Data and Biospecimen Use and Publications Committee

GEOGRAPHIC AND DERIVED CENSUS VARIABLE CODEBOOK

Updated August 2020

TABLE OF CONTENTS

SCCS Geocoding Variables

Address history for SCCS participants was geocoded by a multi-stage process incorporating both batch and interactive processes.[1] The Census 2000 areal unit for the geocoded address was determined by a spatial join to TIGER/Line® Shapefiles[2] using ESRI ArcMap 10.0 software (ESRI, Redlands, CA).

Address and Geographic Location Sources:

Address information was collected for each participant at the baseline interview. Subsequent addresses were added to the SCCS database through information provided by some participants at the time of a follow-up interview, through commercial tracing lookups, through National Change of Address linkages, and through participant contacts with SCCS staff to provide updated information.

Addresses included in the SCCS database do NOT constitute a complete address history but rather are address snapshots gathered periodically to support the mailing of things such as newsletters and follow-up questionnaires to participants.

For SCCS participants enrolled in CMS, annual residential location can also be estimated from administrative enrollment records. Specifically, Medicare and Medicaid enrollment files include annual information on the participant's ZIP code. Medicare data is available for SCCS participants nationally, even if they moved out of a SCCS state. Medicaid data is only available for SCCS participants living in the 12 SCCS states (including participants who moved between SCCS states). Depending on an SCCS participant's enrollment in these programs, retrospective geographic information is available as early as 1999 and as late as 2015.

The Medicare and Medicaid programs have minor differences in how geographic location is reported. Medicare data includes information based on participants' mailing address, which may differ from residential address in some circumstances, such as when participants have a post office box. Medicaid data includes information based on participants' residential address, which is used to determine whether a participant is eligible for Medicaid in a given location.

Medicare and Medicaid ZIP code information cannot be released but linkages involving these fields as well as the production of derived variables using these fields can be performed by SCCS staff for approved data or biospecimens requests.

Geocoding:

All addresses at enrollment that can be geocoded have been geocoded to a specific latitude and longitude. In addition, all geocodable addresses at follow-up where a completed questionnaire was either sent or confirmed via phone have been geocoded.

Use of geographic location data:

Specific addresses or geocoded address locations cannot be released. If approved as a part of an application to the SCCS Data and Biospecimens Use (DBU) committee, geocoded addresses scrambled within a 3km grid may be provided.

Any proposed linkages to exact address components (other than state) or non-scrambled geocode coordinates must be done by SCCS staff. Non-identifiable variables attached to investigator data used by SCCS staff during address linking can be attached to SCCS covariate datasets provided by SCCS staff as defined in approved SCCS DBU requests.

Geocoded Addresses available for linkage by SCCS staff (address at):

Baseline	82,458	Linked to Census 2000 and 2010 block group boundaries
Follow-up 1	47,346	Linked to Census 2000 and 2010 block group boundaries
Follow-up 2	32,504	Linked to Census 2000 and 2010 block group boundaries
Follow-up 3	28,455	Linked to Census 2010 block group boundaries

Variable Name	Description and Coding	Comments
EnrollmentAddress	1 = Indicates the paticipant's address at enrollment.	If the the participant gave both a physical address and a post office (PO) Box address at enrollment, priority was given to the physical address.
Survey1Address	1 = Indicates the address to which the first follow-up survey incentive was mailed. 7777 = Participant did not complete first follow-up survey	
Survey2Address	1 = Indicates the address to which the second follow-up survey incentive was mailed. 7777 = Participant did not complete second follow-up survey	
Survey3Address	1 = Indicates the address to which the third follow-up survey incentive was mailed. 7777 = Participant did not complete third follow-up survey	
AddressYear	Year address was reported to SCCS. For addresses at enrollment, this will be the year of enrollment. Numeric value from 2002 to 2013	

Variable Name	Description and Coding			Comments
STATE	State Federal Information Processing Standard (FIPS) code 01 = Alabama 02 = Alaska 04 = Arizona 05 = Arkansas 06 = California 08 = Colorado 09 = Connecticut 10 = Delaware 11 = District of Columbia 12 = Florida 13 = Georgia 15 = Hawaii 16 = Idaho 17 = Illinois 18 = Indiana 19 = Iowa 20 = Kansas 21 = Kentucky	22 = Louisiana 23 = Maine 24 = Maryland 25 = Massachusetts 26 = Michigan 27 = Minnesota 28 = Mississippi 29 = Missouri 30 = Montana 31 = Nebraska 32 = Nevada 33 = New Hampshire 34 = New Jersey 35 = New Mexico 36 = New York 37 = North Carolina 38 = North Dakota 39 = Ohio	40 = Oklahoma 41 = Oregon 42 = Pennsylvania 44 = Rhode Island 45 = South Carolina 46 = South Dakota 47 = Tennessee 48 = Texas 49 = Utah 50 = Vermont 51 = Virginia 53 = Washington 54 = West Virginia 55 = Wisconsin 56 = Wyoming 72 = Puerto Rico	State is the only geographic identifier that will be released on SCCS datasets
GEOCODE	Indicates quality of geocoded coordinate, in descending order 1 = Address match (best) S = Street centroid 4 = ZIP + 4-digit centroid 2 = ZIP + 2-digit centroid X = 5-digit delivery weighted ZIP Code Centroid (least specific)			Delivery weighted ZIP code centroids were provided by Tele Atlas and weighted toward the highest concentration of addresses within the ZIP code[3].
POBox	1 = Participant address is a PO box			
RR	1 = Participant address is a rural delivery route			
NI_CountyID	Non-informative unique county identifier			
NI_TractID	Non-informative unique census tract identifier			
NI_BlkGrpID	Non-informative unique block group identifier			
NI_ZIPCodeID	Non-informative unique ZCTA identifier			

Linkage to US Census data

Participant addresses are linked to Census 2000 and Census 2010 block-group **boundaries**. Any Census data or the new, rolling American Community Survey (ACS) data is freely available from Census.gov. Investigators may download the Census and/or ACS data and derive any analytic variables they wish (preferably using SAS). Requests approved by the SCCS DBU Committee can then have SCCS staff use investigator code to link to participant addresses at whichever geographic level they prefer (either State, county, tract, or block-group) and provide them the linked data back, *but without the actual Census boundary identifiers*.

Census 2000 Derived Variables

Census 2000 variables have been derived at the State, County, Census Tract, Block Group, and ZCTA level. Be sure to specify the areal level you are requesting. Note that all derived census variables will be missing if: a) TotalPersons for the area is less than 300 or b) the address could not be geocoded. These variables have been derived for previous analyses and do not represent the best or only way to calculate these variables. For additional information or to request additional census variables, email datause@southerncommunitystudy.org

Variable Name	Description	Universe	Calculation (SF3 Variables)[4]
Population			
TotalPersons	Total number of persons within the geographic unit	Total population	P001001
Urban/Rural			
PerUrban	Percentage of persons living within urbanized area or clusters	Total population	(P005002/P005001) * 100
PerRural	Percentage of persons living within rural areas (farm and non-farm)	Total population	(P005005/P005001) * 100
Race			
PerWhite	Percentage of persons that are non-hispanic white	Total population	(P007003/P001001) * 100
PerBlack	Percentage of persons that are non-hispanic black	Total population	(P007004/P001001) * 100
Native/Foreign Born			
PerNativeBorn	Percentage of persons that are native born	Total population	(P021002/P021001) * 100
PerForeignBorn	Percentage of persons that are native born	Total population	(P021013/P021001) * 100
Household Type			
PerPersonsInHHolds	Percentage of persons that reside within households	Total population	(P009002/P001001) * 100
PerPersonsInGQuart	Percentage of persons that reside within group quarters	Total Population	(P009025/P001001) * 100
Housing Unit Value			
PerHUOwnerOccupied	Percentage of occupied housing units that are owner occupied	Occupied housing units	(H007002/H007001) * 100

Variable Name	Description	Universe	Calculation (SF3 Variables)[4]
MedianValue_OOHU	Median value (dollars) for all owner-occupied housing units	Owner-occupied housing units	H085001
Poverty/Income			
PerInPoverty	Percentage of persons with income below the 1999 povery status	Persons for whom poverty staus was determined	(P087002/P087001) * 100
Household Income			
MedHHIncome1999	Median household income in 1999	Households	(P053001)
PerHHLessThan25K	Percentage of households with income <$25,000 per year	Households	(sum(P052002-P052005)/P052001) * 100
PerHH_DivIntRentInc	Percentage of households with dividend, interest, or rental income	Households	(P061002/P061001) * 100
Education			
PerNoHS	Percentage of persons that did not graduate HS (age ≥25)	Persons age ≥25 years	(sum(P037003-P037010,P037020-P037027)/(P037002 + P037019))* 100
PerHSGrad	Percentage of persons that graduated HS or equivalent (age ≥25)	Persons age ≥25 years	(sum(P037011-P037018,P037028-P037035)/(P037002 + P037019)) * 100
PerColGrad	Percentage of persons that are college graduates (age ≥25)	Persons age ≥25 years	(sum(P037015-P037018,P037032-P037035)/(P037002 + P037019)) * 100
Employment			
PerEmployed	Percentage of persons in the labor force (age ≥16)	Persons age ≥16 years	((P043003 + P043010)/P043001) * 100
PerOccProfessional	Percent In management, professional, and related occupations	Employed civilian population age ≥16 years	((P050003 + P050050)/P050001) * 100

Variable Name	Description	Universe	Calculation (SF3 Variables)[4]
Crowding			
PopDensity	Persons per square kilometer of land area within the geographic unit	Total Population	(P001001/AREALAND) * 1,000,000
MeanPersonPerRoom	Mean number of occupants per room within occupied housing units	Persons in occupied housing units	(H015001/H025001) * 100

Deprivation Index Development[5,6]

Deprivation index variables are based on the formulas created by Messer et al.[5] using Census 2000 variables. These variables have been derived at the census tract level, but are additionally available for request at the state, county, block group, or ZCTA level. Be sure to specify the areal level you are requesting.

Variable Name	Description	Universe	Calculation (SF3 Variables)[4]
Education			
PerNoHighSchoolDI	Percentage of persons that did not graduate HS (age ≥25)	Persons age ≥25 years	(sum(of P037003-P037010, P037020-P037027)/P037001) * 100
Employment			
PerUnemployedDI	Percentage of males and females who are unemployed	Civilian population age ≥16 years in the labor force	((P043007 + P043014)/(P043005 + P043012))* 100
PerMalesNILFDI	Percentage of males no longer in work force	Males age ≥16 years	(P043008/P043002) * 100
Housing			
PerOHURentedDI	Percentage of housing units renter occupied	Occupied housing units	(H007003/H007001) *100
PerHUVacantDI	Percentage of housing units vacant	Housing units	(H006003/H006001) * 100
PerOHUwGT1perRoomDI	Percentage of housing units with ≥1 occupant per room	Occupied housing units	(sum(of H020005-H020007) + sum(of H020011-H020013))/H020001 * 100
PerHousingCost50PGreaterDI	Percentage of occupied housing units with renter/owner costs >50% of income	Specified owner-occupied housing units	(H069010+H094011+H094022)/((H069001-H069011)+(H094001-(H094011+H094023))) * 100
MedianValue_OOHUDI	Median HH Value	Owner-occupied housing units	H085001
Occupation			
PerMalesManagOccpDI	Percent males in management	Employed civilian population age ≥16 years	(P050004/P050002) * 100

Variable Name	Description	Universe	Calculation (SF3 Variables)[4]
PerFemalesManagOccpDI	Percent females in management	Employed civilian population age ≥16 years	(P050051/P050049) * 100
PerMalesProfessOccpDI	Percent males in professional occupations	Employed civilian population age ≥16 years	(P050010/P050002) * 100
PerFemalesProfessOccpDI	Percent females in professional occupations	Employed civilian population age ≥16 years	(P050057/P050049) * 100
Poverty			
PerInPovertyDI	Percentage of persons with income below the 1999 povery status	Persons for whom poverty staus was determined	(P087002/P087001) * 100
PerHHwFHHwC18DI	Percent female headed households with dependent children	Households	(P010015/P010001) * 100
PerHHIncomeLT30kDI	Percentage of households with income <$30,000 per year	Households	(sum(of P052002-P052006)/P052001) * 100
PerHHPublicAssistDI	Percentage of households with public assistance income	Households	(P064002/P064001) * 100
PerOHUNoVehicleDI	Percentage of households with no car	Occupied housing units	((H044003 + H044010)/H044001) * 100
Race			
PerNonHispanicBlackDI	Percentage non-Hispanic Black	Total population	(P007004/P001001) * 100
Residential Stability			
PerSameResidence1995DI	Percentage in same residence since 1995	Population age ≥5 years	(P024002/P024001) * 100

Variable Name	Description	Universe	Calculation (SF3 Variables)[4]
PerAge65PlusDI	Percentage age ≥65 years	Total population	(sum(of P008035-P008040)+sum(of P008074-P008079))/P008001 * 100
Final Index			
DeprivationIndex	SCCS-derived deprivation index	SCCS Census Tracts	This variable was constructed through principal conponents analysis based on 11 census tract-level variables: PerNoHighSchoolDI, PerUnemployedDI, PerMalesProfessOccpDI, PerOHUwGT1perRoomDI, PerHousingCost50PGreaterDI, PerInPovertyDI, PerHHwFHHwC18DI, PerHHIncomeLT30kDI, PerHHPublicAssistDI, PerOHUNoVehicleDI, MedianValue_OOHUDI. For additional information, see Messer et al.[5] and Signorello et al.[6]
DeprivationIndexQuart	Deprivation Index, Quartiles 1 = Q1 2 = Q2 3 = Q3 4 = Q4	SCCS Census Tracts	Quartile cutpoints are based on the distribution of DeprivationIndex within census tracts in the 12 SCCS states
DeprivationIndexQuint	Deprivation Index, Quintiles 1 = Q1 2 = Q2 3 = Q3 4 = Q4 5 = Q5	SCCS Census Tracts	Quintile cutpoints are based on the distribution of DeprivationIndex within census tracts in the 12 SCCS states

USDA Rural-Urban Continuum Codes

Available at the county level.

Variable Name	Description and Coding	Comments
USDA_URCC_1993	1993 Rural-urban Continuum Code Metro Counties: 0 = Central counties of metro areas of ≥1 million population 1 = Fringe counties of metro areas of ≥1 million population 2 = Counties in metro areas of 250,000 to 1 million population 3 = Counties in metro areas of <250,000 population Non-Metro Counties: 4 = Urban population of ≥20,000, adjacent to a metro area 5 = Urban population of ≥20,000, not adjacent to a metro area 6 = Urban population of 2,500 to 19,999, adjacent to a metro area 7 = Urban population of 2,500 to 19,999, not adjacent to a metro area 8 = Completely rural or <2,500 urban population, adjacent to a metro area 9 = Completely rural or <2,500 urban population, not adjacent to a metro area	See http://ers.usda.gov/ for additional information.
USDA_URCC_2003	2003 Rural-Urban Continuum Code 1 = County in metro area with ≥1 million population 2 = County in metro area of 250,000 to 1 million population 3 = County in metro area of <250,000 population 4 = Nonmetro county with urban population of ≥20,000, adjacent to a metro area 5 = Nonmetro county with urban population of ≥20,000, not adjacent to a metro area 6 = Nonmetro county with urban population of 2,500-19,999, adjacent to a metro area 7 = Nonmetro county with urban population of 2,500-19,999, not adjacent to a metro area 8 = Nonmetro county completely rural or <2,500 urban population, adj. to metro area 9 = Nonmetro county completely rural or <2,500 urban population, not adjacent to metro area	See http://ers.usda.gov/ for additional information.
USDA_PWNMCommuting	Percent of workers in nonmetro counties commuting to central counties of adjacent metro areas (Census 2000)	See http://ers.usda.gov/ for additional information.

References

1. Sonderman JS, Mumma MT, Cohen SS, Cope EL, Blot WJ, Signorello LB. A multi-stage approach to maximizing geocoding success in a large population-based cohort study through automated and interactive processes. Geospat Health 6:273-284, 2012.
2. US Census Bureau, 2008. 2008 TIGER/Line® Shapefiles (machine-readable data files).
3. Tele Atlas, 2006. USA_Geo_002 (Service Description Document).
4. US Census Bureau, 2000. Census 2000 Summary File 3 (SF3) Sample Data (machine-readable data files).
5. Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, Culhane J, et al. The development of a standardized neighborhood deprivation index. J Urban Health 2006;83(6):1041-62.
6. Signorello LB, Cohen SS, Williams DR, Hargreaves MK, Blot WJ. Socioeconomic Status, Race, and Mortality: A Prospective Cohort Study. Am J Public Health 2014 Oct 16:e1-e10. [Epub ahead of print]