**MESA Neighborhood Study (Ancillary Study AS023)**

**U.S. Census Measures**

Please acknowledge the following grant in manuscripts and abstracts: R01 HL071759 (Diez Roux)

Please include an acknowledgement:

We thank Kari Moore for her contributions to creating and compiling the census measures.

# Overview:

Neighborhood socioeconomic status (SES) and racial/ethnic composition were obtained from the U.S. Census 2000, American Community Survey (ACS) 2005-2009, and ACS 2007-2011. Measures of race/ethnicity, foreign born status, educational attainment, poverty, unemployment, home values, and household income are included. For household income, the measures are available for the raw value and inflation-adjusted income to the dollar value in year 2000 for comparison across years.

A summary score of neighborhood disadvantage was created based on principle factor analysis (PFA) with a varimax orthogonal rotation. The goal was to have equivalent scales for all Census2000, ACS0509, and ACS0711 that can be used in longitudinal analyses as well as scales that can be used in cross-sectional or baseline only analyses. Two versions of PFA were run:

1. Using only measures that represent SES (referred to as “PC2”)
   1. Includes 4 weighted factors
2. Using SES measures plus race/ethnicity and foreign born status (referred to as “PC3”)
   1. Includes 5 weighted factors

For each of the PFA, scales are available calculated in two ways:

1. Weighted scales where all of the variables included in the PFA are standardized (z-score) and multiplied by the factor loading weight and summed together.
2. Based scales where for each factor, the variables with factor loadings of 0.6 and greater are standardized (z-score) and summed together.

For the scales above, a higher score indicates more neighborhood disadvantage.

In addition, a composite measure of neighborhood advantage based on previous work by Dr. Diez Roux from the 1990 census was also calculated as the sum of standardized (z-score) measures of household income, household wealth, education, and percent of employed persons 16 and older in executive, managerial, or professional occupation. For this scale, a higher score indicates more neighborhood advantage.

Time-varying measures were created using data from the US Census 2000, American Community Survey 2005-2009, and American Community Survey 2007-2011. The measures were applied to the MESA Exams based on calendar years and linked to the census tract of residence.

Years 2000-2004 = Census 2000 data (Exams 1 and 2, part of Exam 3)

Years 2005-2007 = ACS 2005-2009 (Part of Exam 3, Exam 4)

Years 2008-2012 = ACS 2007-2011 (Exam 5)

To assess long-term cumulative exposure to the neighborhood environment, we created time-varying cumulative means, defined as the mean across all months from the baseline to each follow-up exam.

# Recommendations:

For most analyses, if you want to adjust for a single measure of overall neighborhood SES, it is recommended to use a factor scale that takes into account – see Table 1 for suggestions on when each factor scale is best.

For analyses where a single measure of SES is of main interest, it is recommended to use either median household income or poverty. These measures are intuitive to explain the meaning and are typically highly correlated with other measures of neighborhood SES.

Table 1: Recommended variables

| **Variable name** | **Description** | **Recommendation for when to use this** |
| --- | --- | --- |
| inc\_medHH | Median Household income | A good proxy for neighborhood SES when only one measure is wanted. |
| inc\_infladj | Inflation adjusted median Household income to U.S. dollars in year 2000 | A good proxy for neighborhood SES when only one measure is wanted. |
| pov | % persons below poverty level | A good proxy for neighborhood SES when only one measure is wanted. |
| F1\_PC2 | Weighted Factor1 scale from the SES only PCA (PC2). This is highly weighted on % Bachelor degree, % managerial occupation, median home value, % HS education, % interest/dividend/rental income, median household income, and % HH income >$50,000. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. | THIS IS THE SCALE MOST OFTEN USED. For use if you are interested in the SES effect independent from the race/ethnicity. |
| F1\_PC3 | Weighted Factor1 scale from the Full Variables (SES+race/ethnicity) PCA (PC3). This is highly weighted on % Bachelor degree, % managerial occupation, % HS education, median home value, %interest/dividend/rental income, median HH income, and % HH income >$50,000. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. | For use if you are interested in the total SES effect including race/ethnicity. |
| factor\_ana | Factor based scale from Ana Diez-Roux 1990 census PCA methods. This is based on the standardized to both datasets combined transformed variables summed together. The variables are: median housing value log, % HS education, % Bachelor’s degree, % managerial occupation, median HH income log, and % interest/dividend income. A higher value indicates a better SES. This will be missing if any variables are missing. | This gives a summary SES effect. Race/ethnicity is not taken into account in this scale. |

# Example Methodology Section for Manuscripts:

Weighted Factor Scores with only SES (PC2):

A neighborhood socioeconomic status (SES) index was developed from U.S. Census and American Community Survey (ACS) data at the census tract level (U.S. Census 2000(1) exams prior to 2005, ACS 2005–2009 for exams between 2005-2007(2), and ACS 2008–2011(3) for exams in 2009-2012). Using principal factor analysis with varimax rotation, four factors were derived from 16 tract-level measures of educational attainment, occupation, income, wealth, poverty, employment status, and housing characteristics. Four factors were kept which reflects 73% of the variance. Weighted scales were created by multiplying the factor weights by the standardized variables, and a higher index indicate lower neighborhood SES (i.e. greater neighborhood deprivation)(4).

1 US Census Bureau. Census 2000 Summary File 1 & Summary File 3 – United States. 2001. <https://www.census.gov/main/www/cen2000.html>. Accessed July 10, 2012.

2 US Census Bureau. 2005-2009 American Community Survey—United States. 2011. http://www.census.gov/acs/www/data\_documentation/2009\_release/. Accessed July 10, 2012.

3 US Census Bureau. 2007-2011 American Community Survey – United States. 2013. https://www.census.gov/newsroom/releases/archives/news\_conferences/20121203\_acs5yr.html. Accessed July 10, 2012.

4 Moore, K., et al., Home and work neighbourhood environments in relation to body mass index: the Multi-Ethnic Study of Atherosclerosis (MESA). Journal of epidemiology and community health, 2013. **67**(10): p. 846-53.

Weighted Factor Scores with race/ethnicity (PC3):

A neighborhood socioeconomic status (SES) index was developed from U.S. Census and American Community Survey (ACS) data at the census tract level (U.S. Census 2000(1) exams prior to 2005, ACS 2005–2009 for exams between 2005-2007(2), and ACS 2008–2011(3) for exams in 2009-2012). Using principal factor analysis with varimax rotation, five factors were derived from 21 tract-level measures of race/ethnicity, prevalence of foreign-born residents, crowding, educational attainment, occupation, income, wealth, poverty, employment status, and housing characteristics. Five factors were kept which reflects 74% of the variance. Weighted scales were created by multiplying the factor weights by the standardized variables, and a higher index indicate lower neighborhood SES (i.e. greater neighborhood deprivation)(4).

1 US Census Bureau. Census 2000 Summary File 1 & Summary File 3 – United States. 2001. <https://www.census.gov/main/www/cen2000.html>. Accessed July 10, 2012.

2 US Census Bureau. 2005-2009 American Community Survey—United States. 2011. http://www.census.gov/acs/www/data\_documentation/2009\_release/. Accessed July 10, 2012.

3 US Census Bureau. 2007-2011 American Community Survey – United States. 2013. https://www.census.gov/newsroom/releases/archives/news\_conferences/20121203\_acs5yr.html. Accessed July 10, 2012.

4 Christine, P.J., Auchincloss, A.H., Bertoni, A.G.,et al.,2015. Longitudinal associations between neighborhood physical and social environments and incident type 2 diabetes mellitus: The Multi-Ethnic Study of Atherosclerosis. J. Am. Med. Assoc. Intern. Med. 175(8),1311–1320.

“Factor\_Ana”

Neighborhood disadvantage was assessed using data collected in the 2000 U.S. Census(1) and American Community Survey (ACS) 2005-2009(2), and 2007-2011(3) estimates. Data was linked to MESA participant data by census tract using Census and ACS estimates for the closest time period. A previously developed composite index: median household income, household wealth (composed of median value of housing units and percent of household with interest, dividend, or net rental income), education (high school degree or higher and Bachelor’s degree or higher), and percent of employed persons 16 and older in executive, managerial, or professional occupation was used to characterize neighborhood socioeconomic disadvantage for each census tract(4). Several items were log-transformed to improve the distribution before being standardized and summed to create the neighborhood disadvantage measure. Higher scores indicate more socioeconomic advantage.

1 US Census Bureau. Census 2000 Summary File 1 & Summary File 3 – United States. 2001. <https://www.census.gov/main/www/cen2000.html>. Accessed July 10, 2012.

2 US Census Bureau. 2005-2009 American Community Survey—United States. 2011. http://www.census.gov/acs/www/data\_documentation/2009\_release/. Accessed July 10, 2012.

3 US Census Bureau. 2007-2011 American Community Survey – United States. 2013. https://www.census.gov/newsroom/releases/archives/news\_conferences/20121203\_acs5yr.html. Accessed July 10, 2012.

4 Diez Roux, AV et al., Neighborhood of residence and incidence of coronary heart disease. The New England Journal of Medicine. 2001. 345(2), 99-106.

# Published MESA Manuscripts Using the Data:

Auchincloss, A.H., et al., *Association of insulin resistance with distance to wealthy areas: the multi-ethnic study of atherosclerosis.* American journal of epidemiology, 2007. **165**(4): p. 389-97.

Echeverria, S., et al., *Associations of neighborhood problems and neighborhood social cohesion with mental health and health behaviors: the Multi-Ethnic Study of Atherosclerosis.* Health & place, 2008. **14**(4): p. 853-65.

Franco, M., et al., *Neighborhood characteristics and availability of healthy foods in Baltimore.* American journal of preventive medicine, 2008. **35**(6): p. 561-7.

Lemelin, E.T., et al., *Life-course socioeconomic positions and subclinical atherosclerosis in the multi-ethnic study of atherosclerosis.* Social science & medicine, 2009. **68**(3): p. 444-51.

Osypuk, T.L., et al., *Are immigrant enclaves healthy places to live? The Multi-ethnic Study of Atherosclerosis.* Social science & medicine, 2009. **69**(1): p. 110-20.

Mair, C., et al., *Is neighborhood racial/ethnic composition associated with depressive symptoms? The multi-ethnic study of atherosclerosis.* Social science & medicine, 2010. **71**(3): p. 541-50.

Murray, E.T., et al., *Trajectories of neighborhood poverty and associations with subclinical atherosclerosis and associated risk factors: the multi-ethnic study of atherosclerosis.* American journal of epidemiology, 2010. **171**(10): p. 1099-108.

Nazmi, A., et al., *Cross-sectional and longitudinal associations of neighborhood characteristics with inflammatory markers: findings from the multi-ethnic study of atherosclerosis.* Health & place, 2010. **16**(6): p. 1104-12.

Allen, N.B., et al., *Association of health professional shortage areas and cardiovascular risk factor prevalence, awareness, and control in the Multi-Ethnic Study of Atherosclerosis (MESA).* Circulation. Cardiovascular quality and outcomes, 2011. **4**(5): p. 565-72.

Do, D.P., et al., *Circadian rhythm of cortisol and neighborhood characteristics in a population-based sample: the Multi-Ethnic Study of Atherosclerosis.* Health Place, 2011. **17**(2): p. 625-32.

Desantis, A.S., et al., *Associations of neighborhood characteristics with sleep timing and quality: the multi-ethnic study of atherosclerosis.* Sleep, 2013. **36**(10): p. 1543-51.

Hajat, A., et al., *Air pollution and individual and neighborhood socioeconomic status: evidence from the Multi-Ethnic Study of Atherosclerosis (MESA).* Environ Health Perspect, 2013. **121**(11-12): p. 1325-33.

Moore, K., et al., *Home and work neighbourhood environments in relation to body mass index: the Multi-Ethnic Study of Atherosclerosis (MESA).* Journal of epidemiology and community health, 2013. **67**(10): p. 846-53.

Tomey, K., et al., *Associations between neighborhood characteristics and self-rated health: A cross-sectional investigation in the Multi-Ethnic Study of Atherosclerosis (MESA) cohort.* Health & place, 2013. **24**: p. 267-74.

Jones, M.R., et al., *Race/Ethnicity, Residential Segregation, and Exposure to Ambient Air Pollution: The Multi-Ethnic Study of Atherosclerosis (MESA).* Am J Public Health, 2014: p. e1-e8.

Le-Scherban, F., et al., *Neighborhood ethnic composition, spatial assimilation, and change in body mass index over time among Hispanic and chinese immigrants: multi-ethnic study of atherosclerosis.* Am J Public Health, 2014. **104**(11): p. 2138-46.

Needham, B., et al., *Neighborhood characteristics and leukocyte telomere length: The Multi-Ethnic Study of Atherosclerosis.* Health & place, 2014. **28**: p. 167-172.

Unger, E., et al., *Association of neighborhood characteristics with cardiovascular health in the multi-ethnic study of atherosclerosis.* Circ Cardiovasc Qual Outcomes, 2014. **7**(4): p. 524-31.

Brenner, A.B., et al., *Associations of Alcohol Availability and Neighborhood Socioeconomic Characteristics With Drinking: Cross-Sectional Results From the Multi-Ethnic Study of Atherosclerosis (MESA).* Subst Use Misuse, 2015. **50**(12): p. 1606-17.

Brenner, A.B., et al., *Longitudinal associations of neighborhood socioeconomic characteristics and alcohol availability on drinking: Results from the Multi-Ethnic Study of Atherosclerosis (MESA).* Soc Sci Med, 2015. **145**: p. 17-25.

Christine, P.J., et al., *Longitudinal Associations Between Neighborhood Physical and Social Environments and Incident Type 2 Diabetes Mellitus: The Multi-Ethnic Study of Atherosclerosis (MESA).* JAMA Intern Med, 2015. **175**(8): p. 1311-20.

Hajat, A., et al., *Examining the cross-sectional and longitudinal association between diurnal cortisol and neighbohrood characteristics: Evidence from the multi-ethnic study of atherosclerosis.* Health & place, 2015. **34**: p. 199-206.

Kershaw, K.N., et al., *Associations of chronic individual-level and neighbourhood-level stressors with incident coronary heart disease: the Multi-Ethnic Study of Atherosclerosis.* J Epidemiol Community Health, 2015. **69**(2): p. 136-41.

Kershaw, K.N., et al., *Neighborhood-level racial/ethnic residential segregation and incident cardiovascular disease: the multi-ethnic study of atherosclerosis.* Circulation, 2015. **131**(2): p. 141-8.

Mair, C., et al., *Change in neighborhood environments and depressive symptoms in New York City: The Multi-Ethnic Study of Atherosclerosis.* Health & Place, 2015. **32**(0): p. 93-98.

Braun, L.M., et al., *Walkability and cardiometabolic risk factors: Cross-sectional and longitudinal associations from the Multi-Ethnic Study of Atherosclerosis.* Health Place, 2016. **39**: p. 9-17.

Moore, K.A., et al., *Neighborhood Social Resources and Depressive Symptoms: Longitudinal Results from the Multi-Ethnic Study of Atherosclerosis.* J Urban Health, 2016. **93**(3): p. 572-88.

# Dataset Description:

Participant inclusion: MESA participants who agreed to participate in the MESA Neighborhood study (N=6191)

Data set-up: Panel (stacked) dataset with 1 row per participant per exam; an exam indicator is included

Notes: Data will be missing for any addresses that were unable to geocode or where census data is unavailable due to data suppression by the U.S. Census Bureau (typically due to low sample counts).

**Data Set Name: MESANBH\_CENSUS07272016**

| **Variable Order** | **Variable name** | **Description** | **Coding** |
| --- | --- | --- | --- |
| 1 | idno | MESA ID number |  |
| 2 | accuracy | Geocoding accuracy indicator. It is recommended that that only those with accuracy to at least zip code + 4 centroid are used in analyses, at least for sensitivity analysis. | 1 = Street level  2 = Zip+4 centroid level  3 = Zip+2 centroid level  4 = Zip code centroid  5 = Unable to geocode |
| 3 | cenid | Fake census tract id for clustering analysis. This should be used when using census tract level data in models. This is from 2000 census since census 2000 and ACS use this geography. |  |
| 4 | EXAM | MESA Exam number |  |
| 5 | race\_hisp | % Hispanic in census tract |  |
| 6 | race\_whiteNH | % non-Hispanic White in census tract |  |
| 7 | race\_blackNH | % non-Hispanic Black in census tract |  |
| 8 | race\_asianNH | % non-Hispanic Asian in census tract |  |
| 9 | birth\_foreign | % foreign born in census tract |  |
| 10 | Educ\_minHS | % Age 25+ with minimum High School education in census tract |  |
| 11 | Educ\_minBA | % Age 25+ with minimum Bachelor degree education in census tract |  |
| 12 | unemployed | % Unemployed among civilians age 16+ in the labor force in census tract |  |
| 13 | Occup\_I | Percent with managerial, professional, or related occupation among those 16+ in the labor force |  |
| 14 | inc\_HHge50k | Percent of households with household income >$50,000 |  |
| 15 | inc\_medHH | Median Household income |  |
| 16 | inc\_IntDivRent | Percent of households with interest, dividend, or net rental income |  |
| 17 | pov | % persons below poverty level |  |
| 18 | HUcost\_medownval | Median value of households (owner units) |  |
| 19 | F1\_PC2 | Weighted Factor1 scale from the SES only PCA (PC2). This is highly weighted on % Bachelor degree, % managerial occupation, median home value, % HS education, % interest/dividend/rental income, median household income, and % HH income >$50,000. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 20 | F2\_PC2 | Weighted Factor2 scale from the SES only PCA (PC2). This is highly weighted on % no vehicle, % owner occupied housing, % poverty, and % unemployed. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 21 | F3\_PC2 | Weighted Factor3 scale from the SES only PCA (PC2). This is highly weighted on % not in labor force and % occupied households. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 22 | F4\_PC2 | Weighted Factor4 scale from the SES only PCA (PC2). This is highly weighted on % in same house. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 23 | F1\_PC3 | Weighted Factor1 scale from the Full Variables (SES+race/ethnicity) PCA (PC3). This is highly weighted on % Bachelor degree, % managerial occupation, % HS education, median home value, %interest/dividend/rental income, median HH income, and % HH income >$50,000. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 24 | F2\_PC3 | Weighted Factor2 scale from the Full Variables (SES+race/ethnicity) PCA (PC3). This is highly weighted on % Black, % owner occupied housing, % no vehicle, % interest/dividend income, % poverty, and % unemployed. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 25 | F3\_PC3 | Weighted Factor3 scale from the Full Variables (SES+race/ethnicity) PCA (PC3). This is highly weighted on % foreign born, % Hispanic, and % crowding. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 26 | F4\_PC3 | Weighted Factor4 scale from the Full Variables (SES+race/ethnicity) PCA (PC3). This is highly weighted on % not in labor force and % occupied housing. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 27 | F5\_PC3 | Weighted Factor5 scale from the Full Variables (SES+race/ethnicity) PCA (PC3). This is highly weighted on % in same house. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 28 | F1\_PC2\_BT | Factor1 based scale from the SES only PCA (PC2). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % Bachelor degree reverse coded, % managerial occupation reverse coded, median home value reverse coded, % HS education reverse coded, % interest/dividend/rental income reverse coded, median household income reverse coded, and % HH income >$50,000 reverse coded. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 29 | F2\_PC2\_BT | Factor2 based scale from the SES only PCA (PC2). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % no vehicle, % owner occupied housing reverse coded, % poverty, and % unemployed. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 30 | F3\_PC2\_BT | Factor2 based scale from the SES only PCA (PC2). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % not in labor force and % occupied households reverse coded. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 31 | F1\_PC3\_BT | Factor1 based scale from Full Variables (SES+race/ethnicity) PCA (PC3). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % Bachelor degree reverse coded, % managerial occupation reverse coded, % HS education reverse coded, % interest/dividend/rental income reverse coded, median home value reverse coded, median HH income reverse coded, and % HH income >$50,000 reverse coded. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 32 | F2\_PC3\_BT | Factor2 based scale from Full Variables (SES+race/ethnicity) PCA (PC3). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % Black, % owner occupied housing reverse coded, % no vehicle, % interest/dividend income reverse coded, % poverty, and % unemployed. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 33 | F3\_PC3\_BT | Factor3 based scale from Full Variables (SES+race/ethnicity) PCA (PC3). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % foreign born, % Hispanic, and % crowding. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 34 | F4\_PC3\_BT | Factor4 based scale from Full Variables (SES+race/ethnicity) PCA (PC3). This is based on the standardized to both datasets combined raw variables summed together. The variables are: % not in labor force and % occupied housing reverse coded. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 35 | factor\_ana | Factor based scale from Ana Diez-Roux 1990 census PCA methods. This is based on the standardized to both datasets combined transformed variables summed together. The variables are: median housing value log, % HS education, % Bachelor’s degree, % managerial occupation, median HH income log, and % interest/dividend income. A higher value indicates a better SES. This will be missing if any variables are missing. |  |
| 36 | PFBLATIN | % foreign born from Latin America including Caribbean, Central America, and South America in census tract |  |
| 37 | PFBCHINA | % foreign born from China including Hong Kong and Taiwan in census tract |  |
| 38 | inc\_infladj | Inflation adjusted median Household income to U.S. dollars in year 2000 |  |
| 39 | popdenmi\_tot | Population density in census tract including land+water (persons per mile square) |  |
| 40 | popdenmi\_nowat | Population density in census tract with only land (no water) (persons per mile square) |  |
| 41 | A\_race\_hisp | Cumulative Average: % Hispanic in census tract |  |
| 42 | A\_race\_whiteNH | Cumulative Average: % non-Hispanic White in census tract |  |
| 43 | A\_race\_blackNH | Cumulative Average: % non-Hispanic Black in census tract |  |
| 44 | A\_race\_asianNH | Cumulative Average: % non-Hispanic Asian in census tract |  |
| 45 | A\_birth\_foreign | Cumulative Average: % foreign born in census tract |  |
| 46 | A\_Educ\_minHS | Cumulative Average: % Age 25+ with minimum High School education in census tract |  |
| 47 | A\_Educ\_minBA | Cumulative Average: % Age 25+ with minimum Bachelor degree education in census tract |  |
| 48 | A\_unemployed | Cumulative Average: % Unemployed among civilians age 16+ in the labor force in census tract |  |
| 49 | A\_Occup\_I | Cumulative Average: Percent with managerial, professional, or related occupation among those 16+ in the labor force |  |
| 50 | A\_inc\_HHge50k | Cumulative Average: Percent of households with household income >$50,000 |  |
| 51 | A\_inc\_medHH | Cumulative Average: Median Household income |  |
| 52 | A\_inc\_IntDivRent | Cumulative Average: Percent of households with interest, dividend, or net rental income |  |
| 53 | A\_pov | Cumulative Average: % persons below poverty level |  |
| 54 | A\_HUcost\_medownval | Cumulative Average: Median value of households (owner units) |  |
| 55 | A\_F1\_PC2 | Cumulative Average: Weighted Factor1 scale from the SES only PCA (PC2). This is highly weighted on % Bachelor degree, % managerial occupation, median home value, % HS education, % interest/dividend/rental income, median household income, and % HH income >$50,000. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 56 | A\_F2\_PC2 | Cumulative Average: Weighted Factor2 scale from the SES only PCA (PC2). This is highly weighted on % no vehicle, % owner occupied housing, % poverty, and % unemployed. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 57 | A\_F3\_PC2 | Cumulative Average: Weighted Factor3 scale from the SES only PCA (PC2). This is highly weighted on % not in labor force and % occupied households. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 58 | A\_F4\_PC2 | Cumulative Average: Weighted Factor4 scale from the SES only PCA (PC2). This is highly weighted on % in same house. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 59 | A\_F1\_PC3 | Cumulative Average: Weighted Factor1 scale from the Full Variables (SES+race/ethnicity) PCA (PC3). This is highly weighted on % Bachelor degree, % managerial occupation, % HS education, median home value, %interest/dividend/rental income, median HH income, and % HH income >$50,000. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 60 | A\_F2\_PC3 | Cumulative Average: Weighted Factor2 scale from the Full Variables (SES+race/ethnicity) PCA (PC3). This is highly weighted on % Black, % owner occupied housing, % no vehicle, % interest/dividend income, % poverty, and % unemployed. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 61 | A\_F3\_PC3 | Cumulative Average: Weighted Factor3 scale from the Full Variables (SES+race/ethnicity) PCA (PC3). This is highly weighted on % foreign born, % Hispanic, and % crowding. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 62 | A\_F4\_PC3 | Cumulative Average: Weighted Factor4 scale from the Full Variables (SES+race/ethnicity) PCA (PC3). This is highly weighted on % not in labor force and % occupied housing. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 63 | A\_F5\_PC3 | Cumulative Average: Weighted Factor5 scale from the Full Variables (SES+race/ethnicity) PCA (PC3). This is highly weighted on % in same house. This is weighted to both datasets combined. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 64 | A\_F1\_PC2\_BT | Cumulative Average: Factor1 based scale from the SES only PCA (PC2). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % Bachelor degree reverse coded, % managerial occupation reverse coded, median home value reverse coded, % HS education reverse coded, % interest/dividend/rental income reverse coded, median household income reverse coded, and % HH income >$50,000 reverse coded. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 65 | A\_F2\_PC2\_BT | Cumulative Average: Factor2 based scale from the SES only PCA (PC2). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % no vehicle, % owner occupied housing reverse coded, % poverty, and % unemployed. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 66 | A\_F3\_PC2\_BT | Cumulative Average: Factor2 based scale from the SES only PCA (PC2). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % not in labor force and % occupied households reverse coded. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 67 | A\_F1\_PC3\_BT | Cumulative Average: Factor1 based scale from Full Variables (SES+race/ethnicity) PCA (PC3). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % Bachelor degree reverse coded, % managerial occupation reverse coded, % HS education reverse coded, % interest/dividend/rental income reverse coded, median home value reverse coded, median HH income reverse coded, and % HH income >$50,000 reverse coded. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 68 | A\_F2\_PC3\_BT | Cumulative Average: Factor2 based scale from Full Variables (SES+race/ethnicity) PCA (PC3). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % Black, % owner occupied housing reverse coded, % no vehicle, % interest/dividend income reverse coded, % poverty, and % unemployed. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 69 | A\_F3\_PC3\_BT | Cumulative Average: Factor3 based scale from Full Variables (SES+race/ethnicity) PCA (PC3). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % foreign born, % Hispanic, and % crowding. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 70 | A\_F4\_PC3\_BT | Cumulative Average: Factor4 based scale from Full Variables (SES+race/ethnicity) PCA (PC3). This is based on the standardized to both datasets combined transformed variables summed together. The variables are: % not in labor force and % occupied housing reverse coded. A higher value indicates a worse SES. This will be missing if any variables are missing. |  |
| 71 | A\_factor\_ana | Cumulative Average: Factor based scale from Ana Diez-Roux 1990 census PCA methods. This is based on the standardized to both datasets combined transformed variables summed together. The variables are: median housing value log, % HS education, % Bachelor’s degree, % managerial occupation, median HH income log, and % interest/dividend income. A higher value indicates a better SES. This will be missing if any variables are missing. |  |
| 72 | A\_PFBLATIN | Cumulative Average: % foreign born from Latin America including Caribbean, Central America, and South America in census tract |  |
| 73 | A\_PFBCHINA | Cumulative Average: % foreign born from China including Hong Kong and Taiwan in census tract |  |
| 74 | A\_inc\_infladj | Cumulative Average: Inflation adjusted median Household income to U.S. dollars in year 2000 |  |
| 75 | A\_popdenmi\_tot | Cumulative Average: Population density in census tract including land+water (persons per mile square) |  |
| 76 | A\_popdenmi\_nowat | Cumulative Average: Population density in census tract with only land (no water) (persons per mile square) |  |

# Details:

## Inflation adjusted household income

Over the 2001-2010 period, inflation has been about 2% per yr (ranged 0%-4%). The Bureau of Labor Statistics estimates an inflation factor, the annual percentage change in the Consumer Price Index (CPI), which can be used to standardize income over time. The CPI can be used to index (i.e., adjust for the effect of inflation) the real value of wages, salaries, pensions, for regulating prices and for deflating monetary magnitudes to show changes in real values. In the United States, the CPI is one of several price indices calculated by most national statistical agencies and is defined by the Bureau of Labor Statistics as "a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services."[[1]](#footnote-1) Sub-indexes and sub-sub-indexes are computed for different categories and sub-categories of goods and services, being combined to produce the overall index with weights reflecting their shares in the total of the consumer expenditures covered by the index. The Bureau of Labor Statistics website publishes the CPI table[[2]](#footnote-2). This does not take into account regional differences in cost of living.

The inflation adjustment was applied to the ACS2005-2009 and 2007-2011 data for median household income. The inflation factor for the year 2007 was used for ACS2005-2009 this since this represents the mid-point year of the ACS2005-2009. The inflation factor for the year 2009 was used for ACS2007-2011 this since this represents the mid-point year of the ACS2007-2011. Table 1 shows the mean for the unadjusted and adjusted median household income.

Table 1: Mean for median household income in ACS05-09 and ACS07-11 both raw and inflation adjusted

| **ACS year** | **Variable** | **N** | **Mean** | **Std** | **Min** | **Max** |
| --- | --- | --- | --- | --- | --- | --- |
| ACS05-09 | Raw HH income | 65632 | 53380.55 | 26631.19 | 2499 | 250001 |
|  | Inflation adjusted HH income | 65632 | 45374.62 | 22637.08 | 2124.20 | 212506.26 |
| ACS07-11 | Raw HH income | 73030 | 56081.50 | 27929.21 | 2499 | 250001 |
|  | Inflation adjusted HH income | 73030 | 45991.01 | 22904.03 | 2409.37 | 205019.47 |

Code for the inflation rate factor:

ACS2005-2009:

inf\_rate\_factor=(1-0.0159)\*(1-0.0227)\*(1-0.0268)\*(1-0.0339)\*(1-0.0324)\*(1-0.0285);

inc\_infladj=inc\_medHH\*inf\_rate\_factor;

ACS2007-2011:

inf\_rate\_factor=(1-0.0159)\*(1-0.0227)\*(1-0.0268)\*(1-0.0339)\*(1-0.0324)\*(1-0.0285)\*(1-0.0385)\*(1+0.0034);

inc\_infladj=inc\_medHH\*inf\_rate\_factor;

## Crosswalk from 2010 tract boundaries to 2000 boundaries (ACS2007-2011)

Boundaries of the census do not remain consistent across all census years. They can split or merge depending on population changes[[3]](#footnote-3). Since the data downloaded from the 2010 census does not necessarily match the boundaries from the 2000 census, a crosswalk which gives weights to each tract in 2010 was downloaded to give data that is related the comparable area definitions. Census 2000 and ACS2005-2009 both use tract boundaries from 2000. ACS2007-2011 uses boundaries from 2010. The crosswalk was applied to ACS2007-2011 to equate the data to 2000 boundaries. a crosswalk to convert 2010 boundaries to 2000 boundaries was downloaded from the website (http://www.s4.brown.edu/us2010/Researcher/LTDB.htm)[[4]](#footnote-4).

## Principal Factor Analysis – Variable selection

Variables were selected a priori from available census measures to reflect the racial/ethnic composition, housing, crowding, residential stability, education, employment, occupation, and income/wealth of the census tracts. This includes a total of 21 variables. See Table 2 for a list of variables used. Interpretation of the variables is designed to indicate a WORSE SES for HIGHER levels (ie: those close to 0 are better SES, close to maximum (usually 1) are a worse SES). To maintain consistency and reduce the number of negative loadings when running the factor analysis, variables that are interpreted as having a higher value being better SES are reverse coded. Factor analysis is based on multivariate normal data distributions. The variables to be included in the factor analysis were checked for skewness. For those that had skewness>1.5, transformations of log, square root, and cube root were created and checked for “best” transformation. The transformation that gives the least amount of skewness (ie: skewness closest to 0) was used for factor analysis. The cube root transformation was used in all cases. Median household income and median home value were not reverse coded despite a higher value indicating better SES due to normality transformations. When these variables are not reverse coded, a transformation with cube root gives skewness <1.5. When they are reverse coded, then a good transformation isn’t possible. We decided normality is more important than reverse coding. They will have negative loadings in the PCA and this was taken into account when creating factor based scales.

Table 2: List of variables used in principal components analysis to create SES scales. In the last column, PC2 refers to SES only PCA, and PC3 refers to Full variables PCA.

| **Domain** | **Variable** | **Description** | **Reverse coded** | **Transformation used** | **Which PCA number was this included in?** |
| --- | --- | --- | --- | --- | --- |
| RACE/ ETHNICITY | race\_hisp | Percent Hispanic | NO | Cube root | PC3 |
| RACE/ ETHNICITY | race\_blackNH | Percent non-Hispanic black | NO | Cube root | PC3 |
| RACE/ ETHNICITY | race\_asianNH | Percent non-Hispanic Asian | NO | Cube root | PC3 |
| RACE/ ETHNICITY | birth\_foreign | Percent foreign born | NO | Cube root | PC3 |
| CROWDING | crowd\_gt1\_ppr | Percent of occupied houses with more than 1 person per room | NO | Cube root | PC3 |
| HOUSING | HU\_sampleocc | Percent of occupied housing units | YES | Cube root | PC2, PC3 |
| HOUSING | ownerocc\_hh | Percent of housing units that are owner occupied out of total housing units | YES | NONE | PC2, PC3 |
| HOUSING | HUcost\_medownval | Median value of occupied housing units | NO (due to transformation issues) | Cube root | PC2, PC3 |
| HOUSING | phone\_none | Percent of housing units without telephone | NO | Cube root | PC2, PC3 |
| HOUSING | vehicle\_none | Percent of housing units without vehicle | NO | Cube root | PC2, PC3 |
| RESIDENTIAL STABILITY | samehouse | Percent living in same house in 1995 (Census 2000) or 1999 (ACS) | YES | Cube root | PC2, PC3 |
| EDUCATION | Educ\_minHS | Percent of person 25 or older with at least high school education | YES | NONE | PC2, PC3 |
| EDUCATION | Educ\_minBA | Percent of persons 25 or older with at least a Bachelor’s degree | YES | NONE | PC2, PC3 |
| EMPLOYMENT | unemployed | Percent unemployed among civilians 16 and over in the labor force | NO | Cube root | PC2, PC3 |
| EMPLOYMENT | NotInLaborForce | Percent of civilians 16 and over not in the labor force | NO | NONE | PC2, PC3 |
| OCCUPATION | Occup\_I | Percent with management, professional, and related occupation | YES | NONE | PC2, PC3 |
| INCOME/ WEALTH | inc\_medHH | Median household income | NO (due to transformation issues) | Cube root | PC2, PC3 |
| INCOME/ WEALTH | inc\_HHge50k | Percent households with household income>$50,000 | YES | NONE | PC2, PC3 |
| INCOME/ WEALTH | inc\_IntDivRent | Percent of households with interest, dividends, or net rental income | YES | NONE | PC2, PC3 |
| INCOME/ WEALTH | inc\_pubass | Percent of households with public assistance | NO | Cube root | PC2, PC3 |
| INCOME/ WEALTH | pov | Percent of persons below the poverty level | NO | Cube root | PC2, PC3 |

## Principal Factor Analysis – Methods and Output

Principal factor analysis (PFA) with a varimax orthogonal rotation was used to create the final scales with prior communalities set to 1 (default in SAS). This type of PFA was used to allow for uncorrelated scales so they can be used together in regression models. The varimax rotation allows for each component to have a small number of large loadings and the others being primarily small loadings. For these analyses, the Census2000, ACS0509, and ACS0711 data were set together as one dataset (each census tract will have 3 observations) and the PFA was performed on this dataset. This decision was made to allow for the ability to have common weights for both years to allow for longitudinal analyses. For all PFA, the number of factors to keep was determined by having at least 70% of the variance explained. Analyses were performed using SAS PROC FACTOR.

### SES Only Principal Factor Analysis

A PFA was performed using only the SES variables that without the measures for Race/Ethnicity and Crowding to have SES scales independent of the race effects. This PCA will be referred to as PC2. Four factors were retained based on the criteria of having at least 70% of the variance explained. With four factors, 73.2% of the variance is explained (see Table 3).

Factor scales variables are created for the 4 weighted factors using SAS PROC SCORE. This gives 4 variables with each one being more highly weighted on the variables that have higher loadings for that factor. This gives the following patterns:

Factor1 = Education/Occupation/Income

Factor2 = Wealth/Poverty/Unemployment/Owner occupied housing

Factor3 = Occupied housing/Labor force

Factor4 = Residential stability

In addition to the weighted scales, “based” scales were also created. These scales take the variables that have loadings of at least 0.60 (when rounded to 2 decimal places), standardizes (z-score) the variables, and sums together these standardized variables. Standardization was done with Census2000, ACS0509, and ACS0711 set together for use in longitudinal analysis. The variables used in each scale are as follows:

Factor1 = sum of standardized percent with at least Bachelor degree (reverse coded), percent with managerial/professional occupation (reverse coded), median housing value (reverse coded), percent with at least HS education (reverse coded), median household income (reverse coded), and percent with household income >$50,000 (reverse coded), percent interest/dividend/rental income

Factor2 = sum of standardized percent with no vehicle, percent owner occupied housing (reverse coded), percent below poverty level, and percent unemployed

Factor3 = sum of standardized percent occupied housing (reverse coded) and percent not in labor force

Factor4 = this was not created since there is only 1 variable (percent in same house) with loading>=0.60

The variables percent with no phone, and percent on public assistance were not included in any of the based scales since they did not load to at least 0.60 on any factor.

For all factor scales created, a higher value indicates a WORSE SES status.

Table 3: Factor loadings (with varimax rotation) and variance explained for 4 factors kept in the SES only variables factor analysis (N=192403). (PC2)

Highlighted values are loadings greater than 0.60 (if rounded to 2 decimal places)

| **Variable** | **Factor1** | **Factor2** | **Factor3** | **Factor4** |
| --- | --- | --- | --- | --- |
| Educ\_minBA\_rev | 0.91039 | 0.11972 | 0.19289 | 0.01767 |
| Occup\_I\_rev | 0.82757 | 0.25302 | 0.14729 | -0.18939 |
| HUcost\_medownval\_cbrt | -0.69239 | 0.09143 | -0.35554 | -0.42313 |
| Educ\_minHS\_rev | 0.65364 | 0.44296 | 0.10779 | 0.13235 |
| inc\_IntDivRent\_rev | 0.61943 | 0.56478 | 0.11447 | -0.26489 |
| inc\_medHH\_cbrt | -0.61775 | -0.46919 | -0.41334 | -0.37501 |
| inc\_HHge50k\_rev | 0.59755 | 0.47362 | 0.43216 | 0.38171 |
| ownerocc\_hh\_rev | -0.01210 | 0.83879 | 0.27073 | 0.22354 |
| vehicle\_none\_cbrt | 0.11875 | 0.80205 | 0.12796 | 0.19323 |
| Pov\_cbrt | 0.44404 | 0.69896 | 0.33983 | 0.11694 |
| Unemployed\_cbrt | 0.39411 | 0.59606 | 0.09907 | -0.26743 |
| HU\_sampleocc\_rev\_cbrt | 0.05247 | 0.22612 | 0.82675 | -0.11744 |
| NotInLaborForce | 0.25106 | 0.11773 | 0.61640 | 0.05708 |
| samehouse\_rev\_cbrt | 0.01205 | 0.18172 | -0.07334 | 0.89357 |
| phone\_none\_cbrt | 0.30273 | 0.47583 | 0.34092 | -0.09425 |
| inc\_pubass\_cbrt | 0.47199 | 0.51637 | -0.08329 | 0.17975 |
|  |  |  |  |  |
| Pre-Rotation Eigenvalue | 7.66074906 | 1.58099444 | 1.41531450 | 1.05392996 |
| Pre-Rotation Difference | 6.07975462 | 0.16567994 | 0.36138454 | 0.13241129 |
| Pre-Rotation Variance Explained | 47.88% | 9.88% | 8.85% | 6.59% |
| Pre-Rotation Cumulative Variance Explained | 47.88% | 57.76% | 66.61% | 73.19% |
| Post-Rotation Variance (Eigenvalue) explained after rotation | 4.2898044 | 3.8281163 | 1.9746230 | 1.6184443 |
| Post-Rotation Variance Explained | 26.81% | 23.93% | 12.34% | 10.12% |
| Post-Rotation Cumulative Variance Explained | 26.81% | 50.74% | 63.08% | 73.20% |

### Full variables (SES+race/ethnicity) Principal Factor Analysis

An additional PFA was performed using only the SES variables and Race/Ethnicity and Crowding to have scales including the race effects. This PFA will be referred to as PC3. Five factors were retained based on the criteria of having at least 70% of the variance explained. With five factors, 73.2% of the variance is explained (see Table 4).

Factor scales variables are created for the 5 weighted factors using SAS PROC SCORE. This gives 5 variables with each one being more highly weighted on the variables that have higher loadings for that factor. This gives the following patterns:

Factor1 = Education/income/occupation

Factor2 = Poverty/wealth/Black race

Factor3 = Foreign born/Hispanic/Crowding

Factor4 = Labor force/occupied housing

Factor5 = Housing stability

In addition to the weighted scales, “based” scales were also created. These scales take the variables that have loadings of at least 0.60 (when rounded to 2 decimal places), standardizes (z-score) the variables, and sums together these standardized variables. Standardization was done with Census2000, ACS0509, and ACS0711 data set together for use in longitudinal analysis. The variables used in each scale are as follows:

Factor1 = sum of standardized percent Bachelor degree (reverse coded), percent Managerial/professional occupation (reverse coded), percent HS education (reverse coded), median home value (reverse coded), median household income (reverse coded), percent household income >$50,000 (reverse coded), and percent with Interest/Dividend/Rent income (reverse coded)

Factor2 = sum of standardized percent Black, owner occupied housing (reverse coded), percent w/o vehicle, percent with Interest/Dividend/Rent income (reverse coded), percent below poverty, and percent unemployed

Factor3 = sum of standardized percent foreign born, percent Hispanic, and percent crowding

Factor4 = sum of standardized percent not in labor force and percent occupied housing (reverse coded)

Factor5 = this was not created since there is only 1 variable (percent in same house) with loading>=0.60

The variables percent w/o phone, percent with public assistance, and percent Asian were not included in any of the based scales since they did not load to at least 0.60 on any factor.

For all factor scales created, a higher value indicates a WORSE SES status.

Table 4: Factor loadings (with varimax rotation) and variance explained for 5 factors kept in the Full variables factor analysis (N=192403). (PC3)

Highlighted values are loadings greater than 0.60 (if rounded to 2 decimal places)

| **Variable** | **Factor1** | **Factor2** | **Factor3** | **Factor4** | **Factor5** |
| --- | --- | --- | --- | --- | --- |
| Educ\_minBA\_rev | 0.91072 | 0.13829 | -0.03347 | 0.14872 | 0.03945 |
| Occup\_I\_rev | 0.82637 | 0.27095 | 0.05361 | 0.11648 | -0.15758 |
| Educ\_minHS\_rev | 0.68048 | 0.25242 | 0.37386 | 0.29375 | 0.16954 |
| HUcost\_medownval\_cbrt | -0.66432 | -0.10229 | 0.41506 | -0.09869 | -0.41406 |
| inc\_IntDivRent\_rev | 0.61779 | 0.62661 | 0.13583 | 0.04260 | -0.20608 |
| inc\_medHH\_cbrt | -0.60830 | -0.46485 | 0.08368 | -0.38969 | -0.40390 |
| inc\_HHge50k\_rev | 0.59227 | 0.45133 | -0.07076 | 0.42693 | 0.40863 |
| race\_blackNH\_cbrt | 0.08382 | 0.76142 | -0.07011 | -0.19429 | 0.12509 |
| ownerocc\_hh\_rev | 0.01121 | 0.70583 | 0.27686 | 0.37827 | 0.27729 |
| vehicle\_none\_cbrt | 0.09907 | 0.66815 | 0.21311 | 0.32951 | 0.25066 |
| Pov\_cbrt | 0.44862 | 0.63726 | 0.11761 | 0.39951 | 0.16927 |
| Unemployed\_cbrt | 0.36821 | 0.62069 | 0.09845 | 0.12899 | -0.20458 |
| birth\_foreign\_cbrt | -0.19718 | 0.08812 | 0.90437 | -0.08938 | -0.02016 |
| race\_hisp\_cbrt | 0.20648 | 0.05217 | 0.85320 | -0.02018 | -0.02131 |
| crowd\_gt1\_ppr\_cbrt | 0.41066 | 0.17314 | 0.64687 | 0.02851 | 0.28652 |
| NotInLaborForce | 0.24180 | 0.03282 | -0.03456 | 0.75961 | 0.04000 |
| HU\_sampleocc\_rev\_cbrt | 0.10844 | 0.31953 | -0.23055 | 0.63839 | -0.12061 |
| samehouse\_rev\_cbrt | 0.00984 | 0.09039 | 0.11060 | -0.04661 | 0.90917 |
| phone\_none\_cbrt | 0.32800 | 0.51043 | -0.04184 | 0.25453 | -0.05731 |
| inc\_pubass\_cbrt | 0.44322 | 0.39125 | 0.15588 | 0.10354 | 0.23816 |
| race\_asianNH\_cbrt | -0.50422 | 0.01886 | 0.57768 | -0.22139 | 0.08800 |
|  |  |  |  |  |  |
| Pre-Rotation Eigenvalue | 8.25929956 | 3.07937723 | 1.58480940 | 1.46117235 | 0.98757017 |
| Pre-Rotation Difference | 5.17992233 | 1.49456783 | 0.12363704 | 0.47360219 | 0.07753673 |
| Pre-Rotation Variance Explained | 39.33% | 14.66% | 7.55% | 6.96% | 4.70% |
| Pre-Rotation Cumulative Variance Explained | 39.33% | 53.99% | 61.54% | 68.50% | 73.20% |
| Post-Rotation Variance (Eigenvalue) explained after rotation | 4.7517164 | 3.8606534 | 2.8876308 | 2.0537565 | 1.8184715 |
| Post-Rotation Variance Explained | 22.63% | 18.38% | 13.75% | 9.78% | 8.66% |
| Post-Rotation Cumulative Variance Explained | 22.63% | 41.01% | 54.76% | 64.54% | 73.20% |

## Additional Factor-Based Scale

An SES score was created using the results of a factor analysis performed by Dr. Ana Diez-Roux using the Census 1990 data. A principal factor analysis with orthogonal rotation on the census block groups was performed (results not available). From this, one factor score was created by summing the standardized variables with transformations described in Table 5. This scale represents a basic SES score as one variable scale. This is only a based scale and is not weighted by any loadings. Note that this scale was performed on the Census 1990 data then applied to Census2000, ACS0509, and ACS0711 data. The interpretation of this scale is a higher value indicates a BETTER SES status.

For reference on how this was created see:

Diez Roux AV, Stein Merkin A, et al. Neighborhood of Residence and Incidence of Coronary Heart Disease. N Engl J Med 2001;345(2):99-106.

Table 5: List of variables used in principal factor analysis to create scale based on Ana Diez-Roux Census 1990 analysis.

| **Domain** | **Variable** | **Description** | **Reverse coded** | **Transformation used** |
| --- | --- | --- | --- | --- |
| HOUSING | HUcost\_medownval | Median value of occupied housing units | NO | Log |
| EDUCATION | Educ\_minHS | Percent of person 25 or older with at least high school education | NO | NONE |
| EDUCATION | Educ\_minBA | Percent of persons 25 or older with at least a Bachelor’s degree | NO | NONE |
| OCCUPATION | Occup\_I | Percent with management, professional, and related occupation | NO | NONE |
| INCOME/ WEALTH | inc\_medHH | Median household income | NO | Log |
| INCOME/ WEALTH | inc\_IntDivRent | Percent of households with interest, dividends, or net rental income | NO | NONE |

## Cumulative Averages

To assess long-term cumulative exposure to the neighborhood environment, we created time-varying cumulative means, defined as the mean across all months from the baseline to each follow-up exam.

A monthly address dataset was created with a row for each month that the participant was in the study starting with their Exam 1 date. The last row in the dataset is the date at the last exam the participant completed. The cumulative average is then calculated as:

Where t = number of months from Exam 1 to follow-up exam (ie: Exam 2, Exam 3, Exam 4, or Exam 5)

For Exam1, the value of the cumulative average variable is the same as the actual value for Exam1. For any months with missing data, these months are excluded from the calculation of the average. Since the census data only has 2 time changes (3 total values), this will be more relevant in the later exams (after 2004) and for those who moved, where the values will change at the point in which they moved to the new address.

These summary measures are intended to be used in a model such as (Cumulative Average Model):

Ykit= βo+ β1Xki0+ β2Tkit+ β3(avgXkit\*Tkit)+ β4Aki0+ β5( Aki0\*Tkit)+ βmCovki0+βn(Covki0\*Tkit)+ βpCovTkit+ηk+αki+(νki\*Tkit)+ekit

Where:

Y*kit* = Outcome of interest

X*ki0* = Neighborhood exposure at Exam 1 (baseline)

T*kit* = Time elapsed since Exam1. This will be 0 for Exam1 (baseline) since no time has elapsed.

avgX*kit* = Average neighborhood exposure from Exam1 (baseline) to time *t*

Aki0 = Age at Exam1 (baseline). This controls for age and cohort effects assuming that period effects are null.

Covki0 = Covariates for adjustment at Exam1 (baseline - not time varying)

CovTkit = Time Varying covariates for adjustment

β0= Intercept

β1= Coefficient for baseline neighborhood exposure

β2= Coefficient for time trend

β3= Coefficient for the cumulative neighborhood exposure with change in outcome over time

β4= Coefficient for baseline age effect

β5= Coefficient for baseline age with change in outcome over time

βm= Coefficients for baseline covariates

βn= Coefficients for baseline covariates with change in outcome over time

βp= Coefficients for time varying covariates

ηk =Random effect for neighborhood intercept

αki= Random effect for person i intercept

νki\*Tkit= Random effect for person i time slope

e*kit* = measurement error associated with outcome

1. BLS. (2007). Chaper 17 The Consumer Price Index (updated 06/2007) http://www.bls.gov/opub/hom/pdf/homch17.pdf. In BLS Handbook of Methods http://www.bls.gov/opub/hom/. U.S. Department Of Labor Bureau of Labor Statistics, Washington, D.C. [↑](#footnote-ref-1)
2. BLS. (2013). Consumer Price Index. all urban consumers - US City averages 1913-2012 ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt. In. U.S. Department Of Labor Bureau of Labor Statistics, Washington, D.C. [↑](#footnote-ref-2)
3. <http://www.census.gov/geo/education/pdfs/CensusTracts.pdf> [↑](#footnote-ref-3)
4. John R. Logan, Zengwang Xu, and Brian Stults. 2012. “Interpolating US Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database” Professional Geographer, forthcoming [↑](#footnote-ref-4)