

# Geographic Variables in the NDB

---

CB EAGYE

# Two Types of Geo Data in NDB

---

- Geo-identification (geo-id) variables aligned with an address (GeocodeResults)
  - Retrieved from U.S. Census website
  - CensusTract, CensusBlock
  - These variables are associated with an address in Form 2
    - An individual subject may have multiple addresses (collected at different follow-up time points) and therefore multiple census tracts/ census blocks
  - At each quarterly submission, new geo-id variables are added to correspond to newly-collected addresses
    - As of Q3 2021, there are 20,823 census tract records in the NDB (out of 72,455 total Form 2 records = 29%)
- Socioeconomic Status (SES) variables aligned with census tracts (GeoSESVariables)
  - Retrieved from IPUMS website that houses American Community Survey data
  - SESIndex2019, PercentUnemployed, PercentSingleHoH, PercentHoHSorGED, PercentBSorUp, PercentBelowPoverty, PercentSNAP, MedHHIncome, MedFamIncome
  - These are census tract-level variables
    - An individual subject with multiple census tracts will have multiple sets of associated SES variables
    - As of Q3 2021, there are 17,557 SES records

# GeocodeResults Data Set

GeocodeResults\_20211006.sav [DataSet9] - IBM SPSS Statistics Data Editor

	Mod2Id	CensusTract	CensusBlock	StateString	CountyString	TractString
1	10	34003009100	2015	34	003	009100
2	1000	48113009605	2002	48	113	009605
3	10000	48439113513	1024	48	439	113513
4	100000	08014030700	1005	08	014	030700
5	100003	04013082023	2015	04	013	082023
6	100004	42101017400	1007	42	101	017400
7	100006	29510112400	3007	29	510	112400

# GeoSESVariables Data Set (1 of 4)

GeoSESVariables\_20211201.sav [DataSet2] - IBM SPSS Statistics Data Editor

	Mod2Id	IntStatus	FollowUpPeriod	Followup	GeoYear	CensusTract	SESIndex2019
1	6	1	15	24-Jul-2017 00:00:00	2013-2017	48439102100	-0.72
2	10	1	1	20-Sep-2017 00:00:00	2013-2017	34003009100	-1.84
3	11	1	2	28-Jan-2016 00:00:00	2012-2016	12011090102	-0.50
4	12	1	1	23-Aug-2016 00:00:00	2012-2016	53033031703	-0.79
5	14	1	5	22-Jun-2012 00:00:00	2008-2012	39049007393	-1.43
6	16	1	2	16-Mar-2015 00:00:00	2011-2015	12086002703	-0.97
7	23	1	10	29-Jun-2017 00:00:00	2013-2017	39153532302	-1.27
8	29	1	1	10-Oct-2011 00:00:00	2007-2011	42017100500	-0.31
9	35	1	1	5-Mar-2018 00:00:00	2014-2018	34027041802	-1.07
10	36	1	1	7-Aug-2015 00:00:00	2011-2015	39027964300	0.55
11	53	1	1	16-Sep-2010 00:00:00	2006-2010	17197881005	-0.90
12	57	1	2	8-Aug-2017 00:00:00	2013-2017	39049006393	-1.30

# GeoSESVariables Data Set (2 of 4)

The most recent SES variables come from the 2015-2019 date range.

It is possible to have a Followup year that is more recent than the GeoYear.

Need to include indicator variable (interim = lags address collection date; final = matches address collection date)

Variable Name	geoYear	CensusTract	geoStatus
<b>Definition</b>	Aggregate Dates of SES Variables from IPUMS	11-digit census tract, matches GeocodeResults	Indicator of whether the date of the SES variables matches the date the subject's address was collected (interim/final)
<b>Examples</b>	2015-2019	01001020100	*Interim (Lags address collection date)
	2012-2016	01003010300	**Final (Matches address collection date)
			*Meaning the ending date in geoYear is 2019, but the date the address was collected (interview date) is greater than that (e.g., 2020)
			**Meaning the ending date in geoYear is 2016, which matches the date the address was collected (interview year = 2016)








# GeoSESVariables Data Set (3 of 4)

GeoSESVariables\_20211201.sav [DataSet2] - IBM SPSS Statistics Data Editor

	PercentUn employed	PercentSingle HoH	PercentNoHS orGED	PercentBS orUp	PercentBelow Poverty	Percent SNAP	MedHH Income	MedFam Income
1	0.01	0.20	0.07	0.53	0.20	0.10	61912	115956
2	0.05	0.04	0.02	0.69	0.03	0.01	149271	160331
3	0.04	0.31	0.01	0.46	0.17	0.06	54833	77202
4	0.06	0.13	0.08	0.33	0.09	0.09	87384	94668
5	0.06	0.09	0.01	0.61	0.03	0.01	99583	116759
6	0.05	0.29	0.04	0.56	0.07	0.06	81250	104037
7	0.01	0.10	0.03	0.57	0.03	0.05	94294	129167
8	0.09	0.18	0.13	0.16	0.09	0.06	55096	68271
9	0.05	0.14	0.09	0.47	0.03	0.03	106186	108750
10	0.09	0.50	0.13	0.15	0.28	0.22	44152	48654
11	0.09	0.09	0.04	0.36	0.00	0.00	100643	105815
12	0.04	0.10	0.02	0.61	0.02	0.01	94205	106063

# GeoSESVariables Data Set (4 of 4)

GeoSESVariables\_20211201.sav [DataSet1] - IBM SPSS Statistics Data Editor

	 Percent White	 Percent Black	 Percent Hispanic	 STATEA	 GeoState	 COUNTYA	 GeoCounty
1	0.86	0.06	0.20	48	Texas	439	Tarrant County
2	0.64	0.00	0.08	34	New Jersey	003	Bergen County
3	0.96	0.04	0.18	12	Florida	011	Broward County
4	0.77	0.05	0.12	53	Washington	033	King County
5	0.83	0.13	0.00	39	Ohio	049	Franklin County
6	0.87	0.06	0.48	12	Florida	086	Miami-Dade County
7	0.96	0.01	0.01	39	Ohio	153	Summit County
8	0.95	0.02	0.02	42	Pennsylvania	017	Bucks County
9	0.62	0.00	0.13	34	New Jersey	027	Morris County
10	0.97	0.00	0.02	39	Ohio	027	Clinton County
11	0.98	0.00	0.04	17	Illinois	197	Will County
12	0.82	0.05	0.03	39	Ohio	049	Franklin County

# Limited Data Set Restrictions

---

- Both the GeocodeResults and GeoSESVariables data sets include PHI. We can give them to you as part of a Limited Data Set.
- Per the HIPAA Privacy Rule, a Limited Data Set is information from which “facial” identifiers have been removed. For example, names, street addresses, telephone numbers, social security numbers, etc.
- Certain health information items may remain in a Limited Data Set:
  - Dates (such as birth, death, admission, discharge, injury, follow-up)
  - Age
  - Certain elements of geographic address smaller than a state (town or city, county, zip code)
  - Other geographic identifiers smaller than a state (census tract, census block)



# How to Request Geo Data

---

- Internal Notification (Form 602bf) / Request for National Database (Form 602cf)
  - Revised in February 2021 to include requests of geo data and information on linking to external datasets
- Data Use Agreement (SOP 602a)
  - Revised in February 2021 to include policies surrounding the linking of NDB data to external dataset
  - Linking of the TBIMS NDB data to any other data is prohibited without the explicit description of the data to be linked and the linking procedures to be used in the request for TBIMS NDB data.
  - If allowed by the source of the data linked to the TBIMS NDB cases, the data linked to TBIMS cases may be redistributed by the original requestor along with the TBIMS NDB variables Mod1Id (a unique case ID assigned by the TBINDSC for Form 1 variables) and Mod2Id (a unique case ID assigned by the TBINDSC for Form 2 variables), but not any other TBIMS NDB data. [This is intended to prohibit the redistribution of the TBIMS NDB data, while still allowing others to re-link the non-TBIMS NDB data to TBIMS NDB data based on the Mod1Id and Mod2Id stored and distributed as a part of the non-TBIMS NDB data.]

# Notification/Data Request

---

**Do you intend to link the TBIMS NDB with any other datasets? If so, please describe the data to be linked to the TBIMS National Database or Module Database data, and the linking procedures/techniques to be used:**

**Do you require any linking variables in the limited dataset? If so, please check those required and include justification of why variable is needed for the purposes of your study:**

- ☐ StateCode
- ☐ CountyCode
- ☒ CensusTract
- ☐ CensusBlock
- ☐ Zip Code
  - ☐ ZipInj (zip code at injury)
  - ☐ ZipDis (zip code at discharge)
  - ☐ ZipF (zip code at follow-up)
- ☒ The full date of any date variable

Please indicate which dates: Birth, Death, Injury

# The Data Generation Process

---

- At each follow-up period, ask for consent to collect an address
- Upon each quarterly data submission:
  - Retrieve census tract/block from U.S. Census website for each new address
  - Create dataset called GeocodeResults that contains geo-id variables
  - Update dataset called GeoSESVariables that contains the SES variables to include new census tracts
- Annually:
  - Download latest SES variables from IPUMS website
  - Create latest set of SES variables called GeoSESVariables
- Every 10 years:
  - Update SES Index variable
- Note: We are still working on putting accurate, thorough definitions in the Data Dictionary
  - Until this happens, feel free to contact the NDSC with any questions