

|  |  |  |
| --- | --- | --- |
|  | Public-Use File Users Guide Report |  |

May 2023

Contents

[1. Introduction 1](#_Toc126914538)

[1.1 Overview of the SED 1](#_Toc126914539)

[1.2 Public Use File Contents 1](#_Toc126914540)

[2. Study Design 2](#_Toc126914541)

[2.1 Target Population 2](#_Toc126914542)

[2.2 Recruitment and Enrollment 2](#_Toc126914543)

[2.3 Randomization 4](#_Toc126914544)

[3. Questionnaire Design 5](#_Toc126914545)

[4. Data Collection 6](#_Toc126914546)

[5. Response rates and non-response weighting 7](#_Toc126914547)

[5.1 Definitions of attrition and analysis of attrition rates 8](#_Toc126914548)

[5.2 Accommodating missed quarterly interviews 9](#_Toc126914549)

[5.3 Participation and non-participation in follow-up survey 13](#_Toc126914550)

[5.4 Weighting adjustments 16](#_Toc126914551)

[5. PUF Data Set Details 18](#_Toc126914552)

[5.1 Data Structure 18](#_Toc126914553)

[5.2 Record Identifier 18](#_Toc126914554)

[5.3 Variable Naming Conventions 18](#_Toc126914555)

[5.4 Value Labels 19](#_Toc126914556)

[5.5 Missing Values 19](#_Toc126914557)

[5.6 Example Code 19](#_Toc126914558)

This page is intentionally blank.

## 1. Introduction

### 1.1 Overview of the SED

The Supported Employment Demonstration (SED) is a multi-component intervention aimed at improving the employment outcomes of Social Security disability applicants alleging a mental impairment who were recently denied benefits. The primary question that SSA seeks to answer is whether offering the Individual Placement and Support (IPS) model of employment services, along with behavioral health and other services fosters employment and clinical recovery that leads to self-sufficiency, improved quality of life, and less demand for disability benefits. Findings from this study can help policymakers improve existing programs that may lead to increased participation of individuals with disabilities in the workforce.

The SED used a randomized controlled trial (RCT) experimental design with three study arms to compare two treatment alternatives, entitled “Full-Service” and “Basic-Service,” against a “Usual Services” control group as the counterfactual. Participants randomized to one of the two treatment arms receive intervention services from one of 30 sites across 20 states. The sites were primarily community mental health agencies (CMHAs; 23 of the 30); the remaining sites were employment centers (2) or social services centers (5).

The duration of participation in the study is three years. Study participants live within the catchment area (i.e., geographical service area) of one of the sites and all expressed a desire to work or, if employed, sustain or seek a better job. The study excludes individuals who do not have the ability to provide informed consent, those who were already receiving employment services from the demonstration site, and those who resided in a nursing home or other custodial institution.

### 1.2 Public Use File Contents

The Public Use File contains the documents listed in Table 1-1.

**Table 1-1 List of documentation included with SED PUF**

|  |  |
| --- | --- |
| **Document Name** | **Description** |
| sed\_puf\_final\_fmt\_upd\_202305.sas7bdat | Public Use Data file (SAS format) |
| Formats.sas7bcat | Public use file format catalogue |
| sed\_puf\_final\_PUFCodebook 202305.pdf | Codebook containing variable labels, frequency distributions / means and ranges |
| SED Baseline Survey annotated \_ 02.16.18\_INSTRUMENT\_English.docx | Baseline survey (administered one time at enrollment) |
| SED Quarterly Survey\_INSTRUMENT\_English.docx | Quarterly survey (administered once every quarter for twelve quarters) |
| Changes in Questionnaire Content Between SED Quarterly Survey.pdf | Additional questions added to end of the quarterly survey regarding COVID (CV) |

## 2. Study Design

### 2.1 Target Population

Study enrollees included recently denied disability applicants alleging a mental impairment residing in the catchment area of a participating community mental health agency. The study targeted individuals whose application received a denial within two months of the study start and those denied in each month going forward until month 11 of recruitment, or until the site enrolls its quota of enrollees.

### 2.2 Recruitment and Enrollment

Recruitment began with Westat home office staff mailing a study invitation packet to potential enrollees. The packet included an introductory letter and brochure, which gave the potential enrollee an overview of the study and informed them that a local research assistant would contact them shortly. To allow enough time for the invitation packet to arrive, the recruiter attempted to contact each potential enrollee by phone 5 days after the mailing was sent. The recruiter followed up with an in-person visit if they were unable to reach the potential enrollee by phone.

During initial contact, the recruiter screened the potential enrollee for study eligibility. The eligibility screener consisted of three questions designed to ascertain whether the potential enrollee had an interest in finding a job or getting a better job, and if and where he or she currently received employment services. Potential enrollees met the study’s eligibility requirements only if they had an interest in working (or getting a better job) and did not currently receive services from the demonstration site (at the time of recruitment). Potential enrollees receiving employment services elsewhere remained eligible for the study. If the screener deemed potential enrollees eligible, the recruiter invited them to attend a Recruitment Information Meeting (RIM), either immediately following the screener or at another time.

#### Sampling

The recruitment process did not involve a formal sampling plan. This means that study participants were not selected for participation based on pre-specified criteria, the sample was not designed to represent the characteristics of a larger universe of denied applicants, and there is no Primary Sampling Unit (PSU). Because study participation was voluntary and participants were drawn from site service areas (i.e., ‘catchment areas’), the study participants are a “convenience” sample of denied applicants willing to participate.

The only “sampling” completed was a simple random sample to reduce the number of attempted contacts for the recruitment. After conducting the programmatic screener to identify ineligible denied applicants, Westat drew a 50% random sample of the 47,007 to release for contact by the recruitment team. Table 2-1 shows the total numbers of denied applicants with a mental impairment after the programmatic screening for each of the thirty sites separately. The two right hand columns in the table show the total from each site who enrolled in the SED study and the percentage of the total denied applicants enrolled out of the records received during the recruitment period.

**Table 2-1. Denied applicants with a mental impairment by site and total enrolled in SED**

|  |  |  |  |
| --- | --- | --- | --- |
| **Site** | **Total Denied Applicants in Catchment Areaa** | **Total Enrolled in SED** | **Percent Enrolled out of total** |
| 1 | 1349 | 121 | 9% |
| 2 | 1410 | 119 | 8% |
| 3 | 2292 | 118 | 5% |
| 4 | 1396 | 71 | 5% |
| 5 | 607 | 60 | 10% |
| 6 | 3881 | 114 | 3% |
| 7 | 483 | 59 | 12% |
| 8 | 1456 | 118 | 8% |
| 9 | 899 | 54 | 6% |
| 10 | 1173 | 123 | 10% |
| 11 | 3769 | 118 | 3% |
| 12 | 1873 | 120 | 6% |
| 13 | 1308 | 120 | 9% |
| 14 | 672 | 60 | 9% |
| 15 | 4839 | 134 | 3% |
| 16 | 563 | 72 | 13% |
| 17 | 1829 | 116 | 6% |
| 18 | 1990 | 120 | 6% |
| 19 | 897 | 120 | 13% |
| 20 | 362 | 59 | 16% |
| 21 | 1520 | 118 | 8% |
| 22 | 1988 | 119 | 6% |
| 23 | 754 | 101 | 13% |
| 24 | 1100 | 115 | 10% |
| 25 | 2728 | 107 | 4% |
| 26 | 1561 | 107 | 7% |
| 27 | 1335 | 71 | 5% |
| 28 | 1692 | 117 | 7% |
| 29 | 468 | 34 | 7% |
| 30 | 813 | 59 | 7% |
| **Total** | 47007 | 2944 | 6% |

*Notes:* a Includes potentially eligible denied applicants after programmatic screening eliminated ineligible applicants. Denied applicant records received from December 2017 through March 2019.

### 2.3 Randomization

The RIM consisted of an hour-long meeting in which the recruiter described the SED in more detail using visual aids (i.e., study flip book and video). This meeting also provided the opportunity for the potential enrollee to ask questions and consider his or her options. For individuals willing to participate in the study, the recruiter arranged the next step of administering a competency screener to determine whether the potential enrollee had the mental capacity to provide informed consent. The recruiter then obtained written informed consent from those who passed the competency screener, conducted an hour-long baseline interview, and provided the enrollee with the results of his or her randomization assignment to either the Full-Service, Basic-Service, or Usual Services (Control) group.

Wrap-up activities included collecting the enrollees’ insurance status information, assigning enrollees a reloadable study debit card, and providing them with other study-related materials. Additionally, for enrollees assigned to either the Full-Service and Basic-Service treatment groups, the recruiter called the demonstration site Team Lead on the enrollees’ behalf to make an introduction. In some cases, the recruiter arranged an in-person meeting to hand off the new enrollee to the demonstration site. Finally, within 2 weeks of enrollment, the recruiter or Westat home office staff scheduled the enrollee for the Composite International Diagnostic Interview (CIDI) (if selected) and conducted the interview, either in person or by phone.[[1]](#footnote-1)

## 3. Questionnaire Design

The questionnaire is divided in to sections by subject area. The survey sections are:

1. Demographics (DM) – baseline demographics including age, gender, race and ethnicity, living situation, education
2. Work history and income (WI) / Employment Outcomes (EO): dates of employment and earnings
3. Health Status: The items of the SF-12 scale used to estimate mental health status (MCS) and physical health status (PCS)
4. Colorado Symptom Index (CSI): The items of the CSI measure psychological or emotional difficulties. These items are used to construct the CSI scale.
5. Satisfaction with Life (SL): Participants are asked to rank their satisfaction with life on a 7-piont scale from 1=Terrible to 7=Delighted
6. Alcohol, Drug, and Tobacco Use (SA): The AUDIT, DAST, and tobacco use of the participant as self-reported.
7. Health Care Coverage and Service Utilization (HC): Participants are asked to indicate their health care coverage and report hospital visits and emergency department visits.
8. Health conditions / comorbidities (CM): Participants are asked about specific diagnoses for comorbidities
9. Prescription Medication (PM): The PM items ask respondents to indicate the conditions for which they take prescription medications
10. Justice Involvement (JI) : Participants self-report arrests and convictions.
11. Digit Symbol Test: At enrollment, participants complete the Digit Symbol Test to assess cognitive function.
12. Work Disability Functional Assessment Battery (WD-FAB): The WD-FAB is a computer adaptive test that assesses work function on several domains.

## 4. Data Collection

All enrollees will complete an in-person survey at baseline (in English or Spanish), and again every 3 months (quarterly) via telephone through the end of study participation at 36 months. The information in the surveys will inform the impact and cost-benefit analysis. All enrollees, regardless of study arm participate in quarterly follow-up interviews, in which they provide updates on their employment and use of services during the previous three months. The annual survey (e.g., quarters 4, 8 and 12) include additional items related to health status and functioning.

The information from the survey informs various outcomes of interest for the impact and cost-benefit analyses, including the following domains: Clinical recovery, employment and earnings; and quality of life. Table 4-1 presents the measures where the data source is the survey.

Table 4-1. Content domains for survey instruments

|  |  |  |  |
| --- | --- | --- | --- |
| Survey domain | Mode | | |
| Baseline | Quarterly | Annually |
| Contact Information | ✓ | ✓ | ✓ |
| Demographics | ✓ |  |  |
| CIDI Modules | ✓ |  |  |
| Work History/Employment Outcomes | ✓ | ✓ | ✓ |
| Individual Income | ✓ | ✓ | ✓ |
| Household Income | ✓ | ✓ | ✓ |
| Health Status (SF-12) | ✓ |  | ✓ |
| Colorado Symptom Index | ✓ |  | ✓ |
| Lehman Quality of Life Scale | ✓ |  | ✓ |
| Substance Use (Alcohol) (AUDIT)\* | ✓ |  | ✓ |
| Substance Use (Drugs) (DAST-10)\*\* | ✓ |  | ✓ |
| Body Mass Index | ✓ |  | ✓ |
| Health Care Coverage | ✓ | ✓ | ✓ |
| Health Care Service Utilization | ✓ | ✓ | ✓ |
| Health Conditions/Comorbidities | ✓ |  | ✓ |
| Involvement with Criminal Justice System | ✓ |  | ✓ |
| Digit Symbol Test | ✓ |  |  |
| Work Disability Functional Assessment Battery | ✓ |  | ✓ |

\* Alcohol Use Disorders Identification Test (AUDIT)

## 5. Response rates and non-response weighting

The SED study team measured employment, health status, health care utilization, and quality of life outcomes using follow-up surveys. Table 5-1 summarizes the quarterly survey response rates for Quarters 1 through 12. The five columns appearing under the label “All” show the number of completes, deceased, skipped, and withdrawn participants for each quarter. The numbers of deceased and withdrawn increased over time; by Quarter 12, there were 83 deceased participants and 50 participants who formally withdrew at some time prior to the interview.

The columns to the right of Table 5-1 show the response rates for each quarter excluding deceased and withdrawn participants. Response rates among eligible participants held above 70 percent for the first two years of study enrollment (quarters 1 through 8). The third year of the study saw a drop-off in completion rates; by Quarter 12, roughly two-thirds (65.3%) of eligible enrollees completed the survey.

Table 5-1. Response rates by quarterly interview for eligible enrollees

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Quarter** | **All** | | | | | **Excluding Deceased and Withdrawn** | | |
| **Complete** | **Deceased** | **Skipped** | **Withdrawn** | **Total** | **Complete** | **Skipped** | **Total** |
| 1 | 2,195 | 4 | 739 | 6 | 2,944 | 2,195 | 739 | 2,934 |
|  | (74.56) | (0.14) | (25.1) | (0.20) |  | (74.81) | (25.19) |  |
| 2 | 2,070 | 11 | 856 | 7 | 2,944 | 2,070 | 856 | 2,926 |
|  | (70.31) | (0.37) | (29.08) | (0.24) |  | (70.75) | (29.25) |  |
| 3 | 2,071 | 17 | 847 | 9 | 2,944 | 2,071 | 847 | 2,918 |
|  | (70.35) | (0.58) | (28.77) | (0.31) |  | (70.97) | (29.03) |  |
| 4 | 2,080 | 27 | 820 | 17 | 2,944 | 2,080 | 820 | 2,900 |
|  | (70.65) | (0.92) | (27.85) | (0.58) |  | (71.72) | (28.28) |  |
| 5 | 2,105 | 33 | 786 | 20 | 2,944 | 2,105 | 786 | 2,891 |
|  | (71.50) | (1.12) | (26.70) | (0.68) |  | (72.81) | (27.19) |  |
| 6 | 2,094 | 38 | 787 | 25 | 2,944 | 2,094 | 787 | 2,881 |
|  | (71.13) | (1.29) | (26.73) | (0.85) |  | (72.68) | (27.32) |  |
| 7 | 2,061 | 50 | 802 | 31 | 2,944 | 2,061 | 802 | 2,863 |
|  | (70.01) | (1.70) | (27.24) | (1.05) |  | (71.99) | (28.01) |  |
| 8 | 2,043 | 56 | 807 | 38 | 2,944 | 2,043 | 807 | 2,850 |
|  | (69.40) | (1.90) | (27.41) | (1.29) |  | (71.68) | (28.32) |  |
| 9 | 1,959 | 62 | 879 | 44 | 2,944 | 1,959 | 879 | 2,838 |
|  | (66.54) | (2.11) | (29.86) | (1.49) |  | (69.03) | (30.97) |  |
| 10 | 1,915 | 67 | 916 | 46 | 2,944 | 1,915 | 916 | 2,831 |
|  | (65.05) | (2.28) | (31.11) | (1.56) |  | (67.64) | (32.36) |  |
| 11 | 1,875 | 76 | 945 | 48 | 2,944 | 1,875 | 945 | 2,820 |
|  | (63.69) | (2.58) | (32.1) | (1.63) |  | (66.49) | (33.51) |  |
| 12 | 1,835  (62.33) | 83  (2.82) | 976  (33.15) | 50  (1.70) | 2,944 | 1,835  (65.28) | 976  (34.72) | 2,811 |

*Notes*: Percentages are in parentheses, summing across each row.

### 5.1 Definitions of attrition and analysis of attrition rates

We measure four different types of outcomes: outcomes in the domains of employment and earnings, 2) outcomes relating to health care utilization, 3) outcomes relating to health status and quality of life, and 4) SSA benefit outcomes (approvals, appeals, and benefits paid). For the first two types of outcomes, we use data from the 12 quarterly follow-up surveys. For the third type of outcomes, we use responses to additional questions included in the quarterly interviews for the end of each full year of the intervention (i.e., quarters 4, 8, and 12). For the fourth type of outcomes we use data from SSA administrative records. The SSA administrative records provide complete data for all SED participants. The follow-up surveys, however, provide data for only those study participants who completed the surveys. Table 5-2 below provides the analytic sample sizes after removing participants due to attrition for each of the analyses, by study arm.

### 5.2 Accommodating missed quarterly interviews

The quarterly follow-up surveys allowed measure of the employment-related and health care utilization outcomes over the entire intervention period. The study team expected, however, that many participants would miss at least some follow up interviews throughout the three-year study period. Therefore, interviewers asked participants to fill in gaps in employment-related and health services outcomes due to missed prior surveys. For example, if a respondent missed the previous quarterly survey, on the next completed survey the interviewer asks the participant to provide work history for the entire 6 months since the last completed survey (not just the current quarter).

Although the survey design reduced missing data due to skipped surveys by asking respondents to fill in information for skipped surveys, there are two limitations to note. First, participants who broke off and never completed the final survey before transitioning off the study have missing data for all quarters after the final completed survey. Second, participants who missed a large number of surveys must provide information such as dates of employment, hospitalizations, and emergency room visits for an extended period of time in order to fill in the gap. A lengthy recall period may lead to inaccurate estimates.

To address these limitations, the study team limited the analysis of outcomes to those participants who completed enough surveys to provide an accurate accounting of their work and health over the course of the study. The outcomes that rely on respondent recall are employment, weeks worked, hours worked, total earnings, hospitalizations, and emergency room visits. For these outcomes, the analysis includes participants who meet the following criteria:

1) Participants must complete the final (quarter 12) follow-up survey, and

2) Participants must not have a gap in completed quarterly surveys of longer than one year.

The rationale behind the second criterion is that research shows that recall periods of less than one year can provide reliable estimates of aggregate numbers of events such as hospitalizations.[[2]](#footnote-2)

The team adopted a different approach for the health status and quality of life data reported in the Quarter 4, 8, and 12 interviews. For these measures, relevant questions asked for single point in time responses and no gaps due to missing any of these interviews could be filled in from other non-missing quarterly interviews (since they did not include the questions relevant to these outcomes).

Thus, health status and quality of life outcomes were measured as the changes between scores measured at baseline and at the Quarter 4, 8 and 12 follow ups. Accordingly, these outcomes include only responses from all participants who completed the relevant follow up survey items from those specific quarterly interviews.

#### 5.2 Completion rates for outcome analyses

Table 5-2 shows completion rates by study year and for the group that finished the final survey and had no gaps in surveys longer than one year.

For measuring outcomes based on recall such as employment, earnings, and health care utilization, 60.19 percent of enrollees completed the final survey and had no survey gaps over one year. Larger percentages completed the surveys needed to measure health status and quality of life annually throughout the study: 71.31, 70.74, and 64.14 percent, completed the year 1, year 2, and year 3 surveys respectively.

The Basic-Service and Full-Service study participants were more likely to complete the surveys than the Usual Services participants. In each year, the Basic-Service study participants had the highest completion rates, followed by the Full-Service and then the Usual Services participants. The differences in completion rates among the study arms are present in each year of the study but are larger in the final year than in the first year.

Table 5-2. Completion by year and recall by study arm

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Completed**  **Q4** | | **Completed Q8** | **Completed Q12** | | **Completed Q12 & no survey gaps over 1 year** | **Total**  **N** |
| **All eligible participants** | |  |  | |  |  |  |
| Percent of non-deceased % | | 71.31 | 70.74 | | 64.14 | 60.19 |  |
| N | | 2,080 | 2,043 | | 1,835 | 1,722 | 2,944 |
| **Study arm** | |  |  | |  |  |  |
| Basic-Service % | | 73.95 | 72.89 | | 66.98 | 62.59 |  |
| n | | 724 | 707 | | 641 | 599 | 987 |
| Full-Service % | | 71.64 | 70.52 | | 65.27 | 60.88 |  |
| n | | 692 | 677 | | 624 | 582 | 976 |
| Usual Services % | | 68.31 | 68.79 | | 60.13 | 57.07 |  |
| n | | 664 | 659 | | 570 | 541 | 981 |

Notes: Percentages exclude participants who were deceased at the time of the survey from the denominator.

Table 5-3 provides a summary of the potential sample size for each analysis by outcome given the response rates among all study participants. Outcomes that rely on recall of dates throughout study enrollment (e.g., employment, health care utilization) utilize data from participants who completed the final interview and had no gaps in surveys over one year (n=1,722). Outcomes measured at specific points in time, such as MCS scores, utilize all the participants who provided those scores on the annual interviews.

Table 5-3. Outcome measures, definitions, and data sources

|  |  |  |  |
| --- | --- | --- | --- |
| **Impact Measure** | **Definition** | **Data Source** | **Sample** |
| **Employment** |  |  |  |
| **Employment Rate** | Percentage of participants who worked during study enrollment | Participant Interviews | Completed Q12 survey with no survey gaps over one year  N=1,722 |
| **Weeks Employed** | Number of weeks of participant-reported employment |
| **Total Hours worked** | Sum of reported hours worked across all jobs |
| **Total Earnings** | Sum of earnings based on reported work and pay |
| **Earnings in the past month** | Sum of earnings reported in the most recent month |
| **SSA Disability Benefits** |  |  |  |
| **Allowance Rate** | Percentage of participants accepted onto the disability rolls | SSA administrative data | All eligible enrollees  N=2,944 |
| **Time to Award** | Number of days to disability award |
| **Benefit appeal attempts** | Number of appeal attempts during study |
| **Total Disability Payments** | Total disability award payments during study period |
| **Health Status** |  |  |  |
| **Mental Health (SF-12)** | MCF difference scores (Study Entry vs. Exit Interview) | Participant Interviews | Completed Q12 survey  N=1,835 |
| **Physical Health (SF-12)** | PCF difference scores (Study Entry vs. Exit Interview) |
| **Colorado Symptom Index** | Difference scores (Study Entry vs. Exit Interview) |
| **Quality of Life** |  |  |  |
| **Satisfaction with Life** | Difference scores (Study Entry vs. Exit Interview) | Participant Interviews | Completed Q12 survey  N=1,835 |
| **Utilization of Services** |  |  |  |
| **Emergency Room Visits** | Total emergency room visits during study | Participant Interviews | Completed Q12 survey with no survey gaps over one year  N=1,722 |
| **Hospital Overnight Stays** | Total overnight hospital stays during study | Participant Interviews |
| **Outpatient visits** | Total outpatient visits during study | Participant Interviews |
| **Routine mental health visits** | Total mental health visits in last month of study enrollment | Participant Interviews | Completed Q12 survey  N=1,835 |
| **Routine general health visits** | Total general health visits in last month of study enrollment | Participant Interviews |
| **Routine employment support visits** | Total employment support visits in last month of study enrollment | Participant Interviews |

### 5.3 Participation and non-participation in follow-up survey

We followed three approaches to test and adjust for attrition patterns when estimating impacts: nonresponse weighting, regression adjustment, and bounding. First, we conducted exploratory regressions to identify characteristics associated with attrition and simultaneously control for multiple characteristics that may relate to attrition. We began with baseline characteristics collected from all study participants and identified characteristics related to completion of the final survey and completion of the necessary surveys to construct recall measures. These regressions identified several variables that the team used in the weighting and regression analyses. Variables related to attrition included:

* Gender,
* Age,
* TANF receipt,
* SNAP receipt,
* Time spent in shelter or on street prior to enrollment,
* Work history (e.g., working at baseline, worked in the past two years prior to enrollment),
* PCS baseline score,
* Number of outpatient visits in year prior to enrollment,
* Arrests prior to enrollment, and
* Race and ethnicity.

Table 5-4 summarizes characteristics of enrollees by subgroups based on completion rates. Generally, women, older participants, those with higher levels of formal education, and black non-Hispanic participants had higher response rates than others. Participants with lower PCS scores also had higher response rates than those with higher PCS scores.

Table 5-4. Completion rates by subgroup among non-deceased SED enrollees

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Completed Q4** | | **Completed Q8** | | **Completed Q12** | | **Completed Q12 & no survey gaps over 1 year** | |
| **n** | **%** | **n** | **%** | **n** | **%** | **n** | **%** |
| **Gender** |  |  |  |  |  |  |  |  |
| Male | 850 | 67.03 | 828 | 66.13 | 738 | 59.23 | 692 | 55.54 |
| Female | 1,230 | 74.59 | 1,215 | 74.27 | 1,097 | 67.93 | 1,030 | 63.78 |
| **Age** |  |  |  |  |  |  |  |  |
| 18-34 | 834 | 67.42 | 816 | 66.07 | 58.96 | 58.96 | 673 | 54.80 |
| 35 and over | 1,246 | 74.17 | 1,227 | 74.23 | 68.03 | 68.03 | 1,049 | 64.24 |
| **Education** |  |  |  |  |  |  |  |  |
| Less than high school | 354 | 65.68 | 349 | 65.48 | 325 | 61.67 | 292 | 55.41 |
| Completed high school | 611 | 69.04 | 619 | 70.42 | 539 | 61.88 | 510 | 58.55 |
| Some college | 754 | 75.40 | 725 | 73.16 | 652 | 66.33 | 619 | 62.97 |
| Associates | 155 | 75.98 | 149 | 73.76 | 132 | 66.33 | 122 | 61.31 |
| Bachelors or higher | 206 | 71.28 | 201 | 71.02 | 187 | 66.55 | 179 | 63.70 |
| **Work history at baseline** |  |  |  |  |  |  |  |  |
| Working when enrolled | 390 | 70.02 | 391 | 70.32 | 361 | 65.05 | 336 | 60.54 |
| Not working when enrolled but worked in past 2 years | 917 | 70.38 | 901 | 69.84 | 800 | 62.60 | 753 | 58.92 |
| Worked but not in the past 2 years | 710 | 73.65 | 686 | 72.21 | 614 | 65.60 | 578 | 61.75 |
| Never worked | 57 | 71.25 | 59 | 74.68 | 54 | 68.35 | 51 | 64.56 |
| Missing | 6 | 46.15 | 6 | 46.15 | 6 | 46.15 | 4 | 30.77 |
| **Race and ethnicity** |  |  |  |  |  |  |  |  |
| White non-Hispanic | 970 | 68.89 | 958 | 68.82 | 856 | 61.98 | 794 | 57.49 |
| Black non-Hispanic | 624 | 75.54 | 600 | 73.35 | 545 | 67.37 | 513 | 63.41 |
| Hispanic | 247 | 68.99 | 253 | 71.07 | 230 | 65.53 | 219 | 62.39 |
| Two or more races non-Hispanic | 183 | 74.39 | 181 | 74.18 | 154 | 63.37 | 148 | 60.91 |
| Other / missing | 56 | 72.41 | 51 | 63.79 | 50 | 61.40 | 48 | 59.65 |
| **Received SNAP** | 1,460 | 72.85 | 1,444 | 72.75 | 1,286 | 65.55 | 1,213 | 61.82 |
| **Received TANF** | 214 | 70.39 | 210 | 69.77 | 181 | 60.74 | 171 | 57.38 |
| **Arrested in past year** | 247 | 64.32 | 245 | 65.16 | 214 | 57.53 | 198 | 53.22 |
| **Mental Component Score (MCS)** |  |  |  |  |  |  |  |  |
| Less than 20 | 353 | 68.28 | 349 | 68.16 | 322 | 63.14 | 296 | 58.04 |
| 21 to 30 | 659 | 72.90 | 639 | 71.48 | 573 | 64.67 | 542 | 61.17 |
| 31 to 40 | 522 | 73.83 | 510 | 72.86 | 445 | 64.68 | 425 | 61.77 |
| 41 to 50 | 316 | 69.45 | 318 | 70.35 | 297 | 66.29 | 276 | 61.61 |
| 51 or greater | 184 | 69.17 | 182 | 68.68 | 164 | 61.89 | 151 | 56.98 |
| Missing | 46 | 67.65 | 45 | 69.23 | 34 | 53.13 | 32 | 50.00 |
| **Physical Component Score (PCS)** |  |  |  |  |  |  |  |  |
| Less than 20 | 159 | 74.65 | 155 | 73.11 | 144 | 68.25 | 135 | 63.98 |
| 21 to 30 | 560 | 74.47 | 540 | 72.87 | 494 | 67.67 | 462 | 63.29 |
| 31 to 40 | 532 | 72.78 | 550 | 75.76 | 486 | 67.59 | 457 | 63.56 |
| 41 to 50 | 412 | 71.65 | 401 | 69.98 | 363 | 63.91 | 336 | 59.15 |
| 51 or greater | 371 | 64.19 | 352 | 61.65 | 314 | 55.18 | 300 | 52.72 |
| Missing | 46 | 67.65 | 45 | 69.23 | 34 | 53.13 | 32 | 50.00 |

Notes: Percentages exclude participants who were deceased at the time of the survey from the denominator.

Table 5-4 Continued. Completion rates by subgroup among non-deceased SED enrollees

### 5.4 Weighting adjustments

We created nonresponse-adjusted weights to estimate outcomes for each study arm. As described above, participants must respond to more surveys to be included in measures that require recall over the entire study period. These recall outcomes require a response to the final (quarter 12) interview and no gaps in surveys over one year throughout the three-year study period. Outcomes that we measure at a single point in time, such as health status at the end of the study, require only the completion of the final (quarter 12) interview.

We also analyzed outcomes at each intermediate anniversary of study enrollment (Year 1 and Year 2). To accommodate these analyses, we used four different definitions of nonresponse to calculate four separate weights:

1. Weight 1: Respondents include any study participants who responded to the final (quarter 12) survey. Weight 1 is suitable for analyses of outcomes measured at the end of the study period that do not require recall over the entire study period (e.g., MCS, PCS, CSI, earnings in the past month). These weights are included in the PUF data set under variable **PNR2W0\_GRP\_B**.
2. Weight 2: Respondents include any study participant who responded to the final (quarter 12) survey and who did not have a gap in interviews longer than one year throughout the study period. Weight 2 is suitable for analyses that rely on recall (e.g., employment, total earnings, total hours worked, number of inpatient hospital stays). These weights are included in the PUF data set under variable **PNR2W0\_GRP\_F**.
3. Weight 3: Includes as a respondent any study participant who responded to the first annual (quarter 4) interview. These weights are included in the PUF data set under variable **PNR2W0\_Q4**.
4. Weight 4: Includes as a respondent any study participant who responded to the second annual (quarter 8) interview. These weights are included in the PUF data set under variable **PNR2W0\_Q8.**

We calculated non-response adjusted weights using data from the baseline survey identified from exploratory regressions of attrition listed above. The team used a CHAID model to identify variables related to nonresponse using each of the four definitions of nonresponse.

All enrolled SED participants received an initial weight of 1.0. The team then applied adjustment factors to increase the initial weights of responding participants upward for the nonresponding participants.

### 5.5 Cleaning and outliers

Survey data are self-reported by participants. For continuous measures such as income, BMI, hours worked, and weeks worked, the evaluation team checked data for outliers and corrected errors when it was clear that an error was made based on reviews by interviewers and responses to other questions. If it was not possible to determine if a mistake was made, and the values were within a plausible range, the participant’s responses were left unchanged. Some instances of high values due to self-report include:

* Total hours worked: Participants were asked how many hours they worked in a typical day, days worked in a typical week, and weeks worked in a typical month. These responses were used to construct total hours worked per quarter / annual / study period. Some participants may have misinterpreted the question and included on-call time or time that they were not necessarily paid in their response, resulting in higher-than-expected values.
* BMI: Participants were asked to report their height and weight, which were then used to calculate BMI. Self-reported height and weight may not be accurate. Responses were checked to ensure that they fell within plausible ranges at each interview; however, extreme values that still fell within the plausible range were left. It was not possible to verify self-reported BMI because participants completed follow-up interviews by telephone. In the PUF, BMI values were top-coded to reduce the risk of identification for those with values on the extreme end of the distribution. The team determined that this top-coding did not significantly reduce the utility of the BMI variable because top-coding did not change the clinical categorization (i.e., obese) of any individual.

## 5. PUF Data Set Details

### 5.1 Data Structure

The SED data are represented in a file that includes:

* A file with one record for every respondent (n=2,944). The data are in SAS format in the file named sed\_puf\_final.sas7bdat. The program FormatCodeBook\_sed\_puf\_final.sas should be run first to define value labels and formats.

### 5.2 Record Identifier

The variable named SMSID serves as the unique identifier for each record in the SED data file.

### 5.3 Variable Naming Conventions

The SED Study variable naming convention was designed to provide context for each variable based on the content being represented by the item and thus to facilitate the identification and classification of items for analyses.

Variables ending in \_qX represent information from a specific quarter of the survey, where X is a number from 0 (baseline) to 12 (end of the third year). Variables ending in \_yX represent annual information, such as employed\_y1 meaning ever employed in year 1 (using variables employed\_q1 through employed\_q4). Finally, variables ending in \_study are versions that comprise the entire time range of the study.

All variables on the data file include a label that briefly explains the content of the item. For clarity, consistency, and usability, labels are assigned to each variable using a standard convention.

### 5.4 Value Labels

The labels associated with variable values are provided. All value labels include both the data value as well as the description of the value. For example, a variable for an item with yes/no responses has “1 = Yes” and “2 = No” as value labels.

### 5.5 Missing Values

The data file includes missing values, as are generally defined in Table 5-1. The user is advised to refer to the descriptive frequencies in the documentation accompanying the data files to ascertain the levels of missing data in variables of interest.

Table 5-1. SED Study missing value codes and their descriptions

| Value | Definition | Applicable  data types | Description |
| --- | --- | --- | --- |
| -5 | Refused | Numeric | This value is coded when a respondent answered a question as “Refused.” |
| -7 | Proper skip | Numeric, Character | This value is coded when a respondent skipped the question due to previous responses and skip patterns. |
| -9 | Don’t know | Numeric, Character | This value is coded when a respondent answered a question as “Don’t know.” |

### 5.6 Example Code

The following SAS code creates tables including the unweighted frequencies of categorical variables *var1*, *var2*, *var3*, and *var4*, and estimates of proportions for each level of those variables (using the nonresponse weight PNR2W0\_GRP\_F) along with the standard errors:

**proc freq** data=*analysis\_data;*

tables var1 var2 var3 var4 / list missing;

weight PNR2W0\_GRP\_F;

**run;**

The following code creates the weighted mean of continuous variable *var5* (using the weight PNR2W0\_GRP\_F).

**proc means** data=*analysis\_data*

mean;

var *var5*;

weight PNR2W0\_GRP\_F;

**run;**

1. A random sample of enrollees completed the CIDI questionnaire. A total of 1,830 participants completed the CIDI out of 2,944 total participants. The variable “TOOK\_CIDI” indicates whether the participant was selected and completed the CIDI modules. [↑](#footnote-ref-1)
2. Kjellsson, G., Clarke, P., Gerdtham, U. “Forgetting to remember or remembering to forget: A study of the recall period length in health care survey questions,” Journal of Health Economics, May 2014, 35: 34-46. [↑](#footnote-ref-2)