our well-being measures. We also are able to capture populations, such as those experiencing homelessness, who are missed from most major household surveys. As a result, we are able to create a much more accurate and complete picture of disadvantage in the United States than has ever before been available.

The CID benefits from broader ongoing efforts in the federal government. The Evidence-Based Policymaking Commission and the Evidence Act sought to make data more secure, more available, and more widely used for evaluation inside and outside of government, furthering efforts like the CID. Specific commission recommendations that have yet to be

implemented, such as those on improved access to state data, would further strengthen research and evaluation efforts.

Ultimately, the evidence generated by the CID will inform the areas of greatest need, so practitioners, communities, and policymakers can tackle the biggest problems. And by providing highly accurate evidence on changes in disadvantage over time, we can measure how individual efforts add up to overall societal progress.

PROBLEMS WITH EXISTING DATA ON DISADVANTAGE

Government surveys are used extensively by federal agencies and researchers to assess the extent of disadvantage and broader measures of well-being in the United States. For example, the U.S. Census Bureau's annual report on the official poverty rate, median income trends, income inequality, and health insurance coverage relies on the Current Population Survey Annual Social and Economic Supplement.

Unfortunately, surveys suffer from several problems that reduce their accuracy and completeness. First, people increasingly fail to respond to surveys, which can lead to difficulty in attaining a sample representative of the U.S. population. Second, those people who continue to respond to surveys may provide inaccurate information. A vast body of research has shown that many categories of income—such as means-tested benefits, social insurance, and private pension income—are greatly underreported in surveys.⁵ Third, particularly vulnerable segments of the population (for example, individuals experiencing homelessness and unauthorized immigrants) are either under-surveyed or missing from surveys altogether. Fourth, some surveys completely exclude certain types of income, such as housing assistance and capital gains. These issues are likely to bias any survey-based analyses of income distribution, poverty, inequality, and the effects of government taxes and transfers. As a result, existing evidence on the most disadvantaged Americans is biased and incomplete, limiting the ability of service providers and policymakers to target programs to those most in need.

The problems with survey data have led some researchers to turn to administrative data, instead. But administrative data sources on their own do not capture the full set of resources available to individuals, and they do not contain the rich demographic information available in surveys that enables a focus on vulnerable groups.

THE COMPREHENSIVE INCOME DATASET

We are building the CID to overcome the inaccuracies in our basic understanding of economic well-being in the United States. The fundamental insight of the CID is that no single data source on its own can provide a full or accurate measure of income or well-being. But when multiple data sources are linked, the strengths of each data source can be harnessed while overcoming its limitations. The CID relies on three main types of data sources—household surveys, tax records, and federal and state administrative program data on government benefits. Each data source has unique strengths. Surveys provide rich demographic information that allows for the construction of families and analysis by race, educational attainment, and other characteristics. Tax data contain highly accurate information on certain income sources, such as earnings, and have near universal coverage, including many non-filers whose tax forms are supplied by employers and government agencies. Administrative data from government programs provide income information that is not captured well or at all by surveys or tax data.

We link all data sources at the individual level using anonymized identification codes created by the Census Bureau to ensure the confidentiality of personal data. We conduct rigorous research and apply cutting-edge statistical techniques to impute missing data, and also to inform broader conceptual decisions about how to optimally combine data sources. For example, we are pioneering a new methodology that uses a novel set of dozens of material hardship measures—such as housing quality problems, food insecurity, and mortality patterns—to validate decisions on how to construct a comprehensive measure of income. This evidence-based approach for constructing income measures represents a major step forward for the income measurement field, and it will ensure our comprehensive income and poverty measures are as accurate as possible.

The CID Project has, to date, linked together four household surveys with an extensive set of tax records and twelve sources of federal and state administrative program data—to our knowledge, the most comprehensive set of linked income-related data ever created for the United States.

EVIDENCE GENERATED BY THE COMPREHENSIVE INCOME DATASET PROJECT

As we continue to build and improve the CID, we already are demonstrating its power to transform our understanding of poverty, income, and well-being in the United States. To this end, our two major strands of research to date focus on: a) identifying those who are most disadvantaged and their levels of deprivation, and b) understanding a particularly vulnerable population missed by most surveys—individuals experiencing homelessness.

As previewed in the beginning of this chapter, one of the earliest papers using the CID examined the prevalence of households in the United States living on less than \$2 per person, per day (that is, “extreme poverty”).⁶ We focused on extreme poverty in our early research because the results starkly demonstrate the capacity of the linked data to change our understanding of poverty; in this case, due to the presence of survey outliers. We find that more than 90 percent of households with survey-reported cash incomes below \$2/person/day are not in extreme poverty once we include in-kind transfers, replace survey reports of earnings and transfer receipts with administrative records, and account for the ownership of substantial assets. Contrary to widely cited findings in the prior literature that over 3.5 million children live on less than \$2 per day in the United States, we find no children in our surveys falling below such an extremely low standard when using the CID and making the aforementioned improvements.

In research in progress, we extend our CID-based analysis of poverty to standards more applicable to the modern United States, a project directly relevant for informing improvements to the widely monitored official poverty measure estimates. In particular, we examine how poverty rates change when using the CID to correct for measurement error in pre-tax money income and when we incorporate tax liabilities and credits, in-kind transfers, and other non-cash income sources. In addition to more accurately estimating the level of poverty (holding the original poverty line constant), we can identify a more accurate picture of the poor population in terms of family composition, demographics, and material well-being.

In related ongoing work, we use the CID to study the best way to compare the economic well-being of families in different geographic regions. While families who live in high-cost areas may need to spend more resources to meet their basic needs, they may, at the same time, have access to higher-quality public services, more job opportunities, and a

cleaner environment. We pair the CID with a battery of deprivation measures (for example, material hardships, availability of appliances, home quality problems, long-term income, and mortality) to provide an evidence-based test for whether geographically adjusting poverty thresholds leads to a more deprived poor population compared to what one would see without applying geographic adjustments. This project helps illuminate the geographic distribution of disadvantage in the United States and, thus, informs where new efforts to innovate new solutions may be most needed.

We also have worked to improve our understanding of individuals experiencing homelessness, who—despite being one of the most deprived populations in the United States—are largely omitted from household surveys and, therefore, not reflected in official poverty statistics and the extreme poverty literature. We overcome this limitation by linking a census of the entire (sheltered and unsheltered) homeless population conducted as part of the 2010 Decennial Census with administrative data on tax records and government program benefits. Our research to date has shed new light on this highly vulnerable but poorly understood population.⁷ For example, we find that 53 percent of sheltered homeless adults under age 65 in 2010 had formal earnings during the year, and a substantial 40 percent rate for those found on the streets. In addition, the vast majority of individuals who experience homelessness receive government benefits—89 percent of sheltered homeless adults under 65 and 78 percent of unsheltered homeless adults under 65 received benefits from the Supplemental Nutrition Assistance Program, veterans benefits, housing assistance, Medicaid, or Medicare at some point during the year. While homeless individuals have greater connection to the formal labor market and government benefits than sometimes thought, they still face low levels of well-being that improve little over time. Among all non-elderly adults who experienced sheltered (unsheltered) homelessness in 2010, less than half had more than \$2,000 (\$200) of annual earnings in any year between 2005 and 2015.

FUTURE OF THE COMPREHENSIVE INCOME DATASET PROJECT

While the CID already represents the most comprehensive and accurate income-related dataset ever created for the United States, we are committed to pushing the frontier as far as possible. We are linking new administrative data sources, extending the CID back in time to cover more than two decades, and developing new statistical and conceptual methods for combining

survey and administrative data to maximize the accuracy of income measures.

In addition, we will use the CID to produce better evidence on disadvantage in the United States. Examples include measuring poverty over more than two decades, validating survey-based, real-time measures of poverty using the CID, examining the effects of government programs on well-being, identifying holes in the safety net, and exploring new projects on vulnerable groups under-covered by surveys, including individuals experiencing homelessness and unauthorized immigrants.

In each of these areas, the unprecedented accuracy and richness of the CID will transform our understanding of deprivation in the United States and break new ground on overlooked segments of the population. The new evidence we generate will be essential to inform practitioners, communities, and policymakers seeking to improve the well-being of the most disadvantaged members of society. Already, our research reevaluating extreme poverty has shifted narratives on deprivation in the media, and our methodological work combining survey and administrative data sources heavily informed the recommendations of a recently concluded federal interagency technical working group tasked with developing new poverty measures for the United States.⁸ As we learn more about poverty, and especially vulnerable groups including people experiencing homelessness, it will be imperative to ensure both policymakers and those on the frontlines serving these groups are able to use our research to inform their decisions. Ultimately, knowing where disadvantage is most prevalent will provide the big picture needed for service providers to lead the next generation of evidence building as they innovate their services to deliver better outcomes to the people they serve.

NOTES

1. Kathryn J. Edin and H. Luke Shaefer, *\$2.00 a Day: Living on Almost Nothing in America* (Mariner Books, 2015).

2. Human Rights Council of the United Nations, "Report of the Special Rapporteur on Extreme Poverty and Human Rights on His Mission to the United States of America," 2018, <https://digitallibrary.un.org/record/1629536?ln=en>.

3. Bruce D. Meyer, Derek Wu, Victoria Mooers, and Carla Medalia, "The Use and Misuse of Income Data and Extreme Poverty in the United States," *Journal of Labor Economics* 39, no. S1 (2001), pp. S5–S58.

4. For further information, see the CID Project website, <https://cid.harris.uchicago.edu/>.

5. For example, see Bruce D. Meyer, Wallace K. C. Mok, and James X. Sullivan, “Household Surveys in Crisis,” *Journal of Economic Perspectives* 29, no. 4 (2015), pp. 199–226.

6. Bruce D. Meyer, Derek Wu, Victoria Mooers, and Carla Medalia, “The Use and Misuse of Income Data and Extreme Poverty in the United States,” *Journal of Labor Economics* 39, no. S1 (2021), pp. S5–S58.

7. Bruce D. Meyer, Angela Wyse, Alexa Grunwaldt, Carla Medalia, and Derek Wu, “Learning about Homelessness Using Linked Survey and Administrative Data,” Working Paper 28861, NBER, 2021.

8. See *Final Report of the Interagency Technical Working Group on Evaluating Alternative Measures of Poverty*, U.S. Census Bureau, www.census.gov/content/dam/Census/library/publications/2021/demo/EvaluatingAlternativeMeasuresofPoverty_08Jan2021.pdf.