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

**Address and Moving Indicators**

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

# Overview:

Home addresses for the Multi-Ethnic Study of Atherosclerosis (MESA) were obtained at initial enrollment visit and during follow-up calls. These addresses were compiled, cleaned, and geocoded by the MESA Neighborhood Study at the University of Michigan (UM). The address data covers each month from July 2000-February 2012 which represents the time frame that the MESA study is active for Exam1 – Exam5. The dates that the address are available are from the first contact at Exam1 (as determined by date of Exam1) until the last known contact date as provided by the MESA Coordinating Center. Indicators for moving status, when the moves occurred, and number of moves are included in the dataset.

Geocoding was completed using TeleAtlas. Geocoding accuracy was good for these addresses with over 96% being geocoded to the street level. Census tracts were provided using Census 2000 boundaries. A fake census tract ID was created to use to keep data confidential. This ID can be used for analysis with clustering by census tract.

# Recommendations:

It is recommended in most analyses to restrict to those who have a geocoding accuracy of street level or Zip+4 centroid. These provide the most accurate location when applying the neighborhood measures.

# 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

**Data Set Name: MESANBH\_ADDRESSIND01192017**

| **Variable Order** | **Variable name** | **Description** | **Coding** |
| --- | --- | --- | --- |
| 1 | idno | MESA ID number |  |
| 2 | EXAM | MESA Exam number |  |
| 3 | add\_number | Address number. Indicates the order of addresses in the study. Ranging from 1-9. |  |
| 4 | 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 level2 = Zip+4 centroid level3 = Zip+2 centroid level4 = Zip code centroid5 = Unable to geocode |
| 5 | 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. |  |
| 6 | naddr | Total number of addresses during study period. |  |
| 7 | move | Did participant ever move during study period? | 0 = No1 = Yes |
| 8 | nmove | Total number of moves during study period. |  |
| 9 | Exammove | Exam number when first moved | 0 = Never moved2 = Exam 23 = Exam 34 = Exam 45 = Exam 56 = After Exam 5 |
| 10 | movenow | Did participant move before this exam? | 0 = No1 = Yes |
| 11 | nmovenow | Total number of moves before this exam |  |
| 12 | MOVELAST | Did participant move since last exam? | 0 = No1 = Yes |
| 13 | moveco | Did participant move outside of the main MESA counties? | 0 = Never moved1 = Moved, not outside MESA counties2 = Moved, outside MESA counties |
| 14 | d\_miles | Distance between current address and first (exam 1) address in miles |  |

# Details:

## Address File Received

The MESA addresses data covers each month from July 2000-February 2012 which represents the time frame that the MESA study is active for Exam1 – Exam5. Address updates were obtained by MESA as part of the follow-up phone calls. The data was sent to UM from UW in multiple batches over time in 2006, 2010, and 2012. These files contain the MESA id, address, city, state, zip code, date of entry (tracking date), and move in and move out dates if available. Addresses were cleaned by hand as necessary to fix any typos (ie: I or L instead of 1 or O instead of 0) before geocoding. The first batch of addresses contained the addresses for all participants. After that initial batch, each batch only contained new or updated addresses. There were a total of 9270 addresses across time for the 6191 participants who agreed to participate in the neighborhood study.

## Geocoding

Geocoding for the MESA home addresses was performed using TeleAtlas. Addresses received at UM before 2007 and before were sent directly to TeleAtlas for in house geocoding. For addresses received 2010 and later, these were geocoded remotely through TeleAtlas EZ-Locate web based software (v2.47; <http://www.geocode.com>). TeleAtlas EZ-Locate software allows for uploading of the files of addresses directly from any personal computer over the internet so no files need to be sent directly to TeleAtlas and the server returns a specified set of values including converting the street addresses into geographic coordinates (latitude/longitude) in WGS84 projection. In both cases, in addition to providing latitude/longitude coordinates, it also provides detailed information on census tract id, centroid type and Match Type Code (accuracy level) of the addresses as well as associated summary report of geocoding accuracy. These geocodes use a 5 foot offset from major roadways. TeleAtlas uses the TeleAtlas proprietary street database (Dynamap) and Matchmaker® SDK Professional software to standardize addresses and geocode them. Any unmatched or ungeocodeded addresses were edited individually and manually with the help of internet mapping services such as Google Maps, Yahoo Maps and MapQuest. This was done to increase the rate of matching and geocoding. Table 1 shows the geocoding accuracy of the addresses in the study. Geocoding accuracy was good for these addresses with over 96% being geocoded to the street level.

A fake census tract ID was created to use to keep data confidential. This ID can be used for analysis with clustering by census tract.

**Table 1. Geocoding accuracy of final follow-up address data through Feb 2012**

|  |  |
| --- | --- |
| **Accuracy Code** | **N (%)** |
| 1 = Street Level | 8932 (96.35%) |
| 2 = Zip+4 Centroid | 17 (0.18%) |
| 3 = Zip+2 Centroid | 36 (0.39%) |
| 4 = Zip Centroid | 241 (2.60%) |
| 5 = Unable to geocode | 44 (0.47%) |
|  | 9270 |

## Determining Dates

Since the address files received included the addresses with tracking dates (entry date), move in, and/or move out dates, “start” and “stop” dates for each address needed to be determined for length of time a participant was living at each address. Based on the order of the start and stop date, an address number is linked to each address.

For addresses received 2007 and prior, the actual move in and move out dates were not available. These only contained a tracking date for when the address was entered. For participants with more than one address (moved during 2000-2007), then the mid-point between the tracking dates were used as the move-in date.

For addresses received 2010 and later, move in, move out, tracking dates, and last contact dates were used to determine when the start and stop dates are for each address. The move in, move out, and tracking dates were included in the excel files with the addresses received from UW. The “last contact” referred is the maximum of the last contact date sent by UW or the tracking date in the address file indicating the last time the participant was contacted. For any participants who were still active in the study but there were no new address updates, the last known address was used and extended until the last contact date. For addresses that had a move in and/or move out dates, these dates were used as the “start” and “stop” dates. For addresses where the move in and move out dates were unavailable, we used the mid-point of the tracking dates as the move in date.

Once these dates were determined, the addresses were expanded into a monthly dataset and the month/year where each exam occurred was extracted as the address for that exam.

## Moving Indicators

Indicators for moving across time were created as follows:

**Total number of addresses (naddr)**: This counts the total number of addresses reported throughout the study period (Exam 1 to Feb 2012). The value is the same across all exam records in the dataset.

**Moving Indicators**: There are multiple indicators for moves.

**Any move during study period (move)**: A yes/no indicator if the participant moved at any point during the study period (Exam 1 to Feb 2012). The value is the same across all exam records in the dataset.

**Total number of moves (nmove)**: Counts the total number of times the participant moved during the study period (Exam 1 to Feb 2012). The value is the same across all exam records in the dataset.

**Exam number when first moved (exammove):** This indicates the exam when the first moved occurred.

 0 = Never moved during the study period (Exam 1 to Feb 2012)

2 = Exam 2 – First moved between Exam 1 and Exam 2

3 = Exam 3 – First moved between Exam 2 (or Exam 1 if did not attend Exam 2) and Exam 3

4 = Exam 4 – First moved between Exam 3 (or last exam attended if did not attend Exam 3) and Exam 4.

5 = Exam 5 – First moved between Exam 4 (or last exam attended if did not attend Exam 4) and Exam 5.

6 = After Exam 5 - For persons who attended Exam 5 before Feb 2012, the move may be between their Exam 5 date and Feb 2012.

**Indicator if moved before this exam (movenow)**: Yes/No indicator if a person moved from their Exam 1 address at this point in time. For example, if a person moved between Exam 1 and Exam 3, then this will be “no” for Exam 1 and Exam 2 and “yes” for Exam 3, Exam 4, and Exam 5.

**Total number of moves before this exam (nmovenow)**: Counts the number of moves that occurred from Exam 1 to that point in time.

**Moving indicator if the participant moved between visits (MOVELAST):** Yes/No indicator if the participant moved between the exam and the last exam they attended. For example, if a participant moved between Exam 2 and Exam 3 then again between Exam 4 and Exam 5, this participant would have records of Exam 1 = ‘no’, Exam 2 = ‘no’, Exam 3 = ‘yes’, Exam 4 = ‘no’, Exam 5 = ‘yes’.

**Indicator if participant moved outside of the original MESA counties (moveco)**: This indicates if the move occurred within the counties most represented by MESA participants in Exam 1. This indicator is for the current address at the time of the exam.

 0 = Never moved during the study period (Exam 1 to Feb 2012)

1 = Moved, not outside MESA counties: Los Angeles County, CA; Cook County, IL; Baltimore County, MD; Baltimore City, MD; Dakota County, MN; Ramsey County, MN; Bronx County, NY; New York (Manhattan) County, NY; Forsyth County, NC

2 = Moved, outside MESA counties

**Distance between current address and Exam 1 (d\_miles):** Straight line distance between the current address and Exam 1 address in miles.