Social Security

SOCIAL SECURITY BULLETIN

Vol. 82, No. 1, 2022

IN THIS ISSUE:

- Where Is Information About Social Security Retirement Benefits Provided to the Public? An Initial Assessment
- Improving County-Level Earnings Estimates with a New Methodology for Assigning Geographic and Demographic Information to U.S. Workers

The *Social Security Bulletin* (ISSN 1937-4666) is published quarterly by the Social Security Administration, 250 E Street SW, 8th Floor, Washington, DC 20254.

The *Bulletin* is prepared in the Office of Retirement and Disability Policy, Office of Research, Evaluation, and Statistics. Suggestions or comments concerning the *Bulletin* should be sent to the Office of Research, Evaluation, and Statistics at the above address. Comments may also be sent by e-mail to ssb@ssa.gov.

Errata Policy: If errors that impair data interpretation are found after publication, corrections will be posted as errata on the web at https://www.ssa.gov/policy/docs/ssb/v82n1/index.html.

SSA Publication No. 13-11700
Produced and published at U.S. taxpayer expense

Social Security Bulletin

Volume 82 • Number 1 • 2022

Articles

Research has documented how knowledgeable Americans are about certain aspects of Social Security programs and benefits. As a result, researchers have identified information gaps in both the types of knowledge that individuals optimally should have and the demographic groups who would most benefit from informational outreach. However, research has not investigated the settings in which Social Security information is or could be provided to the public by sources other than the Social Security Administration. This article explores the presence and extent of Social Security information provided to employees in workplaces, servicemembers in military facilities, students in secondary and postsecondary schools, and participants in religious and community organization settings. The authors visited seminars and classes; interviewed trainers and educators; and reviewed textbooks, other publications, and Internet content to assess where this important information is—or could be—provided to the public.

This article describes a new methodology developed by the Office of Research, Evaluation, and Statistics (ORES) of the Social Security Administration (SSA) to indicate the state and county of residence, date of birth, and sex of nearly all workers for whom tax records provide earnings data in a given year. Applying these geographic and demographic indicators will enable ORES to use a vastly larger sample of workers to generate annual earnings estimates. The current methodology assigns state and county codes and demographic information only to workers in SSA's 1-percent Continuous Work History Sample—fewer than 1.7 million workers in 2017. The new methodology would assign state and county codes and demographic information to more than 178 million workers for 2017. A much larger sample of workers mitigates the limitations associated with the current estimation process and enables ORES to generate more comprehensive and accurate county-level earnings estimates.

Where Is Information About Social Security Retirement Benefits Provided to the Public? An Initial Assessment

by Lois A. Vitt and Barbara A. Smith*

Research has documented how knowledgeable Americans are about certain aspects of Social Security programs and benefits. As a result, researchers have identified information gaps in both the types of knowledge that individuals optimally should have and the demographic groups who would most benefit from informational outreach. However, research has not investigated the settings in which Social Security information is or could be provided to the public by sources other than the Social Security Administration. This article explores the presence and extent of Social Security information provided to employees in workplaces, servicemembers in military facilities, students in secondary and postsecondary schools, and participants in religious and community organization settings. We visited seminars and classes; interviewed trainers and educators; and reviewed textbooks, other publications, and Internet content to assess where this important information is—or could be—provided to the public.

Introduction

Research has documented that Americans know basic aspects of Social Security, such as the existence of the retirement and disability programs, but they know little about the future benefits they can expect to receive from those programs (Greenwald and others 2010; Smith and Couch 2014; Yoong, Rabinovich, and Wah 2015; Alattar and others 2019). For example, Alattar and others, using data from the University of Southern California's Understanding America Study, found that 80 percent of respondents know about the Social Security Disability Insurance program. However, only 21 percent of respondents can correctly answer a multiple-choice question about the way Social Security benefits are calculated, and only 63 percent are aware that a widow(er) need not be caring for children to qualify for survivor benefits. These studies also show that knowledge about Social Security benefits is low across most demographic groups, whether categorized by race, sex, income, or education. The one exception is age: Knowledge of Social Security increases with respondent age. In 2020, the Social Security Administration (SSA) commissioned a survey by Ipsos Public Affairs to measure the effectiveness of the *Social Security Statement* in informing the public about the Social Security programs and benefits. The Ipsos survey found, for example, that respondents aged 62–69 are more knowledgeable than those aged 25–34 about the adjustment of Social Security benefits for inflation (60 percent versus 22 percent) and the existence of survivor benefits (78 percent versus 38 percent).

Could the gaps in Social Security knowledge be related to the settings in which information about retirement benefits is provided—or not provided?

Selected Abbreviations

DOD	Department of Defense
HR	human resources
SSA	Social Security Administration
TSP	Thrift Savings Plan

* *Lois A. Vitt is the director of the Institute for Socio-Financial Studies. When this research was conducted, she was on a temporary procurement with the Office of Research of the Office of Research, Evaluation, and Statistics (ORES), Social Security Administration (SSA) under the Intergovernmental Personnel Act mobility program and Barbara A. Smith, now retired, was a senior economist with ORES, SSA.*

The low levels of knowledge about benefits among young people suggest that providing Social Security information to students in high schools, colleges, and universities could close some of the gaps. The lack of knowledge about benefits regardless of sex or race suggests that an overlapping variety of providers such as cultural- and minority-support organizations, religious institutions, and military trainers should be employed to provide Social Security information to most segments of society. The low levels of knowledge across income groups suggests that the workplace can be a primary venue for providing information on Social Security, especially because benefits are tied to employment and earnings.

A brief review of the literature indicates that researchers have not widely studied the question of where and how information about Social Security retirement benefits is presently provided. Little is known about which organizations provide information about Social Security, what specific kinds of information they provide, and to whom it is provided. This article explores how information about Social Security retirement benefits is disseminated, who provides that information, and where it is made available. We believe that identifying organizations that do or could provide information on Social Security might be of interest to SSA. The agency could use our findings to develop effective collaborations for disseminating information about Social Security as it creates outreach strategies involving pamphlets, videos, social media posts, blogs, online applications, and newspaper and magazine articles.

Of course, the primary source of information about Social Security programs and benefits is SSA itself. The agency provides this information through its website, www.ssa.gov; through personal *my* Social Security accounts, for which individuals can sign up at https://www.ssa.gov/myaccount/; and through annual mailings of printed *Social Security Statements* to individuals aged 60 or older who are not receiving benefits and have not signed up for a *my* Social Security account.

However, SSA is not the only source of information about Social Security. In addition to obtaining financial advice from family and friends, individuals can acquire information from public, private, or nonprofit providers in various settings. People can also seek information from financial advisors, online trading platforms, social media, television, books, and periodicals.

This article consists of four sections beginning with this introduction. The next section briefly describes our research methods and limitations. The third section highlights our findings and discusses the implications of where we found Social Security information currently being provided. The fourth section concludes, with a brief discussion about possible next steps.

Research Methods and Limitations

We conducted a qualitative study to identify, document, and assess where, when, and how information about Social Security's programs and benefits is provided by employers, the military, religious institutions, secondary and postsecondary schools, and community organizations. We selected these providers based on extensive prior research that surveyed the state of personal financial education offerings nationwide.[1]

Our research took place primarily in 2015. We conducted literature searches and other informational reviews, analyzed online and print materials, observed classes and courses (both in person and online), and spoke with sponsors and educators to identify the methods used to inform participants about Social Security programs and benefits. We also reviewed websites featuring Social Security information and benefit calculators, a sample of textbooks on personal finance, and similarly themed books published in the popular press.

More specifically, we collected information using a combination of the following:

- Discussions with thought leaders, program sponsors, policymakers, financial service vendors, and educators who have broad knowledge of the financial-education content of programs offered in the settings under study.
- Attendance at a purposive sample of seminars, courses, webinars, and workshops about Social Security.
- Informal talks with participants at the seminars, courses, and workshops and, in some cases, a review of a sample of evaluations submitted by attendees.
- Assessment of Social Security–related literature, materials, and practices from academic, commercial, and popular sources disseminated via broadcast media, the Internet, and in print.

Our qualitative review provides illustrative examples of how and where Social Security information is provided or not provided, but it is not meant to be definitive or exhaustive. Rather, it indicates the types of information provided. The thought leaders and educators with whom we spoke conduct or sponsor their personal finance training sessions in major cities, small towns, and rural areas. We also talked to individuals with direct knowledge of the financial education programs provided by large and small organizations that include financial institutions, multinational corporations, universities, military service branches, a nonprofit employee-benefit research group, and a labor union. The webinars, seminars, classes, and other meetings we attended are few in number but diverse in terms of geographic location, program type, and the ages and interests of their attendees.[2] Although the research for this study was conducted primarily in 2015, we have updated the information where possible. Note that the individuals who provided information about Social Security to their clients, students, or attendees hold divergent opinions about the longevity or viability of the Social Security program, which may have inadvertently skewed the findings we report.

Findings: Where and What Information About Social Security Is Provided to the Public

In this section, we highlight the settings where the information about Social Security benefits is, or could be, provided.

Workplaces

Americans typically are introduced to Social Security through the Federal Insurance Contributions Act (FICA) payroll-tax deductions from their paychecks. Thus, workplaces are appropriate venues for providing information about Social Security to working Americans. For example, information about Social Security could be offered along with information about employer-provided retirement plans.

We found, however, that Social Security is not addressed in most of the workplace financial education programs we investigated. In the 1990s, as employer-sponsored financial education became increasingly available to employees, programs emphasized a "three-legged stool" metaphor for retirement security, with the legs consisting of employer-sponsored pension plans, personal savings, and Social Security (DeWitt 1996). However, from recent discussions with the chief executive of a major employee-benefit research organization, human resources (HR) managers, financial educators, and retirement-benefit consultants, we learned that that approach began to disappear in the mid-2000s and that now, Social Security education is missing entirely from nearly all workplace financial education programs. Further, we found that when information about Social Security is included in employer-sponsored financial wellness programs, the topic is usually focused narrowly on the timing of claiming retirement benefits.

From the 1990s through the 2000s, employers increasingly discontinued offering defined benefit pensions in favor of 401(k) and other defined contribution retirement plans. As 401(k) plans grew in numbers and importance, company-sponsored education programs often focused exclusively on their plan offerings and ignored Social Security entirely. This trend coincided with the emergence of retirement plan vendors rather than the company's HR department as the most likely providers of the saving and investment guidance offered to a company's employees. In a June–July 2015 survey, for example, only 16 percent of consumers reported that a current or former employer had been a source of information about Social Security retirement benefits (Perron 2015).

Although we recognize the importance of employer-sponsored retirement saving plans, many American workers have little or no access to such plans (Bureau of Labor Statistics 2021) and will likely rely more heavily on, and need to understand, their Social Security retirement benefits. As a consequence of the unmet need, many workers base critical Social Security claiming decisions on limited information (Consumer Financial Protection Bureau 2015).

Fortunately, employers appear to have begun more recently to again offer Social Security information to their employees. As a result, benefits consultants and financial advisors hoping to be retained by employers to provide employee education have greater incentives to include Social Security claiming advice, tools, and calculators in their educational offerings. In particular, training providers may refer participants to online retirement-planning resources available from government, proprietary, or independent organizations.[3] However, claiming advice alone is insufficient; workers should be apprised of the purposes, importance, and breadth of Social Security benefits.

A broader trend of employers providing information about Social Security benefits in the workplace would give workers the knowledge they need to make informed decisions about their future retirement benefits and their financial needs in general. However, our research indicates that few employers provide comprehensive information and suggests that employers and retirement plan providers should add information about Social Security to the material they present to employees about company retirement plans. This information should include details about

Social Security benefits, including eligibility requirements; benefit calculations; delayed claiming effects; and disability, dependent, and survivor benefit availability. Providers could also enhance their presentations by directing participants to the Department of Labor's "Retirement Toolkit" and SSA's "Retirement Checklist." [4] Employers, plan sponsors, or third-party financial educators could show workers how to sign up not only for the company's retirement plan(s) but also for an online *my* Social Security account. Employers report their workers' earnings to SSA each year. By checking a *my* Social Security account or a *Social Security Statement* received by mail, workers can ensure that their earnings are correctly reported.

Military Facilities

The U.S. armed forces have a long record of providing financial education to servicemembers. Prompted in part by congressional hearings, the army established personal financial management programs in the mid-1960s. Informal personal finance education for navy personnel began in the early 1970s in wives' clubs and ombudsman training programs. The navy started a formal program in 1979 and the air force initiated a similar program in the early 1980s. The navy revised its formal program in the 1990s and the marine corps likewise revised a prior formal program. Under Department of Defense (DOD) leadership, all service branches consistently reappraise, upgrade, and support personal financial management education and counseling programs (Vitt 2010).

Since 1950, servicemembers have contributed to and been covered by Social Security. The service branches provide personal finance training, required by DOD regulation and policy, to servicemembers during boot camp and at intervals throughout their military careers until they leave active service. Servicemembers aged 30 or younger constitute nearly three-quarters of active-duty personnel. The average ages for enlisted personnel and officers are 26.9 and 34.4, respectively (DOD 2019, Exhibits 2.41 and 2.45). Given the age range of most servicemembers and their families, the programs concentrate on the financial needs of younger individuals, covering topics such as credit cards, borrowing, buying a car, payday loans, deciding whether to rent or buy a home, and postservice career choices.[5] Social Security is addressed only briefly.

In all workplace settings but especially for young military servicemembers, Social Security topics covered in training should include the number of credits needed for benefit eligibility, how benefit levels are calculated, and the importance of retirement saving to complement Social Security benefits. The training should also correct a common misconception that a spouse will receive 50 percent of the primary earner's benefit in virtually all cases; the actual percentage will depend on the spouse's earnings history and claiming age. Financial educators and counselors must be knowledgeable about the circumstances and life events that affect Social Security income in retirement and be able to discuss the many intricacies involved in long-term Social Security planning.

Social Security information should become a standard component of the personal finance education provided during boot camps. As part of that training, educators could avail themselves of SSA information that is targeted to younger audiences, such as the "Retirement Ready Fact Sheet for Workers Ages 18–48" (https://www.ssa.gov/myaccount/assets/materials/workers-18-48.pdf).[6]

To learn more about the ways the armed forces could provide information about Social Security, we attended a meeting of the Financial Education Roundtable of Hampton Roads, comprising military financial educators and counselors in greater Norfolk, Virginia—home to approximately 83,000 active duty personnel in all service branches—and participated in a discussion about the role of Social Security in servicemembers' financial futures.

We learned that roundtable educators did not address Social Security in personal finance classes and counseling sessions, and that they knew little about the program themselves. Interest in learning about Social Security was strong among roundtable educators, especially among those approaching retirement age. By the end of our 2-hour discussion, educators acknowledged that Social Security education was needed for all servicemembers regardless of age. Yet a subsequent follow-up with our roundtable contacts disclosed that key Social Security issues and concepts are still offered in very few money-management classes for servicemembers or training courses for financial educators, with the possible exception of information provided when a servicemember transitions from the armed services into civilian life, from active to reserve components, or, especially, into retirement. Each of these transition training sessions provides military educators with a timely opportunity to increase servicemembers' knowledge of Social Security programs and benefits.

Results of the 2020 Ipsos survey commissioned by SSA show that age is closely linked to knowledge about Social Security.[7] For example, 60 percent of respondents aged 62–69 know that benefits are adjusted for inflation while only 22 percent of those aged 25–34 do. Seventy-eight percent of the older group know that Social Security provides survivor benefits compared with 38 percent of the younger group. This split suggests that many younger individuals might also not be aware of other important information about Social Security, such as the earnings requirements for benefit eligibility and the relationship between earnings and benefit calculations.

Because our investigation indicates that information on Social Security is not widely offered to military servicemembers at present, we infer that DOD could play a much more important role in informing its young personnel. The army implemented a mandatory 8-hour financial education course for new soldiers in 2007–2008. After the sessions, which included information on the Thrift Savings Plan (TSP—the government equivalent of a 401(k) plan), TSP participation increased by 125 percent and contributions increased by 115 percent among program attendees (Skimmyhorn 2016). Such behavioral response to information about the TSP suggests that providing information about Social Security would similarly increase soldiers' knowledge about the benefits they could expect from the program.

Religious Institutions

Our research found that financial education programs sponsored by religious groups generally fall into one of two categories. The first is developed specifically for members of a particular denomination. These courses are offered within the congregational community and often align with a denominational message. The second includes efforts that are sponsored, initiated, or developed by the religious organization for the benefit of the greater community, generally to assist particular groups by offering a variety of financial-skills training or identifying where and how to find potential financial resources.

Thrivent Financial, a Lutheran nonprofit membership organization, is an example of the first type of program. Thrivent provides insurance products and services for its member-owners as a fraternal organization under the lodge system, which means a member belongs to a local chapter of the larger society. Thrivent's purpose is "to help people achieve financial clarity, enabling lives full of meaning and gratitude" (Thrivent 2020). It provides financial planning, which includes courses on investing and government programs. Thrivent also provides seminars on Social Security and retirement. We attended two seminars led by a Social Security expert who addressed groups of 30–35 attendees, most of whom appeared to be preretirees.

An example of the second type of program is provided by the Unity church, which also sponsors classes on Social Security for groups of church members, mostly preretirees. Unlike the Thrivent program, however, the Unity church ministry independently determines that a need exists in its congregation (for example, for a course on Social Security) and then invites members of the greater community to attend. Other religious organizations making a similar determination have turned to community-based programs, or Internet offerings such as the "Benefits Checkup" sponsored by the National Council on Aging, to provide needed information.

A 2020 Gallup poll shows that religious identification is strongly correlated with age (Jones 2021). For example, 66 percent of respondents born before 1945 report that they are members of a congregation, compared with 58 percent of baby boomers (born 1946–1964), 50 percent of generation X (born 1965–1979), and 36 percent of millennials (born 1980–2000). These results suggest that religious organizations might be particularly effective in reaching adults aged 40 or older with information on Social Security. Religious groups could thereby fill a gap for individuals who do not receive such information from their employers. This avenue has a significant limitation, however: These sessions are usually organized only when a particular need is identified by one or more members and when potential attendees already know the person who would conduct the program.

High Schools

We investigated secondary school–level personal finance curricula and found little available information that might interest students in, or prepare them for, understanding the Social Security program. However, one encouraging exception might foretell a change. At a Washington, DC, charter school, a social studies teacher of 11th graders instructed the students to research Social Security reform. Teams of students took turns over several days presenting their research on the funding needs and the policies that will likely affect their future Social Security income (and perhaps that of their parents). The students advocated for policy initiatives to sustain Social Security benefits, having become aware during their research of their parents' and grandparents' experienced or anticipated reliance on the retirement program.

We also reviewed personal-finance textbooks used in high school (and college) courses. Among those we chose was *Foundations of Personal Finance, Eighth Edition* (Campbell 2010). We selected this textbook because it aligns with national standards, developed by the Jump$tart Coalition for Personal Financial Literacy, "to ensure that high-school graduates have the knowledge necessary to become financially responsible adults."

Social Security is introduced in the Campbell textbook under "Federal Government Spending" as the "largest entitlement program" funded by payroll taxes. In a later section headlined "Paying Social Security Taxes," the topic is expanded with brief descriptions of Social Security retirement, disability, and survivor benefits. No social or financial history is cited, nor are the societal purposes for which the programs exist. The "story" of Social Security is omitted along with any account of its accomplishments over more than 7 decades. Social Security plays a lifelong role in any young person's saving habits and future financial well-being. We believe young people would be well served if they were provided with details about Social Security—its history and societal context as well as its benefits.

One book providing an informative discussion of Social Security programs and benefits is *A Young Person's Guide to Social Security* (Edwards, Turner, and Hertel-Fernandez 2012).[8] Other publications we reviewed—not necessarily targeting young readers—contained scant information about Social Security. Of 150 general-audience personal-finance books found in one public library, only 19 mentioned Social Security. Of 75 new general-audience titles on personal finance reviewed at a book retailer, only 12 had information about Social Security. In a separate search, 41 "money books" for youths were examined at two different public libraries; of them, four briefly mentioned Social Security. Although a detailed analysis of these books and their Social Security content is beyond the scope of this article, such an analysis could be highly informative.

Information on Social Security is lacking in high school personal finance course offerings and textbooks.[9] Providing that information to students is important—critically so for the retirement security of those who do not continue their education beyond high school. The Ipsos 2020 survey on public knowledge about Social Security found that individuals with no more than a high school diploma are less knowledgeable about Social Security programs and benefits than are those with higher education levels. Yet the individuals with less education are also less likely to work in jobs that offer retirement plans, less likely to have the resources to save on their own, and more likely to rely on Social Security benefits for their retirement income. Information about Social Security retirement benefits must therefore be provided to them before they leave high school.

Colleges and Universities

We found personal finance courses that include information about Social Security programs and benefits offered in business schools and as electives in other 2- and 4-year institutions. Social Security information is also taught to students who specialize in HR management or related fields. Apart from those who have career or business interests in such topics, though, college students have little opportunity to be taught about Social Security.

On a more encouraging note, National Institute of Food and Agriculture teachers and advisors from land-grant universities work with partnering organizations, community groups, employers, and financial service providers to support efforts to show individuals and families how to obtain and use money, time, human capital, material resources, and community services. Because land-grant university faculty tend to work in classroom environments and are well trained as professional family economists, they are knowledgeable and are often asked to participate in community-based educational programs. However, the positive effect of these efforts is small in scope and will remain so unless the existing relationships between individuals and families on one side and local and larger economies on the other can be broadened.

Improving the provision of information in higher education settings is important because our review of surveys indicates that even many college degree-holders lack adequate knowledge about Social Security. For example, only 39 percent of Ipsos 2020 survey respondents with at least a college degree knew that Social Security retirement benefits increase with the cost of living and only 62 percent knew that the program provides survivor benefits.

Community Organization Settings

Community organizations serve diverse populations of lower- and moderate-income individuals—those who rely most heavily on Social Security—so the need for awareness about the programs and their benefits is high. Community-based organizations tend to join coalitions of other public, private, and nonprofit organizations to broaden economic opportunities for program participants. Because financial education is often already embedded in those coalitions' overall purpose and functions, Social Security education is a logical fit.

We investigated Stand By Me, a coalition of community partners led by Delaware state officials, the United Way, and private foundations. Launched in 2011 in New Castle County, the program now offers services statewide. It enables participating employers to provide their employees with an information toolkit and a personal finance coach to guide them toward financial stability and future economic opportunities.

Stand By Me was highlighted in *Financial Wellness at Work Report: A Review of Promising Practices and Policies* (Consumer Financial Protection Bureau 2014) as an effective partnership that supports employer efforts to educate their workers.[10] Financial education is an essential step in achieving the program's goals of enabling individuals to make informed economic decisions, increasing self-sufficiency, building community networks, improving the accessibility of financial benefits, and revitalizing and stabilizing neighborhoods and regions.

Many community-based programs are distinguished by supporting longer-term objectives for their participants: building credit, saving for specific purposes, buying a house, taking advantage of job training opportunities, and so forth. Such forward-looking goals align with individuals' present imperatives to save and to recognize the connection between Social Security contributions and future benefits.

Discussion

A lack of knowledge about Social Security could negatively affect the well-being of many Americans in retirement. We therefore examined personal finance education programs and other resources sponsored or provided by employers, the armed forces, religious groups, high schools, colleges and universities, and community organizations or partnerships. We also reviewed print and online books and articles on personal finance. We observed financial courses, participated in seminars and webinars, spoke with personal finance educators, interviewed policymakers and thought leaders, and reviewed academic literature and the popular press.

We also sought to identify potential new venues for educating the public about Social Security. For instance: Although many employers already inform employees about the retirement plans they offer, information on Social Security benefits is generally omitted and could easily be added to the retirement-plan material. As a specific example, workers could be shown how to sign up for an online *my* Social Security account when they are shown how to sign up for the company's retirement plan. This simple enhancement could significantly increase the number of *my* Social Security accountholders.

Providing military servicemembers with information on Social Security benefits could expand awareness among young adults (73 percent of enlisted personnel are aged 30 or younger). Social Security information could be included in the personal financial management programs already offered by the service branches. A study of financial education provided during army boot camp found that it significantly increased participation in and contributions to the TSP, hinting at promising outcomes for the inclusion of similar content about Social Security.

The religious denominations we investigated engage experts to conduct seminars on Social Security fundamentals for midcareer workers and near-retirees. Our research suggests that these programs might complement similar ones offered in workplace settings. However, the Social Security information sessions provided by religious groups occur irregularly, depending on perceived need and the availability of a known presenter.

Schools at both the secondary and postsecondary levels are underused potential venues. Our investigation found few references to Social Security either in classes or textbooks. Information about Social Security could either be presented in standalone courses or incorporated into history or civics classes. Textbooks should feature the Social Security programs and their societal and historical contexts. Exploring ways of informing students about Social Security is a particularly promising topic for further research.

Our research revealed that financial advisors increasingly include content on Social Security claiming decisions and related issues in the seminars they offer to current and prospective clients. Although the addition of any source of sound information about Social Security, Medicare, and Medicaid is a positive development, lower-income individuals generally are not included in this outreach.

Our research shows that the level of Social Security information currently being provided is less than adequate, but this crucial material could be made available by employers, the military, religious and community groups, and schools at the secondary and postsecondary levels. We have provided examples of where and how Social Security information could be taught.

This study is an initial assessment of the settings where information about Social Security retirement benefits is or can be provided. Detailed research into any of the settings highlighted here is beyond the scope of this article. Future research could focus on one or more of these settings (and on the detailed types of Social Security information to be provided) and identify ways this information could be more effectively delivered to the public.

Notes

Acknowledgments: We are indebted to Dallas Salisbury for his overall views on Social Security education in the workplace. We appreciate the contributions of Sally Hass, Andy Landis, Dean Brassington, and Bruce Brunson, who provided valuable help for our initial and follow-up research for this article. We are grateful as well to Karen McMahon for her technical skills and to Laith Alattar, Kevin Whitman, and Joni Lavery, our technical reviewers at the Social Security Administration, for their thoughtful and substantive comments and suggestions.

[1] For extensive background information, case studies, histories, and topics covered in prior research on the state of consumer financial education nationwide, see Vitt and others (2000) and (2005).

[2] Because of resource and time limitations (and Paperwork Reduction Act concerns), we held informal discussions with our contacts instead of conducting formal surveys.

[3] Martin and Kintzel (2016) provide information about a sample of these tools and calculators.

[4] The Retirement Toolkit is available at https://dol.gov/sites/dolgov/files/EBSA/about-ebsa/our-activities/resource-center/publications/retirement-toolkit.pdf. The Retirement Checklist is available at https://ssa.gov/pubs/EN-05-10377.pdf.

[5] The navy employs the Career Options and Navy Skills Evaluation Program, which includes personal financial planning and financial literacy skills and is designed to engage young participants. Other service branches use similarly designed programs.

[6] This is one of several age-specific fact sheets that supplement the *Social Security Statement* and the *my* Social Security account.

[7] Ipsos surveyed 1,400 respondents aged 25 or older drawn from its nationally representative online panel.

[8] A revised edition is available.

[9] This situation is beginning to change. Twenty-one states now require high school students to complete a course in financial literacy to graduate (Council for Economic Education 2020).

[10] The report points out that the coaches "are employed by independent nonprofit organizations…This separation from the employer's HR department is seen by [Stand By Me] as critical as it supports the employees' need for complete confidentiality in their relationship with the coach."

References

Alattar, Laith, Matt Messel, David Rogofsky, and Mark A. Sarney. 2019. "The Use of Longitudinal Data on Social Security Program Knowledge." *Social Security Bulletin* 79(4): 1–9.

Bureau of Labor Statistics. 2021. "67 Percent of Private Industry Workers Had Access to Retirement Plans in 2020." TED: The Economics Daily. https://www.bls.gov/opub/ted/2021/67-percent-of-private-industry-workers-had-access-to-retirement-plans-in-2020.htm.

Campbell, Sally R. 2010. *Foundations of Personal Finance, Eighth Edition (Teacher's Edition).* Tinley Park, IL: Goodheart-Wilcox Co, Inc.

Consumer Financial Protection Bureau. 2014. *Financial Wellness at Work: A Review of Promising Practices and Policies.* Washington, DC: CFPB. https://www.consumerfinance.gov/data-research/research-reports/financial-wellness-at-work/.

———. 2015. "Issue Brief: Social Security Claiming Age and Retirement Security." Washington, DC: CFPB, Office for Older Americans. https://www.consumerfinance.gov/data-research/research-reports/issue-brief-social-security-claiming-age-and-retirement-security/.

Council for Economic Education. 2020. *Survey of the States 2020: Economic and Personal Finance Education in Our Nation's Schools.* https://www.councilforeconed.org/wp-content/uploads/2020/02/2020-Survey-of-the-States.pdf.

Department of Defense. 2019. *2019 Demographics: Profile of the Military Community*. https://download.militaryonesource.mil/12038/MOS/Reports/2019-demographics-report.pdf.

DeWitt, Larry. 1996. "Origins of the Three-Legged Stool Metaphor for Social Security." Research Notes & Special Studies from the Historian's Office, Research Note #1. https://www.ssa.gov/history/stool.html.

DOD. *See* Department of Defense.

Edwards, Kathryn Anne, Anna Turner, and Alexander Hertel-Fernandez. 2012. *A Young Person's Guide to Social Security.* Washington, DC: Economic Policy Institute and National Academy of Social Insurance.

Greenwald, Mathew, Arie Kapteyn, Olivia S. Mitchell, and Lisa Schneider. 2010. "What Do People Know About Social Security?" Pension Research Council Working Paper No. WR-792-SSA. Philadelphia, PA: University of Pennsylvania Wharton School.

Jones, Jeffrey M. 2021. "U.S. Church Membership Falls Below Majority for First Time." Gallup News: Politics. https://news.gallup.com/poll/341963/church-membership-falls-below-majority-first-time.aspx.

Martin, Patricia P., and Dale Kintzel. 2016. "A Comparison of Free Online Tools for Individuals Deciding When to Claim Social Security Benefits." Research and Statistics Note No. 2016-03. Washington, DC: SSA.

Perron, Rebecca. 2015. *Social Security Planning in 2015 & Beyond: Perspectives of Future Beneficiaries and Financial Planners.* Washington, DC: AARP and the Financial Planning Association.

Skimmyhorn, William. 2016. "Assessing Financial Education: Evidence from Bootcamp." *American Economic Journal: Economic Policy* 8(2): 322–343.

Smith, Barbara A., and Kenneth A. Couch. 2014. "How Effective Is the *Social Security Statement*? Informing Younger Workers About Social Security." *Social Security Bulletin* 74(4): 1–19.

Thrivent. 2020. "How a Clear Purpose Impacts All We Do." https://www.thrivent.com/posts/from-our-leadership/how-a-clear-purpose-impacts-all-we-do.html.

Vitt, Lois A. 2010. *Evaluation of the FINRA Investor Education Military Financial Education Campaign 2008–2010.* Unpublished report to the Board of Directors of the Financial Industry Regulatory Authority's Financial Education Foundation. Middleburg, VA: Institute for Socio-Financial Studies.

Vitt, Lois A., Carol Anderson, Jamie Kent, Deanna M. Lyter, Jurg K. Siegenthaler, and Jeremy Ward. 2000. *Personal Finance and the Rush to Competence: Financial Literacy Education in the U.S.* Middleburg, VA: Institute for Socio-Financial Studies. https://www.isfs.org/documents-pdfs/rep-finliteracy.pdf.

Vitt, Lois A., Gwen M. Reichbach, Jamie L. Kent, and Jurg K. Siegenthaler. 2005. "Goodbye to Complacency: Financial Literacy Education in the U.S. 2000–2005." Middleburg, VA: Institute for Socio-Financial Studies. https://www.isfs.org/documents-pdfs/GoodbyetoComplacency-nocover.pdf.

Yoong, Joanne, Lila Rabinovich, and Saw Htay Wah. 2015. "What Do People Know About Social Security?" CESR-Schaffer Working Paper No. 2015-022. Los Angeles, CA: University of Southern California Center for Economic and Social Research.

Improving County-Level Earnings Estimates with a New Methodology for Assigning Geographic and Demographic Information to U.S. Workers

by Michael Compson*

This article describes a new methodology developed by the Office of Research, Evaluation, and Statistics (ORES) of the Social Security Administration (SSA) to assign a state and county of residence code and identify the date of birth and sex of nearly all workers for whom tax records provide earnings data in a given year. The current methodology assigns state and county codes and demographic information only to workers in SSA's 1-percent Continuous Work History Sample—fewer than 1.7 million workers in 2017. The new methodology assigns state and county codes and demographic information to more than 178 million workers for 2017. Applying these geographic and demographic indicators will enable ORES to use a vastly larger sample of workers to generate annual earnings estimates and mitigate the limitations associated with the current estimation process.

Introduction

The Office of Research, Evaluation, and Statistics (ORES) of the Social Security Administration (SSA) compiles earnings data from administrative records and provides them to internal staff for research and policy analysis, as well as to other federal agencies and outside research organizations via data exchange agreements. ORES also produces annual statistical publications containing approximately 140 tables presenting earnings data in the context of the Social Security and Medicare programs. The primary administrative data source for earnings tables used in research, policy analysis, and program evaluation is the Continuous Work History Sample (CWHS). In use since the 1930s, the CWHS is a system of files that contains annually updated earnings and benefits data, as well as demographic and geographic information for a random 1-percent sample of all Social Security numbers (SSNs) ever issued.[1]

One of ORES' key statistical publications is *Earnings and Employment Data for Workers Covered Under Social Security and Medicare, by State and County* (hereafter, *Earnings and Employment*; see https://www.ssa.gov/policy/docs/statcomps/eedata_sc/index.html). This annual publication presents the worker counts, earnings amounts, and Social Security and Medicare payroll-tax contribution amounts for all workers, wage and salary workers, and self-employed individuals for each county or county equivalent in the United States. ORES has identified three significant limitations associated with the current methodology for generating earnings estimates at the county level.

The first limitation is that many estimates are suppressed to comply with SSA's rigorous data protection rules, which prevent the direct or indirect disclosure of any information that could identify individuals.

Selected Abbreviations

10-CWHS-HE	Ten-percent CWHS with high earners subsample
CWHS	Continuous Work History Sample
EIN	employer identification number
FIPS	Federal Information Processing Standards
IRS	Internal Revenue Service

* *Michael Compson is a senior economist with the Office of Statistical Analysis and Support, Office of Research, Evaluation, and Statistics, Office of Retirement and Disability Policy, Social Security Administration.*

Selected Abbreviations—*Continued*

MEF	Master Earnings File
MGD	Master Geographic-Demographic
OEIS	Office of Enterprise Information Systems
ORES	Office of Research, Evaluation, and Statistics
SCC	state and county code
SSA	Social Security Administration
SSN	Social Security number

In the *Earnings and Employment* 2017 edition, more than one-half of the county-level estimates were suppressed, primarily because the CWHS includes relatively few individuals with self-employment income. The current tabulation process, based on a 1-percent sample, allocates its sample of about 186,000 self-employed individuals across 3,215 counties in the United States and Puerto Rico, an average of only 58 self-employed persons per county.

The second limitation is that the current methodology occasionally produces relatively large year-to-year variances in some county-level earnings estimates, especially for Medicare-taxable earnings. In most cases, this problem emerges when relatively few workers reside in the affected county and one or several of them experience a large change in earnings. A single worker who experiences such a change in a given tax year can significantly affect the amount of aggregate earnings reported in a county because ORES assigns each worker a weight of 100 to reflect national estimates. Because all earnings are subject to the Medicare payroll tax, while earnings subject to the Social Security payroll tax are capped at an annual maximum amount ($127,200 in 2017), the year-to-year increase in an individual's Medicare-taxable earnings can be substantially larger than the increase in Social Security–taxable earnings.

The third limitation is that the state and county codes (SCCs) used in *Earnings and Employment* differ from the Federal Information Processing Standards (FIPS) SCCs that are used by ORES in other statistical publications and by other federal agencies.

The solution to the problems of county-level data suppression and the occasional large variance in the year-to-year earnings estimates is to expand the sample size that is used to generate the estimates. However, despite its limited sample size, only the CWHS currently provides the earnings, geographic, and demographic information necessary to generate the annual earnings tables ORES publishes. To enable the use of a larger sample, ORES has developed a systematic and automated process for assigning the SCC and demographic information (birth year and sex) for nearly all wage and salary workers and self-employed individuals in a given year. With the new methodology, ORES can assign a geographic location for nearly every worker based on the complete address reported each year on Internal Revenue Service (IRS) Forms W-2 and W-2c (filed by employers) and Form 1040 Schedule SE (filed by the self-employed).[2] ORES can then determine a worker's birth year and sex by cross-referencing SSA's Numerical Identification System (Numident) master file.[3] The new methodology enables ORES to assign a single SCC to 99.89 percent of the 178,863,694 workers whose earnings were reported on a W-2, W-2c, or Schedule SE for tax year 2017. It generates a standalone data file that contains the SSN, SCC, date of birth, and sex for nearly all wage and salary workers and self-employed individuals in a given tax year. ORES is developing a new process that will generate the annual earnings estimates by matching the data in the standalone file with a much larger sample of earnings records. That larger sample will replace the 1-percent CWHS that ORES currently uses for earnings estimation. SSA expects the new process and the expanded data set to be finalized and fully implemented within 2 years.

The effect of the new process for assigning geographic and demographic information and using the larger sample of workers extends well beyond improving the annual earnings estimates. The dramatic expansion of the number of workers with geographic and demographic information will open many new avenues, and improve existing ones, for using earnings data in research and policy analysis. For example, assigning SCCs to all workers will allow ORES to generate more accurate estimates of the size and characteristics of the U.S. workforce, identify and analyze worker migration patterns, evaluate SSA procedures for assigning SCCs to other administrative data, and provide valuable data and insights to other federal agencies that generate labor-force estimates.

This article discusses key aspects and limitations of the methodology SSA currently uses for generating county-level earnings estimates. Then it describes the new methodology for assigning SCCs and demographic information to records for workers who had earnings reported on a Form W-2, W-2c, or 1040 Schedule SE for a given tax year.

Limitations of the Current Methodology for Assigning Geographic Codes

ORES developed the current methodology for assigning SCCs to worker records in the CWHS file in the early 1990s and it became operational with the estimates using 1993 data.[4] Until then, ORES had used only the employer's location to assign worker SCCs. Now, ORES would shift to a hybrid approach, using the location of the worker's residence when that information was available, and the employer's location when it was not. In general, ORES used the employer's location only for workers whose employers submitted paper Forms W-2 and W-2c, which was still relatively common at the time. SSA would scan the paper forms to capture all the earnings information needed for program operations—information that did not include the employee's address. Beginning in the mid-1990s, however, the use of paper forms declined dramatically in favor of electronic filing, and basing a worker's SCC on employer location similarly declined. As a result, most of the SCCs assigned in the most recent versions of the CWHS reflect employee addresses.

In the early 1990s, computer storage capacity was limited. In developing the current methodology for assigning SCCs, ORES conserved storage space by using only the first five letters of a city's name and the 5-digit (rather than the 9-digit) ZIP Code to assign a county code to a worker in the CWHS. This approach saved storage space, but it limited the data available for assigning SCCs and may have led to occasional imprecision. For example, approximately 10 percent of 5-digit ZIP Codes lie in more than one county.[5]

The current methodology of assigning SCCs is prone to inaccuracy for other reasons as well. For example, the current SCCs, as noted earlier, are not the same as the FIPS-based codes used in other ORES statistical publications and by other federal agencies.[6] Additionally, some of the programming that assigns the current SCCs is hard-coded—that is, it is embedded in the source code, for which documentation may not exist. As a result, there is no way to verify the accuracy of these SCC assignments. Thus, a new methodology that uses the complete street addresses, city names, and 9-digit ZIP Codes reported on tax forms can only improve the accuracy of the SCCs assigned to each worker.

State-Level Estimates

Most of the earnings tables that ORES produces appear in two annual publications: the *Annual Statistical Supplement to the Social Security Bulletin* (with 21 earnings tables; see https://www.ssa.gov/policy/docs/statcomps/supplement/index.html) and *Earnings and Employment* (containing 106 tables). ORES also produces 11 earnings tables strictly for internal and interagency research.

Only two of the *Annual Statistical Supplement* tables present geographic detail for their earnings estimates, and both of them show data at the state level. Four of the 106 *Earnings and Employment* tables present earnings estimates at only the state level (the other 102 present estimates at both the state and county levels).

Errors in assigning state codes are rare because very few ZIP Codes cross state lines. Moreover, with only 53 geographic divisions (50 states, the District of Columbia, Puerto Rico, and a catch-all "other and unknown" category[7]) in which to allocate approximately 1.6 million wage and salary workers and 186,000 self-employed workers in the 2017 CWHS, no state-level earnings estimates require suppression.

County-Level Estimates

Earnings and Employment includes 102 tables with county-level data. Fifty-one tables (one for each state and one for Puerto Rico) contain estimates for Social Security–taxable earnings and 51 contain estimates for Medicare-taxable earnings.

The 50 states and Puerto Rico contain 3,215 counties or county equivalents. Each of the 51 county-level tables on Social Security–taxable earnings in *Earnings and Employment* also includes estimates that account for all residents of unknown locations within the state. Thus, these 51 tables together contain discrete estimates for 3,266 locations (3,215 counties + 51 unknown-location categories).[8] Each table includes nine columns of estimates (number of workers, taxable earnings, and Social Security contributions, each shown separately for wage and salary workers, self-employed workers, and total workers). For each location, those nine estimates are shown separately for all, male, and female workers: thus, three rows cross-tabulated by nine columns, for 27 estimates per location. Those 27 estimates multiplied by 3,266 locations yield 88,182 computations, or 9,798 estimates per column (88,182 ÷ 9). The 51 county-level tables on Medicare-taxable earnings likewise contain a total of 88,182 discrete county-level estimates, 9,798 for each type of computation.

Effect of Data Disclosure Standards on County-Level Estimates

In publishing earnings estimates, SSA follows strict data disclosure standards. If the unweighted count of workers in a county in the 1-percent CWHS is lower than 10, SSA suppresses the estimates for that county. Recall that the county-level tables show estimates broken down by sex, increasing the frequency of cell suppression, even if only one category has fewer than 10 workers (unweighted). *Primary cell suppression* refers to nondisclosure of a cell with fewer than 10 workers. *Secondary cell suppression* refers to the necessary suppression of the estimates for *both* male and female workers if the count of *either* is fewer than 10.

For example: For a county that has eight workers in the 1-percent CWHS file, ORES would suppress all estimates. For another county, with 17 workers, ORES would publish the "all workers" estimates but would suppress the estimates by sex, because one or both of the estimates would not have the requisite 10 workers to meet SSA disclosure standards. If 12 of this county's workers were women, and thus above the threshold for primary cell suppression, ORES would nevertheless suppress that estimate to prevent the number of male workers from being deduced. This example illustrates the principle of secondary cell suppression.

Secondary cell suppression may also be required for estimates broken down by type of worker (all, wage and salary, self-employed). Consider a county with 25 total workers, of whom 20 are wage and salary workers and 5 are self-employed. These categories are not necessarily mutually exclusive; a worker with both wage and salary earnings and self-employment income in a given year would be counted in both categories. Nevertheless, ORES would disclose only the estimate for that county's total workers and would suppress the estimates for self-employed individuals and for wage and salary workers.

Rigorous data disclosure requirements complicate the allocation of approximately 186,000 self-employed individuals in the 2017 CWHS across 3,215 counties (plus 51 state-level "unknown" locations). Table 1 reveals that in the county-level data of *Earnings and Employment* for 2017 (SSA 2019), 58.4 percent of the 9,798 estimates for the number of individuals with Social Security–taxable self-employment income required suppression. Secondary cell suppression required SSA to withhold the corresponding estimates for wage and salary workers. As a result, only 41.6 percent of the 19,596 county-level cells for the numbers of self-employed individuals and wage and salary workers with Social Security–taxable earnings contain estimates.

Table 2 shows the distribution of the 2017 *Earnings and Employment* county-level data cells for the number of self-employed individuals with Social Security–taxable income by the number of unweighted records that existed in the CWHS. The CWHS data for 53.2 percent of those cells had fewer than 10 unweighted records and an additional 39.1 percent of the cells contained estimates based on 10 to 99 unweighted records. This table underscores the problem of trying to allocate relatively few self-employed individuals across 3,215 counties while adhering to data disclosure standards: Only 7.6 percent of U.S. counties were represented by 100 or more self-employed individuals in the CWHS.

Table 1.
Suppression of *Earnings and Employment* county-level data cells to comply with data disclosure standards: Number of self-employed workers, tax year 2017

Status	Number	Percent
Total	9,798	100.0
Suppressed	5,726	58.4
Not suppressed	4,072	41.6

SOURCE: Author's calculations based on SSA (2019, Table 3).

Table 2.
Number and percentage of *Earnings and Employment* data cells showing county-level estimates of number of self-employed workers, by number of unweighted records for such workers in the underlying CWHS, tax year 2017

Number of unweighted records	Number	Percent	Cumulative: Number	Cumulative: Percent
Total	9,798	100.0	. . .	. . .
0	555	5.7	555	5.7
1–9	4,662	47.6	5,217	53.2
10–99	3,835	39.1	9,052	92.4
100–199	365	3.7	9,420	96.1
200–299	146	1.5	9,575	97.6
300–399	73	0.7	9,653	98.3
400–499	47	0.5	9,708	98.8
500 or more	115	1.2	9,798	100.0

SOURCE: Author's calculations based on SSA (2019, Table 3).

NOTES: Rounded components of percentage distribution do not all sum precisely to cumulative percentages.

. . . = not applicable.

Tables 3 and 4 repeat Tables 1 and 2 for all workers in the 2017 CWHS and they present a much different picture regarding the extent of cell suppression. Less than 9 percent of the estimates for all workers were suppressed (Table 3). In absolute terms, a 9-percent suppression rate is still high, but it is in stark contrast with the 58.4 percent of cells suppressed for self-employed workers (Table 1). In addition, only 6.9 percent of the estimates for all workers were based on fewer than 10 unweighted records (Table 4), compared with 53.2 percent of those for self-employed individuals (Table 2). More than 43 percent of the estimates of numbers of all workers were based on records for more than 100 workers in the CWHS, compared with only 7.6 percent for the estimates for self-employed individuals.

Table 3.
Suppression of *Earnings and Employment* county-level data cells to comply with data disclosure standards: Number of all workers, tax year 2017

Status	Number	Percent
Total	9,798	100.0
Suppressed	838	8.6
Not suppressed	8,960	91.4

SOURCE: Author's calculations based on SSA (2019, Table 3).

Table 4.
Number and percentage of *Earnings and Employment* data cells showing county-level estimates of total number of workers, by number of unweighted records for such workers in the underlying CWHS, tax year 2017

Number of unweighted records	Number	Percent	Cumulative	
			Number	Percent
Total	9,798	100.0	. . .	. . .
0	36	0.4	36	0.4
1–9	637	6.5	673	6.9
10–99	4,897	50.0	5,570	56.8
100–199	1,577	16.1	7,147	72.9
200–299	701	7.2	7,848	80.1
300–399	380	3.9	8,228	84.0
400–499	270	2.8	8,498	86.7
500 or more	1,300	13.3	9,798	100.0

SOURCE: Author's calculations based on SSA (2019, Table 3).

NOTES: Rounded components of percentage distribution do not sum to 100.0.

. . . = not applicable.

The reason for the dramatic differences in the number of estimates subject to cell suppression is the gulf between the numbers of wage and salary workers and self-employed individuals in the CWHS. For 2017, the CWHS contained almost 1.6 million wage and salary workers to allocate across the 3,215 counties, nearly nine times the number of self-employed individuals in the CWHS.[9]

Potential High Variance in Year-to-Year Estimates

ORES has investigated some very large differences in the estimates of taxable Social Security and Medicare earnings amounts from one year to the next and determined that they tend to result from large changes in the taxable earnings of one or two individuals in counties with a relatively small number of workers. Steep year-over-year changes in earnings are more prevalent for Medicare-taxable earnings because there is no cap on the amount of Medicare earnings subject to the payroll tax, and thus no ceiling on the countable amount by which an individual's earnings may increase.

In summary, ORES has identified several limitations associated with the current process for generating the annual earnings estimates at the county level:

- The relatively small number of self-employed individuals in the 1-percent CWHS combines with SSA's data disclosure rules to suppress more than one-half of the cells that would otherwise contain estimates at the county level.
- The limited number of self-employed individuals sometimes yields large variances in the year-to-year earnings estimates for counties with relatively few workers.
- *Employment and Earnings* uses SSA-developed SCCs instead of the standard FIPS-based SCCs; additionally, the current tabulation process truncates the address information used to assign its codes, which can diminish SCC accuracy.
- The process for assigning some of the SCCs via hard coding is not documented and cannot be verified as correct.

The New Methodology, Step by Step

As the first step in attempting to assign geographic and demographic codes to records for all wage and salary workers and all self-employed persons, ORES and SSA's Office of the Chief Actuary (OCACT) asked the SSA Office of Enterprise Information Systems (OEIS) to extract the address information from Forms W-2, W-2c, and 1040 Schedule SE—or "the tax forms" for short. OEIS uses Pitney Bowes' Finalist software to assign a single SCC based on the address data shown in each of the tax forms processed in a given calendar year.[10]

ORES and OCACT also asked OEIS to develop a larger sample of earners for *Earnings and Employment* that would address many of the limitations imposed by using the 1-percent CWHS. After assigning SCCs and demographic information for all workers in a given tax year, ORES will match that information with the records for the larger sample of earners and thereby develop a new process for generating annual earnings estimates. The new process will dramatically reduce the need for cell suppression and the year-to-year variance in the county-level earnings estimates, improve the accuracy of the SCC assignments, and use SCCs that are consistent with those used in other ORES publications and by other agencies.

The new methodology consists of five steps:

1. Assigning a single SCC for *each job* held by each worker, as recorded on a worker's tax form(s) for a given year, using the Finalist software.
2. Extracting those results and housing them on tape files.
3. Using the Numident master file to identify valid and invalid SSNs and to provide each worker's demographic information (sex and date of birth).
4. Assigning a single SCC to *each worker* based on data in the tape file or, if needed, on various imputations, described below.
5. Merging the resulting SCC and demographic data files into a single standalone output file.

Preliminary Steps

Every year, SSA and the IRS share tax-form information for their respective programmatic needs. SSA receives and processes the earnings information contained in hundreds of millions of Forms W-2 and W-2c and millions of Schedules SE as part of the annual wage reporting process. This information is critical to ensure the proper tracking of workers' earnings histories for program operations.[11] To begin SSA's annual wage reporting process, OEIS extracts all of the information reported on the tax forms and stores it on an administrative data file. The ensuing steps of the wage reporting process include data cleaning and validating procedures. OEIS initiates a separate process for the more limited purpose of extracting only the address information for each job in that data file, which ORES can then use to assign SCCs.

In 2010, OEIS began using the Finalist software each year to extract the full address information reported on Forms W-2, W-2c, and 1040 Schedule SE. ORES then worked with a contractor to consolidate the thousands of OEIS data tapes containing the extracted data. However, developing a methodology for assigning a single SCC and attaching demographic information for each SSN in a given tax year was delayed until 2018. An unfortunate consequence of this delay was the loss of the data generated by the OEIS process for tax years 2009, 2010, and 2011, as file retention limits expired. However, ORES has retained the data for 2012–2019.

The new methodology enables ORES to extract state and county data either on a flow basis throughout the year or all at once at the end of the processing year. Extracting state and county data on a flow basis offers two key benefits. First, any problems with the data could be addressed in real time, without delays. Second, a flow basis would allow ORES to smooth out the uneven timing of the arrival of the tax forms, with the bulk of the previous year's wage data received early in the calendar year: ORES would simply adjust the processing schedule to account for the uneven distribution of incoming records throughout the calendar year. Presently, ORES is still processing a backlog of prior-year data, but upon completion, it expects to process the extracted data on a flow basis. In that scenario, ORES would run only the jobs that extract the state and county data from the thousands of data files that OEIS creates throughout the year. Once OEIS completed its processing for a given calendar year, ORES would begin the steps to assign a single SCC (as needed) and to attach date of birth and sex information to the record for each worker in the target tax year.

The discussion in the following subsections uses data from tax year 2017 to illustrate the steps ORES takes to extract, process, and merge the data it receives from OEIS.

The OEIS Process

OEIS uses the Finalist software to capture information from the complete address reported on the administrative data file that is derived from the individual's tax form(s). The software cross-references the address information with an underlying database that contains every U.S. postal delivery address and assigns the

FIPS SCCs. The software thereby enables ORES to improve the accuracy of its geographic coding and assign SCCs that are consistent with those used in other ORES publications and by other federal agencies.

As noted earlier, the data processed in a given calendar year are generally for earnings in the prior year. For example, in 2018, OEIS primarily processed tax-form data for 2017 earnings. At the end of each calendar year, OEIS provides ORES with a report that summarizes the processing results for each type of data source (W-2, W-2c, Schedule SE, W-3, W-3c, and SS-4). In 2018, OEIS processed 256,574,346 W-2 input records and generated 254,788,713 output records (Table 5). The difference—1,785,633—constitutes the 0.7 percent of all W-2 records processed in 2018 for which Finalist was unable to assign geographic information. OEIS also processed 3,682,466 W-2c input records and produced 3,452,217 output records in 2018. Finalist was unable to assign a geographic identifier to slightly more than 6 percent (230,249) of the W-2c records processed in 2018. OEIS processed 21,194,793 Schedule SE input records and generated the same number of output records. The records that OEIS processed in 2018 generated 1,055 tapes on SSA's mainframe computer.

In the next step, ORES separates the target-year records from the records for other tax years. Table 5 presents the number of earnings records extracted from the data tapes and identifies the number of output records OEIS processed for 2017 and non-2017 tax years. In 2018, ORES extracted 276,006,937 records for 2017 earnings. Of those, 90.7 percent were W-2s, 7.2 percent were Schedule SEs, and 0.9 percent were W-2cs. Table 5 also reveals that nearly 99 percent of the records processed in 2018 were for earnings in 2017. The 276,006,937 records for 2017 earnings represent 178,863,694 unique SSNs (workers).

Validating SSNs

After ORES extracts the data resulting from the OEIS process, the task of assigning a single SCC and demographic data to worker-level earnings records begins. ORES merges the records from the three tax-form sources into a single data table to generate a "finder" file that contains all unique SSNs in the population of workers for a given year. The finder file is then compared with the Numident master file to identify valid and invalid SSNs (SSNs that exist in both the finder and the Numident are considered valid). For all valid SSNs, an algorithm extracts data identifying the worker's sex, date of birth, date of death, and the date when the death date was posted to the Numident.[12] ORES can assign SCCs, but cannot assign demographic information, for the invalid SSNs in the file. ORES creates a standalone data file to contain the linked SSN and demographic information. That file is ultimately merged with a file that contains the results of ORES' multistep process for assigning a single SCC for each worker in a given year.

The next step groups earners into one of the following mutually exclusive categories based on the tax forms that provide their geographic data:

- W-2 only
- W-2c only
- Schedule SE only
- W-2 and W-2c
- W-2 and Schedule SE
- W-2c and Schedule SE
- W-2, W-2c, and Schedule SE

Table 6 shows the distribution of earnings records and unique SSNs (workers) by source of geographic data. Nearly 89 percent of U.S. workers in 2017 had wage and salary income only (W-2, W-2c, or both).

Table 5.
Output records after OEIS processing in calendar year 2018, by tax year and tax-form data source

Source	Total		Tax year: 2017		Tax year: Other	
	Number	Percent	Number	Percent	Number	Percent
Total	279,435,723	100.0	276,006,937	98.8	3,428,786	1.2
W-2	254,788,713	91.2	253,365,171	90.7	1,423,542	0.5
W-2c	3,452,217	1.2	2,591,048	0.9	861,169	0.3
Schedule SE	21,194,793	7.6	20,050,718	7.2	1,144,075	0.4

SOURCE: Author's calculations based on SSA data processing audit reports.

Table 6.
Total and worker-level output records for 2017 processed in 2018, by type of tax-form data source

Source	Earnings records processed: Number	Earnings records processed: Percent	Unique SSNs (workers) processed: Number	Unique SSNs (workers) processed: Percent
Total	276,006,937	100.0	178,863,694	100.0
One form only				
W-2	234,677,110	85.0	156,748,520	87.6
W-2c	34,886	(L)	32,189	(L)
Schedule SE	10,720,522	3.9	10,720,024	6.0
Two forms only				
W-2 and W-2c	6,712,867	2.4	2,032,979	1.1
W-2 and Schedule SE	23,275,253	8.4	9,202,656	5.2
W-2c and Schedule SE	6,062	(L)	2,887	(L)
All three forms	580,237	0.2	124,439	0.1

SOURCE: Author's calculations based on SSA data processing audit reports.

NOTES: Rounded components of percentage distributions may not sum to 100.0.

(L) = less than 0.05 percent.

Slightly more than 5 percent had both wage and salary and self-employment income (Schedule SE and any combination of W-2 and/or W-2c), and 6 percent reported self-employment income only (Schedule SE).

The number of earnings records SSA processes always widely outnumbers the number of wage and salary and self-employed workers because some workers may have multiple tax forms for a single job, or forms for multiple jobs, in a given year. Therefore, some of the workers have multiple SCC fields populated for a given job (or for multiple jobs) in the OEIS process.[13] Note that "job" in this context is defined as one or more tax forms that report a specific worker/employer combination—indicated by the worker's SSN and the employer identification number (EIN). For tax year 2017, SSA processed an average of 1.54 earnings records for each of the 178,863,694 workers (SSNs) in the OEIS file.

Initial SCC Revisions

For any job, the presence of a W-2c presents ORES with an opportunity to improve the accuracy of the SCC assigned by the OEIS process if certain conditions are met. By definition, a Form W-2c corrects data reported on a prior Form W-2 for the same job. Therefore, if the SCCs assigned to a W-2 and a W-2c for a given job differ, it may be appropriate to assign the SCC associated with the W-2c for that job. The opportunity to revise the SCCs reported on a W-2 applies to two groups: the 2,032,979 workers whose earnings were reported on both a W-2 and a W-2c and the 124,439 workers whose earnings were reported all three tax forms.

Consider the workers who had earnings and geographic information for 2017 reported on both a Form W-2 and a W-2c (and not a Schedule SE). Table 7 presents the number of records and the number of jobs for these workers, broken out by the type of SCC information present in the data tape files after OEIS processing. In those tapes, a worker's earnings record may include SCC fields for each job or tax form, and, for various reasons, one or more of those SCC fields may be empty after the initial address extraction process. Note the different numbers of jobs associated with the Forms W-2 (3,669,556) and W-2c (2,092,796). From this information, we know that many (about 43 percent) of the jobs represented in a Form W-2 are not reflected in a corresponding W-2c.[14] For tax year 2017, the overwhelming majorities of both records and jobs have a single populated SCC field after OEIS processing.

In general, three conditions must be met to revise the SCC that the OEIS process assigns based on a W-2 with an SCC based on a W-2c. First, there must be a matching W-2c for the W-2 (that is, both forms must refer to the same job). Second, there must be a single SCC based on the Form W-2c that differs from the SCC based on the W-2. Third, if a W-2 has more than one matching W-2c, the SCCs based on the W-2cs must all be the same.

Table 8 illustrates the process using the 3,669,556 job-level W-2 records for 2017 shown in Table 7.[15] Approximately 55 percent of the job-level Forms W-2

Table 7.
OEIS process results for workers with both a W-2 and a W-2c (and no Schedule SE): Earnings records extracted, and number of jobs represented, by type of SCC information derived from the tax forms, tax year 2017

Type of information	Extracted records			Jobs [a]	
	Total	W-2	W-2c	W-2	W-2c
Total	6,712,867	4,326,838	2,386,029	3,669,556	2,092,796
Single SCC field					
Populated	6,513,097	4,205,939	2,307,158	3,630,751	2,019,478
Empty	26,079	23,331	2,748	7,130	1,151
Multiple SCC fields					
All empty	112,727	38,092	74,635	30,992	72,139
Mixed [b]	60,964	59,476	1,488	683	28

SOURCE: Author's calculations based on SSA data processing audit reports.

a. Because some jobs are reflected on both a W-2 and a W-2c and some are not, the unduplicated number of total jobs is unknown at this stage of the process.

b. At least one populated SCC field and at least one empty SCC field.

Table 8.
Updating SCC fields using address information on Form W-2c, if applicable, to OEIS-process results based on job-level Form W-2: Workers with both a W-2 and a W-2c (but no Schedule SE) in tax year 2017

Outcome	Number	Percent
Total	3,669,556	100.0
W-2 with a matching W-2c	2,010,274	54.8
SCC not updated using the W-2c because matching the forms resulted in—		
The same single populated SCC field (no update needed)	1,901,509	51.8
Other outcomes (a single SCC could not be identified)	69,008	1.9
SCC updated using the W-2c	39,757	1.1
W-2 without a matching W-2c	1,659,282	45.2
OEIS process resulted in—		
A single populated SCC field	1,639,733	44.7
Other outcomes	19,549	0.5

SOURCE: Author's calculations based on SSA data processing audit reports.

were also represented by a corresponding W-2c, with which SCC information can potentially be updated. The W-2s for the remaining 45 percent of the jobs did not have a matching W-2c and the SCCs assigned for those jobs remained unchanged. ORES updated the SCCs of only 39,757 of the 2,010,274 jobs represented by both a W-2 and a matching W-2c. The largest group of jobs that were not updated had the same address information reported on both the W-2 and the W-2c.

The same technique was applied for the 580,237 earnings records for which state and county data were provided on all three tax forms,[16] which enabled ORES to update the SCC based on the address reported on the W-2c for 2,237 additional jobs. In total, updating an assigned SCC using the address on a corresponding W-2c resulted in SCC changes for 41,994 jobs. Although the number of jobs for which ORES updated the SCC using W-2cs is very small relative to the full sample of almost 179 million workers for tax year 2017, this step illustrates the extent to which ORES strives to maximize the accuracy of the new methodology for assigning a single SCC to each SSN.

Determining a Single SCC

At this stage, the focus shifts from jobs to workers, as ORES aims to determine, wherever possible, the SCC of the worker's residence. Assigning the state and county data begins with sorting workers into three mutually exclusive categories based on the number of SCCs that were assigned to them by the OEIS process: a single SCC, multiple SCCs, or no SCC (an empty SCC field or fields) assigned for any of their jobs. Table 9 presents the distribution of workers among these categories for tax year 2017. OEIS assigned a single SCC to 168.3 million workers, or 94.1 percent of the workers in the file. Another 9.3 million workers (5.2 percent) were assigned multiple SCCs, and the remaining 1.2 million workers (0.7 percent) were not assigned an SCC.

For geographic data, the objective of the new methodology is to assign a single SCC to the maximum number of workers. That process is complete for the 168.3 million workers with a single SCC assigned by OEIS using the Finalist software—hereafter, the "gold-standard file." Because this group represents an overwhelming majority of workers in 2017, it broadly reflects the geographic distribution of SCCs for the U.S. workforce in that year. For the 9.3 million workers to whom the software has assigned multiple SCCs, ORES imputes the one SCC that is deemed the most statistically likely to be accurate, and for the 1.2 million workers with no assigned SCC, ORES determines whether one can be imputed from other available data. The geographic distribution of the gold-standard file is used in imputing a single SCC for members of both the other groups; those processes are described below.

Imputing a Single SCC for Workers Outside the Gold-Standard File

ORES developed a multistep imputation process that uses the information in the extracted OEIS data to assign a single SCC to as many of the multiple-SCC and no-SCC workers as possible. The tax forms for some workers with no SCC assigned by the OEIS process nevertheless include a 5- or 9-digit ZIP Code.[17] In those cases, ORES uses the frequency distribution of ZIP Code/SCC combinations in the gold-standard file to impute an SCC. Table 10A shows the first step of the process. It presents frequency distributions for various hypothetical ZIP Code/SCC combinations to represent the actual distributions from the gold-standard file. Table 10B shows how a randomly generated number from 0 to 1, applied to each job with a ZIP Code but no assigned SCC, enables ORES to place the job within a frequency band (from Table 10A) and thereby assign SCCs in a pattern that follows the distribution in the gold-standard file.

Chart 1 diagrams the post-OEIS processing of all files with tax year 2017 earnings data. In Panel A, the records for the 178.9 million unique SSNs are grouped according to the number of SCCs that OEIS assigned using the Finalist software (one, two or more, none). The subsections that follow describe the methods with which ORES imputed a single SCC for most of the records to which OEIS assigned either zero or multiple SCCs.

Table 9.
Worker-level earnings records by number of SCCs assigned after the OEIS process and type of tax-form data source, tax year 2017

Source	Unique SSNs		Workers assigned— A single SCC (gold-standard file)		Multiple SCCs		No SCC (empty field)	
	Number	Percent	Number	Share of unique SSNs	Number	Share of unique SSNs	Number	Share of unique SSNs
Total	178,863,694	100.0	168,338,342	94.1	9,304,745	5.2	1,220,607	0.7
One form only								
W-2	156,748,520	87.6	147,455,296	82.4	8,117,494	4.5	1,175,730	0.7
W-2c	32,189	(L)	29,241	(L)	520	(L)	2,428	(L)
Schedule SE	10,720,024	6.0	10,686,954	6.0	108	(L)	32,962	(L)
Two forms only								
W-2 and W-2c	2,032,979	1.1	1,851,438	1.0	173,632	0.1	7,909	(L)
W-2 and Schedule SE	9,202,656	5.2	8,206,088	4.6	995,008	0.6	1,560	(L)
W-2c and Schedule SE	2,887	(L)	2,489	(L)	397	(L)	1	(L)
All three forms	124,439	0.1	106,836	0.1	17,586	(L)	17	(L)

SOURCE: Author's calculations based on SSA data processing audit reports.

NOTE: (L) = less than 0.05 percent.

Table 10A.
Simulated SCC imputation, step 1: Applying the frequency distribution of ZIP Code/SCC combinations in the gold-standard (single-SCC) file

ZIP Code/SCC combination	Hypothetical instances of ZIP Code/SCC linkage in the gold-standard file		Frequency band	
	Number	Percent	Lower bound	Upper bound
ZIP Code 1 and—				
SCC A	10	100.0	0.0001	1.0000
ZIP Code 2 and—				
SCC B	20	66.7	0.0001	0.6667
SCC C	10	33.3	0.6668	1.0000
ZIP Code 3 and—				
SCC D	10	33.3	0.0001	0.3333
SCC E	10	33.3	0.3334	0.6667
SCC F	10	33.3	0.6668	1.0000

SOURCE: Author's illustration.

Table 10B.
Simulated SCC imputation, step 2: Applying a randomly generated number to the frequency bands to impute the SCC for four hypothetical workers

Scenario	Worker's ZIP Code on tax form	Randomly generated number	Imputed SCC
Worker W: ZIP Code not available on tax form	None	0.6395	Unknown
Worker X: Had one job and residential address appears on tax form	ZIP Code 1	0.9051	SCC A
Worker Y: Had two jobs and a different residential address associated with each job			
Job 1	ZIP Code 1	0.3654	SCC A
Job 2	ZIP Code 2	0.3816	SCC B
Worker Z: Had three jobs and the same residential address associated with each job			
Job 1	ZIP Code 3	0.1275	SCC D
Job 2	ZIP Code 3	0.2491	SCC D
Job 3	ZIP Code 3	0.8374	SCC F

SOURCE: Author's illustration.

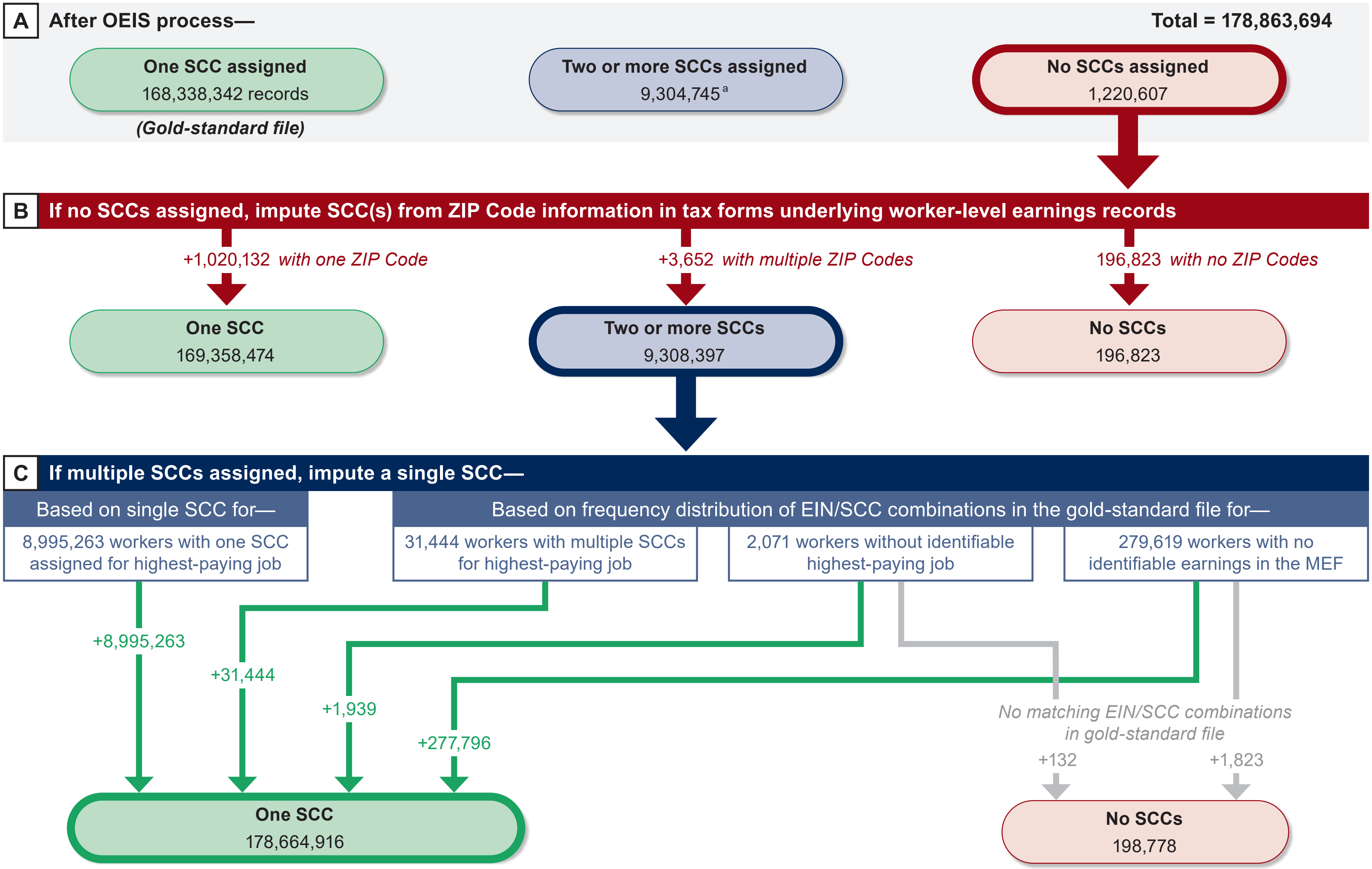
NOTE: SCC is imputed by identifying the randomly generated number's placement within the frequency bands shown in Table 10A.

Workers with No SCCs Assigned by the OEIS Process: ZIP Code Imputation

The OEIS process assigned no SCC to the records for 1.2 million workers. Using ZIP Code information on the tax forms, ORES imputed a single SCC for nearly 99 percent (1,020,132) of those workers (Chart 1, Panel B). For another 3,652 of those workers, the tax forms indicated multiple ZIP Codes and ORES therefore imputed multiple SCCs; those workers were added to the group to which OEIS assigned multiple SCCs. The remaining 196,823 records did not include ZIP Code information and could not be moved out of the "no SCC" group.

ORES uses the ZIP Code imputation method for the workers to whom OEIS assigns no SCCs and does not use it for workers to whom multiple SCCs are assigned—except for one subgroup. Of the 9,304,745 workers with multiple SCCs assigned for tax year 2017, 9,140,974 had no empty SCC fields associated with their jobs; but 163,771 had an empty SCC field for at least one job. ORES was able to impute a single SCC for 156,173 of the workers in the latter group using available ZIP Code information but could not impute an SCC for the other 7,598 workers based on ZIP Codes (Chart 2). Instead, ORES used EIN imputation, described below, for those workers.

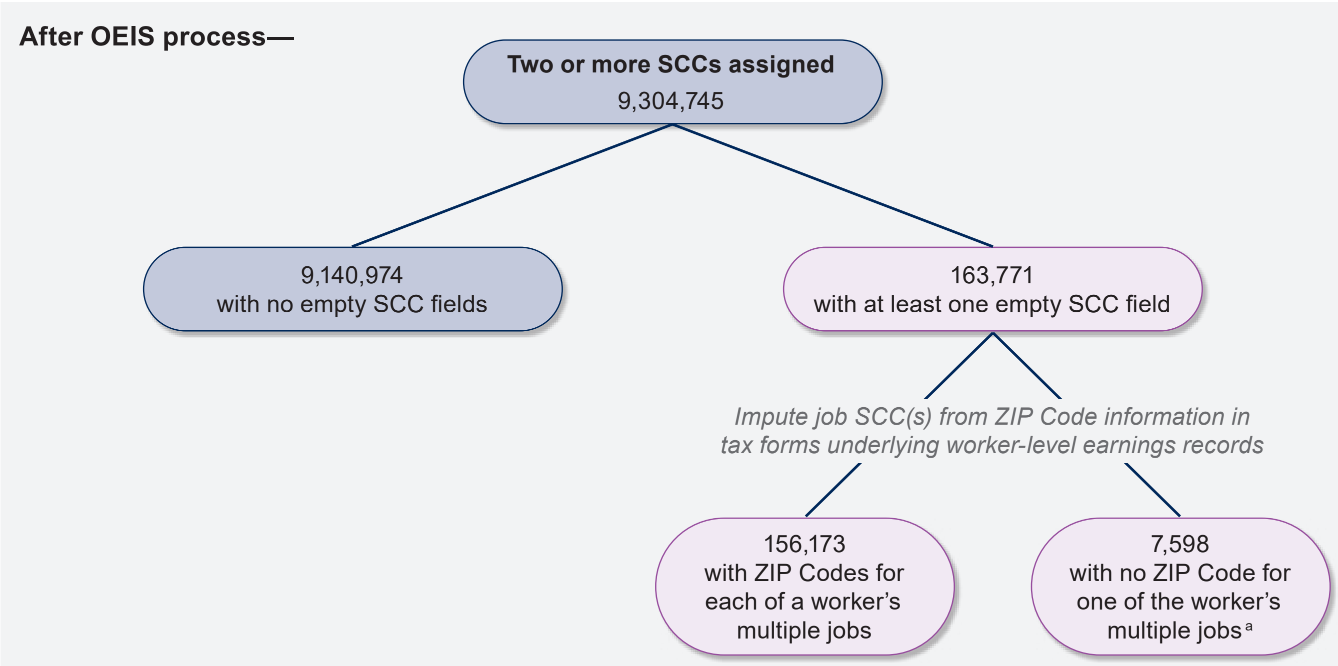
Chart 1.
New ORES methodology for assigning a single SCC to worker-level earnings (illustrated with tax year 2017 earnings records)



SOURCE: Author's calculations based on SSA data processing audit reports.

a. See Chart 2.

Chart 2.
Composition of worker-level earnings records with multiple SCCs assigned in the OEIS process, tax year 2017



SOURCE: Author's calculations based on SSA data processing audit reports.

a. ORES does not impute an SCC for the job with the missing ZIP Code but it imputes an SCC for that worker based on the ZIP Codes associated with the worker's other job(s).

Workers with Multiple SCCs Assigned by the OEIS Process: Imputing a Single SCC Based on Location of Highest-Paying Job

The next step is to determine which SCC to assign to a worker who was assigned multiple SCCs after the OEIS process or ZIP Code imputation. Using data from SSA's Master Earnings File (MEF),[18] ORES determines a worker's highest-paying job and assigns the SCC associated with the worker's address on the tax form for that job, with the rationale that the highest-paying job likely indicates where the worker resided for the longest time in that year. Chart 1, Panel C diagrams the process.

For the tax year 2017 records, ORES first examined the SSNs among the 9,308,397 workers whose records had multiple SCCs after the OEIS process and ZIP Code imputation and identified 9,250,150 that were valid and 58,247 that were not.[19] The MEF provided data on the earnings amount for each job these workers held in 2017. ORES assigned the single SCC corresponding with the location of the highest-paying job for 8,995,263 workers, or almost 97 percent of those with multiple SCCs assigned to their records by the OEIS process.

The earnings records for the remaining 313,134 workers either indicated multiple SCCs for the highest-paying job, indicated multiple SCCs but no highest-paying job identified, or had no earnings recorded in the MEF. ORES imputed a single SCC for most of these workers by comparing their assigned SCCs with the geographic distribution of workers in the gold-standard file who had the same employer, according to the EIN on their tax forms. In other words, ORES uses the method, described above and illustrated in Tables 10A and 10B, of aligning the EIN/SCC combinations in the tax records for these workers with the frequency distribution of those combinations for workers in the gold-standard file. Specifically:

- The records for 31,444 workers indicated multiple SCCs for the highest-paying job.[20] Imputations based on EIN/SCC combinations enabled ORES to assign a single SCC for each of these workers.
- The records for 2,071 workers did not indicate a single highest-paying job. Most of these workers had two different jobs with the same amount of reported earnings. Imputations based on EIN/SCC combinations enabled ORES to assign a single SCC

for 1,939 of these workers. The reported EIN/SCC combination for the remaining 132 workers in this group had no matches in the gold-standard file and ORES moved the records for those workers into the no-SCC group.

- The records for the remaining 279,619 workers did not have a matching EIN/SSN combination in the MEF. Imputations based on EIN/SCC combinations enabled ORES to assign a single SCC for 277,796 of the workers in this group. The reported EIN/SCC combination for the remaining 1,823 workers in this group had no matches in the gold-standard file and ORES moved the records for these workers to the no-SCC group.

Chart 1, Panel C summarizes the redistributions resulting from each of these imputations and for each type of record reflecting multiple SCCs. It shows that ORES ultimately was able to assign a single SCC to 99.89 percent of the 178,863,694 workers in the tax year 2017 file.

After assigning a single SCC for as many workers as possible, ORES adds a data field to the earnings record that identifies how the single SCC was assigned to each worker. Table 11 shows the distribution of workers by the type of information or imputation that generated their assigned SCC. For example, the OEIS process assigned a single SCC to 94.1 percent of all workers (the gold-standard file). Using the location of the worker's highest-paying job during the tax year was the next most common method of assigning an SCC. Combined, these two methods accounted for 99.1 percent of all workers in tax year 2017.[21]

ORES then joins the data files containing the demographic and the geographic data into a single Master Geographic-Demographic (MGD) data file. Table 12 presents the distribution of workers by each combination of information in the four primary data fields of the MGD file.[22] Note that 99.1 percent of the workers in the 2017 file had a valid SSN. Although ORES could not attribute any demographic information to the 1,524,401 individuals with invalid SSNs, an SCC was assigned for 1,521,552 (99.8 percent) of them.

With its new process, ORES attached date of birth, sex, and SCC data to 176,254,369 worker-level earnings records, or 98.5 percent of all the individuals in the MGD file. ORES assigned a value for all three variables to 99.4 percent of individuals with a valid SSN.

Next Steps

The MGD file permits ORES to generate annual earnings estimates using a data file that is substantially larger than the 1-percent CWHS. As noted earlier, OEIS has been creating a larger sample (10-CWHS-HE) that includes a 10-percent version of the CWHS' sampling frame plus all "high earners," defined as those whose earnings exceeded the Social Security maximum taxable amount in any year from 1978 forward. ORES has been working with a contractor to develop a new process for generating the annual earnings estimates using the 10-CWHS-HE and MGD files. Expanding the underlying sample tenfold and

Table 11.
Number and percentage distribution of SCCs assigned for tax year 2017, by source of information or type of imputation

Source or type	Number	Percent
Total	178,863,694	100.0
Initial OEIS process generates—		
Single SCC (gold-standard file)	168,338,342	94.1
No SCCs; ORES imputes worker's SCC using ZIP Code information	1,020,132	0.6
Multiple SCCs		
A single SCC is associated with the highest-paying job	8,995,263	5.0
ORES imputes a single SCC based on the frequency distribution of matching EIN/SCC combinations in the gold-standard file for—		
Workers with multiple SCCs for highest-paying job	31,444	(L)
Workers without an identifiable highest-paying job	1,939	(L)
Workers without earnings data in the MEF	277,796	0.2
Missing data or SCC not assigned	198,778	0.1

SOURCE: Author's calculations based on SSA data processing audit reports.

NOTE: (L) = less than 0.05 percent.

Table 12.
Worker records for tax year 2017 in the MGD data file, for each combination of types of information contained

Combination	Type of information				All individuals		Valid SSNs	
	SSN status	Date of birth	Sex	FIPS SCC	Number	Percent	Number	Percent
Total	. . .	. . .	. . .	. . .	178,863,694	100.0	177,339,293	100.0
1	Invalid	Missing	Missing	Missing	2,849	(L)	. . .	. . .
2	Invalid	Missing	Missing	Provided	1,521,552	0.9	. . .	. . .
3	Valid	Missing	Missing	Missing	650	(L)	650	(L)
4	Valid	Missing	Missing	Provided	767,401	0.4	767,401	0.4
5	Valid	Missing	Unknown	Missing	4	(L)	4	(L)
6	Valid	Missing	Unknown	Provided	3,188	(L)	3,188	(L)
7	Valid	Missing	Male	Missing	14	(L)	14	(L)
8	Valid	Missing	Male	Provided	32,812	(L)	32,812	(L)
9	Valid	Missing	Female	Missing	19	(L)	19	(L)
10	Valid	Missing	Female	Provided	22,963	(L)	22,963	(L)
11	Valid	Provided	Unknown	Missing	56	(L)	56	(L)
12	Valid	Provided	Unknown	Provided	62,763	(L)	62,763	(L)
13	Valid	Provided	Male	Missing	123,756	0.1	123,756	0.1
14	Valid	Provided	Male	Provided	90,628,653	50.7	90,628,653	51.1
15	Valid	Provided	Female	Missing	71,298	(L)	71,298	(L)
16	Valid	Provided	Female	Provided	85,625,716	47.9	85,625,716	48.3

SOURCE: Author's calculations based on SSA data processing audit reports.

NOTE: (L) = less than 0.05 percent; . . . = not applicable.

including high earners will significantly reduce the number of estimates that must be suppressed under data disclosure standards. Including all high earners in the sample should also mitigate the large year-to-year earnings fluctuations in counties with relatively few workers.

ORES will evaluate the new process for generating its earnings estimates by comparing them with other estimates from its own statistical publications and those of other federal agencies. Specifically, ORES will evaluate worker counts by sex, age, and type of earnings (wage and salary, self-employment). The evaluation will also assess the results of the new process for assigning SCCs.

Caveats

Researchers wishing to use the MGD file generated using the new methodology should be mindful of several issues when using the data:

1. The data files ORES receives from OEIS contain only the geographic information produced by the software; they do not contain any of the earnings information reported on the tax forms.
2. The data files used to assign SCCs are not part of the annual wage reporting process and thus are not subject to the standard cleaning and validating procedures that SSA executes before posting records to the MEF. As a result, the data contain some SSNs that are assumed to be invalid because they are not present in the Numident file. The *Annual Statistical Supplement* and *Earnings and Employment* focus on workers covered under the Social Security or Medicare Hospital Insurance programs. As such, only workers with valid SSNs and earnings reported on the MEF are included in those publications' earnings estimates. However, the W-2, W-2c, and Schedule SE provide valuable geographic information about all U.S. workers. For this reason, ORES processes records for all SSNs (valid and invalid) to provide data for the U.S. workforce. ORES cannot assign demographic information for the invalid SSNs.
3. Some individuals have deferred-compensation distribution amounts rather than current-year earned wage and salary income reported in Box 1 ("Wages, tips, other compensation") on Form W-2.[23]

Individuals who receive deferred-compensation distributions as their only earnings in a given tax year are not considered to be workers in the traditional sense. This situation applies to very few earners. For now, ORES includes these individuals in the MGD file and will reexamine these instances as it refines the new SCC-assignment methodology.

4. The OEIS process assigns multiple SCCs to a single worker in millions of instances, raising the critical question of how to determine a single correct SCC. As described above, ORES uses earnings data from the detailed segment of the MEF (containing job-level data) to determine the individual's highest-paying job during the year and assigns the SCC corresponding with the worker's address on the tax form submitted by that employer. This method reflects the assumption that the year's highest paying job indicates where the worker resided for the longest time during the year. However, the resulting SCC assignment may not necessarily reflect the current location of a worker who relocated during the earnings year, as the tax records do not specify a worker's most recent address.
5. The overwhelming majority, but not quite all, of the data SSA processes in a given calendar year are for earnings in the previous calendar year. For the time being, the new methodology does not use data received in a calendar year that are not for the previous calendar year's earnings. If the timing of the earnings data is unknown, those data may reflect earnings from before the previous calendar year, and the worker may have relocated at some point in the interim. Once ORES has processed all available historical geographic data, it will reassess whether these "off-year" data can be used.

Summary

This article identifies several limitations of ORES' current process for generating annual earnings estimates at the county level. A primary concern is the relatively small number of self-employed individuals in the 1-percent CWHS used to generate county-level estimates. ORES currently suppresses more than one-half of the county-level estimates for self-employed individuals to comply with data disclosure standards. Attempting to allocate approximately 186,000 self-employed individuals across 3,215 U.S. counties is problematic under the strict standards of primary and secondary cell suppression that are required to maintain confidentiality. In addition, reported earnings may vary widely from year to year in counties with small workforces.

The current methodology also truncates available address information to accommodate data storage restrictions dating from the 1990s, compromising the accuracy of some assigned SCCs. Additionally, the SCCs currently assigned for the earnings estimates are not compatible with the standard FIPS-based codes used in other ORES statistical publications and by other federal agencies. Further, the current process for assigning SCCs employs some hard-coded instructions that were never fully documented, meaning that their validity cannot be confirmed.

ORES is working toward replacing the 1-percent CWHS with the 10-CWHS-HE data file. The larger sample will reduce the incidence of cell suppression and the frequency of large year-over-year changes in earnings reported in counties with small workforces. Until then, the 1-percent CWHS is the only data set available for policy analysis and research that contains the earnings, geographic, and demographic data necessary to produce statistical publications containing earnings estimates.

This article describes the new process for assigning SCCs and demographic information for nearly all individuals with earnings reported on a Form W-2, Form W-2c, or Form 1040 Schedule SE in a given tax year. Testing the new methodology with records for tax year 2017, ORES increased the number of workers with SCC, date of birth, and sex information from the 1.5 million represented in the CWHS to more than 178 million, representing essentially the entire U.S. labor force. In addition to expanding the number of workers with geographic and demographic data assigned, the new methodology makes three other key changes: (1) SCCs are based on the complete address information reported in the W-2, W-2c, or Schedule SE; (2) standard FIPS-based SCCs are assigned; and (3) SCCs determined by information on the tax forms replace hard-coded SCC assignments.

With the ability to assign SCCs and demographic information for nearly all workers in a given tax year, ORES can develop a process to use the 10-CWHS-HE and MGD files to generate annual county-level earnings estimates. This work is ongoing and will mitigate or eliminate the problems identified with the current process for generating those estimates.

Notes

Acknowledgments: I would like to thank Greg Diez for his hard work and dedication related to the earnings data throughout his career at SSA and for sharing his knowledge with me. Greg was instrumental in getting ORES access to the geographic data that made this project possible. I would also like to thank Pat Purcell, Brad Trenkamp, Richard Chard, and Ben Pitkin for their comments on the draft and editorial assistance. I would also like to thank Cherice Jefferies, Theresa Wolf, Lewis Gaul, Dinesh Dasari, and Yikang Li for their assistance in developing the new methodology; and Athar Shaik, Jithender Gundawar, Rajitha Bandi, Kai Liu, and Kevin Beck for generating all the code for the project. I dedicate this article to my mom, my dad, and my brother. Although they are no longer with us, their love, support, and encouragement live on.

[1] Smith (1989) describes the CWHS.

[2] IRS Form W-2 is the annual wage and tax statement that employers file on behalf of employees. Form W-2c, "Corrected Wage and Tax Statement," is filed when a worker's original W-2 contained any errors or needs to be updated.

[3] The Numident contains records for all SSNs ever issued. The information is derived from SSA Form SS-5, the application for an SSN, which contains the individual's name, place and date of birth, and sex.

[4] The need to develop the current methodology for assigning geographic codes is discussed in Dill, Enis, and Williams (1991). The methodology is described in Dill, Bye, and Williams (1994).

[5] The purpose of ZIP Codes is to speed the flow of mail, not to designate county location.

[6] The SCCs currently used for the *Earnings and Employment* county-level earnings estimates predate the development of the FIPS codes.

[7] "Other and unknown" includes persons employed in American Samoa, Guam, the Northern Mariana Islands, and the U.S. Virgin Islands; U.S. citizens employed abroad by U.S. employers; persons employed on U.S. oceanborne vessels; and workers with unknown residence.

[8] Each of the county-level tables also includes estimates for the entire state and for the entire United States (including Puerto Rico). Those estimates duplicate figures shown in the state-level tables. This discussion disregards the duplicative state and U.S. totals to focus on the unique county-level estimates.

[9] Note that approximately 80,000 workers had both wage and salary earnings and self-employment income in the CWHS for 2017. These individuals are counted in both worker categories. The unduplicated count of total workers in the 2017 CWHS is slightly less than 1.7 million.

[10] There is a 1-year lag between the tax year (which is essentially equivalent to the earnings year) and the calendar year in which OEIS processes the data. For example, OEIS processed most of the tax year 2017 data in calendar year 2018.

[11] Although SSA also receives employer addresses from IRS Form W-3 ("Transmittal of Wage and Tax Statements"), IRS Form W-3c ("Transmittal of Corrected Wage and Tax Statements"), and SSA Form SS-4 ("Application for Employer Identification Number"), that information is not used in assigning SCCs.

[12] The algorithm was developed for the CWHS.

[13] Having more than one job during a given year is the most common reason the OEIS process assigns multiple SCCs to a worker based on that year's tax forms. Other reasons include the worker relocating while retaining a job, resulting in tax forms potentially indicating different addresses for the same job; errors in filing or processing the tax forms; and OEIS' use of address data that have not been subjected to the data cleaning procedures associated with SSA's annual wage reporting process.

[14] It may seem counterintuitive that the numbers of W-2s and W-2cs for workers who have both tax forms differ so widely. The numbers differ for several reasons; for example, a worker who holds a job that generates both a W-2 and a W-2c may also hold one or more jobs that generate only W-2s in that year.

[15] Technically, all W-2s are "job-level." Table 7 distinguishes between the W-2 records extracted in the OEIS process, which may include multiple jobs for a single worker or multiple forms for a single job, and the unduplicated number of discrete SSN/EIN combinations reported on the W-2s (labeled "jobs" in the table).

[16] The unduplicated number of jobs with records from all three types of tax forms is unavailable.

[17] The Finalist software attempts to assign SCCs using full addresses rather than relying only on ZIP Codes (which sometimes cross county lines). Therefore, after OEIS processing, some records may have an empty SCC field—for various reasons—even if a ZIP Code appears on the tax form.

[18] Olsen and Hudson (2009) describe the MEF.

[19] As noted earlier, ORES cannot assign demographic information to the records for workers with an invalid SSN but it can impute an SCC for them.

[20] For most of these workers, multiple W-2s (with differing addresses and ZIP Codes) were associated with the highest-paying job.

[21] Researchers who use the resulting data file and have concerns about any of the ORES imputation methods can identify the workers whose SCCs were based on each of these techniques and substitute "unknown" for the imputed values.

[22] Although the MGD file also includes information on date of death and date of death posting, Table 12 omits those fields to focus on the demographic data needed for the published SSA earnings estimates.

[23] Some deferred-compensation plans consider the distributions to be taxable as self-employment income. In these cases, the distributions may be subject to income tax as well as to Social Security and Medicare taxes under the Self-Employment Contributions Act.

References

Dill, Linda M., Barry V. Bye, and Cheryl I. Williams. 1994. "The Development of a New Geographic Coding System for the Continuous Work History Sample." *Social Security Bulletin* 57(4): 34–48.

Dill, Linda M., Adah D. Enis, and Cheryl I. Williams. 1991. "The Decline in Establishment Reporting: Impact on CWHS Industrial and Geographic Data." *Social Security Bulletin* 54(1): 2–20.

Olsen, Anya, and Russell E. Hudson. 2009. "Social Security Administration's Master Earnings File: Background Information." *Social Security Bulletin* 69(3): 29–46.

Smith, Creston M. 1989. "The Social Security Administration's Continuous Work History Sample." *Social Security Bulletin* 52(10): 20–28.

[SSA] Social Security Administration. 2019. *Earnings and Employment Data for Workers Covered Under Social Security and Medicare, by State and County, 2017.* Publication No. 13-11784. Washington, DC: SSA.