

1101 Geogl. Author manuscript, available in rivic 2013 November 20

Published in final edited form as:

Prof Geogr. 2012 April 1; 64(2): . doi:10.1080/00330124.2011.583586.

A Nationwide Comparison of Driving Distance Versus Straight-Line Distance to Hospitals

Francis P. Boscoe, New York State Cancer Registry

Kevin A. Henry, and New Jersey Cancer Registry

Michael S. ZdebUniversity at Albany School of Public Health

Abstract

Many geographic studies use distance as a simple measure of accessibility, risk, or disparity. Straight-line (Euclidean) distance is most often used because of the ease of its calculation. Actual travel distance over a road network is a superior alternative, although historically an expensive and labor-intensive undertaking. This is no longer true, as travel distance and travel time can be calculated directly from commercial Web sites, without the need to own or purchase specialized geographic information system software or street files. Taking advantage of this feature, we compare straight-line and travel distance and travel time to community hospitals from a representative sample of more than 66,000 locations in the fifty states of the United States, the District of Columbia, and Puerto Rico. The measures are very highly correlated ($t^2 > 0.9$), but important local exceptions can be found near shorelines and other physical barriers. We conclude that for nonemergency travel to hospitals, the added precision offered by the substitution of travel distance, travel time, or both for straight-line distance is largely inconsequential.

Keywords

accessibility; detour index; proximity; travel distance

Numerous geographic studies have analyzed the distances between residential locations and locations of features in an effort to identify risks, gaps, shortages, and disparities. These include emergency services, such as distances from fire stations or ambulance response times (Jones and Bentham 1995; Lyon et al. 2004; Liu, Huang, and Chandramouli 2006; Nicholl et al. 2007; Schuurman et al. 2009); medical services, such as distances and travel times to primary care physicians, hospitals, or specialists (Luo 2004; Wang and Luo 2005; Patel, Waters, and Ghali 2007; Ludwick et al. 2009); and proximity to amenities such as schools, playgrounds, and greengrocers (Pearce, Witten, and Bartie 2006; Pearce et al. 2007; Larsen and Gilliland 2008; Sharkey 2009). There are also a host of studies that look at spatial proximity to adverse features such as gambling centers, liquor stores, and pollution sources (Burdette and Whitaker 2004; Cradock et al. 2005; Pearce et al. 2008; Hart et al. 2009; Hay et al. 2009; Kearney and Kiros 2009). The preponderance of these studies has found that distance is a relevant explanatory variable, with shorter distance corresponding to higher utilization or exposure.

In many of these studies, geographic distance is the single measure of accessibility or exposure. Others incorporate additional measures involving population size, density of features, and choice among competing options (Guagliardo 2004), but even for these more complex analyses, distance is a necessary component. A majority of studies define distance as the straight-line distance between locations, using either Euclidean distance with projected coordinates or spherical distance with latitude and longitude coordinates. Locations are often first aggregated to some geographic unit for which the data have been collected, such as postal codes or census-defined areas. An advantage of this approach is that calculations are straightforward, not necessarily requiring specialized geographic information systems (GIS) software. A smaller number of studies measure distance using actual road network distances or automobile travel times. This approach offers greater sophistication and precision, although traditionally at the expense of purchasing and managing specialized GIS software or street network data. In our experience, we have found that driving distances and times are perceived to be substantially more precise than straightline distance. For example, this perception was a major impetus for the recent development of a shortest path calculator by the North American Association of Central Cancer Registries and the University of Southern California GIS Research Laboratory (2009).

Recent technological advances have essentially eliminated the cost of using street-network distance in analyses. There are now at least five commercial Web sites offering precise driving directions between nearly all locations in the developed world (Google, Yahoo!, Mapquest, Bing, and Rand McNally). Simple programs written in open-source programming languages such as Python can be used to make repeated calls to these sites to obtain the travel time and distance information for any number of locations. In this article, we make use of this functionality to conduct a large-scale comparison of straight-line distances and travel times and distances for the fifty states of the United States, District of Columbia, and Puerto Rico. Our aim is to assess the extent to which using travel time or distance confers a genuine advantage over straight-line distance and to identify locations where differences between the two are most pronounced.

Interest in this question dates at least to the 1960s and research on network models in geography (Haggett 1967). Cole and King's (1968) *Quantitative Geography* defined the ratio of travel distance to straight-line distance as the "detour index" and reported typical values of 1.2 to 1.6 for rural areas in various parts of Britain, with the calculations done by having students trace roadways on paper maps. This ratio has been applied in other fields ranging from ecology to computer science where travel over networks is measured, often under different names such as the "index of circuitry" or "route factor" (Cardillo et al. 2006; Bebber et al. 2007; Buhl et al. 2009).

Several studies have found that the correlation between straight-line distance and roadnetwork distance or time is extremely high and that substituting one for the other is unlikely
to have a substantial impact on analytic results (Martin et al. 2002; Wood and Gatrell 2002;
Jordan et al. 2004; Fone, Christie, and Lester 2006; Apparicio et al. 2008). In particular, a
New York State study considering travel from postal ZIP codes to hospitals via state
highways found a near-perfect correlation between straight-line distance and road-network
distance (Phibbs and Luft 1995). In contrast, a study of access to renal units in England
reported the use of road-network distance represented a "significant improvement" (Martin
et al. 1998), and a Spanish study found that travel distance offered better predictors of transit
ridership than straight-line distance (Gutierrez and Garcia-Palomares 2008).

It is sensible that the correlation between straight-line distance and road-network distance would be very high in the United States given the overall density of roads. Of course, it is not a perfect correlation. Islands, points along an irregular coastline or lakeshore, and

locations separated by uncrossable lakes, rivers, and mountains would be expected to have higher-than-expected travel times. This was the case for many locations in a study of hospital access in rural British Columbia (Schuurman et al. 2006). In the New York State study just cited, points on opposing sides of the Hudson River where there were no nearby bridges were among the largest outliers. In this article, we evaluate the magnitude of these deviations throughout the United States and the extent to which they argue for the standard use of travel distance in social scientific research.

Methods

We developed a population-based nationwide sample of travel paths by selecting one point from each census tract as origins and locations of community hospitals as destinations. Census tracts are designed to be demographically homogeneous and roughly equal in population, with an average of about 4,000 people per tract. There were 66,125 census tracts containing at least one road in the fifty states, District of Columbia, and Puerto Rico according to the 2000 Census. The geographic centroid of each tract, derived from the Census cartographic boundary files, was snapped to the nearest vertex of the nearest road, and this was taken to be a representative location in the tract. ¹

The straight-line distance, travel time, and travel distance between these points and the nearest community hospital were then calculated. Community hospitals consisted of the set of hospitals categorized as general acute care hospitals (N= 5,111) by the Centers for Medicare and Medicaid Services (2009) as of March 2009. These are nonfederal, publicly accessible hospitals that mirror the population distribution generally, with 40 percent located outside metropolitan statistical areas. Geocoding of hospital locations was done using QualityStage Geolocator software, version 2.0.1, and the Dynamap street reference file, version 9. Hospitals that could not be geocoded (< 1 percent) were manually reviewed and geocoded using Google Maps (Google 2009).

The nearest hospital was found by first identifying all candidate tract—hospital pairs within one degree of latitude and longitude of each other, calculating the straight-line distance for each pair, and retaining the minimum distance. For tracts not within one degree of latitude and longitude of a hospital, the search was expanded to two degrees, and so on. This method reduced the number of potential calculations by 98 percent. Straight-line distances were computed as great circles assuming a spherical earth. Results summarizing the difference between the predicted and actual driving distances for all tracts were viewed on a scatterplot, basic statistics were calculated, and the magnitude and locations of substantial outliers were noted.

Travel distance and travel time were obtained through repeated calls to the Google Maps Web page using the SAS FILENAME URL method in SAS version 9.1 (Helf 2005; Zdeb 2009). Each call generates a vast amount of HTML code from which the travel distance and travel time can be extracted using character string functions. (Walking time and distance can also be obtained in this manner, along with driving time during peak traffic periods and public transportation time for selected metropolitan areas. These were explored but not used in this study.) The SAS code is available from the authors.

Once straight-line distances, driving distances, and driving times were obtained for all tracts, the linear relation and correlations between these three measures were assessed using ordinary least-squares regression. The model was fitted using a zero intercept given that a

¹The alternative of choosing a random location in the tract does not impact the results. The average resulting displacement from the centroid is about 1 kilometer and is independent on the hospital locations.

trip of zero distance requires zero travel time. The ratio of driving distance to straight-line distance was defined as the detour index. The difference between the actual driving distance and the predicted driving distance as derived from the regression equation was calculated for all tracts and used to examine outliers. The "predicted" travel distance or travel time was defined as the straight-line distance multiplied by the slope of the regression line. Except where noted previously, all analyses were performed using SAS version 9.1 and ArcGIS version 9.3.

Results

There were 66,011 census tracts with a valid driving route to a hospital. The 114 remaining tracts mainly consisted of islands without ferries or bridges, along with a very small number where the selected point fell within a gated residential community.² Straight-line distance predicted travel distance very well in nearly all locations, with the r^2 for the United States as a whole equal to 0.94. The largest outliers were disproportionately located in Alaska, which has significant roadless areas and locations connected by ferry, but excluding Alaska did not alter the r^2 . The detour index was 1.417 for the entire data set. Straight-line distance also predicted travel time very well, with $r^2 = 0.91$ for the United States as a whole—reasonable given that travel distance and travel time are themselves highly correlated (Table 1).

The r^2 values are presented merely to establish that they are extremely high; their exact interpretation is confounded by the spatial autocorrelation of the observations. Of greater interest is identifying the number and location of tracts where the straight-line distance is a poor predictor of driving distance. Both the absolute and relative differences between the actual and predicted driving distances were used to measure this (Table 2). Over 90 percent of the tracts have good agreement using thresholds within 10 percent or 5 kilometers. For the remainder, positive relative errors represent locations where the actual driving distance exceeds the predicted driving distance; negative values represent the converse. Large positive relative errors are found near irregular shorelines, on islands, in very low-population-density wilderness areas, adjacent to other impassable physical features, or some combination of these. Large negative relative errors are in locations that are not close to a hospital but that have a very straight drivable route to the nearest one. The lowest possible relative error is 41.7 percent, the situation when a route follows exactly the shortest straight-line distance.

The most extreme difference between straight-line distance and travel distance is found between Grand Marais, Minnesota, and Houghton, Michigan (Figure 1). Here, an 8.5-hour drive through three states is required to cover a distance that via straight line is just 155 kilometers, yielding a detour index of 3.4. This example also misassigns the hospital that is truly closest, as the hospital in Duluth, Minnesota, would be reached before the one in Houghton. Other large differences are found at other locations in the western Great Lakes; between the eastern end of Long Island, New York, and Connecticut; and in remote parts of western states such as Utah and Idaho. The same type of pattern can be found within urban areas, albeit on a much reduced scale. In New York City, there is a section of Queens close to the East River where the closest hospital is 1.1 km across the river in Manhattan. With no bridge immediately nearby, driving this route requires traveling 9.4 km, a detour index of 8.5, but there are hospitals in Queens closer than this.

²The Google Maps database structure does not allow these points to be connected to the greater road network. This database characteristic was introduced into the Google Maps database during the course of our research, and it is unclear whether this was an intended feature or a bug. We considered using a competing database, but the number of problematic locations was small enough to have no impact on our results.

The differences between straight-line distance and travel distance are further illustrated by mapping the outliers in Nevada, a state with some of the largest outliers (Figure 2). The map reveals longer-than-expected travel distances in the mountainous suburbs west of Reno, where roads are sparse and serpentine. Meanwhile, in the small town of Elko, travel distances are shorter than expected owing to a very direct route to the nearest hospital—albeit one that is in an adjacent state, roughly four hours away. Overall, though, the two measures agree to within 10 kilometers for over 90 percent of the tracts.

In terms of computational complexity, few studies involving geographic distance use as many points as we have used here (66,000 origins and 5,000 destinations). For studies larger than this, processing time could become an issue. In our study, identifying the nearest hospitals from each sample point and calculating the straight-line distances took about 1.5 hours using a desktop computer with dual 3.16 Ghz processors, 3.3 GB of RAM, and 250 GB of free drive space. Finding the travel times via the repeated calls to Google Maps took about five hours. Our approach would be inappropriate for the calculation of travel-distance or travel-time buffers, where very large numbers of travel routes would need to be evaluated. In contrast, the calculation of buffers based on straight-line distance is trivial.

Discussion and Conclusions

In nearly all locations within the United States, the straight-line distance is an adequate proxy for travel distance, after applying a detour index of about 1.4. Exceptions are limited to areas located near uncrossable physical features such as lakes, rivers, and mountains and in wilderness areas of the western United States and Alaska. If errors up to 5 kilometers or 10 percent are tolerated, then the two distance measures are equivalent for over 90 percent of the population; relaxing the threshold to 10 kilometers or 10 percent raises this figure to over 96 percent. These are conservative tolerances in the area of nonemergency medical care, where variations in travel of less than thirty minutes generally do not pose significant barriers (Lee 1991).

These results strongly suggest that the many past studies where straight-line distance was used remain valid, and they contradict the widespread perception that travel distance or time represent a tremendous improvement in precision that should be pursued even at significant cost. But because the cost of obtaining travel distance and travel time has become negligible, we do recommend incorporating their small added precision into future studies that relate residential location to geographic features. In the area of emergency response, where results are sensitive to even small differences, such inclusion is essential. Although we focused on community hospitals, we do believe that our findings logically extend to other geographic destinations of varying densities and spatial scales such as parks, schools, and shopping centers that typically involve car travel. They would not necessarily extend to fine spatial scales where driving is unlikely (Okabe and Kitamura 1996). Our results assume the accuracy of the route choices and drive-time estimates using Google's database. Although we find these data reliable, we are unaware of any formal evaluation of this. The many comments that have been posted online on this subject tend to describe inaccuracies of several hundred meters at most.

Finally, we call attention to the observation that the nationwide detour index of 1.417 is virtually equal to the diagonal of a unit square (1.414). This means that, on average, traveling from an arbitrary address in the United States to the nearest community hospital is equivalent to the maximum possible Manhattan distance between those two points (that is, the distance measured along the two equal axes of an isosceles right triangle). We leave as a future project the determination of whether this is merely an interesting coincidence or a theoretically meaningful result.

Literature Cited

Apparicio P, Abdelmajid M, Riva M, Shearmur R. Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. International Journal of Health Geographics. 2008; 7:7. [PubMed: 18282284]

- ArcGIS, version 9.3. Redlands, CA: ESRI;
- Bebber DP, Hynes J, Darrah PR, Boddy L, Fricker MD. Biological solutions to transport network design. Proceedings of the Royal Society B. 2007; 274(1623):2307–15. [PubMed: 17623638]
- Buhl J, Hicks K, Miller E, Persey S, Alinvi O, Sumpter D. Shape and efficiency of wood ant foraging networks. Behavioral Ecology and Sociobiology. 2009; 63(3):451–60.
- Burdette HL, Whitaker RC. Neighborhood playgrounds, fast food restaurants, and crime: Relationships to overweight in low-income preschool children. Preventive Medicine. 2004; 38(1): 57–63. [PubMed: 14672642]
- Cardillo A, Scellato S, Latora V, Porta S. Structural properties of planar graphs of urban street patterns. Physical Review E. 2006; 73(6):066107.
- Centers for Medicare and Medicaid Services. [last accessed 10 January 2010] National Provider Identifier Standard: Overview. 2009. http://www.cms.hhs.gov/NationalProvIdentStand/
- Cole, JP.; King, CAM. Quantitative geography. London: Wiley; 1968.
- Cradock AL, Kawachi I, Colditz GA, Hannon C, Melly SJ, Wiecha JL, Gortmaker SL. Playground safety and access in Boston neighborhoods. American Journal of Preventive Medicine. 2005; 28(4): 357–63. [PubMed: 15831341]
- Dynamap/Transportation Street Layer, version 9.0. Lebanon, NH: Tale Atlas;
- Fone DL, Christie S, Lester N. Comparison of perceived and modelled geographical access to accident and emergency departments: A cross-sectional analysis from the Caerphilly Health and Social Needs Study. International Journal of Health Geographics. 2006; 5:16. [PubMed: 16613601]
- Guagliardo MF. Spatial accessibility of primary care: Concepts, methods, and challenges. International Journal of Health Geographics. 2004; 3:3. [PubMed: 14987337]
- Google. [last accessed 1 October 2009] Google Maps. 2009. http://maps.google.com
- Gutierrez J, Garcia-Palomares JC. Distance-measure impacts on the calculation of transport service areas using GIS. Environment and Planning B. 2008; 35(3):480–503.
- Haggett, P. Network models in geography. In: Chorley, RJ.; Haggett, P., editors. Integrated models in geography. London: Methuen; 1967. p. 609-68.
- Hart JE, Laden F, Puett RC, Costenbader KH, Karlson EW. Exposure to traffic pollution and increased risk of rheumatoid arthritis. Environmental Health Perspectives. 2009; 117(7):1065–69. [PubMed: 19654914]
- Hay GC, Whigham PA, Kypri K, Langley JD. Neighbourhood deprivation and access to alcohol outlets: A national study. Health and Place. 2009; 15(4):1086–93. [PubMed: 19540790]
- Helf, G. Extreme Web access: What to do when FILENAME URL is not enough. In: Nelson, GS., editor. Proceedings of the Thirtieth Annual SAS Users Group International Conference. Cary, NC: SAS Institute, Inc; 2005. http://www2.sas.com/proceedings/sugi30/toc.html [last accessed 23 May 2011]
- Jones AP, Bentham G. Emergency medical-service accessibility and outcome from road traffic accidents. Public Health. 1995; 109(3):169–77. [PubMed: 7597129]
- Jordan H, Roderick P, Martin D, Barnett S. Distance, rurality and the need for care: Access to health services in South West England. International Journal of Health Geographics. 2004; 3:21. [PubMed: 15456514]
- Kearney G, Kiros GE. A spatial evaluation of socio demographics surrounding National Priorities List sites in Florida using a distance-based approach. International Journal of Health Geographics. 2009; 8:33. [PubMed: 19531266]
- Larsen K, Gilliland J. Mapping the evolution of "food deserts" in a Canadian city: Supermarket accessibility in London, Ontario, 1961–2005. International Journal of Health Geographics. 2008; 7:16. [PubMed: 18423005]

Lee RC. Current approaches to shortage area designation. Journal of Rural Health. 1991; 7(4):437–50. [PubMed: 10116034]

- Liu N, Huang B, Chandramouli M. Optimal siting of fire stations using GIS and ANT algorithm. Journal of Computing in Civil Engineering. 2006; 20(5):361–69.
- Ludwick A, Fu RW, Warden C, Lowe RA. Distances to emergency department and to primary care provider's office affect emergency department use in children. Academic Emergency Medicine. 2009; 16(5):411–17. [PubMed: 19388919]
- Luo W. Using a GIS-based floating catchment method to assess areas with shortage of physicians. Health and Place. 2004; 10(1):1–11. [PubMed: 14637284]
- Lyon RM, Cobbe SM, Bradley JM, Grubb NR. Surviving out of hospital cardiac arrest at home: A postcode lottery? Emergency Medicine Journal. 2004; 21(5):619–24. [PubMed: 15333549]
- Martin D, Roderick P, Diamond I, Clements S, Stone N. Geographical aspects of the uptake of renal replacement therapy in England. International Journal of Population Geography. 1998; 4(3):227–42.
- Martin D, Wrigley H, Barnett S, Roderick P. Increasing the sophistication of access measurement in a rural healthcare study. Health and Place. 2002; 8(1):3–13. [PubMed: 11852259]
- Nicholl J, West J, Goodacre S, Turner J. The relationship between distance to hospital and patient mortality in emergencies: An observational study. Emergency Medicine Journal. 2007; 24(9):665– 68. [PubMed: 17711952]
- Okabe A, Kitamura M. A computational method for market area analysis on a network. Geographical Analysis. 1996; 28(4):330–49.
- Patel AB, Waters NM, Ghali WA. Determining geographic areas and populations with timely access to cardiac catheterization facilities for acute myocardial infarction care in Alberta, Canada. International Journal of Health Geographics. 2007; 6:47. [PubMed: 17939870]
- Pearce J, Mason K, Hiscock R, Day P. A national study of neighbourhood access to gambling opportunities and individual gambling behaviour. Journal of Epidemiology and Community Health. 2008; 62(10):862–68. [PubMed: 18791042]
- Pearce J, Witten K, Bartie P. Neighbourhoods and health: A GIS approach to measuring community resource accessibility. Journal of Epidemiology and Community Health. 2006; 60(5):389–95. [PubMed: 16614327]
- Pearce J, Witten K, Hiscock R, Blakely T. Are socially disadvantaged neighbourhoods deprived of health-related community resources? International Journal of Epidemiology. 2007; 36(2):348–55. [PubMed: 17182634]
- Phibbs CS, Luft HS. Correlation of travel times on roads versus straight line distance. Medical Care Research and Review. 1995; 52(4:5):32–42.
- QualityStage GeoLocator, version 2.0.1. Ascential Software; Westboro, MA:
- SAS, version 9.1. Cary, NC: SAS Institute;
- Schuurman N, Bell NJ, L'Heureux R, Hameed SM. Modelling optimal location for pre-hospital helicopter emergency medical services. International Journal of Health Geographics. 2009; 9:6.
- Schuurman N, Fiedler RS, Grzybowski SCW, Grund D. Defining rational hospital catchments for non-urban areas based on travel-time. International Journal of Health Geographics. 2006; 5:43. [PubMed: 17018146]
- Sharkey JR. Measuring potential access to food stores and food-service places in rural areas in the U. S American journal of preventive Medicine. 2009; 36(4 Suppl):S151–S155.
- University of Southern California GIS Research Laboratory. [last accessed 1 October 2009] Shortest path. 2009. https://webgis.usc.edu/Services/ShortestPath/Default.aspx
- Wang F, Luo W. Assessing spatial and nonspatial factors for healthcare access: Towards an integrated approach to defining health professional shortage areas. Health and Place. 2005; 11(2):131–46. [PubMed: 15629681]
- Wood, DJ.; Gatrell, AC. Equity of geographical access to inpatient hospice care within North West England: A geographical information systems (GIS) approach. Lancaster, UK: Institute for Health Research, Lancaster University; 2002.

Zdeb, MS. [last accessed 1 October 2009] Driving distances and drive times using SAS and Google Maps. 2009. http://www.sascommunity.org/wiki/ Driving_Distances_and_Drive_Times_using_SAS_and_Google_Maps

Biographies

FRANCIS P. BOSCOE is a Research Scientist at the New York State Cancer Registry, 150 Broadway, Suite 361, Menands, NY 12204. : fpb01@health.state.ny.us. His research interests include cancer and chronic disease epidemiology, medical geography, environmental health, spatial methods, and data visualization.

KEVIN A. HENRY is an Associate Director and Research Scientist at the New Jersey Cancer Registry, The Cancer Institute of New Jersey, UMDNJ-Robert Wood Johnson Medical School, 120 Albany Street, Tower II 5th Floor, New Brunswick, NJ 08901.: kevin.henry@doh.state.nj.us. His research interests include geographic analysis of disease, applied methods in health geography, and socioeconomic disparities in cancer stage at diagnosis, treatment, and survival.

MICHAEL S. ZDEB is an Assistant Professor in the Department of Epidemiology and Statistics at the University at Albany School of Public Health, Rensselaer, NY 12144.: msz03@albany.edu. His research interests include biostatistical methods, data visualization, and analysis of vital records data

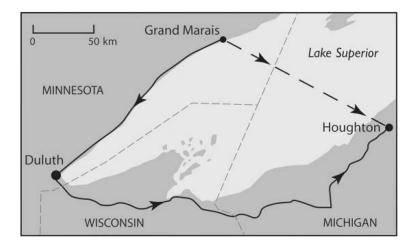


Figure 1. Example of incorrectly determined nearest hospital using straight-line distance.

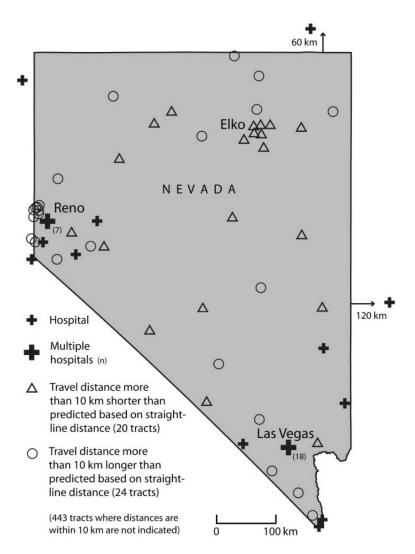


Figure 2. Outliers in predicted travel distance based on straight-line distance, state of Nevada.

Table 1

Correlations between straight-line distance (km), travel distance (km), and travel time (min) for the fifty states of the United States, District of Columbia, and Puerto Rico (N=66,011)

	r ²	Slope
Travel distance and straight-line distance	0.94	1.417
Travel time and straight-line distance	0.91	1.501
Travel time and travel distance	0.96	1.056

 $\label{eq:Table 2}$ Differences between actual and predicted travel distance for the fifty states of the United States, District of Columbia, and Puerto Rico (N = 66,011)

Difference	N	%
Absolute difference of less than 5 km or relative difference within 10 percent		90.5
Absolute difference of more than 5 km and relative difference		
> 50 percent	688	1.0
25 percent to 50 percent	1,342	2.0
10 percent to 25 percent	1,018	1.5
−10 percent to −25 percent	1,155	1.8
-25 percent to -41.7 percent	2,045	3.1

Note: Predicted Travel Distance = Straight-Line Distance * 1.417.