HEALTH AND RETIREMENT STUDY

Child Proximity Measures

2004, 2006, 2008

Data Description

Version 1.0, May 2010

Table of Contents

1. Introduction

The University of Michigan Health and Retirement Study (HRS) surveys more than 22,000 Americans over the age of 50 every two years. Supported by the National Institute on Aging (NIA U01AG009740) and the Social Security Administration, the HRS is a large-scale longitudinal project that studies the labor force participation and health transitions that individuals undergo toward the end of their work lives and in the years that follow.

Since its launch in 1992, the study has collected information about income, work, assets, pension plans, health insurance, disability, physical health and functioning, cognitive functioning, and health care expenditures. Through its unique and in-depth interviews with a nationally representative sample of adults over the age of 50, the HRS provides an invaluable, growing body of multidisciplinary data to help researchers address the challenges and opportunities of aging.

The Health and Retirement Study began with two separate studies: the original HRS cohort, born between 1931 and 1941, first interviewed in 1992 (at 51-61 years of age); and the AHEAD (Asset and Health Dynamics among the Oldest Old) cohort, born before 1923 and first interviewed in 1993 (ages 70 and above). Both samples included spouses or partners, regardless of age.[1] In 1998, after two follow-ups of HRS (1994, 1996), and one of AHEAD (1995), the studies were merged and combined with new respondents born 1942-1947 (War Baby -- WB) and 1924-1930 (Children of the Depression Age -- CODA) to become a complete panel of the population over age 50. In 2004 a new sample cohort (Early Baby Boomer – EBB) of individuals born in 1948-53 (age 51-56 in 2004) was introduced. Future plans call for a new birth cohort to be enrolled every 6 years. All cohorts are drawn as nationally representative samples, with those based on the 1992 screen containing over-samples of African-Americans and Hispanic Americans.

This document summarizes the procedures used to create proximity measures for HRS family respondents and 'children and step-children' (children in further references) for 2004, 2006 and 2008. The proximity measures contained in this data set are calculated for those cases where the child lives more than 10 miles from the family respondent

2. Setup

By downloading this freely provided data set, you agree to use its contents for research and statistical purposes only and to make no effort to identify the respondents. You also agree to provide information regarding research based on these data by sending a copy of any papers or publications to:

---

[1] The HRS sample came from a screening of 69,336 households conducted in 1992 and generated using a multi-stage, clustered area probability frame. The AHEAD sample was generated for what began as a separate study. Individuals born between 1914 and 1923, and about half of those born in 1913 or before, were identified through the same household screening used to identify the original HRS sample. The other half of those born in 1913 or before were identified using the Medicare enrollment files maintained by the Health Care Financing Administration (HCFA, since renamed the Centers for Medicare, Medicaid Services, or CMS).

```
Health and Retirement Study
Attn: Papers and Publications
426 Thompson Street, 3050 ISR
Ann Arbor, MI 48104
```

Alternatively, you may contact us by e-mail at *hrsquest@isr.umich.edu*. Include the phrase "Attn: Papers and Publications" on the subject line.

2a. Distribution Set

The *Child Proximity Measure* data set is packaged for distribution in a .ZIP file, `cprox.zip`. Extract the data file, the program statement file that matches your analysis environment, the data description (this file), and the codebook file. If you have problems when downloading this data set or in extracting its contents, please contact the HRS Help Desk. See Table 1 (below) for a description of the data set contents as well as a suggested subdirectory structure.

2b. Program Statements

Each data file comes with associated SPSS, SAS, or Stata program statements to read the data. Files containing SPSS statements are named with an .sps extension, those with SAS statements with an .sas extension, and those with Stata statements with .do and .dct extensions.

2b-1. Using the Files with SAS

To create a SAS system file for a particular data set, load the .sas program statement files into the SAS Program Editor and reference the appropriate .da data files. If the *.sas file is located in 'c:\cprox\sas'and the data file is located in 'c:\cprox\data', you can run the file as is. A SAS system file will be saved to directory 'c:\cprox\sas'. If the files are not located in the specified directories, you will need to edit the SAS program file to reflect desired path names prior to running the file.

2b-2. Using the Files with SPSS

To create an SPSS system file for a particular data set, load the .sps program statement files into the SPSS syntax editor window, reference the appropriate .da data files, and select the *Run>All* option. If the *.sps file is located in 'c:\cprox\spss'and the data file is located in 'c:\cprox\data', you can run the file as is. An SPSS system file (*.sav) will be saved to directory 'c:\cprox\spss'. If the files are not located in the specified directories, you will need to edit the SPSS program file to reflect desired path names prior to running the file.

2b-3. Using the Files with Stata

To use Stata with a particular data set, three file types must be present for that data set: .dct, .do, and .da. Files with the suffix ".da" contain the raw data for Stata to read. Files with the suffix ".dct" are Stata dictionaries used by Stata to describe the data. Files with the suffix ".do" are short Stata programs ("do files") which you may use to read in the data. Load the .do file into Stata and then submit it. If the *.do and .dct files are located in 'c:\cprox\stata'and the data file is located in 'c:\cprox\data', you can run the .do file as is. If the files are not located in these directories, you must edit the *.do and *.dct files to reflect desired path names before you run the files.

3. Methodology

This document summarizes the procedures used to create proximity measures for HRS family respondents and 'children and step-children' (children in further references) for 2004, 2006 and 2008. The HRS asks family respondents for information on the ZIP Code or city/state of their children, provided that those children live further than 10 miles from the household. To track potential missing responses and attrition, we first identify all family respondents. Children who live with the respondents or live within 10 miles of the respondents are identified from household information. All other children (i.e., those who are not identified as children living within 10 miles) are considered as the baseline data sample for distance calculation. Table 2 summarizes sample restrictions and frequencies.

Pairs of family respondent and child are the unit of observation. However there have been some household splits over time, so duplicate child records can appear in the data due to the relationships with their separated parents.

In 2004 (see Table 2 for other years)
- 12,289 respondents with 40, 871 children are identified (column (2) in Table 2).
- 28% (3,462 out of 12,289) of family respondents whose children are identified have at least one co-resident child. 12% (4,712 out of 40,871) of children live with at least one parent (column (3) in Table 2).

---

[2] When using HRS data products, you should feel free to create the directory structure that is most suitable for your needs. By using the suggested directory structure (or a Unix equivalent), you will not have to change the path name references in the data descriptor files. If you want to use a different structure, just change the directory references in the program files.

- 49% (5,999 out of 12,289) of family respondents whose children are identified have at least one non-coresident child living within 10 miles. 23 % (9,498 out of 40,871) of children do not live with parent/s but live within 10 miles of at least one parent (column (4) in Table 2).
- 86% of family respondents (10,533 out of 12,289) have at least one child living further than 10 miles. 66% (26,846 out of 40,871) of children are living further than 10 miles from at least one parent (column (5) in Table 2).

In this summary, we refer the sample of respondent-child pairs living further than 10 miles to each other (column (5) in Table 2) as the baseline sample.

3a. Assigning Family Respondent Coordinate Information

The coordinate information for 2004, 2006 and 2008 family respondents is based on the address where the respondent was interviewed. We used the Tele-Atlas North America[3] EZ-Locate service to obtain centroid (latitude/longitude) data by merging on respondent street address, city, state and ZIP Code.

In 2004, 99% (10,465 of 10,533) of the family respondents in the baseline sample have valid centroid information (column (6) in table 2).

3b. Assigning Child Coordinate Information

Child coordinate information was obtained from the *Child ZIP Codes* data set. This restricted data product is made up of three separate files: 2004, 2006 and 2008. In 2006 and 2008, HRS collected geographic information only for those children:
- Who had moved since the last interview
- Whose geographic locations were newly identified by respondents
- Whose respondents' households had split since the last interview

In situations where the child ZIP Codes for 2006 and 2006 were not obtained (no move reported or missing data) the information for previous years was carried forward: 2004 data for 2006 ZIP Codes; 2004 and 2006 data for 2008 ZIP Codes.

The centroids of children with valid ZIP Codes are obtained by merging with a ZIP-centroid cross-reference table. [4] The resulting file is keyed on household identifier (HHID), sub-household identifier (SUBHH) and other person identifier (OPN).

 In 2004, 86% (9,095 out of 10,533) of family respondents in the baseline sample have at least one child's coordinate information. 78% (20,868 out of 26,846) of children have coordinate information (column (7) in Table 2).

3c. Calculating Distance between Family Respondents and Children

For distance calculation, only those observations with valid coordinate information for both family respondent and child are used: 9,038 family respondents and 20,737 children in 2004

---

[3] http://www.teleatlas.com

[4] The latitude/longitude information for child ZIP Codes was obtained by matching against the Zip-Codes.com (http://www.zip-codes.com) database.

(column (8) in table 1). Using the Great Circle Distance formula[5] described in Appendix B we calculated distance between two coordinates for respondent and child. As a check on these results, we used ArcGIS[6] to perform an alternative distance calculation (See Appendix C). Both results are included in the public file.

3d. Summary

The public proximity data files for 2004, 2006 and 2008 contain 9 variables as summarized in Table 1. The distributions of family respondents (Frequency and %) over the distance categories in 2004 are presented in table 3A and that of children in table 3B.

In terms of family respondents as the unit of analysis (table 3A):
- One of the *exclusion* criteria for this proximity data file is children living within 10 miles. However, in the final proximity data, we have 20% of family respondents (1,800 out of 9,038) in the final data that have at least one child within 10 miles[7] (46% within 30 miles) in 2004.
- 8 % of family respondents (709 out of 9,038) have mean distances to their children less than 10 miles (9% in terms of median distance).

In terms of children as the unit of analysis (Table 3B):
- Approximately 13% (2,615 out of 20,737) of children in the proximity data have at least one parent within 10 miles (33% within 30 miles).
- Note that distributions in terms of mean and median distance to parents are identical because each child in the proximity data file is usually associated with one family respondent. Only in the case of a household split can a child be associated with multiple family respondents.

3e. Discussion

1. The proximity data file contains measurements of the distance between family respondents and their children. The proximity of respondents other than family respondents (e.g., spouses/partners of family respondents) and children can be also identified using this proximity data file in conjunction with relationship information in the HRS core files.

2. 22% of children (5,978 out of 26,846) in the baseline sample have missing information for ZIP Code coordinates. As a result about 14% of family respondents in the baseline sample are dropped from of the final proximity data file due to missing child geographic information. Although the final proximity data files have sizable observations (family respondents=9,038; children=20,737; all responses=20,910 in 2004), it is important to consider exploring characteristics of missing responses on children's geographic information.

---

[5]ArcGIS uses Equidistant Conic projection in its calculations. For a good overview of map projection systems, see: http://egsc.usgs.gov/isb/pubs/MapProjections/projections.html

[6] http://www.esri.com/software/arcgis/index.html

[7] Of those family respondents with any non-resident children (11,623), 58% (6,772 family respondents) were identified to have children living within 10 miles in 2004, based on i) direct response on whether children living with or within 10 miles and ii) measured distance from geocodes.

3. In general, distance measurements based on the ZIP Code centroids can overstate or understate the true residential distance. For example, suppose both respondent and his/her child live in the same ZIP Code area, the measured distance value in this proximity data will be zero (Case 1 in Appendix A) because they have the same centroid. Suppose respondent and his/her child live in neighboring ZIP Code areas close to the boundary between them, the measured distance based on the centroids of the two ZIP Code areas is likely to overstate proximity of the respondent and child (Case 3 in Appendix A). There are many other cases that can overstate or understate the proximity of respondent and children. Although it is difficult to illustrate all cases, it might be important to understand potential bias, especially when researchers are interested in relatively close proximity (e.g. 10 miles) because the measurement error can be well over a few miles.

4. More specifically, in the final 2004 proximity file (which is supposed to include only respondent-child pairs more than 10 miles away from each other), 680 cases with zero proximity and 2,008 cases with the proximity greater than zero but less than 10 miles are found using calculations based on Equidistant Conic projection[8].

Table 1: Child Proximity Variables

Variable Name	Type	Label
HHID	str6	Household Identification Number
SUBHH	str1	Sub Household Identification Number
PN	str3	Family Response Person Number
OPN	str3	Other Person Number
rel_r	byte	People Relation to R
year	float	Data collection year
dist_gcd	float	Great Circle Distance (miles)
dist_gis	float	Distance based on Equidistant Conic projection (ArcGIS, miles)

---

[8] Based on the Great Circle Distance, 2,626 such cases (respondent-child pairs) are found (306 of zero proximity and 2,320 of proximity between 0 and 10).

Table 2: Summary of Number of Respondents and Children

		Reference Sample				Baseline Sample	Non-Missing Coordinate Information		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		All Family Respondents	Family Respondents with Children	Resident children	(Any) non-resident children living within 10 miles	(Any) children living further than 10 miles	Respondent Geographic Information File	Child Geographic Information File	From Both
2004	FR	13,514	12,289	3,462	5,999	10,533	10,465	9,095	9,038
	C	-	40,871	4,712	9,498	26,846	-	20,868	20,737
	FR-C		41,357	4,721	9,534	27,105	-	21,043	20,910
2006	FR	12,513	11,429	2,990	5,631	9,971	9,951	8,476	8,463
	C	-	38,290	3,915	8,830	25,719	-	18,715	18,690
	FR-C		38,800	3,924	8,878	25,999	-	18,868	18,843
2008	FR	11,813	10,824	2,674	5,313	9,653	9,626	8,156	8,142
	C	-	36,499	3,413	8,198	25,084	-	17,732	17,704
	FR-C		37,124	3,425	8,239	25,460	-	17,873	17,845

FR=Family Respondent
C=Child
FR-C=Family Respondent - Child

Table 3A: Distributions of Family Respondents – Proximity to Children

(HRS Proximity Data, N=9,038, Year=2004, Great Circle Distance)

	Respondents' proximity to children					
	Minimum		Mean		Median	
	N	%	N	%	N	%
< 10 miles	1800	20	709	8	827	9
< 30	2389	26	1,367	15	1,644	18
< 50	671	7	576	6	676	7
<100	971	11	910	10	999	11
<500	1799	20	2,906	32	2,543	28
<1000	792	9	1,553	17	1,327	15
1000+	616	7	1,017	11	1,022	11
Total	9,038	100	9,038	100	9,038	100

Minimum: distance to the closest child
Mean distance: mean distance to children
Median distance: median distance to children

Table 3B: Distributions of Children – Proximity to Parents

(HRS Proximity Data, N=20,737, Year=2004, Great Circle Distance)

	Children's proximity to parent/s					
	Minimum		Mean		Median	
	N	%	N	%	N	%
< 10 miles	2,615	13	2601	13	2601	13
< 30	4,055	20	4,040	19	4,040	19
< 50	1,446	7	1444	7	1444	7
<100	2,084	10	2094	10	2094	10
<500	4,952	24	4,957	24	4,957	24
<1000	2,812	14	2,821	14	2,821	14
>=1000	2,773	13	2,780	13	2,780	13
Total	20,737	100	20,737	100	20,737	100

Minimum: distance to the closest parent
Mean distance: mean distance to parent
Median distance: median distance to parent

4. If You Need to Know More

This document is intended to serve as a brief overview of the *Child Proximity Measures* data set. If you have questions or concerns that are not adequately covered here or on our Web site, or if you have any comments, please contact us. We will do our best to provide answers.

4a. HRS Internet Site

Health and Retirement Study public release data and additional information about the study are available on the Internet. To access public data and other relevant information, point your Web browser to the HRS Web site at http://hrsonline.isr.umich.edu/.

Visit the HRS Restricted Data Web pages at http://hrsonline.isr.umich.edu/rda if you want to find out more about restricted data products and procedures.

4b. Contact Information

If you need to contact us, you may do so by one of the methods listed below.

```
Internet: Help Desk at the HRS Web site (http://hrsonline.isr.umich.edu)

E-mail: hrsquest@isr.umich.edu

Postal service:
     Health and Retirement Study
     The Institute for Social Research
     426 Thompson Street, 3050 ISR
     Ann Arbor, Michigan  48104

FAX: (734) 647-1186
```
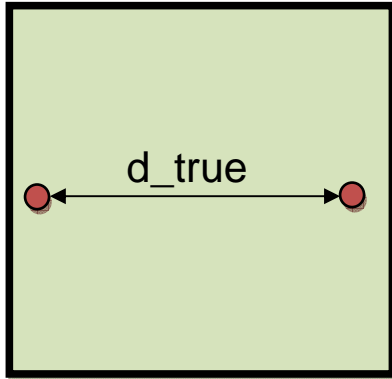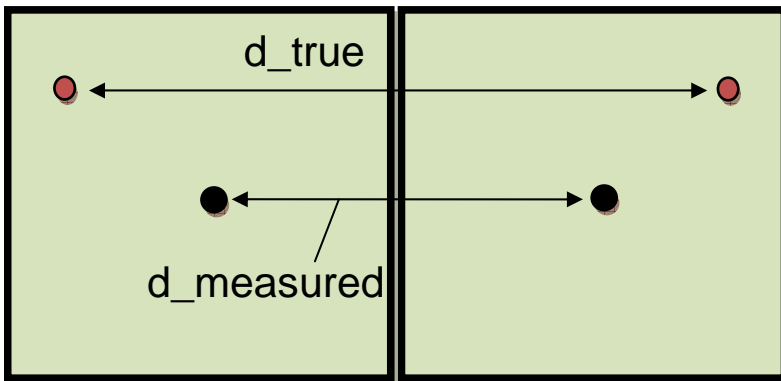
Appendix A: Example Cases of Measurement Error

Example Case 1: Both parent and child are in the same ZIP Code

d_true

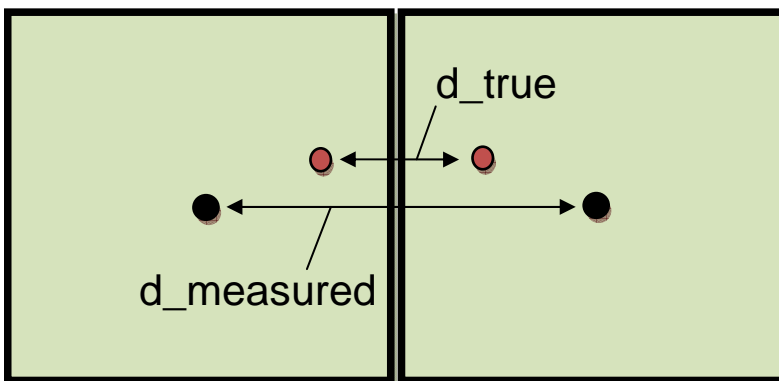True Distance > 0
Measured Distance=0

Example Case 2: Different ZIP Code (beyond centroid)

d_true

d_measured

Measured Distance < True Distance

Example Case 3: Different ZIP Code (within centroid)

d_true

d_measured

Measured Distance > True Distance

Appendix B: Great Circle Distance Calculation Program

```
/*=============================================================*/
/*   DistMoveAll.sas -- calculate distance between two points */
/*   Monday, November 13, 2006;  revised:  12/07/2006         */
/*=============================================================*/

libname in    'C:\GeoWork\SasData\FromMdbAll\';
libname out   'C:\GeoWork\SasData\DistMove\';
libname out2  'C:\GeoWork\SasData';

* macro to calculate distance;
%macro dist(lib1=in, lib2=in, fi1=, fi2=, yrs=);

*--------------------------------------------------------------------;
* first year;
data da1;
     set &lib1..&fi1;
     keep hhid pn mat_lon mat_lat StateUSPS;
     rename mat_lon=lon1 mat_lat=lat1 StateUSPS=StateUSPS1;
run;

*--------------------------------------------------------------------;
* second year;
data da2;
     set &lib2..&fi2;
     keep hhid pn mat_lon mat_lat StateUSPS;
     rename mat_lon=lon2 mat_lat=lat2 StateUSPS=StateUSPS2;
run;

*--------------------------------------------------------------------;
* merge two years;
proc sort data=da1; by hhid pn; run;
proc sort data=da2; by hhid pn; run;

data da3;
     merge da1(in=a) da2(in=b);
     by hhid pn;
     if a and b;
run;
*--------------------------------------------------------------------;
* calculate distance;

/*  SAS Customer Support Resources
        1. Sample 24897: Calculate distances between cities
           (http://support.sas.com/kb/24/897.html)
        2. Usage Note 3973 Calculating the distance between points from
           map coordinates (http://support.sas.com/kb/3/973.html)

    Great Circle Distance Formula – calculate the distance in miles
    between two pairs of latitude/longitude values specified in radians:

        Dist = 3949.99 * arcos(sin(Y1) * sin(Y2) +
        cos(Y1) * cos(Y2) *
        cos(X1 – X2));

    This formula more closely compensates for the curvature of the Earth.
*/
```

```
data out.DistMove&yrs;
     set da3;

     if lon1 ne . and lon2 ne . and lat1 ne . and lat2 ne . then do;
        * Convert degrees to radians;
          FromX=atan(1)/45 * lon1;
          FromY=atan(1)/45 * lat1;
          drop FromX FromY;

          ToX=atan(1)/45 * lon2;
          ToY=atan(1)/45 * lat2;
          drop ToX ToY;

        * Calculate distance using Great Circle Distance Formula ;
          if lon1 eq lon2 and lat1 eq lat2 then Dist&yrs = 0;
            else Dist&yrs = round(3949.99 * arcos( sin( FromY ) * sin( ToY )
                            + cos( FromY ) * cos( ToY ) * cos( FromX – ToX ) ));
          end;
run;

%mend dist;
*********************************************************************;
%dist(fi1=export_all1998, fi2=export_all2000, yrs=9800)
%dist(fi1=export_all2000, fi2=export_all2002, yrs=0002)
%dist(fi1=export_all2002, fi2=export_all2004, yrs=0204)


*********************************************************************;
*      combine files;
data out2.distmove;
     merge
     out.DISTMOVE9800 (drop=stateusps1 stateusps2 lat1 lat2 lon1 lon2)
     out.DISTMOVE0002 (drop=stateusps1 stateusps2 lat1 lat2 lon1 lon2)
     out.DISTMOVE0204 (drop=stateusps1 stateusps2 lat1 lat2 lon1 lon2)
     ;
     by hhid pn;
run;

proc contents data=out2.distmove varnum noprint;
run;
```

Appendix C: Procedure for Calculating Distances Using ArcGIS

1) Prepare data: convert Stata data to dBase IV format

2) In ArcMap, open dBase table and display it (display XYdata and specify latitude and longitude)

3) Export the table to shape file (points_rc*YY*: Geographic coordinate system: GCS_WGS_1984)

4) XTool Pro[9]: →Feature Conversions→Make One Poly line from Points (Specify group ID as HHIDSUBOPN ⇒ lines_rcYY: Geographic coordinate system: GCS_WGS_1984)

5) Convert projections:
ArcTool box: Data management Tools→Projections & Transformations → Feature → Project: Select a projected coordinate (not geographic coordinate system) such as Projection: Equidistant Conic system as an output coordinate system ⇒ lines_rcYY_Project.

6) Add distance variable:
Open table → option → Add field/s → Type: double → Calculate Geometry (Property: length, Units: miles)

---

[9] http://www.xtoolspro.com/