MATHEMATICA Policy Research

REPORT

FINAL REPORT

Data on Earnings: A Review of Resources for Research

September 28, 2018

John L. Czajka
Ankita Patnaik
Marian Negoita

Submitted to:
U.S. Department of Labor
Chief Evaluation Office
200 Constitution Avenue, NW
Washington, DC 20210
Project Officer: Janet Javar

Contract Number: DOLQ129633249/DOL-OPS-15-U-00183

Submitted by:
Mathematica Policy Research
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director: Sarah Dolfin
Reference Number: 50174

This report was prepared for the U.S. Department of Labor (DOL), Chief Evaluation Office by Mathematica Policy Research, under contract number DOLQ129633249. The views expressed are those of the authors and should not be attributed to DOL, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

CONTENTS

TABLES

FIGURES

This page has been left blank for double-sided copying.

I. INTRODUCTION

Earnings are a key outcome in evaluating the impact of job training and other employment-oriented assistance programs. Adequate and sustained earnings provide a reliable measure of success for these interventions and contribute to other positive outcomes as well. Accordingly, high quality data on the income earned from employment—including self-employment—are an essential component of any analysis that seeks to understand the consequences of participating in an employment-assistance program.

This brief report describes several sources of data on wages or earnings and is addressed to researchers in a variety of settings who wish to incorporate reports of earned income in an evaluation or analysis. These sources include administrative records, custom surveys of study participants, and existing general population surveys.

We focus primarily on administrative sources, as these data have the potential to be linked to the participants in an evaluation, whether a randomized controlled trial or an observational study. Collecting wage data for the study participants from a survey designed and administered solely for the evaluation is also an option and one that affords the researcher the ability to tailor the data collection to the exact project needs. However, collecting original data can be costly and require significant lead time to develop instruments and obtain the approvals that might be necessary before collecting the data. Depending on the length of the survey reference period, respondents could have difficulty recalling details of earlier employment. Obtaining access to administrative data might also require a long lead time, and the data themselves will have an associated lag, but the timing might be less critical, and there is no issue of respondent recall affecting the quality of the data.

Household survey data such as those collected by the U.S. Census Bureau provide another source of data on earnings, but because these data are collected from samples of the population, they are unlikely to exist for any of the participants in a study. Their value lies in their ability to support analyses at the level of the entire population or large, well-defined subpopulations. Survey data can also be linked, potentially, to administrative data on earnings to enable analyses that correct for measurement error and nonresponse in the reporting of earned income.

Chapter II provides an overview of the principal sources of administrative data on earnings. Chapter III presents a brief discussion of a custom survey as an alternative or supplement to administrative records and reviews recent research findings comparing the two with respect to estimates of program outcomes. Chapter IV provides an overview of three major household surveys that collect data on earnings from representative national samples and can be linked to administrative data at the U.S. Census Bureau or, remotely, from Federal Statistical Research Data Centers (FSRDCs) located around the country.

II. ADMINISTRATIVE DATA ON EARNINGS

There are two primary sources of administrative data on earnings: the federal tax system, which obtains separate reports of earnings paid by employers and received by employees, and the state Unemployment Insurance (UI) system, which requires quarterly reports of wages paid to individual employees by employers whose jobs are covered by UI. The tax system encompasses all wage and salary employment and self-employment, in principle (that is, excepting whatever may be unreported due to noncompliance), whereas the state UI system excludes federal employment, the military, state and local government employment, self-employment, and a small amount of additional employment that is not covered by UI as defined by each state (Table II.1).[1] However, the wages reported for jobs in the UI data include nontaxable components of earnings that are not captured in the wages reported for these same jobs in the tax system. Neither source captures informal or off-the-books employment that is not already exempted for reporting by federal and state law. Lastly, both sources can document employment, but neither source can establish that an individual was not working or had no earnings.[2]

[1] There are additional federal exemptions to coverage that are defined by the Federal Unemployment Tax Act (FUTA) that states are obligated to include in their state UI laws. These include employment for a religious, charitable, educational, or other organization that is exempt from income tax under section 501(a); railroad employment; student and student spouse employment defined as financial aid; and others. On top of these, states can exempt additional categories of employment. For example, California excludes service performed in the employ of a foreign government or an international organization, and other states have exceptions. Some of these are discussed in Stevens (2007). In addition, in what manner and how quickly states deal with employer noncompliance also contributes to coverage differences across states.

[2] For example, if a state has no wage data for an individual residing in the state, the individual could be working in a job not covered by the UI system or working in another state. The IRS has broader coverage of employment, but if an individual has no reported earnings, that individual could be working off the books.

Table II.1. Types of earnings captured in tax data and state UI wage data

Source of earnings	Tax data	State UI wage data
State UI-covered employment[a]	X	X
Federal employment	X	
Military employment	X	
Other non-UI-covered employment	X	
Self-employment	X[b]	
Informal or off-the-books employment		

[a] Exactly what employment is covered by the UI system is defined in state law, which reflects both federal requirements and state exemptions and, therefore, varies to some degree across the states. State practice with respect to employer noncompliance accounts for additional state variation.

[b] Substantially underreported; see below.

The Internal Revenue Service (IRS) holds all federal tax data, but the Social Security Administration (SSA) also holds data on earnings—both wage and salary and self-employment—as it requires this information to administer its programs. Employers submit their employees' annual wage and tax statements to SSA, which processes the forms and transmits the data to the IRS. Taxpayers report their self-employment income on the annual tax returns that they file with the IRS, which sends to SSA the self-employment data along with the self-employment taxes that it collects. Other federal and state agencies can obtain selected tax data for applications specified in the Internal Revenue Code, but only the U.S. Census Bureau and the Statistics of Income Division of the IRS currently provide research access to tax data, as discussed in detail below. States collect and maintain UI wage data, and the states remain the ultimate source for obtaining access to these data. However, under a program (the National Directory of New Hires [NDNH]) created by the 1996 welfare reform law, the states must submit data from their UI wage reporting systems to the Office of Child Support Enforcement (OCSE) in the Administration for Children and Families in the U.S. Department of Health and Human Services (HHS). In addition, the Census Bureau negotiated with the states (and was successful with all but Massachusetts) to obtain UI wage data to develop an extensive database linking

employees' wage data with employers' data collected by the Bureau of Labor Statistics and a number of less inclusive data sources. These linked data form the Longitudinal Employer-Household Dynamics (LEHD) program database.

The heart of this report is a set of five matrices that describe five sources of administrative data on earnings:

1. The earnings data collected by the IRS for tax purposes
2. The earnings data maintained by the SSA in its Master Earnings File
3. The earnings data collected by the state UI agencies
4. The NDNH obtained from states by OCSE at HHS
5. The LEHD program at the U.S. Census Bureau

Figures II.1 through II.5 describe the five data sources and the earnings data they hold. Each matrix includes the following elements:

- Identification of the data source
- An associated web address, if applicable
- A description of the data and their origin
- The universe from which the data are collected and any known coverage issues (both population and amounts)
- Information included
- Options for data linkage
- Procedures and practices surrounding access
- Timeliness of the data—from collection to availability and from application to receipt
- Key strengths and limitations for analysis
- Two citations of research using the data, preferably from seminal or cutting-edge literature
- References for descriptive information about the data

Next, we provide a few summary observations culled from the information presented in the five matrices.

Access is an issue with regard to all five data sources. For the tax data (IRS and SSA) and the LEHD, data cannot be removed and must be analyzed on agency computers—and for data held by SSA, only by SSA staff. For data held by the Census Bureau, which include some IRS and SSA data, research can be conducted remotely from one of several FSRDCs located around the country. Regardless of where the data are analyzed, the products of any analysis, whether tabulations or model parameter estimates, must be reviewed for possible disclosure before they can be removed from the site. With regard to UI data, the restrictions with regard to access and use depend on the source. For UI data obtained from the NDNH, there are provisions for obtaining data linked to a data set that the researcher provides, but the data returned to the researcher is de-identified, which limits potential uses. For UI wage data obtained from one or more states, the terms of usage depend on what the researcher is able to negotiate with the state(s) providing the data.

Coverage is an important consideration in choosing among these alternative data sources. State UI wage data exclude federal and military employment plus a number of smaller sources. For studies of training program participants, excluding federal employees and the military might not be as important as in other contexts. The tax data collected by the IRS and SSA do not have these limitations, but the earnings amounts collected are taxable, not gross earnings. For example, depending on the source of tax data used, the earnings from employment could include pre-tax contributions to retirement accounts and will definitely not include other pre-tax deductions, such as for dependent care, health insurance premiums, and flexible spending plans, to name the largest. On the other hand, taxable earnings can include employee benefits that are not paid out in wages, whereas the earnings reported to the UI system will tend not to include such items. Of course, all of the administrative sources will miss informal or off-the-

books employment that survey data could potentially capture if respondents are willing to report more than they report to the IRS.

Earnings data from self-employment are highly problematic. First, the state UI data do not capture self-employment. Second, the self-employment income reported to the IRS is subject to substantial under-reporting—on the order of two-thirds, based on the agency's audits of random samples of tax returns (IRS 2016). In addition, survey estimates vary widely, owing in part to definitional differences across surveys and variation in how respondents understand what they are being asked to provide. Furthermore, unlike earnings from employment, most self-employment income does not correspond to amounts reported on pay stubs, so respondents lack a readily accessible reference to guide their responses.

Periodicity of the data can be important to consider as well. The UI data are quarterly, whereas the tax data are only annual. If annual data are sufficient, the tax data are easier to work with, but quarterly data enable better measurement of employment rates over time. They also offer greater flexibility in aligning the data with the start of an evaluation and in differentiating between the in-program and post-program periods.

Figure II.1. Key features of earnings data collected by the Internal Revenue Service

Data source	Earnings data from Form 1040, the Employer Wage and Tax (W-2) and Miscellaneous Income (Form 1099-MISC) Statements
Agency	Internal Revenue Service (IRS)
URL	https://www.IRS.gov includes tax forms and instructions and links to extensive statistics derived from tax returns.
Key data	Annual taxable earnings by employer, with separate pre-tax contribution to retirement income; combined annual taxable earnings of filer and spouse, if present; total self-employment income separately for filer and spouse
Data description	IRS collects data on earnings directly from taxpayers via the Individual Income Tax Return (Form 1040) and from employers who file the Wage and Tax Statement (W-2) documenting the earnings paid to their employees and the Miscellaneous Income statement (Form 1099-MISC) documenting the income paid to non-employees. Employers submit W-2 data to the Social Security Administration (SSA), which processes the data for its own use and transmits the data to the IRS. All other documents are filed directly with the IRS. Form 1040 includes the earnings received by the filing unit (single filer or married couple) and the net profit or loss from a business or farm owned by the taxpayer (a sole proprietorship). Taxpayers with net business income of $400 or more are required to file Schedule SE to calculate the self-employment taxes that they owe as their contributions to Social Security and Medicare. Box 7 of Form 1099-MISC is used to report nonemployee compensation of $600 or more. Non-employee compensation is treated by the IRS as self-employment income, which taxpayers are required to report on Schedule C (as net profit or loss from a business) and Schedule SE. **Wage data sources:** W-2, 1099-MISC, and Form 1040 **Data duration:** no limit on retention **Frequency:** annual; latest data are from 2016 tax year (data collected and processed in 2017)
Universe and any known coverage issues	• Forms W-2 and 1099-MISC are required to be filed by employers for all individuals to whom they paid earnings and, therefore, will include earnings for individuals regardless of whether they file tax returns. • Compliance with the reporting of wages, salaries, and tips is very high. The most recent compliance estimates from the IRS, for the tax years 2008 to 2010, place the net misreporting of these earnings at just 1 percent (IRS 2016). • Compliance with the reporting of self-employment income is much lower. The net misreporting of nonfarm and farm income was estimated at 64 and 71 percent, respectively (IRS 2016). These sources are subject to very little third-party reporting to serve as a check on the accuracy of self-reporting. • Individuals and married couples whose annual income from all taxable sources exceeds the filing threshold for that year must file Form 1040. For the 2017 tax year, the filing thresholds were $10,400 for single individuals younger than 65 and $20,800 for couples younger than 65. • People with incomes below these thresholds might be required to file to obtain premium credits for health insurance coverage under the Affordable Care Act or to obtain refunds for excess withholding of federal taxes. • Larrimore et al. (2018) determined that 91 percent of the 2010 U.S. resident population appeared on a tax return (as primary filer, secondary filer, or dependent), and 99.5 appeared on either a tax return or an information document (for example, a W-2 or 1099). Cilke (2014), using a similar methodology, obtained the same 99.5 percent figure for 2011.

Figure II.1 (*continued*)

	• The earnings reported by employers in box 1 of Form W-2 and by taxpayers on Form 1040 exclude amounts deducted pre-tax for contributions to retirement plans, health insurance premiums, dependent care, and, depending on the employer, certain other benefits. However, pre-tax contributions to retirement plans are reported separately in box 12 and are included in Medicare wages reported in box 5. • The W-2 and 1040 earnings include taxable employee benefits such as the cost of life insurance in excess of $50,000.
Information included (for example, the availability and quality of information on self-employment)	• IRS record layouts are agency-confidential, but key information can be inferred from tax returns and the contents of the information documents filed by third parties. • All of these documents include the taxpayer's Social Security number (SSN). • Form W-2 includes the employee's annual taxable wages, Medicare-taxable wages (which adds to taxable wages the pre-tax contribution to a retirement account, which is also reported separately), and amounts of federal and state withholding. • Form 1099-MISC includes several types of income, but we focus on the annual earnings paid to a non-employee in box 7. • Form 1040 includes the annual earnings of the filer (and spouse, if a joint return), self-employment income from a business and farm (including the spouse, if a joint return), and total self-employment income for each filer (reported on Schedule SE). • All other sources of income included in Adjusted Gross Income are included as well.
Data linkage options	• The tax data held by the Census Bureau have been assigned Protected Identification Keys (PIKs). The PIK assigned to each record enables linking the earnings data at the person level to any file to which the Census Bureau has assigned this same unique identifier. This includes Census Bureau and selected other agency surveys and selected administrative data. • Permission to link these files must be secured from the responsible agency. • In theory, the user could also provide an external file with sufficient personally identifiable information for the Census Bureau to assign a PIK. These data could then be linked to the earnings data. Appropriate permissions would have to be obtained and documented before this could occur. • If performed at the IRS, linkage to external data without SSNs would be more challenging and possibly less complete and less accurate. • With either linkage option, the linked data must be analyzed on site; they cannot be removed. Estimates made from the data must undergo a disclosure review before they can be taken out of a secure environment.
Procedures and practices surrounding access	• Access to individual-level tax data from outside the IRS and the Treasury Department is governed by Section 6103 of the Internal Revenue Code (Title 26). Section 6103(j) provides for uses by employees of the U.S. Census Bureau "for the purpose of, but only to the extent necessary, in the structuring of censuses and national economic accounts and conducting related statistical activities authorized by law" (Internal Revenue Code). • There is no provision in the law for the IRS to provide microdata to people outside the agencies with access entitlements spelled out in Section 6103. • IRS contractors can be approved for access, but outside of being hired by the IRS, there are no standard procedures for obtaining access directly from the IRS. • Recently, the Statistics of Income (SOI) Division has provided research access to a small number of researchers following an application process.

Figure II.1 (*continued*)

	Like contractors, the researchers can obtain staff-like access following a clearance process. • The most common avenue used to gain access to IRS data is through the U.S. Census Bureau, which obtains selected fields from Form 1040 and information documents. • Researchers can obtain access to the IRS data held by the Census Bureau under Section 6103(j) by working through one of several Federal Statistical Research Data Centers (FSRDCs) distributed around the country. The administrator of the FSRDC at which the research is to be undertaken can provide information on access fees and the required form and content of a proposal. Information on how to apply for access may be found at: https://www.census.gov/ces/rdcresearch/howtoapply.html. In general, proposals must provide benefit to Census Bureau programs, demonstrate scientific merit, require nonpublic data, be feasible given the data, and pose no risk of disclosure. The following document describes the criteria that the Census Bureau and the IRS will apply in reviewing and approving projects that request the use of federal tax information: https://www2.census.gov/foia/ds_policies/ds002.pdf • Another avenue that has been used fruitfully, including by Mathematica, is to partner with an IRS or Treasury employee on a research project of mutual interest. The participants outside the two agencies do not obtain direct access to the data and are limited to viewing only those tabulations that have passed a disclosure review, but IRS and Treasury staff have access to far more data than are provided to the Census Bureau. Because it depends on the identification of individuals who may be interested in collaborating, this is clearly the least reliable means of access to tax data. • The following are major users with unrestricted access to federal tax information: Research, Applied Analytics, and Statistics within the IRS; the Office of Tax Analysis in the Treasury Department; the Joint Committee on Taxation of the U.S. Congress.
Timeliness of data, both typical length of time between data collection and potential availability, and (if possible) typical wait time from application to obtaining data	• Employers are required to file W-2s with SSA by the end of January of each year. Forms 1099-MISC are due to the IRS by that date as well. • Most tax returns are filed by the mid-April deadline, but processing of these and late returns extends through the end of the calendar year. • Research access to the full year of data on the IRS computer system is possible early the following year, with part-year data available during processing; the Census Bureau is now receiving monthly updates of tax data during processing, as having the most current data possible is critical for planned uses of tax data in conjunction with the decennial census
Key strengths and limitations of earnings data for analysis	• The principal strengths are the broad coverage (no segment of the workforce is excluded) and overall quality of the wage and salary data and their inclusion of income from self-employment. • The principal limitations are the exclusion of nonfederally taxable earnings, the absence of information on job characteristics (such as industry, occupation, and hours worked), the uncertain time lag for data accessed through an FSRDC, the substantial underreporting of self-employment income, and the possibility that if used at an FSRDC the data might exclude an individual's self-employment income below $400. • If accessing the data through the IRS, self-employment income without this exclusion can be obtained from the business income and farm income lines on Form 1040. • It is possible that the business and farm income fields are included among the data that the Census Bureau receives as well, but these data may not be available to outside researchers. The Census Bureau receives more extensive IRS data as an IRS contractor than it receives under Section 6103(j), and these data are used for joint IRS-Census Bureau research supporting tax administration.

Figure II.1 (*continued*)

Relevant law	• Internal Revenue Code (Title 26), Section 6103; available at https://www.law.cornell.edu/uscode/text/26/6103. Accessed August 23, 2018.
1 or 2 citations of research or that have us the data set, either seminal or cutting-edge work	• Bee, Adam, and Joshua Mitchell. "Do Older Americans Have More Income Than We Think?" SEHSD Working Paper 17-39. Washington, DC: U.S. Census Bureau, 2017. Available at https://www.census.gov/content/dam/Census/library/working-papers/2017/demo/SEHSD-WP2017-39.pdf. Accessed August 29, 2018. • Using an array of administrative data drawn from tax returns and information documents linked to a Current Population Survey Annual Social and Economic Supplement sample, the authors found that replacing survey responses with administrative data increased the estimate of median household income for householders 65 and older from $33,800 to $44,400 and reduced the poverty rate from 9.1 to 6.9 percent. The discrepancy was due mainly to survey underreporting of retirement income from defined benefit pensions and withdrawals from retirement accounts. The analysis used wages and self-employment income from SSA's Detailed Earnings Record, dividends and interest from Form 1040, and pension payments and retirement account withdrawals from Form 1099-R, which retirement accounts file with the IRS. • Manoli, Day, Marios Michaelides, and Ankur Patel. "Long-Term Effects of Job-Search Assistance: Experimental Evidence Using Administrative Tax Data." Unpublished manuscript, March 2018. Available at http://www.daymanoli.com/portfolio/nvrea/. Accessed August 29, 2018. • This paper used tax data to evaluate the long-term effects of an experimental job-search assistance program that operated in Nevada in 2009. Unemployed workers who had just started to receive UI were randomly assigned to receive or not receive personalized job-counseling services. Participation in the program produced substantial short-term reductions in UI receipt and persistent, long-term increases in employment and earnings, as measured by W-2. The analysis also found that program participants were more likely to file tax returns and had higher total family income and tax liability relative to the control group, but they were no more likely to engage in self-employment, as measured by Form 1099-MISC or reported self-employment on Form 1040.
References for descriptive information about database	Forms and instructions are available at https://www.irs.gov. Internal Revenue Code (Title 26), Section 6103. Abowd, John M., and Martha H. Stinson. "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data." *Review of Economics and Statistics,* vol. 95, December 2013, pp. 1451–1467. Cilke, James. "The Case of the Missing Strangers: What We Know and Don't Know about Non-Filers." Proceedings of the 107th Annual Conference of the National Tax Association, 2014. Available at https://www.ntanet.org/wp-content/uploads/proceedings/2014/029-cilke-case-missing-strangers-know-don.pdf. Accessed August 29, 2018. Internal Revenue Service. "Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2008-2010." Publication 1415. Washington, DC, May 2016. Available at https://www.irs.gov/pub/irs-soi/p1415.pdf. Accessed August 23, 2018. Larrimore, Jeff, Jacob Mortenson, and David Splinter. "Household Incomes in Tax Data: Using Addresses to Move from Tax Unit to Household Income Distributions." Paper presented at the Federal Committee for Statistical Methodology Research Conference, Washington, DC, January 2018.

Figure II.2. Key features of Social Security Administration earnings data

Data source	Summary Earnings Record (SER) and Detailed Earnings Record (DER)
Agency	Social Security Administration
URL	https://www.ssa.gov/dataexchange/
Key data	Annual federal taxable earnings, Medicare taxable earnings, and pre-tax contribution to retirement income, by employer; total self-employment income
Data description	The Social Security Administration (SSA) requires earnings data for individuals to administer its programs. Currently, these data are obtained from two sources. Employers file W-2 documents for their employees directly with the SSA, which processes these documents and transmits the data to the Internal Revenue Service (IRS), which uses these data to validate taxpayers' reports of earnings and tax withholding. The W-2 does not include self-employment, but as most self-employment is covered by Social Security, the SSA needs information on self-employment as well. This information is provided to SSA by the IRS in the form of data from Schedule SE, which the self-employed file with their individual tax returns (Form 1040). Unlike Form 1040, which collects the combined income for joint filers, Schedule SE is filer-specific. That is, Schedule SE identifies the self-employed individual by name and Social Security number (SSN). A joint return may have two Schedules SE if both filers had self-employment income during the tax year. Schedule SE reports not only the amount of self-employment income but the Social Security taxes owed on that income. These taxes are paid through the 1040 filing, as there is no mechanism in place for withholding taxes from self-employment income as it is received. The SSA compiles the W-2 and Schedule SE data in its Master Earnings File (MEF), which is a longitudinal record of the annual lifetime earnings of individuals in the United States. The SSA produces two extracts from the MEF: the Summary Earnings Record (SER) and the Detailed Earnings Record (DER). The SER contains annual earnings covered by Social Security, which are capped at the level of earnings subject to the Social Security tax. The SER includes earnings back to 1951. Before 1978, the SSA received quarterly earnings reports from employers. Starting in 1978, SSA began to receive earnings reports annually, on Form W-2, and the MEF began to include earnings above the taxable maximum. Self-employment income, subject to limitations discussed below, began to be included as well. The DER, which has annual earnings back to 1978, includes earnings not only for jobs covered by Social Security, but for jobs not covered by Social Security as well. The DER earnings are not capped. **Wage data sources:** W-2; Form 1040, Schedule SE **Data duration:** no limit on retention **Frequency:** annual; latest data are from 2016 tax year (data collected and processed in 2017)
Universe and any known coverage issues	• Some of the earnings reports received by SSA cannot be posted to the MEF because the SSNs cannot be validated against SSA's records of SSN issuances. In 2009, this fraction was estimated at 4 percent (Olsen and Hudson 2009). • If net earnings from self-employment (less 7.65 percent) are less than $400, an individual is not required to file Schedule SE and pay the Self-Employment Contributions Act (SECA); SSA will not receive any information from the IRS on those earnings. Such earnings are still subject to federal income tax but would be reported on line 12 of Form 1040, which SSA does not receive. • Before 1991, an individual paid no Social Security taxes on self-employment income that, in combination with earnings from employment, exceeded the taxable maximum, and the IRS sent only taxable self-employment income to SSA. Beginning in 1991, the earnings cap for payment of the Medicare tax was increased beyond the cap for the Social Security tax, and in 1994 the

Figure II.2 (*continued*)

	Medicare cap was eliminated entirely. In 1991, SSA began to receive data on a higher percentage of self-employment income, and since 1994 SSA has received all self-employment income reported on Form SE. • With the exception of contributions to tax-deferred retirement accounts, pre-tax deductions from earnings are excluded from both the SER and the DER, as these deductions are not reported on the W-2. Contributions to retirement accounts are excluded from the SER because they are not subject to the Social Security tax, but such contributions are reported on the W-2 under Medicare wages and tips because they are subject to the Medicare tax. They are included in the DER.
Information included (for example, the availability and quality of information on self-employment)	• Each record in the DER represents an individual worker's taxable calendar year earnings from a single employer or total self-employment. • Each record from an employer contains the worker's SSN, whether the job is covered by the Federal Insurance Contributions Act (FICA), federally taxable earnings, and pre-tax contributions to a retirement account. • Self-employment records do not contain retirement contributions. • Records with the same SSN can be summed within a given calendar year to obtain the individual's total earnings for the year. • Earnings reported in the SER are capped at the Social Security taxable maximum, which has increased over time
Data linkage options	• The presence of an SSN makes it possible to link these data to any other data source for which SSNs are available or can be obtained. • Linkage and analysis can be performed at SSA or through a Federal Statistical Research Data Center (FSRDC), as the Census Bureau obtains DER and SER data from SSA. • There are established procedures for obtaining access to DER and SER data through an FSRDC, which requires obtaining permission from both the Census Bureau and SSA. • Another advantage of obtaining access through an FSRDC is that the Census Bureau has developed and implemented procedures to facilitate record linkage using a Protected Identification Key (PIK)—a unique identifier created by the Census Bureau for record linkage within the Bureau. The Census Bureau uses personally identifiable information (PII) to assign the PIK and then removes the PII from the files that it makes available for linkage. Through the PIK a researcher working at an FSRDC can link DER and SER earnings data at the person level to any file to which the Census Bureau has assigned this same unique identifier. This includes Census Bureau and selected other agency surveys and selected administrative data. Permission to link these files must be secured from the responsible agency. • In theory, a user could also provide an external file with sufficient PII for the Census Bureau to assign a PIK. These data could then be linked to the earnings data. Appropriate permissions would have to be obtained and documented before this could occur.
Procedures and practices surrounding access	• Section 6103 of the Internal Revenue Code, which provides for access by SSA and Census Bureau employees, governs disclosure of tax return data. • SSA contractors may obtain access through SSA, but there do not appear to be established procedures for noncontractor access; data must be used on site at SSA and may be accessed only by an SSA employee. • Partnering with an SSA employee on a project of mutual interest may provide an alternative avenue of access; direct access to the data is restricted to the SSA employee. • Proposals to use these data at the Census Bureau must be submitted to the FSRDC at which the work will be performed. Information on how to apply for research access to DER and SER data at an FSRDC is available at https://www.census.gov/ces/rdcresearch/howtoapply.html.

Figure II.2 (*continued*)

	• The administrator of the FSRDC at which the research is to be undertaken can provide information on access fees and the required form and content of a proposal. In general, proposals must provide benefit to Census Bureau programs, demonstrate scientific merit, require nonpublic data, be feasible given the data, and pose no risk of disclosure. Guidance on defining the benefit to Census Bureau programs when proposing to use federal tax information is available at https://www2.census.gov/foia/ds_policies/ds002.pdf. • The linked data must be analyzed on site; they cannot be removed. Estimates made from the data must undergo a disclosure review before they can be taken from the FSRDC. This can take one to three weeks, ordinarily.
Timeliness of data, both typical length of time between data collection and potential availability, and (if possible) typical wait time from application to obtaining data	• Employers are required to file W-2s with SSA by the end of January of each year, but the IRS cannot provide self-employment data from Schedule SE until the returns filed during a calendar year (for the prior tax year) have been received and processed. • Except for returns filed after the year in which they were due, MEF updates will be completed during the first few months of the second year after the earnings were received. That is, the MEF will be fully updated with 2018 earnings in early 2020. • Transmitting DER and SER data to the Census Bureau follows with a lag that can vary from year to year. For example, the Census Bureau did not receive 2013 DER and SER data until late 2017. This was exceptionally long. • Another factor affecting timely access to DER and SER data through an FSRDC is the turnaround between the submission of a research proposal and the announcement of a decision, which can vary with the volume of submissions. • When a proposal is approved, members of the research team who will access the data or view results before the Census Bureau's disclosure review must apply for Special Sworn Status at the Census Bureau. Ordinarily this process will be completed within a few weeks, but it can take longer. It should be possible to complete this process while waiting for the data to arrive at the Census Bureau.
Key strengths and limitations of earnings data for analysis	• The principal strengths are the overall quality of the earnings data, their inclusion of self-employment, and their availability for a very large fraction of the employed and self-employed universe. • The principal limitations are the exclusion of nonfederally taxable earnings and self-employment earnings below $400, the substantial underreporting of self-employment income, the absence of information on job characteristics (such as industry, occupation, and hours worked), the uncertain time lag for data accessed through an FSRDC, and the lack of established procedures and the more limited linkage options if working with the data at SSA.
Relevant law	Internal Revenue Code (Title 26), Section 6103; available at https://www.law.cornell.edu/uscode/text/26/6103. Accessed August 23, 2018.
1 or 2 citations of research that have used the data set, either seminal or cutting-edge work	Abowd, John M., and Martha H. Stinson. "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data." *Review of Economics and Statistics,* vol. 95, December 2013, pp. 1451–1467. This paper compared survey reports of annual earnings from the Survey of Income and Program Participation (SIPP) to linked earnings data from the DER and differed from other such studies in that it did not assume that either source was error free. SIPP-reported jobs were linked to jobs in the DER using survey-reported employer characteristics and employer information obtained from the Census Bureau's Business Register using the employer identification number recorded in the DER. Comparisons were also made at the person level. DER wages can exclude pre-tax deductions (for example, for health insurance premiums) that the survey respondent would include in gross earnings, but DER

Figure II.2 (*continued*)

	wages can also include employee benefits that the employee does not receive as cash and that a survey respondent would be unlikely to include in gross earnings. Reliability statistics for the survey and administrative sources were comparable except when the survey data contained an imputation. The results indicate a need to allow for measurement error in both the survey and administrative sources when conducting validation studies. Bollinger, Christopher R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak. "Trouble in the Tails? What We Know about Earnings Nonresponse Thirty Years after Lillard, Smith, and Welch." *Journal of Political Economy*, forthcoming. This paper examines the consequences of nonresponse to the earnings questions in the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) on estimates of earnings gaps and inequality. The analysis uses CPS ASEC data linked to administrative data on earnings from SSA's DER. The authors reject the common assumption that earnings are missing at random, finding instead that nonresponse across the earnings distribution is U-shaped, highest in the left and right tails. Nonresponse is found to bias earnings differentials by race, gender, and education, particularly in the tails. Measures of inequality differ between household survey data and administrative data due in part to nonresponse and how it is compensated for by imputation.
References for descriptive information about database	Internal Revenue Service, instructions for Forms 1040, W-2, and Schedule SE. Available at https://www.irs.gov. Bridges, Benjamin, Linda Del Bene, and Michael V. Leonesio. "Evaluating the Accuracy of 1993 SIPP Earnings through the Use of Matched Social Security Administrative Data." 2002 Proceedings of the Joint Statistical Meetings. Alexandria, VA: American Statistical Association, 2003. Hotz, V. Joseph, and John Karl Scholz. "Measuring Employment and Income for Low-Income Populations with Administrative and Survey Data." In Michele Ver Ploeg, Robert A. Moffitt, and Constance F. Citro, eds., *Studies of Welfare Populations: Data Collection and Research Issues*. Washington, DC: National Academy Press, 2002. Available at https://www.nap.edu/catalog/10206/studies-of-welfare-populations-data-collection-and-research-issues. Accessed August 29, 2018. Olsen, Anya, and Russell Hudson. "Social Security Administration's Master Earnings File: Background Information." *Social Security Bulletin*, vol. 69, no. 3, March 2009, pp. 29–45. Available at https://www.ssa.gov/policy/docs/ssb/v69n3/v69n3p29.html. Accessed August 23, 2018.

Figure II.3. Key features of earnings data collected by state Unemployment Insurance agencies

Data source	State Unemployment Insurance (UI) quarterly wage reports
Agency	State UI agencies
URL	Various; for example, the website for Michigan is https://www.michigan.gov/uia/
Key data	Calendar quarter earnings by employer
Data description	By law, most employers are subject to a state UI tax and must report quarterly, to the state UI agency, the earnings of most of their employees. These records, which are maintained in machine-readable format, determine workers' eligibility for UI benefits when they apply for them. All states collect a small number of data elements, generally including employees' Social Security numbers (SSNs), the amount of wages paid to them in each quarter, and the employer's federal employer identification number (FEIN). Beginning in the early 1990s, there has been ongoing discussion about the possibility of including supplementary data elements to the wage records. By 2015, 13 states were offering enhanced data systems with up to ten data fields (such as occupational code and hourly pay rate[a]) in addition to the typical ones—SSN, quarterly earnings, and employer identification number.[b] **Wage data source:** state UI wage records **Data duration:** typically 6 or 7 quarters after each reporting quarter, after which data are archived **Frequency:** quarterly
Universe and any known coverage issues	• UI wage records do not cover all forms of employment. FUTA exempts certain kinds of employment from reporting, such as the employment of self-employed individuals, most independent contractors, military personnel, federal government workers, and railroad employees. States also have the ability to exempt additional types of employment from reporting to the state UI agency. Therefore, wage earnings from these types of employment are not contained in state UI wage records. • Generally, it is believed that a large majority[c] of workers in the U.S. economy are in jobs covered by the UI system. However, several studies find that employers failed to report many cases of casual or part-time workers, and did not report tips, bonuses, or other types of irregular compensation (see, for example, Blakemore et al., 1996, and Burgess et al., 1998). Employers have financial incentives to underreport earnings to state UI programs, because these earnings represent the basis for assessing a payroll tax that finances UI benefit payments. • The inability of UI data to fully capture workers in flexible staffing arrangements could lead to serious underreporting of the income and employment of low-income workers in studies that rely on UI data to measure these outcomes. • Underreporting of earnings in the UI data can also occur because the earnings of workers residing in one state and working in another are unlikely to be reported by the employer to the UI system in the state of the employee's residence (although a few states, such as Missouri and Kansas,[d] have arrangements that allow interstate data exchanges).
Information included (for example, the availability and quality of information on self-employment)	• Wage records contain calendar-quarter earnings by employer. • A few states report the number of hours or weeks worked during a reference quarter, and some states provide a North American Industry Classification System (NAICS) code for each reporting employer.

Figure II.3 (*continued*)

Data linkage options	• Data are linkable with any other administrative data source that contains SSNs, and an SSN is typically the only information required by state UI agencies to perform a data match. • If no other variables (such as name) exist, a match validation might not be possible.
Procedures and practices surrounding access	• State UI authorities control access to UI wage records and the SSNs necessary to link these data to other data sources for individuals. • States often require written consent before releasing individual UI wage records. However, many states have adopted regulations to permit more widespread access to individual-level UI data, without compromising confidentiality. For example, several states passed laws mandating the use of state UI wage records for documenting employment and earnings outcomes of participants in welfare and workforce development programs. • Ease of access began to improve in the 1980s and 1990s as more researchers—including members of the Administrative Data Research and Evaluation (ADARE)[e] project, as well as many others—began to develop ongoing partnerships with state agencies for data access that featured carefully structured agreements with confidentiality protections. • Currently, the DOL-funded Workforce Data Quality Initiative (WDQI)[f] continues ADARE's legacy in aiming to build state longitudinal data systems that integrate education and workforce data (which in many states includes UI data). Procedures for accessing these systems for research and evaluation differ by state (here is an example for Indiana[g]).
Timeliness of data, both typical length of time between data collection and potential availability, and (if possible) typical wait time from application to obtaining data	• Employers are expected to provide data to the state UI agency within one month of the end of a quarter. The state UI agency then has two additional months to process the data before they can be available for internal administrative use in managing the state's unemployment compensation program. Therefore, in principle, the state UI wage records could be available for use within a quarter after submission. However, agencies have many data processing demands that affect their capacity to respond to external requests for data. Thus, delays often occur in the release of these files. • The combined effect of these processes means that wage records for a given calendar quarter are generally not available for about six months (two quarters) after submission. • Because the data needed to determine UI eligibility cover a period that is usually defined as the first four of the last five quarters, most states maintain direct access to their UI wage records for at most six or seven quarters following the quarter in which wages are measured. After that, the data are archived, making their retrieval considerably more difficult.
Key strengths and limitations of earnings data for analysis	• An advantage of state UI wage data is that they are relatively inexpensive to collect compared to individual surveys and thus can be collected for a large sample for a long period of time. • UI earnings data are quarterly, providing more flexibility to align them with the date of enrollment in a program. • They also are fairly uniform across states and over time, a characteristic that facilitates analysis. • Compared to survey data, UI data are not subject to the potential biases that can occur because of recall error and survey nonresponse. • However, as noted above, although UI wage records cover a large majority of workers, some important categories of workers are excluded, including federal employees, military staff members, and the self-employed. • Also, the records will miss earnings if the SSN is reported incorrectly by the worker or employer (state UI agencies do not verify reported SSNs) or because of data-handling errors.

Figure II.3 (*continued*)

	• the absence of information on job characteristics (such as industry, occupation, and hours worked), • Finally, the UI wage records in any given state do not cover an individual's earnings in another state. • Another disadvantage of the UI wage records data is that they are not stored centrally across states; instead, they must be collected separately from each state. Making requests to every state and dealing with individual state access protocols can make acquisition costs expensive for studies that involve many states, albeit much less than the cost of designing and conducting surveys to obtain earnings data.
Relevant Law	Federal Unemployment Tax Act (https://www.law.cornell.edu/uscode/text/26/subtitle-C/chapter-23) defines the conditions under which employers must pay UI. Individual states adopt legislation that conforms to the Federal Unemployment Tax Act (FUTA). For example, California Unemployment Insurance Code Section 1088 describes reporting requirements for employers: (https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=UIC&sectionNum=1088).
1 or 2 citations of research that have used the data set, either seminal or cutting-edge work	Schochet, Peter Z., John Burghardt, and Sheena McConnell. "Does Job Corps Work? Impact Findings from the National Job Corps Study." *American Economic Review,* vol. 98, no. 5, 2008, pp. 1864–1886. This paper presented findings from an experimental evaluation of Job Corps, the nation's largest training program for disadvantaged youths. The study used survey data collected over four years and earnings data over nine years on a nationwide sample of 15,400 treatments and controls. The Job Corps model was shown to have promise; program participation increased educational attainment, reduced criminal activity, and increased earnings for several post-program years (as measured by UI earnings data). Based on data from the MEF maintained by SSA, however, the earnings gains were not sustained except for the oldest participants. Nonetheless, Job Corps is the only federal training program that has been shown to increase earnings for this population. The study relied in part on quarterly wage records reported by employers to state UI agencies in 22 randomly selected states. Heinrich, Carolyn J., Peter R. Mueser, Kenneth R. Troske, Kyung-Seong Jeon, and Daver C. Kahvecioglu. "Do Public Employment and Training Programs Work?" *IZA Journal of Labor Economics,* vol. 2, no. 1, 2013, p. 6. This paper estimated impacts on earnings and employment of the two primary adult workforce support and training programs under the U.S. Workforce Investment Act (WIA) using administrative data on 160,000 participants from 12 states for up to four years following program entry. The authors found that participants in the WIA Adult program, who typically enter with poor work histories, realized improved employment levels and increased average quarterly earnings of several hundred dollars. Earnings gains for Dislocated Worker program participants were appreciably smaller, although these participants did experience employment gains.
References for descriptive information about database	Blakemore, A. E., Burgess, P. L., Low, S. A., & St. Louis, R. D. "Employer Tax Evasion in the Unemployment Insurance Program." *Journal of Labor Economics*, vol. 14, no. 2, 1996, pp. 210-230. Burgess, P. L., Blakemore, A. E., & Low, S. A. "Using Statistical Profiles to Improve Unemployment Insurance Tax Compliance." *Research in Employment Policy*, vol. 1, 1998, pp. 243-271. Hotz, V. Joseph, and John Karl Scholz. "Measuring Employment and Income for Low-Income Populations with Administrative and Survey Data." *Studies of Welfare Populations: Data Collection and Research Issues.* Washington, DC: National Academy Press, 2002. Available at https://www.nap.edu/catalog/10206/studies-of-welfare-populations-data-collection-and-research-issues. Accessed August 29, 2018.

Figure II.3 (*continued*)

	Kornfeld, Robert, and Howard S. Bloom. "Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Reports from Employers Agree with Surveys of Individuals?" *Journal of Labor Economics,* vol. 17, no. 1, 1999, pp. 168–197.

[a] Indiana, Louisiana, and Washington are examples of states that have implemented enhanced data collection.

[b] Administrative Wage Record Enhancement Study Group. "Enhancing Unemployment Insurance Wage Records: Potential Benefits, Barriers, and Opportunities". Washington, D.C.: Workforce Information Council, 2015. Available at https://www.bls.gov/advisory/bloc/enhancing-unemployment-insurance-wage-records.pdf.

[c] Kornfeld and Bloom (1999) estimate coverage at 90 percent.

[d] Wallace, Geoffrey L., and Robert Haveman. "Work and Earnings of Low-Skilled Women: Do Employee and Employer Reports Provide Consistent Information?" *Journal of Economic and Social Measurement,* vol. 32, nos. 2, 3, 2007, pp. 149–176.

[e] http://www.ubalt.edu/jfi/adare/

[f] https://www.doleta.gov/performance/workforcedataquality.cfm

[g] https://www.in.gov/mph/935.htm

Figure II.4. Key features of the National Directory of New Hires data

Data source	National Directory of New Hires (NDNH)
Agency	U.S. Department of Health and Human Services (HHS) Office of Child Support Enforcement (OCSE)
URL	https://www.acf.hhs.gov/css/resource/overview-of-national-directory-of-new-hires
Key data	New hire file • Employee date of hire • Employee state of hire Quarterly wage file • Quarterly employee wage amount Unemployment Insurance file • Benefit amount
Data description	The NDNH is a repository of employment, unemployment insurance (UI), and quarterly wage data maintained by the Office of Child Support Enforcement (OCSE). It combines three types of data: employment information on all newly hired employees as reported by employers, quarterly wage information on individual employees, and UI information on individuals who received or applied for unemployment benefits. The information is derived from W-4 records submitted by employers to the State Directory of New Hires, quarterly wage and unemployment insurance data from the state workforce agencies, and new hire and quarterly wage data from federal agencies. **Data source:** state UI wage records, State Directory of New Hires, data from federal agencies. **Data duration:** 12 to 24 months **Frequency:** Quarterly, with a lag of up to 4.5 months
Universe and any known coverage issues	• The universe is formal employment of individuals, but excludes self-employed individuals, most independent contractors, railroad employees, some part-time employees of nonprofit institutions, employees of religious orders, and some students employed by their schools. • The NDNH does not include employment data for people without valid Social Security numbers. However, 97% of all records submitted to the NDNH are available for matching (Thompson 2012). • Quarterly wage and unemployment insurance data derived from the state UI agencies are susceptible to universe and coverage issues due to under-reporting (see Figure II.5 for details)
Information included (for example, the availability and quality of information on self-employment)	• The new hire file contains information on the employee's date and state of hire, as well as some information about the employer, including the employer's federal employer identification number (FEIN) and city and state. • The quarterly wage file includes information on the gross quarterly wage amount and the employer's FEIN; when an individual works more than one job during the reporting period, separate records are established for each job. • The UI file contains quarterly information on all UI claim applications to state workforce agencies (even if the claim is rejected, denied, or suspended), and includes information on the claimant's gross benefit amount and the reporting period.
Data linkage options	• The structure and content of the files enables combining the quarterly earnings from an employee's multiple employers.

Figure II.4 (*continued*)

	• In theory, the user can also provide an external pass-through file with sufficient personally identifiable information for OCSE to match the external file to individuals' records in the NDNH, thus enabling external data to be linked to the new hire, quarterly wage, or UI files. • OCSE has strict requirements about the content of external pass-through files, with the aim of limiting disclosure from data linkage. After the matching occurs, OCSE returns the pass-through file, stripped of all identifying information and including a masked identifier generated by OCSE, that can be used to link to the de-identified NDNH file given to the user that includes he same masked ID.
Procedures and practices surrounding access (as of publication)	• Access to NDNH data are explicitly limited by statute. Title IV-D of the Social Security Act specifies that researchers outside covered federal agencies may be given access only to de-identified NDNH information "*to conduct research found by the Secretary of Health and Human Services (HHS) to be likely to contribute to achieving the purposes of part A or part D of the Social Security Act*" (42 U.S.C. §653(j)(5)), that is, to contribute to achieving the mission of Temporary Assistance for Needy Families and Child Support Enforcement programs. • Before accessing NDNH information, users must sign an agreement or memorandum of understanding with OCSE that describes the purpose, legal authority, justification, expected results of the match, description of the records, retention and disposition of information, reimbursement, and user's performance reporting requirements, as well as a security addendum that details the security requirements and safeguards that users must have in place before receiving NDNH information. • OCSE also requires NDNH users to provide a written description of the performance outputs and outcomes attributable to using NDNH information. • Users must reimburse OCSE for the costs of obtaining, verifying, maintaining, and comparing the information.
Timeliness of data, both typical length of time between data collection and potential availability, and (if possible) typical wait time from application to obtaining data	• There can be a lag of up to 4.5 months from the end of a calendar quarter before data become available in the NDNH.[a] • All information entered into the NDNH is purged within 24 months; quarterly wage and UI data that do not result in a child support match are often purged within 12 months. However, when a memorandum of understanding is signed, OCSE can retain data that is part of an ongoing research sample for specific users. • Obtaining authorization for the data and working out the details of the implementation process (such as the fee structure and testing and validation of the match) can take several months.
Key strengths and limitations of earnings data for analysis	• The key strengths of the NDNH are the quality of the quarterly wage data, the uniqueness of the new hires data, the availability of data for a large fraction of the employed universe, their capture of gross earnings, and the information on the employers. • A key advantage of using the NDNH instead of the State Directory of New Hires or state UI quarterly wage data is that information is available about people who have obtained work or claimed UI in another state. This is particularly helpful for multistate evaluations or studies targeting populations that are highly mobile, likely to commute across state lines, or work for employers operating in multiples states.[b] • An advantage over the LEHD and state UI records is that the NDNH covers those employed by federal or military agencies. • The principal limitation of the NDNH is the inability to obtain identified data for research purposes. Because OCSE provides researchers only de-identified files, it is impossible for researchers to later incorporate additional years of data, link to new sources of data, or even correct problems with prior linkages after the de-identified file with NDNH data has been returned.

Figure II.4 (*continued*)

	• Other limitations include the tight time frame restrictions around data deletion, requests, and usage, which can complicate the research process; and the absence of information on job characteristics (such as industry, occupation, and hours worked). • Because data are purged every 12 to 24 months, data of interest can sometimes be deleted before a research agreement can be reached, and the NDNH is a poor source of historical data as it contains only up to two years of data at the time of application for access. • The quarterly nature of the Quarterly Wage data and the sparseness of the New Hire data complicate the measurement and analysis of outcomes such as time to re-employment.
1 or 2 citations of research that have used the data set, either seminal or cutting-edge work	Fein, D., and J. Hamadyk. "Bridging the Opportunity Divide for Low-Income Youth: Implementation and Early Impacts of the Year Up Program." OPRE Report #2018-65. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2018. This paper combined quarterly wage data from the NDNH with college enrollment records from the National Student Clearinghouse and data from an 18-month follow-up survey to examine the impacts of a national sectoral training program for urban young adults ages 18 to 24 with a high school diploma or equivalent education. The study used a randomized controlled trial, and the NDNH data enabled authors to track youth's quarterly earnings for the 12 quarters following their random assignment to either the treatment or control group. The authors found that youth's earnings were expectedly lower in the first year following random assignment when the treatment group was likely to prioritize the training program over work, but the program generated an increase of 53 percent in average earnings in the sixth and seventh quarters following random assignment. Impacts remained large (about 40 percent) over the following year, with large positive impacts persisting in every quarter through the end of the third year. Martinson K., J. Williams, K. Needels, Laura Peck, Shawn Moulton, Nora Paxton, Annalisa Mastri, Elizabeth Copson, Hiren Nisar, Alison Comfort, and Melanie Brown-Lyons. "The Green Jobs and Health Care Impact Evaluation: Findings from the Impact Study of Four Training Programs for Unemployed and Disadvantaged Workers." Princeton, NJ: Mathematica Policy Research, 2016. This paper combined quarterly wage data from the NDNH with baseline intake information and data from an 18-month follow-up survey to study four grantees funded by two job training initiatives administered by the Employment and Training Administration (ETA) at the U.S. Department of Labor (DOL). The evaluation used a random assignment research design and cumulative earnings in the fifth and sixth calendar quarters (13 to 18 months) after random assignment, as measured in NDNH data, was designated as the confirmatory outcome of program effectiveness. The study used four quarters of NDNH data collected before random assignment, and six quarters of NDNH data collected after random assignment, covering almost the entire study sample—all but about 3 percent of individuals. For three of the grantees, the authors found no significant differences in earnings between the treatment and control groups over the follow-up period. For one grantee, the authors found the program increased earnings during the fifth and sixth calendar quarters after random assignment by about 20 percent. Using NDNH data, the authors showed that earnings decreased during the first quarter after random assignment, reflecting enrollment and participation in the training program, peaked in the fourth quarter after random assignment, and then declined.
Relevant Law	The Social Security Act [42 U.S.C. 653, §453(i)]. The Personal Responsibility and Work Opportunity Reconciliation Act of 1996: vol. II, P.L. 104-193, §316(h).
References for descriptive information about database	"Federal Parent Locator Service. A Guide to the National Directory of New Hires." Administration for Children and Families, Office of Child Support

Figure II.4 (*continued*)

	Enforcement. Washington, DC, 2017. Available at https://www.acf.hhs.gov/sites/default/files/programs/css/a_guide_to_the_national_directory_of_new_hires.pdf. Accessed August 23, 2018. "Federal Parent Locator Service. National Directory of New Hires: Guide for Data Submission. Version 13.2." Washington, DC: Administration for Children and Families, Office of Child Support Enforcement, 2016. Available at https://www.acf.hhs.gov/sites/default/files/ocse/ndnh_guide_for_data_submission.pdf. Accessed August 23, 2018. Durham, C., and Laura Wheaton. "Investigating Alternative Sources of Quarterly Wage Data: An overview of the NDNH, LEHD, WRIS, and ADARE." New York: Urban Institute, 2012. Solomon-Fears, Carmen. "The National Directory of New Hires." Congressional Research Service Report RS 22889, prepared for Members and Committees of Congress, February 24, 2014. Available at https://greenbook-waysandmeans.house.gov/sites/greenbook.waysandmeans.house.gov/files/RS22889_gb.pdf. Accessed August 23, 2018. Thompson, Lynette. "New Hire Reporting: What Happens With All That Data?" Presentation by the Federal Office of Child Support Enforcement. Available at https://www.naswa.org/assets/utilities/serve.cfm?gid=D61DC16C-C1B1-46B4-A61D-7FC49F0C4496&save=1&dsp_meta=0. Accessed September 23, 2018.

[a] For example, if an employer sends its quarterly wage data late (in a subsequent quarter), the state workforce agency sends the delinquent data to the NDNH, which processes all delinquent data sent along with the current data.

[b] An employer operating in multiple states, such as many retail or fast food enterprises, has the choice of reporting employment in the state where each employee works or reporting on all employees to one state employment security agency.

Figure II.5. Key features of Longitudinal Employer-Household Dynamics data

Data source	Longitudinal Employer-Household Dynamics program (LEHD)
Agency	U.S. Census Bureau
URL	https://lehd.ces.census.gov/
Key data	Quarterly and annual earnings by employer—identified by state Federal Information Processing Standard (FIPS) code and state employer ID; a U.S. Census Bureau-assigned identifier to enable linkage to other Census Bureau files
Data description	The LEHD program combines quarterly earnings data obtained from state Unemployment Insurance (UI) system records, quarterly employment and wage information obtained from the Quarterly Census of Employment and Wages (QCEW), and other information obtained from the decennial census and Census Bureau surveys. The earnings data are derived from the quarterly reports that eligible employers submit to their state employment security offices. These reports provide earnings by month of the quarter for each employee, who is identified by Social Security number and name. The focus of the program is to produce a wide range of data describing local labor market conditions at a very detailed geographic level, but restricted-use microdata from the program are made available as well. This makes the LEHD one of three potential sources of state UI wage data. **Wage data source:** state UI wage records, QC EW, Census Bureau surveys **Data duration:** no limit on retention, but data are not current **Frequency:** quarterly and annual; latest year at present is 2011
Universe and any known coverage issues	• The universe is employment covered by the UI systems of the states. This generally excludes federal and military employment, self-employment, independent contractors, and other employment not covered by the state UI system. Coverage of private wage and salary employment has been estimated at 96 percent (Stevens 2007). • Employment is captured in the state in which a job is located, not the state in which the employee lives. • Data from Massachusetts are not included. • There is some potential additional loss of coverage due to how the Census Bureau designed the LEHD. Employment data for people without valid Social Security numbers could not be included in the LEHD.
Information included (for example, the availability and quality of information on self-employment)	• The variables provided are the employee's annual earnings from that employer, the calendar year, a Protected Identification Key (PIK) unique to the employee, the earnings in each quarter, the state FIPS code, and the state employer ID. • Data include quarterly earnings at the employee and job levels. • The microdata files available through the LEHD program (the Employment History Files) are organized by employee, employer, and year. There is a separate file for each state and the District of Columbia, except for Massachusetts, which did not participate in the LEHD.
Data linkage options	• The PIK included in each record enables linking the earnings data at the person level to any file to which the Census Bureau has assigned this same unique identifier. This includes Census Bureau and selected other agency surveys and selected administrative data. • Permission to link these files must be secured from the responsible agency. • In theory, the user could also provide an external file with sufficient personally identifiable information for the Census Bureau to assign a PIK. These data could then be linked to the earnings data. Appropriate permissions would have to be obtained and documented before this could occur.

Figure II.5 (*continued*)

Procedures and practices surrounding access (as of publication)	• From the LEHD website: "Access to these data will only be granted to qualified researchers on approved projects with authorization to use specific data sets. All research access to restricted-use data occurs at one of the secure Federal Statistical Research Data Centers (FSRDCs)." • Proposals to use these data must be submitted to the FSRDC at which the work will be performed. Information on how to apply for research access to the LEHD data at an FSRDC is available at https://www.census.gov/ces/rdcresearch/howtoapply.html. • The administrator of the FSRDC at which the research is to be undertaken can provide information on access fees and the required form and content of a proposal. In general, proposals must provide benefit to Census Bureau programs, demonstrate scientific merit, require nonpublic data, be feasible given the data, and pose no risk of disclosure.
Timeliness of data, both typical length of time between data collection and potential availability, and (if possible) typical wait time from application to obtaining data	• The LEHD is best viewed as a source of historical rather than current data. The data made available to researchers are updated only periodically—that is, not on a regular basis. The restricted use microdata included in the LEHD history files cover the years 1985 to 2011, reflecting the most recent update. • The turnaround between submitting a research proposal and the announcement of a decision varies with the volume of submissions. • After receiving a decision to grant access, the applicant must apply for Special Sworn Status at the Census Bureau to access the data through an FSRDC. This process can take several weeks to several months.
Key strengths and limitations of earnings data for analysis	• The principal strengths are the quality of the earnings data, their availability for a large fraction of the employed universe, their capture of gross earnings, the information on employers' characteristics that has been linked to the earnings data, and the number of years for which these data are available. • The principal limitation is the absence of regular updates, which means that the data are not current and, therefore, cannot be used to evaluate current programs. • Other limitations are the universe exclusions (federal and military employment, self-employment, other noncovered or unreported jobs and, for LEHD specifically, employment in Massachusetts); and the absence of information on job characteristics (such as industry, occupation, and hours worked).
Relevant law	Social Security Act, Section 303, available at: https://www.ssa.gov/OP_Home/ssact/title03/0303.htm.
1 or 2 citations of research that have used the data set, either seminal or cutting-edge work	Abowd, John, Kevin L. McKinney, and Nellie L. Zhao. "Earnings Inequality and Mobility Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data." Center for Economic Studies Discussion Paper CES 17-24. Washington, DC: U.S. Census Bureau, March 2017. Available at https://ideas.repec.org/p/cen/wpaper/17-24.html. Accessed August 23, 2018. Using earnings data and linked employer characteristics from the LEHD, this paper analyzed the role of the employer in explaining the rise in earnings inequality in the United States. The authors found, first, that the trends in earnings inequality in the LEHD were inconsistent with those for likely UI-eligible workers in other data sources unless they removed LEHD records with invalid or fraudulently used SSNs, as determined from a validity check against SSA data. Second, they found that including inactive spells for workers who were otherwise active was necessary to properly capture changes in earnings at the lower end of the distribution. Third, they found that although workers of all skill types benefitted from working at a middle-paying versus low-paying firm, the gains from working at a top-paying firm were greatest. Andersson, Fredrik, Harry J. Holzer, Julia I. Lane, David Rosenblum, and Jeffrey Smith. "Does Federally-Funded Job Training Work? Nonexperimental Estimates of WIA Training Impacts Using Longitudinal Data on Workers and Firms." Center

Figure II.5 (*continued*)

	for Economic Studies Discussion Paper CES 18-02. Washington, DC: U.S. Census Bureau, January 2018. Available at https://ideas.repec.org/p/cen/wpaper/18-02.html. Accessed August 23, 2018. This paper examined the impact of training provided under the WIA using LEHD data for two unnamed states. The paper estimated impacts of WIA training versus nontraining services conditional on WIA participation. Outcomes included not just earnings and employment, but the types of firms at which WIA participants obtained jobs. The authors found modest positive impacts of training among those served under the adult funding stream but small negative impacts for those served under the dislocated worker funding stream. Expanding the set of variables on which the receipt of WIA training was conditioned (variables based on the last employer and four additional quarters of lagged earnings) did not alter the findings.
References for descriptive information about database	Stevens, David. "Employment That Is Not Covered by State Unemployment Insurance Laws." LEHD Technical Paper No. TP-2007-04. Washington, DC: U.S. Census Bureau, 2007. Available at https://econpapers.repec.org/paper/centpaper/2007-04.htm. Accessed August 29, 2018. Vilhuber, Lars, and Kevin McKinney. "LEHD Infrastructure Files in the Census RDC—Overview." Center for Economic Studies Discussion Paper CES 14-26. Washington: U.S. Census Bureau, June 2014. Available at https://ideas.repec.org/p/cen/wpaper/14-26.html. Accessed August 29, 2018.

III. THE CUSTOM SURVEY OPTION: AN EXAMPLE

Researchers conducting program evaluations with random assignment frequently conduct their own surveys of participants to compare the treatment and control group members with respect to outcomes of interest, which often include variables for which there are no administrative data. Even for outcomes for which administrative data exist, there might be reason to collect such data through a survey, either as an alternative to obtaining administrative data or to complement the administrative data, given that each source has comparative strengths and limitations. We discuss a recent survey conducted as part of an evaluation.

Designing and administering a survey to collect data on the employment and earnings of participants in an impact evaluation affords researchers the opportunity to tailor their data collection to the full range of issues addressed in their evaluation. Conducting a survey is a costly alternative to obtaining data from administrative records, but often the richer data that can be obtained from a custom survey provide sufficient reason to incur the added cost. Additional data elements that can be captured in a survey but are not available from administrative data include hours worked, wage rates, and job characteristics.

Recent research has drawn attention to another consideration, which is that outcomes measured with surveys instead of administrative records yield different estimates of program impacts. In an extensive review of the literature on estimates of program impacts, Barnow and Greenberg (2015) found that estimates based on surveys tended to be larger and more likely to be statistically significant than estimates based on administrative data. These differences could reflect any of a number of factors, including differential coverage of jobs, incomplete or erroneous matching of administrative data to participants, survey nonresponse, and measurement error (for example, recall error, rounding, deliberate misreporting, or proxy responses).

To shed further light on this issue, Mastri et al. (2018) performed a reanalysis of program impacts from the recent evaluation of the Workforce Investment Act (WIA) Adult and Dislocated Worker programs conducted for the U.S. Department of Labor. Participants were assigned at random to receive one of three tiers of services and were then surveyed at 15 and 30 months after random assignment to collect information on their work experience and earnings as well as other possible outcomes during the interim.

The employment section from the 15-month follow-up survey is reproduced in the Appendix. These questions cover employment during the 15 months following random assignment. For the current job and up to four additional jobs the respondent may have held during this period, the survey captured:

- Starting and ending dates
- The approximate fraction of time working while holding that job
- The usual number of hours worked per week
- The average number of days worked per week
- The kind of business or industry
- What the respondent did on the job
- The type of employment (for example, as a regular employee, on-call employee, temporary or contract employee, independent contractor, or day laborer)
- Most recent rate of pay (average amount and period)
- Types of benefits provided
- Union membership

The employment detail captured during a reference period specific to the study contrasts with the more limited employment information collected in the Current Population Survey Annual Social and Economic Supplement (CPS ASEC), discussed in the next chapter, for the prior calendar year. For the longest job and/or longest business during that period the CPS ASEC captures total earnings, industry and occupation, class of worker, and average hours worked. For

the calendar year as a whole the survey captures weeks worked full-time, part-time and total; weeks unemployed and weeks not in labor force; and total earnings from all other jobs and self-employment. The additional detail captured in the WIA survey permits a much richer analysis of program outcomes than would have been possible with the more limited data collected with the CPS ASEC instrument. Arguably, there is reason to be skeptical about the ability of respondents to provide uniformly accurate responses to questions about earnings and hours for as many as five jobs over a 15-month period. At the same time, however, respondents who had multiple jobs during the period may be less challenged in being asked to report a separate pay rate for each job than by having to recall and mentally add the aggregate earnings across these same jobs.

The reanalysis conducted by Mastri et al. compared results using data from the 15- and 30-month follow-up surveys and two sources of administrative data: the NDNH and information drawn from W-2s and Forms 1099-MISC, although the tax data ultimately played a very limited role in the reanalysis owing to logistical limitations.

Findings from the study provide some perspective on the relative advantages and disadvantages of survey versus administrative data for estimates of employment and earnings. Although both the survey and NDNH data showed a positive impact of WIA-funded intensive services on the earnings of customers, the magnitude of the impact was greater with the survey data than the administrative data (Mastri et al. 2018). This difference derived in large part from three factors:

1. Within a given job, survey-reported earnings exceed those reported in the NDNH early in the follow-up period.
2. Survey respondents often underreported the number of jobs held early in the follow-up period.
3. Many survey respondents worked in jobs that the NDNH does not cover, including self-employment, independent contracting, informal work, day labor, and other alternative work arrangements.

Recall error by survey respondents was suggested as a factor in the first two, whereas coverage limitations of the NDNH were clearly a factor in the third. These findings demonstrate how including both survey and administrative data can enhance an impact evaluation if they are used in ways that reflect their relative strengths and limitations.

IV. SURVEY DATA ON EARNINGS

This review includes selected federal household survey data primarily because of their relevance to population-based analyses of earnings. Generally, evaluations of training programs cannot use household survey data because the survey samples are not large enough to include sufficient numbers of participants in these programs. However, survey data linked to administrative data on wages can enhance the analytical value of the administrative data for analyses targeting large segments of the population.

We include matrices on three surveys conducted by the U.S. Census Bureau:

1. The CPS ASEC, a supplement to the monthly labor force survey conducted in February through April of each year and which serves as the official source of statistics on household income and poverty for the United States
2. The Survey of Income and Program Participation (SIPP), a longitudinal survey that follows a sample of households over a period of two to four years and collects detailed monthly data on employment, income from a wide variety of sources, participation in major benefit programs, and household composition
3. The American Community Survey (ACS), an annual survey that replaced the decennial census long form following the 2000 census and collects demographic, housing, and income data from 2.2 million households plus a sample of group-quarters residents each year

Figure IV.1 provides an overview of the CPS ASEC, Figure IV.2 does the same for the SIPP, and Figure IV.3 does so for the ACS. For each of these three surveys the matrix provides the following information:

- Responsible agency
- Survey web address
- Purpose
- Summary of design
- Universe
- Sample size
- Response rate
- People covered in work and income data

- Data on work activity and earnings
- Percentage of earnings dollars imputed
- Income data
- Job search and training
- Link to technical documentation and survey contents
- References for descriptive information about survey

Each of these surveys can be linked to administrative data held by the Census Bureau using a unique identifier, the Protected Identification Key, which the Census Bureau assigns to individual records in the internal versions of these surveys and to many of the administrative data sets that it obtains from other federal agencies and the states. Researchers must perform the linkage work at the Census Bureau, but they can access servers that store the data remotely from any of several FSRDCs located around the country. The administrative data matrices provide a link to a Census Bureau website with information on using the FSRDCs. Obtaining access to such data requires submitting a proposal and receiving approval from the relevant agencies.

Figure IV.1. Key features of the Current Population Survey Annual Social and Economic Supplement (CPS ASEC)

Responsible agency	U.S. Census Bureau and the Bureau of Labor Statistics
Survey URL	https://www.census.gov/programs-surveys/cps.html
Purpose	Primary source of detailed information on income and work experience in the United States; source of official household income and poverty statistics and most widely cited estimates of the uninsured; used extensively for policy analysis and legislative cost estimates and as the basis for major micro-simulation models; underlying monthly CPS is the source of official labor force, unemployment, and wage rate estimates
Summary of design	Annual supplement conducted in February through April (formerly known as the March supplement when it was conducted solely in March; a 2002 sample expansion added interviews in February and April) collecting detailed income and employment data for the prior calendar year and current demographic and labor force data; monthly CPS interview is conducted first, followed by the supplement, which some of the monthly respondents decline to complete
Universe	Resident civilian non-institutionalized population of the United States plus military living with civilian family members on or off base
Sample size	2017 CPS ASEC: 70,000 households interviewed
Response rate	2017 basic monthly CPS: 86.5 percent 2017 basic CPS and supplement: 74.4 percent
People covered in work and income data	All persons ages 15 or older
Data on work activity and earnings	Detail on longest job and longest business in prior calendar year, including earnings from longest job, from other work, and from all jobs plus self-employment earnings; weeks worked full-time, part-time and total; weeks unemployed and weeks not in labor force; industry and occupation codes, class of worker, and usual hours worked for longest job in prior year and current job; earnings per week, hours last week and wage rate for current job
Percentage of earnings dollars imputed	2014 CPS ASEC: 37.5% Wages and salaries 47.9% Self-employment
Income data	More than 50 sources and up to 24 amounts of income (prior calendar year), by person
Job search and training	In prior calendar year for each person
Key strengths and limitations of earnings data for analysis	Key strengths are the survey's status as the official source of monthly labor force, unemployment, and wage rate estimates; the overall quality of the annual earnings data; and the availability of information on hours worked, occupation, and industry. Key limitations are the significant nonresponse at the unit level; the high imputation rate for both wages and salaries and self-employment; and the fact that, as a sample, the data will not include the participants in program evaluations.
Link to technical documentation and survey contents	https://www.census.gov/programs-surveys/cps/technical-documentation/complete.html
References for descriptive information about survey	Czajka, John L., and Gabrielle Denmead. "Income Data for Policy Analysis: A Comparative Assessment of Eight Surveys." Final report. Washington, DC: Mathematica Policy Research, 2008. Available at https://www.mathematica-mpr.com/our-publications-and-findings/publications/income-data-for-

Figure IV.1 (*continued*)

	policy-analysis-a-comparative-assessment-of-eight-surveys. Accessed August 23, 2018. National Academies of Sciences, Engineering, and Medicine. *The 2014 Redesign of the Survey of Income and Program Participation: An Assessment.* Washington, DC: The National Academies Press, 2017.Available at https://doi.org/10.17226/24864 U.S. Census Bureau. "Current Population Survey: 2017 Annual Social and Economic Supplement." Available at https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar17.pdf. Accessed August 24, 2018.

Figure IV.2. Key features of the Survey of Income and Program Participation (SIPP)

Responsible agency	U.S. Census Bureau
Survey URL	https://www.census.gov/sipp/
Purpose	To provide information on income, program eligibility, and participation to help formulate public policy; to measure within-year dynamics of income and program participation and a broad range of data relevant to health, retirement, and income security programs
Summary of design	Longitudinal panel survey collecting 2.5 to 4.0 years of detailed monthly income and other data through interviews at 4-month intervals for all people in initial sample households and those added through household formation or change; recent prior panels started in 2001, 2004, and 2008 Redesigned beginning with 2014 panel to four annual interviews conducted February through May and collecting data from the start of the prior calendar year
Universe	Resident civilian non-institutionalized population of the United States plus military living with civilian family members on or off base
Sample size	2008 panel: 42,000 households interviewed in Wave 1 2014 panel: 29,700 households interviewed in Wave 1
Response rate	2008 panel: 80.8 percent in Wave 1 2014 panel: 70.1 percent in Wave 1
People covered in work and income data	All people ages 15 or older
Data on work activity and earnings	2008 and earlier panels: Detail on up to 2 jobs and 2 businesses in prior 4 months, including start and end dates and monthly earnings from job or monthly draw and net profit from business, 3-digit industry and occupation codes for each job, major industry and 3-digit occupation codes for each business; class of worker, wage rate and usual hours per week for each job; weekly employment status 2014 panel: Detail on up to 7 jobs or businesses in prior calendar year
Percentage of earnings dollars imputed	2008 panel: 17.1% Wages and salaries 38.2% Self-employment 2014 panel: 20.1% Wages and salaries 42.2% Self-employment
Income data	Up to 60 sources and amounts of income (monthly), by person
Job search and training	2008 and earlier panels: In prior four months for each person 2014 panel: In prior calendar year, by month, for each person
Key strengths and limitations of earnings data for analysis	SIPP's key strengths in comparison with all other surveys and administrative data are the amount of information collected on each job and each business; the longitudinal nature of the data; and, at least prior to the 2014 redesign, the greater capture of earnings at the low end of the earnings distribution. Key limitations are the attrition from each panel (higher with the 2014 redesign), the high imputation rate for self-employment; the comparatively weaker estimates of annual earnings relative to the CPS and ACS above the low end of the earnings distribution (prior to the 2014 redesign); and the fact that, as a sample, the data will not include the participants in program evaluations.

Figure IV.2 (*continued*)

Link to technical documentation and survey contents	https://www.census.gov/programs-surveys/sipp/tech-documentation/complete-technical-documentation.html
References for descriptive information about survey	Czajka, John L., and Gabrielle Denmead. "Income Data for Policy Analysis: A Comparative Assessment of Eight Surveys." Final report. Washington, DC: Mathematica Policy Research, 2008. Available at https://www.mathematica-mpr.com/our-publications-and-findings/publications/income-data-for-policy-analysis-a-comparative-assessment-of-eight-surveys. Accessed August 23, 2018. National Academies of Sciences, Engineering, and Medicine. *The 2014 Redesign of the Survey of Income and Program Participation: An Assessment.* Washington, DC: The National Academies Press, 2017. Available at https://doi.org/10.17226/24864.

Figure IV.3. Key features of the American Community Survey

Responsible agency	U.S. Census Bureau
Survey URL	https://www.census.gov/programs-surveys/acs
Purpose	To provide small area demographic and income data, including poverty estimates; replaces the decennial census long form with a continuous survey designed to support annual estimates for areas of 60,000 or larger and five-year estimates down to the tract level
Summary of design	Implemented as a monthly cross-sectional household survey collecting prior 12 months' income (rolling reference period) and current demographics including health insurance coverage; income estimates are inflation adjusted to represent a calendar year; like the decennial census but unlike other household surveys, participation is mandatory; nonrespondents to the initial mailing are subsampled for intensive follow-up by telephone and in-person
Universe	Resident household and group-quarters population (including institutions) of the United States and Puerto Rico
Sample size	2.2 million households and 130,000 group-quarters residents annually https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/sample-size/index.php
Response rate	2016 ACS weighted response rates: Households: 94.7 percent; group quarters residents: 95.7 percent https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/response-rates/index.php
People covered in work and income data	All persons age 15 or over
Data on work activity and earnings	Weeks worked in prior 12 months; usual hours per week; wage and salary and self-employment earnings in prior 12 months; current employment status; class of worker, 3-digit and major industry codes for current main job or most recent job
Percentage of earnings dollars imputed	2002 American Community Survey: 17.2% Wages and salaries 23.1% Self-employment
Income data	Up to 8 sources and amounts of income (prior 12 months), by person
Job search and training	Not collected
Key strengths and limitations of earnings data for analysis	The key strengths of the ACS are its very large sample size, high unit response rate, level of geographic detail (areas of 100,000 or greater in population are identified in public use files, and published estimates cumulated over five years go down to the block group level), and the comparative quality of reported annual earnings. Key limitations include the earnings data being aggregated across all jobs, the rolling reference period for annual earnings (the past 12 months); the more limited job characteristics compared to the CPS ASEC and SIPP; and the fact that, as a sample, the data will not include the participants in program evaluations.
Link to technical documentation and survey contents	https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html

Figure IV.3 (*continued*)

References for descriptive information about survey	Czajka, John L., and Gabrielle Denmead. "Income Data for Policy Analysis: A Comparative Assessment of Eight Surveys." Final report. Washington, DC: Mathematica Policy Research, 2008. Available at https://www.mathematica-mpr.com/our-publications-and-findings/publications/income-data-for-policy-analysis-a-comparative-assessment-of-eight-surveys. Accessed August 23, 2018.

REFERENCES

Barnow, Burt S., and David Greenberg. "Do Estimated Impacts on Earnings Depend on the Source of the Data Used to Measure Them? Evidence from Previous Social Experiments." *Evaluation Review,* vol. 39, no. 2, 2015, pp. 179–228.

Internal Revenue Service. "Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2008-2010." Publication 1415. Washington, DC, May 2016. Available at https://www.irs.gov/pub/irs-soi/p1415.pdf. Accessed August 23, 2018.

Mastri, Annalisa, Dana Rotz, and Elias Sanchez-Eppler. "Comparing Job Training Impact Estimates Using Survey and Administrative Data." Draft report. Washington, DC: Mathematica Policy Research, May 14, 2018.

Stevens, David. "Employment That Is Not Covered by State Unemployment Insurance Laws." LEHD Technical Paper No. TP-2007-04. Washington, DC: U.S. Census Bureau, 2007. Available at https://econpapers.repec.org/paper/centpaper/2007-04.htm. Accessed August 29, 2018.

This page has been left blank for double-sided copying.

APPENDIX

EMPLOYMENT QUESTIONS FROM THE WIA 15-MONTH FOLLOW-UP SURVEY

This page has been left blank for double-sided copying.

D22. (We are finished talking about the job you had at ([fill COMPANY NAME FROM SRF20 OR D2] / your self-employment].) Now I'd like to ask you about your current employment status. Are you . . .

CODE ALL THAT APPLY

Currently employed for someone other than yourself, ... 1
Self-employed, ... 2
Not employed, ... 3
Not employed outside the home, ... 4
Retired, ... 5
A student, or 6
Something else? (SPECIFY) 99

_____

ONLY TEMPORARILY LAID OFF, SICK, OR MATERNITY LEAVE 7
DISABLED, PERMANENTLY OR TEMPORARILY 8
DON'T KNOW ..... d
REFUSED ..... r

ALL

D23. Are you currently looking for work?

INTERVIEWER: IF NEEDED, SAY: **Some people look for work even when they have a job.**

YES 1
NO 0
DON'T KNOW ..... d
REFUSED ..... r

ALL

D24. Including any current job(s), how many different paid jobs have you had since [fill RA MO/YR]?

PROBE: How many different full-time or part-time jobs have you had since you sought services from [fill LWIA ONE-STOP NAME]?

INTERVIEWER: TREAT A JOB INTERRUPTED BY TWO OR MORE UNPAID WEEKS AS SEPARATE JOBS, EVEN IF IT IS WITH THE SAME EMPLOYER. IF SEPARATION IS LESS THAN TWO WEEKS, TREAT AS ONE JOB.

|__|__| NUMBER OF JOBS

ZERO 00 E1
DON'T KNOW ..... d D25
REFUSED ..... r D25

NOTE: SPACE FOR 3RD, 4TH, AND 5TH JOB WILL BE IN CATI PROGRAM.

	JOB 1	JOB 2
(D24 NE 00) D25. Please tell me the name of the companies, organizations, or people you've worked for. Start with your current job or jobs, then the most recent jobs that you had. **PROBE:** What was the job before that? INTERVIEWER: IF DON'T KNOW OR REFUSED AT D24, ASK D25 AND PROBE FOR ANY ADDITIONAL PREVIOUS JOBS. RETURN TO D24 AND ENTER THE CORRECT NUMBER OF JOBS BEFORE CONTINUING.	COMPANYNAME(SPECIFY) ..... 1 _____ SELF-EMPLOYED ..... 2 DON'T KNOW ..... d REFUSED ..... r	COMPANY NAME (SPECIFY) ..... 1 _____ SELF-EMPLOYED ..... 2 DON'T KNOW ..... d REFUSED ..... r
(D24 NE 00) D26. It is important that we get information on every job you have had since [fill RA MO/YR]. Let me verify that since [fill RA MO/YR] you worked at [fill D25 NAMES]. Is this correct, or are there any other jobs you may have had, including your current job? INTERVIEWER: IF CORRECT, ENTER "1" AND CONTINUE. IF IT IS NOT CORRECT, GO BACK TO D24 AND D25 TO ENTER CORRECT NAMES AND NUMBER OF JOBS HELD.	CORRECT ..... 1 DON'T KNOW ..... d REFUSED ..... r	CORRECT ..... 1 DON'T KNOW ..... d REFUSED ..... r
(D24 NE 00) D27. When did you <u>start</u> working for [fill D25_JOB_1 – D25_JOB_5]? INTERVIEWER: ENTER MONTH AND YEAR. **PROBE IF RETURNED TO JOB**: We may have talked about this job previously, but I need to ask some more questions about since you went back.	|___| __| / |___|___|___|___| D29 MONTH YEAR DON'T KNOW ..... d REFUSED ..... r	|___| __| / |___|___|___|___| D29 MONTH YEAR DON'T KNOW ..... d REFUSED ..... r
(D27= d OR r) D28. Do you recall what year you started working there? INTERVIEWER: ENTER YEAR.	|___|___|___|___| YEAR DON'T KNOW ..... d REFUSED ..... r	|___|___|___|___| YEAR DON'T KNOW ..... d REFUSED ..... r

	JOB 1	JOB 2
D29. Are you still working there?	YES ..... 1 D31 NO 0 D29a	YES ..... 1 D31 NO 0 D29a
(D24 NE 00 and D29=0) D29a. When did that job end? INTERVIEWER: ENTER MONTH AND YEAR.	|__|__| / |__|__|__|__| D31 MONTH YEAR DON'T KNOW ..... d REFUSED ..... r	|__|__| / |__|__|__|__| D31 MONTH YEAR DON'T KNOW ..... d REFUSED ..... r
(D29 = d OR r) D30. Do you recall what year that job ended? INTERVIEWER: ENTER YEAR. SOFT CHECK: IF D29a/D30 < D27/D28: **I'm sorry. I need to verify your response. REPEAT D29a/D30.**	|__|__|__|__| YEAR DON'T KNOW ..... d REFUSED ..... r	|__|__|__|__| YEAR DON'T KNOW ..... d REFUSED ..... r
(D24 NE 00) D31. Apart from vacations, holidays, or sick leave, would you say you worked for all or nearly all of the time between when that job started and (when that job ended/now) or was there some time that you were not working? **PROBE:** Between [fill (D27/D28 MO/YR)] and [fill D29/ D30 MO/YR)]/now.	CODE ONE ONLY WORKED ALL OR NEARLY ALL OF THE TIME ..... 1 D33 SOME TIME NOT WORKING ..... 2 DON'T KNOW ..... d REFUSED ..... r	CODE ONE ONLY WORKED ALL OR NEARLY ALL OF THE TIME ..... 1 D33 SOME TIME NOT WORKING ..... 2 DON'T KNOW ..... d REFUSED ..... r
(D31 NE 1) D32. About how many weeks would you say you worked during that time? **PROBE:** Between [fill (D27/D28 MO/YR)] and [fill D29/ D30 MO/YR)]/now.	CODE ONE ONLY Most but not all, ... 1 About half, ... 2 Less than half but more than a few, or ..... 3 Almost none? ..... 4 DON'T KNOW ..... d REFUSED ..... r	CODE ONE ONLY Most but not all, ... 1 About half, ... 2 Less than half but more than a few, or ..... 3 Almost none? ..... 4 DON'T KNOW ..... d REFUSED ..... r
(D24 NE 00) D33. How many hours per week, including regular overtime hours (do/did) you usually work at [fill D25_JOB_1 – D25_JOB_5]?	|__|__| HOURS PER WEEK D35 DON'T KNOW ..... d REFUSED ..... r	|__|__| HOURS PER WEEK D35 DON'T KNOW ..... d REFUSED ..... r
(D33 = d OR r) D34. Would you say you work(ed) . . . **PROBE:** Between [fill (D27/D28 MO/YR)] and [fill D29/ D30 MO/YR)]/now.	CODE ONE ONLY Less than 20 hours per week, ... 1 Between 20 and 29 hours per week, ... 2 Between 30 and 39 hours per week, ... 3 Between 40 and 49 hours per week, or ..... 4 50 or more hours per week? ..... 5 DON'T KNOW ..... d REFUSED ..... r	CODE ONE ONLY Less than 20 hours per week, ... 1 Between 20 and 29 hours per week, ... 2 Between 30 and 39 hours per week, ... 3 Between 40 and 49 hours per week, or ..... 4 50 or more hours per week? ..... 5 DON'T KNOW ..... d REFUSED ..... r

	JOB 1	JOB 2
(D24 NE 00) D35. How many days per week (do/did) you usually work? **PROBE:** How many days in an average week? **PROBE:** Just before you left.	|___|___| DAYS PER WEEK DON'T KNOW ..... d REFUSED ..... r	|___|___| DAYS PER WEEK DON'T KNOW ..... d REFUSED ..... r
(D24 NE 00) D36. What kind of company is [fill D25_JOB_1 – D25_JOB_5]—what do they make, do, or sell? **PROBE:** What kind of business or industry is this? **INTERVIEWER: IF RESPONDENT RETURNED TO JOB, SAY:** You may have told me this information about when you worked for [fill COMPANY NAME] before.	KIND OF BUSINESS OR INDUSTRY (SPECIFY) ..... 1 _____ DON'T KNOW ..... d REFUSED ..... r	KIND OF BUSINESS OR INDUSTRY (SPECIFY) ..... 1 _____ DON'T KNOW ..... d REFUSED ..... r
(D24 NE 00) D37. What (do/did) you do there—what (is/was) your job? **PROBE:** What were your most important duties at that job? INTERVIEWER: TRY TO GET A VERB.	JOB DUTIES (SPECIFY) ..... 1 _____ DON'T KNOW ..... d REFUSED ..... r	JOB DUTIES (SPECIFY) ..... 1 _____ DON'T KNOW ..... d REFUSED ..... r
(D25 NE 2) D38. Which of the following best describes your employment at [fill D25_JOB_1 – D25_JOB_5]? (Are/Were) you working . . . **PROBE:** Between [fill (D27/D28 MO/YR)] and [fill D29/ D30 MO/YR)]/now.	CODE ONE ONLY as a regular full-time or part-time employee, ... 1 for a temporary help agency, ... 2 for a company that contracts out you or your services, ... 3 as an independent contractor, independent consultant, free-lance worker, or self-employed, ... 4 as a day laborer, or ..... 5 as an on-call employee? ..... 6 DON'T KNOW ..... d REFUSED ..... r	CODE ONE ONLY as a regular full-time or part-time employee, ... 1 for a temporary help agency, ... 2 for a company that contracts out you or your services, ... 3 as an independent contractor, independent consultant, free-lance worker, or self-employed, ... 4 as a day laborer, or ..... 5 as an on-call employee? ..... 6 DON'T KNOW ..... d REFUSED ..... r

	JOB 1	JOB 2
(D24 NE 00) D39. What (was/is) your (most recent/ current) rate of pay, before taxes and deductions, at that job? **PROBE:** If your pay (varies/varied), please provide an average amount. D39a. ACCEPT MOST CONVENIENT PAY PERIOD. **PROBE:** I am not asking how often you get paid, but how much you make, for example, an hour, per week, per year. SOFT CHECK: IF D39>(73/hour or 2,885/week or 150,001/year or 5,770 every two weeks or 6251 twice per month, or 12,501 per month, or 411 per day): **I'm sorry. I need to verify your response. REPEAT D39.**	$ |___|___|___|,|___|___|___|.|___|___| AVERAGE AMOUNT D39=DK/R D40 PER HOUR ..... 1 PER WEEK ..... 2 ONCE EVERY TWO WEEKS ..... 3 TWICE A MONTH ..... 4 PER YEAR ..... 5 DAY/DAILY ..... 6 MONTH ..... 7 PER EVENT/ACTIVITY/UNIT/JOB ..... 8 PLUS TIPS/COMMISSION/BONUS ..... 9 OTHER (SPECIFY) ..... 99 DON'T KNOW ..... d REFUSED ..... r	$ |___|___|___|,|___|___|___|.|___|___| AVERAGE AMOUNT D39=DK/R D40 PER HOUR ..... 1 PER WEEK ..... 2 ONCE EVERY TWO WEEKS ..... 3 TWICE A MONTH ..... 4 PER YEAR ..... 5 DAY/DAILY ..... 6 MONTH ..... 7 PER EVENT/ACTIVITY/UNIT/JOB ..... 8 PLUS TIPS/COMMISSION/BONUS ..... 9 OTHER (SPECIFY) ..... 99 DON'T KNOW ..... d REFUSED ..... r
(D25 NE 2) D40. Which of the following benefits (are/were) available to you on your job, even if you (are/were) not receiving them (READ EACH ITEM) . . . SELECT IF AVAILABLE, BUT NOT USED.	CODE ALL THAT APPLY Health insurance or membership in an HMO or PPO plan? ..... 1 Paid vacation? ..... 2 Paid holidays? ..... 3 Paid sick leave? ..... 4 Retirement or pension benefits? ..... 5 Tuition assistance or reimbursement? ..... 6 NONE ..... 7 DON'T KNOW ..... d REFUSED ..... r	CODE ALL THAT APPLY Health insurance or membership in an HMO or PPO plan? ..... 1 Paid vacation? ..... 2 Paid holidays? ..... 3 Paid sick leave? ..... 4 Retirement or pension benefits? ..... 5 Tuition assistance or reimbursement? ..... 6 NONE ..... 7 DON'T KNOW ..... d REFUSED ..... r
(D25 NE 2) D41. (Do/Did) you belong to a union on this job?	YES ..... 1 NO ..... 0 DON'T KNOW ..... d REFUSED ..... r	YES ..... 1 NO ..... 0 DON'T KNOW ..... d REFUSED ..... r

This page has been left blank for double-sided copying.

This page has been left blank for double-sided copying.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

Mathematica® is a registered trademark
of Mathematica Policy Research, Inc.