



## Data Collection Worksheet

**Please Note:** The Data Collection Worksheet (DCW) is a tool to aid integration of a PhenX protocol into a study. The PhenX DCW is not designed to be a data collection instrument. Investigators will need to decide the best way to collect data for the PhenX protocol in their study. Variables captured in the DCW, along with variable names and unique PhenX variable identifiers, are included in the PhenX Data Dictionary (DD) files.

### *Accessing and Understanding the American Community Survey (ACS) Data*

The ACS data used in this protocol can be accessed by using Excel to read the Summary Files at the U.S. Census Bureau's [data.census.gov](https://data.census.gov) website (<https://data.census.gov>) or using SAS programs to read the files. Users can find additional information on these tools at the following locations:

**Using Excel to Access Summary Files:** [https://www2.census.gov/programs-surveys/acs/summary\\_file/2020/documentation/tech\\_docs/ACS\\_SF\\_Excel\\_Import\\_Tool.pdf](https://www2.census.gov/programs-surveys/acs/summary_file/2020/documentation/tech_docs/ACS_SF_Excel_Import_Tool.pdf)

**Using SAS to Access Summary Files:** <https://www.census.gov/programs-surveys/acs/library/handbooks/summary-file.html>

The technical documentation for the American Community Survey (ACS) summary files is available online at <http://www.census.gov/programs-surveys/acs/technical-documentation.html>. Select the "Summary File Documentation" link, and then select the data set of interest. Users not familiar with Census data should consult the technical materials.

If the user is interested in additional variables beyond those included in the neighborhood concentrated disadvantage protocol, they should be aware that not all ACS estimates are available for all geographies. These missing estimates are due to data suppression techniques by which the U.S. Census Bureau limits disclosure of individual data and does not release estimates with poor statistical reliability. Additional information about data suppression and the specific estimates it impacts can be found at <http://www.census.gov/programs-surveys/acs/technical-documentation/data-suppression.html>.

Although block group data have long been available from the Census File Transfer Protocol site, not all tables have block groups available for download at [data.census.gov](https://data.census.gov). Information about the types of geographies that are available are in the Appendix Tables as detailed in the technical documentation at <https://www.census.gov/programs-surveys/acs/library/handbooks/summary->

[file.html](#).

### *Calculating Neighborhood Concentrated Disadvantage*

Concentrated disadvantage is derived from six Census variables:

1. Percent of Individuals Below the Poverty Line (derived from ACS Table C17002)
2. Percent of Households Receiving Public Assistance (derived from ACS Table B19057)
3. Percent Female-Headed Families (derived from ACS Table B11001)
4. Percent Unemployed (derived from ACS Table B23025)
5. Percent Less Than Age 18 (derived from ACS Table B01001)
6. Percent Black (derived from ACS Table B02001)

Concentrated disadvantage is calculated for all subareas within a study area.

While some commercial data products may include the derivation of some of these variables, the detailed material below is based on the assumption that the user will go to the U.S. Census Bureau (original source) for all the raw data counts needed to calculate the individual variables that create the measure Concentrated Disadvantage. The protocol text uses the unique ID of individual variables. These descriptions can be found in the “Table Shells” download on the Summary File Technical Documentation (available here <https://www.census.gov/programs-surveys/acs/technical-documentation/table-shells.html>). *Note: users may download tables as Excel files from <https://data.census.gov>. The tables do not use the unique ID of the variables presented in the summary files but do contain header data that describe the variable.*

**1: "Percent of Individuals Below the Poverty Line" is derived from data in ACS 5-Year “Table C17002: Ratio of Income to Poverty Level in the Past 12 Months.”**

Table C17002: Ratio of Income to Poverty Level in the Past 12 Months

Universe: Population for whom poverty status is determined.

There are eight variables included in table C17002 (see line 14188 of the ACS2020\_Table\_Shells.xlsx file available in the Technical Documentation).

Table C17002: Ratio of Income to Poverty Level in the Past 12 Months is reproduced below:

Variable Code	Variable Name
C17002001	Total:
C17002002	Under .50
C17002003	.50 to .99
C17002004	1.00 to 1.24
C17002005	1.25 to 1.49
C17002006	1.50 to 1.84
C17002007	1.85 to 1.99
C17002008	2.00 and over

The percent of individuals below the poverty line= $[(C17002002 + C17002003) / C17002001] \times 100$ .

**2: "Percent of Households Receiving Public Assistance" is derived from ACS "TableB19057: Public Assistance Income in the Past 12 Months for Households."**

Table B19057: Public Assistance Income in the Past 12 Months for Households

Universe: Households.

There are three variables included in Table B19015. Table B19015 is reproduced below:

Variable Code	Variable Name
B19057001	Total:

B19057002	With public assistance income
B19057003	No public assistance income

The “percent of households on public assistance”=(B19057002/B19057001) \* 100.

From the ACS Summary File Subject Definitions, public assistance income “includes general assistance and Temporary Assistance to Needy Families (TANF). Separate payments received for hospital or other medical care (vendor payments) are excluded. This does not include Supplemental Security Income (SSI) or noncash benefits such as Food Stamps” (p. 87 of 2020 Subject Definitions document).

**3: "Percent Female-Headed Families" is derived from ACS “Table B11001: Household Type (Including Living Alone).”**

There are nine cells in Table B1101. The table is reproduced below:

Variable Code	Variable Name
B11001001	Total:
B11001002	Family households:
B11001003	Married-couple family
B11001004	Other family:
B11001005	Male householder, no wife present
B11001006	Female householder, no husband present
B11001007	Nonfamily households:
B11001008	Householder living alone

B11001009	Householder not living alone
-----------	------------------------------

The “percent of female-headed families”=(B11001006/B11001001) \* 100.

**4: "Percent Unemployed" is derived from ACS "Table B23025: Employment Status for the Population 16 Years and Over."**

Table B23025: Employment Status for the Population 16 Years and Over

Universe: Population 16 years and over.

From the 2020 Subject Definitions document (p. 68), the U.S. Census Bureau definition of being unemployed is the following:

"All civilians 16 years old and over are classified as unemployed if they (1) were neither rsquotat worksquot nor rsquotwith a job but not at worksquot during the reference week,and (2) were actively looking for work during the last 4 weeks,and (3) were available to start a job. Also included as unemployed are civilians who did not work at all during the reference week, were waiting to be called back to a job from which they had been laid off, and were available for work except temporary illness. Examples of job-seeking activities are: registering at a public or private employment office; meeting with prospective employers; investigating possibilities for starting a professional practice or opening a business; placing or answering advertisements; writing letters of application; being on a union or professional register"

Table B23025 contains seven cells. The table is reproduced below.

Variable Code	Variable Name
B23025001	Total:
B23025002	In labor force:
B23025003	Civilian labor force:
B23025004	Employed
B23025005	Unemployed

B23025006	Armed Forces
B23025007	Not in labor force

The "percent unemployed"=( $[B23025005 + B23025003] P43001$ ) × 100.

**5: "Percent Less Than Age 18" is derived from ACS "Table B01001: Sex by Age."**

Table B01001: Sex by Age

Universe: Total Population.

There are 49 cells in ACS Table B01001 (ACS2020\_Table\_Shells.xlsx).

Users need to combine the counts for both males and females. Thus, the sum of males under age 18 years old (from under 5 years old to 15-17 years old) equals the sum of all cells B01001002 through B01001006 and for females the sum of all cells B01001027 through B01001030.

The "percent less than age 18"=

$$(((B01001002: B01001006) + (B01001027: B01001030))/B01001001) * 100$$

**6: "Percent Black" is derived from ACS "Table B02001: Race."**

Table B02001: Race

Universe: Total population.

There are 10 cells in Table B02001 (reproduced below):

Variable Code	Variable Name
B02001001	Total:
B02001002	White alone
B02001003	Black or African American alone

B02001004	American Indian and Alaska Native alone
B02001005	Asian alone
B02001006	Native Hawaiian and Other Pacific Islander alone
B02001007	Some other race alone
B02001008	Two or more races:
B02001009	Two races including Some other race
B02001010	Two races excluding Some other race, and three or more races

The "Percent Black"=(B02001003/B02001001) \* 100

Protocol source: <https://www.phenxtoolkit.org/protocols/view/211302>