Euclidean distances and ZIP code imputations in healthcare research:

Is good enough really good enough?

Stephen G. Jones, Ashby A.J., Momin S.R., Naidoo A., Mandel R.

Department of Medical Informatics, BlueCross BlueShield of Tennessee

BlueCross BlueShield of Tennessee
BlueCross BlueShield of Tennessee Inc., an Independent Licensee of the BlueCross BlueShield Association

BlueCross BlueShield of Tennessee
BlueCross BlueShield of Tennessee Inc., an Independent Licensee of the BlueCross BlueShield Association

Research Objective:
To determine the combinatory effects of using straight-line Euclidean measurements and ZIP code centroid geo-imputation compared to more precise spatial analytical techniques (i.e. drive distance and street-level residential geocoding) in healthcare research

Background:
In geographically based healthcare research, the distance between two points is often measured to assess differential access to care, hospital service area, or patient travel times to a facility. Distance measurements and geographic placements of patients can be accomplished in multiple ways.

Measurement Technique
Distance can be calculated as:
•Euclidean ("straight-line" distance)
•Network or Drive Distance

Geographic Placement
Patients can be placed at:
•Residential street address
•ZIP code centroid

Each of these practices can introduce quantitative bias within the metric. As the use of GIS and spatially oriented data increase in healthcare research, it is important to understand the implications that may exist in using these methodologies. The intent of our research is to determine if significant differences in distance values exist using the different methods. The results of this study can be applied to future research efforts within health services research regardless of outcome.

Study Design:
Members with an inpatient claim for any reason during October 2005- September 2006 were extracted for study (n=66,492). Using a geographic information system (GIS), latitude/longitude coordinates were obtained for 1) the member's residential geocoded address, 2) their corresponding geographic ZIP code centroid and 3.) the facilities actual location.

Distance from the admitting inpatient facility to the member was calculated using two different measurement techniques:
1)Euclidean straight-line and
2)Shortest-path drive distance.

Using non-parametric Wilcoxon signed-rank tests, linear differences between geographic placement (Centroid vs. Address) and measurement techniques (Euclidean vs. DriveDistance) were examined. Lastly, using simple correlation analysis, we compare the most precise methodology (residential geocoded address-drive distance) to the least precise method (Euclidean-ZIP centroid). (Figure 1)

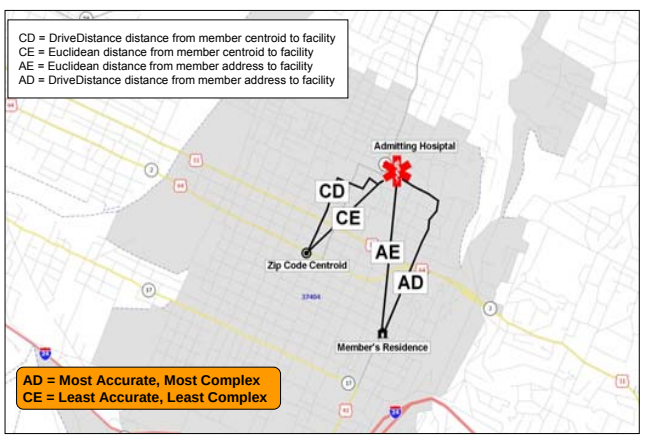

Figure 1. Example of 4 different scenarios of geographic placement of member (Centroid, Address) and measurement techniques (Euclidean, DriveDistance).

Results
Measurement technique produced larger actual differences in linear distance to a facility compared to geographic placement of the member. Differences were greater for rural members compared to urban members. Regardless of geographic placement, DriveDistance measurements to the admitting facility were statistically greater ($p<0.0001$) than Euclidean distances for rural and urban members. Distance values were statistically higher when members were placed at their centroid versus their residential address, although actual median values were low (i.e. 0.8 miles or less) for urban and rural members (Table 1). Distance values using the most precise method were highly correlated ($r=0.99$) to values using the least precise method.

Table 1. Comparing linear distances from patients to admitting facility using Euclidean straight-line measurements and DriveDistance measurements with member origins at residential street address and zip code centroid

		Median Distance from Member to Facility when Member is Placed at:				Difference in Median Values (miles)				
		Centroid		Address		Most Desired Method (AD) vs Least Desired Method (CE)	Measurement Type		Geographic Placement	
	N	Euclidian (CE)	Drive Distance (CD)	Euclidian (AE)	Drive Distance (AD)	AD vs CD	AD vs AE	CD vs CE	CE vs AE	CD vs AD
Rural	27,732	14.8	18.9	14.1	18.2	4.5*	4.1*	4.1*	0.7*	0.7*
Urban	38,760	8.1	9.8	7.6	9.0	1.1*	1.4*	1.7*	0.5*	0.8*
Overall	66,492	9.5	11.9	9.0	11.1	1.9	2.1	2.4	0.5	0.8

* statistically significant difference at α = 0.05 (overall not tested)

A – Compares Euclidean and DriveDistance measurements with member placed at Address (AE vs AD)

B – Compares Euclidean and DriveDistance with member placed at ZIP centroid (CD vs CE)

C – Compares differences of member placement using Euclidean measurements (AE vs CE)

D – Compares differences of member placement using DriveDistance measurements (AD vs CD)

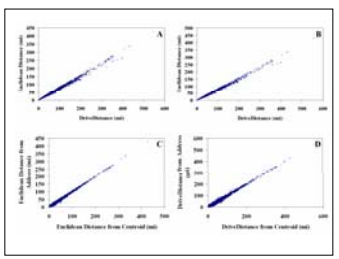

Figure 2. Scatter plot of distances from member to admitting facility

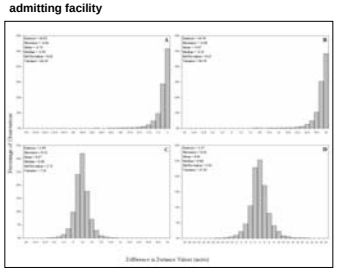

Figure 3. Distribution analysis of measurement differences

Conclusions:
Overall, differences were greater in Rural measures compared to Urban. Actual differences were relatively small. Researchers without capabilities to produce drive distance measurements and/or address geocoding techniques could rely on simple linear regressions to estimate correction factors with a high degree of confidence.

Jones SG, Ashby AJ, Momin SR, Naidoo A. Spatial Implications Associated with using Euclidean Measurements and Zip Code Centroid Geoimputation Methods in Healthcare Research. 2010. Health Services Research, 45(1):316-327

Contact: For more information, please contact Stephen Jones at stephen_jones@bcbst.com