

A conceptual framework for quality healthcare accessibility: a scalable approach for big data technologies

Miloslava Plachkinova¹ · Au Vo² · Rahul Bhaskar³ · Brian Hilton²

Published online: 23 November 2016
© Springer Science+Business Media New York 2016

Abstract Healthcare accessibility research has been of growing interest for scholars and practitioners. This manuscript classifies prior studies on the Floating Catchment Area methodologies, a prevalent class of methodologies that measure healthcare accessibility, and presents a framework that conceptualizes accessibility computation. We build the Floating Catchment Method General Framework as an IT artifact, following best practices in Design Science Research. We evaluate the utility of our framework by creating an instantiation, as an algorithm, and test it with large healthcare data sets from California. We showcase the practical application of the artifact and address the pressing issue of access to quality healthcare. This example also serves as a prototype for Big Data Analytics, as it presents opportunities to scale the analysis vertically and horizontally. In order for researchers to perform high impact studies and make the world a better place, an

overarching framework utilizing Big Data Analytics should be seriously considered.

Keywords Geographic information systems · Big data analytics · Healthcare analytics · Healthcare accessibility · Big data for decision support

1 Introduction

Measuring and predicting access to quality healthcare services are important for decision makers so they can plan more efficiently the distribution of facilities and practitioners to enable access for all individuals in the community. This is especially imperative for the underserved populations who frequently have greater needs but more limited access to social programs. While the latest changes in US healthcare regulations emphasize improving health outcomes,¹ little work has been done to understand what the needs of the population are in terms of obtaining better quality of care. The efforts on a national level demonstrate the pressing need to provide better health outcomes and improved care for patients by utilizing technology in novel and advanced forms.

Previous studies have identified the importance of healthcare accessibility and the proximity of medical facilities to clusters of the population (W. Luo and Qi 2009; W. Luo and Wang 2003; Wang and Luo 2005; Delamater 2013; Fransen et al. 2015; Guagliardo 2004; McGrail and Humphreys 2014). Yet, a universal solution to the problem of defining and measuring quality healthcare accessibility has not been proposed. These studies discuss the advantages and/or disadvantages of

✉ Miloslava Plachkinova
mplachkinova@ut.edu

Au Vo
au.vo@cgu.edu

Rahul Bhaskar
rbhaskar@fullerton.edu

Brian Hilton
brian.hilton@cgu.edu

¹ Department of Information and Technology Management, Sykes College of Business, University of Tampa, Tampa, FL, USA

² Center for Information Systems and Technology, Claremont Graduate University, Claremont, CA, USA

³ Department of Information Systems and Decision Sciences, Mihaylo College of Business and Economics, California State University Fullerton, Fullerton, CA, USA

¹ <http://www.healthit.gov/providers-professionals/how-attain-meaningful-use>, accessed on August 18, 2016

a particular method, providing a ground for a conceptualization opportunity to create a comprehensive framework to guide measuring accessibility in the future.

This study summarizes information on the Floating Catchment Area methodologies (FCA) family, which are vector-based algorithms that utilize concepts in geography, econometrics, and applied physics. Since the inception of the first method in the family in 2003 (W. Luo and Wang 2003), there have been many variations, suggesting different improvement approaches. These enhancements resolve various aspects of measuring accessibility. However, there is no universal classification that encompasses these disparate methodologies and incorporates other aspects of healthcare accessibility such as travel time behavior or quality of care.

Hence, this study provides a framework to conceptualize the process of computing quality healthcare accessibility utilizing the Design Science Research (DSR) paradigm. The presented framework addresses both DSR cycles: relevance – by presenting a solution to a real-world problem – and rigor – by synthesizing the FCA methodologies (Hevner and Chatterjee 2010). We build the framework by extending previous work done by Vo et al. (2015) and we evaluate the utility, usability, application, and impact of the artifact using healthcare accessibility in California as a case study. Our study addresses three important research questions:

- (1) How can we generalize healthcare accessibility index measurements and include other variables in addition to travel time, supply, and demand?
- (2) Can we improve existing FCA methodologies by incorporating other relevant variables?
- (3) Can we provide a scalable approach to compute healthcare accessibility?

To answer the first question, we propose an artifact: the Floating Catchment Area Method General Framework. This framework depicts the overarching processes residing in the FCA methodologies. The artifact is utilized to classify previous research improvements, organizing past research. Furthermore, the framework can also be used as a guideline upon what aspects further research can improve.

To answer the second question, we take into account other pertinent variables that have not been utilized in healthcare accessibility. The cost of the travel variable depicts the pairwise distance between the population and the provider, which is a concept not previously explored. The second variable we take into consideration is quality of care referred as the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). It is a patient satisfaction survey required by the Centers for Medicare and Medicaid Services (CMS) for all hospitals in the United States.

And lastly, to answer the third research question, we demonstrate the utility of the framework and make a case for its

scalability. By showcasing the applicability of the artifact in a particular context, we confirm that our work can be extended to other sites or countries using much larger data sets. The goal of this study is to present a possible application of big data in the context of healthcare but following the principles we discuss, others can extend our work to a wider scope and scale as well as to different fields.

We analyze the access to quality healthcare in California as a demonstration of our work. We selected this research site because its demographics are as diverse as that of the US.² This analysis on California serves as a prototype to demonstrate that a scalable analysis could be performed to the entire US. The scalability is imperative to our practical implication, as the interplay between geographic information systems (GIS) and healthcare presents a novel application of Big Data (Shah and Pathak 2014; Graham and Shelton 2013). Large geographic regions contain enormous data points that include but are not limited to road networks, location points, and areas of coverages. Adding detailed healthcare data increases the complexity of the analysis and calls for substantial processing capabilities (Musa et al. 2013; Barrett et al. 2013; Canlas 2009).

The results of the analysis highlight a disparity in access to quality healthcare. For easier interpretation, we rescale the accessibility index from 0 to 100, 0 meaning no access to quality healthcare and 100 meaning very high accessibility to quality care. According to our score, there are 3.1 million or 16.5% Californians with no access to quality healthcare. Quality healthcare regions cluster together in the big metropolitan areas such as San Francisco, Los Angeles, and San Diego. There are also isolated clusters throughout California that have adequate access. It is intriguing to see the polarity of quality healthcare access as soon as we deviate away from major cities.

The current study makes contributions to knowledge in several aspects. First, it provides a comprehensive literature review of the FCA methodologies family. Second, it proposes an artifact— Floating Catchment Method General Framework—grounded in Design Science Research. As a framework, this artifact presents a comprehensive view on FCA methodologies. The framework processes can be used to enhance quality healthcare accessibility algorithm development and guide future research. Third, we utilize our framework to incorporate two relevant variables (travel behavior and quality of care) that have not been explored by researchers in the past. In addition, we rescale the accessibility score, so practitioners and scholars can interpret the results more easily. Fourth, we model travel behavior to more accurately reflect the travel tendency to and from a healthcare facility. Fifth, through the artifact evaluation, we expound on the state of quality

² <http://quickfacts.census.gov/qfd/states/06000.html>, accessed on August 18, 2016

healthcare accessibility in California: there is a need for improving healthcare access and health outcomes. Finally, our work provides a prototype research that could lend to Big Data Analytics in larger geographic regions and more complex population health issues.

2 Literature review

2.1 Big data

The Affordable Care Act marked the beginning of a new technological reality that required healthcare professionals and hospitals to digitize all patient records. This led to the problem of how to process the enormous amount of data and reap benefits from it. Big Data has been successfully used in a wide variety of disciplines such as astronomy (e.g., the Sloan Digital Sky Survey of telescopic information), retail sales (e.g., Walmart's expansive number of transactions), search engines (e.g., Google's customization of individual searches based on previous web data), and politics (e.g., a campaign's focus of political advertisements on people most likely to support their candidate based on web searches). Following other fields, Murdoch and Detsky (2013) point out that it is inevitable for the healthcare industry to adopt Big Data. The Center for US Health System Reform Business Technology Office even referred to it as “the ‘Big Data’ revolution in healthcare” (Groves et al. 2013).

The promise and potential of Big Data Analytics for healthcare is unparalleled. Raghupathi and Raghupathi (2014) assert that it is “evolving into a promising field for providing insight from very large data sets and improving outcomes while reducing costs” (p. 1). Studies on the impact of Big Data for healthcare in other countries such as Korea support this notion (Jee and Kim 2013). They demonstrate the generalizability of healthcare trend and the global needs for health services. One of the immediate concerns is the conceptualization of the access to health services. Big Data is an essential part of the solution to quality healthcare access.

2.2 Quality healthcare accessibility

There are a number of definitions related to accessibility in the healthcare context. For the purpose of this paper, we use the definition proposed by Khan (1992), which is based on the interaction between the individual and the healthcare system. More specifically, as suggested by Higgs (2004), we employ the notion that accessibility is the “availability of a service moderated by space, or the distance variable” (p. 275) and this information is represented using travel time, road, or map distances. In the following sections, we investigate measurements of accessibility in more detail, focusing in particular on the role of geographical factors.

In addition, Aday and Andersen (1974) propose a theoretical model for health care access. They emphasize on healthcare services utilization and consumer satisfaction by using health policy, characteristics of a health delivery system, and healthcare consumers as proxies. The ultimate objective of the health policy is to “improve access” to health care. We consider these to be especially important features of a healthcare accessibility index, as we not only consider the distance to a medical facility but also the quality of the services provided there.

Access to quality healthcare also means having access to medical professionals. Radke and Mu (2000) suggest a method to delineate the service area of providers delivering social services and produce a probability metric that maps the equity of services for each community. The method they propose measures access to social services for each household and makes adjustments among service providers to accommodate under-served regions. Radke and Mu (2000) also consider the problem of location-allocation and decomposing service regions to predict access and generate equity. Overall, their model computes the ratio of suppliers to residents within a service area centered at the supplier's location and sums up the ratios of residents living in areas where different providers overlap (W. Luo and Wang 2003). This approach gave rise to a family of methodologies, collectively known as the Floating Catchment Area (FCA) methods.

FCA depicts an area that layers, or “floats”, on top of another. Since each location generates an area that depicts its reachability, the close proximity of each location dictates that all generated areas have to float on top of one another. Should that not be the case, the amalgamation of areas created several large regions, inhibiting location-specific analysis. In addition, this area essentially is an overarching layer, capturing the data that are spatially located within it. This is the catchment feature of the method. These features, floating and catchment, depict the FCA methods in its entirety. Researchers use FCA to calculate a quality healthcare accessibility index for a geographic unit, typically a ZIP code or a Census tract. The practical implication of FCA are noticeable: it has been used to inform the US Department of Health and Human Service to designate the Health Professional Shortage Areas (J. Luo 2014; W. Luo and Wang 2003).

Since their inception, FCA methodologies have proliferated and increased in both sophistication and rigor. However, the adoption of the methodologies remains fragmented. There are no uniform application criteria. This review is intended to inform researchers and practitioners about different types of FCA methodologies and to provide a comprehensive summary of the advantages and disadvantages of each method. Table 1 below summarizes briefly prominent FCA methodologies and how they calculate the quality healthcare accessibility index.

Table 1 A Comparison of FCA Methods

Name	Research Location	Author	Highlights
2SFCA	U.S.	W. Luo and Wang (2003)	The first FCA Methodology. It is easy to interpret. It also considers supply and demand of healthcare. Oversimplification in the drive-time analysis, fail to capture rural areas.
Enhanced 2SFCA	U.S.	W. Luo and Qi (2009)	One of the most popular FCA Methodology. It dissects the original 30-min drive time into 0–10, 10–20, and 20–30 min drive time and applies Gaussian weights for each catchment. The weights create other assumptions, which have not been elucidated by researchers.
Optimized 2SFCA	Canada	Ngui and Apparicio (2011)	This method uses weighted calculation of travel time. In addition, the demand is calculated with a proportioning of medical users. This research is only accurate for distances that are less than 550 yards (500 m).
Three-step FCA	U.S.	Wan et al. (2012)	The method accounts for surrounding supply sites as competition by computing a selection weight. The selection weight was implemented into both catchment calculations. This follows the existing two-steps of the FCA Methodology.
Variable Catchment sizes for 2SFCA (A)	U.S.	W. Luo and Whippo (2012)	This research utilizes a fixed provider-to-population ratio and then modifies the catchment accordingly.
Multi-transportation mode 2SFCA (A)	U.S.	Mao and Nekorchuk (2013)	This variation of the Enhanced 2SFCA utilizes different types of transportation as weights. It assigns a weight for each mode of transportation in the calculation of the provider-to-population ratio and the spatial accessibility index.
Modified 2SFCA	U.S.	Delamater (2013)	This method claims to address the overestimation and underestimation of the Three-step FCA method. Using a pairwise comparison between each supply and demand, modified 2SFCA creates catchments that potentially depict suboptimal supplier location.
Variable Catchment sizes for 2SFCA (B)	Australia	McGrail and Humphreys (2014)	This research dissects population into 5 different segments and provides a different catchment size for each. It also assigns distant decay within each segment.
Huff Model-based 2SFCA	U.S.	J. Luo (2014)	The Huff Model is a probabilistic gravity model that was designed to predict consumer behavior within competing retail locations. This method calculates the probability for both supply and demand. To make this model works, there are additional mathematical assumptions.
Kernel Density 2SFCA	Portugal	Polzin et al. (2014)	This research attempts to create a distance decay for the FCA method using Kernel density. The catchment sizes are 30, 45, and 90 min, based on services. It also creates a community index using Principle Component Analysis as a comprehensive assessment of the population.
Domain-based 2SFCA	Germany	Siegel et al. (2016)	The research uses Enhanced 2SFCA as the basis for accessibility measurement. It creates different weights for each health service and accumulates the measurements of the final accessibility index.
Multi-transportation mode 2SFCA (B)	U.K.	Langford et al. (2016)	This method incorporates public and private road networks to improve upon catchment creation and create indices for each mode of transportation. The research concludes that bus-riding has much lower accessibility than car-driving.

The myriad methods in the FCA family that have proliferated in recent years have not paved the way for a better assessment or more accurate measures of quality healthcare accessibility. Rather, they have actually hampered the robustness of the newly developed FCA methods. All improvements within the FCA family merit the need for quality healthcare access. As a result, researchers and practitioners should be aware of framework that guides the computation of quality healthcare accessibility index at a high level. Unfortunately, this high level view does not exist in the current literature. We offer a framework that unifies the FCA methodologies and takes into consideration the quality of care and availability

of medical professionals. This framework can be used to consolidate prior work in the field and provide a more comprehensive approach to quality healthcare accessibility.

3 Research framework design

This study utilizes DSR principles defined by Hevner and Chatterjee (2010) and Hevner et al. (2004) to create an IT artifact. We ground our research in a solid knowledge base incorporating various theories, models, and methods to develop a comprehensive framework to guide quality healthcare

accessibility in general and we take into consideration the notion of relevance to the needs of the environment. Our DSR artifact is thoroughly evaluated using healthcare and spatial data for the state of California. We perform these steps in order to design and build an artifact, which can be successfully utilized by both practitioners and researchers.

We aim to answer the following questions: (1) How can we generalize healthcare accessibility index measurements and include other variables in addition to travel time, supply, and demand? (2) Can we improve existing FCA methodologies by incorporating other relevant variables? (3) Can we provide a scalable approach to compute healthcare accessibility? These three questions are critical to consider as they directly reflect on the ability of hospitals to provide more adequate care for patients, to manage better existing resources, and to suggest a solution for decision makers regarding site suitability analysis of future facilities. Our research is grounded in theory and it also addresses a problem in the real world.

Figure 1 is the general framework to which most, if not all, FCA methods adhere. In congruent with Vo et al. (2015), the Floating Catchment General Framework depicts specific areas in the methods that need improvement.

The framework starts from left to right and from top to bottom. The two preceding processes, “Determine Supply Determinants” and “Determine Demand Determinants”, are located outside of the actual FCA steps. The result of the first and the second step is the “Supply Index” and the “Health Spatial Accessibility Index” respectively. The first step pertains to the supply size while the second step pertains to the demand size. The intricacies between supply and demand are depicted at the final process of the two steps: demand data is fed into the provider-to-population ratio calculation and supply data is a part of the healthcare accessibility index calculation.

All FCA methods are preceded by a processing step consisted of two processes, “Determine Supply” and “Determine Demand”. Past research has not focused on these two preceding processes; rather, researchers use default values for supply and demand of healthcare. Supply of healthcare involves either hospitals or physicians while demand of healthcare is determined by population. Even though these generalizations are adequate, there are research opportunities if researchers and practitioners configure demand and supply appropriately. For instance, in a specific study for childcare, narrowing down physicians’ expertise to pediatrics and utilizing appropriate population age demographics will be suitable for an in-depth analysis.

After completing the preceding step, researchers and practitioners should begin conducting an FCA method. In the first step, they should focus on determining the supply’s catchment sizes. The catchment sizes can follow a process similar to W. Luo and Whippo (2012): incrementally increasing the catchment size until it reaches a predetermined provider-to-

population ratio. Other researchers opt for a fixed catchment size while Dony et al. (2015) employs different modes of transportation (Mao and Nekorchuk 2013; W. Luo and Qi 2009). This process offers research improvement potentials: to figure out the appropriate catchment sizes and the number of catchments. It is imperative to choose catchment sizes correctly since they create a dichotomous boundary for access.

The creation of catchments is a computing-intensive process: it deals with vast and complex road networks and a sizeable number of location points. In our study, we computed the catchment sizes using approximately 1 billion road segments, created 30 catchments for each supply and demand data point, and computed travel time pairwise for all supply and demand. Even though adding more catchment sizes and location points does not exponentially increase the computing power needed, running catchment sizes for a larger geographical region such as the entire US requires significantly more computing capabilities. This is a potential application of Big Data Technologies to streamline the method and to reduce processing time.

The Supply Index in our framework includes a provider-to-population ratio, an output from past FCA research (Delamater 2013; Dony et al. 2015; Fransen et al. 2015; Langford et al. 2016; W. Luo and Qi 2009; W. Luo and Whippo 2012; Ngui and Apparicio 2011; Polzin et al. 2014), and other pertinent variables. This is the final output of the first step. The calculation can impose weights for each variable to represent their importance in relation to other variables.

The second step is similar to the first, with the only difference being the use of demand instead of supply. While the determination of demand catchment sizes is generally similar to that of supply, demand catchment sizes can be different. The Two-step Floating Catchment method produces final quality healthcare spatial accessibility indices for the study site. A choropleth map of the study area is produced to depict the results.

4 Framework evaluation

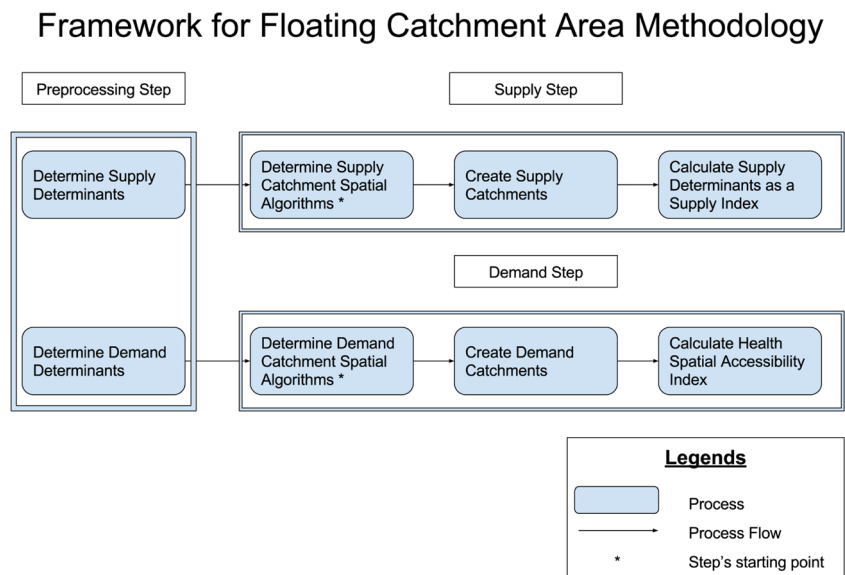
4.1 Case study

In order for us to assess the utility and usability of the framework, we created an instantiation and along with the proposed algorithm we tested it with secondary data on hospitals and spatial data for the State of California. We chose California as its demographics are representative of the US population as a whole.³ We obtained the California hospital data from the Office of Statewide Health Planning and Development (OSHDP).⁴

³ <http://quickfacts.census.gov/qfd/states/06000.html>, accessed on August 18, 2016

⁴ <http://www.oshpd.ca.gov/>, accessed on August 18, 2016

Fig. 1 Floating Catchment Method General Framework



California is a large state with a diverse urban-suburban population. There are three larger metropolitan areas: San Francisco, Los Angeles, and San Diego. All of these metropolitan areas reside on the west of the State. In addition to these areas, there are smaller urban areas separated by large suburban areas. In 2014, California's population is estimated at around 38.8 million and they are served by a total of 435 hospitals (Fig. 2), resulting in a ratio of one hospital per 86,000 Californians. Furthermore, understanding a quality of access is also an important aspect of the study. These present a pressing need to investigate to what extent quality healthcare services are accessible to Californians.

4.2 Algorithm design

Since our framework depicts a high level view of the FCA methods, we evaluate the framework by creating an instantiation of it. This allows us to also test whether the proposed new algorithms can calculate quality healthcare accessibility and to obtain a better understanding of its multifaceted nature. The improvements provide a much more comprehensive calculation because we include travel cost and quality of healthcare. Further, we present the results in a straightforward manner.

We utilize the Floating Catchment Method General Framework to assess the quality healthcare accessibility using the large data sets from the state of California by creating an algorithm. As an instantiation, the algorithm would serve as an evaluation of the framework artifact. This approach demonstrates the utility of the artifact and provides a methodology for others to follow. In the following segment, we delineate the algorithm through the processes depicted in the Floating Catchment Method General Framework.

At the *Preprocessing Step* of the framework, we focus on establishing supply and demand determinants. Generally,

supply of medical services encompasses hospitals, clinics, physicians' offices, hospices, nursing homes, pharmacies, and laboratory centers. The determination of the supply depends on the nature of the research. Past literature primarily uses physicians and hospitals as the supply determinant. For physicians, each physician is a single supply element. For hospitals, licensed beds are the standard supply determinants. As a result, licensed beds are chosen as supply determinants. When establishing demand determinants, we use ZIP code-based census data.

During the *Supply Step*, we propose the following calculation for the Supply Index:

$$SI_j = \beta_1 * R_j + \beta_2 * St_j + \beta_3 * TC_j$$

The Supply Index at a site j (SI_j) is calculated through three supply determinants: R_j , which is the provider-to-population ratio from previous literature, St_j pertains to the hospital's quality and TC_j is the travel cost function. The weight, β , depicts the relationships among the supply determinants. The sum of all weights should equal to 1. In our study formula, we set all weights equivalently.

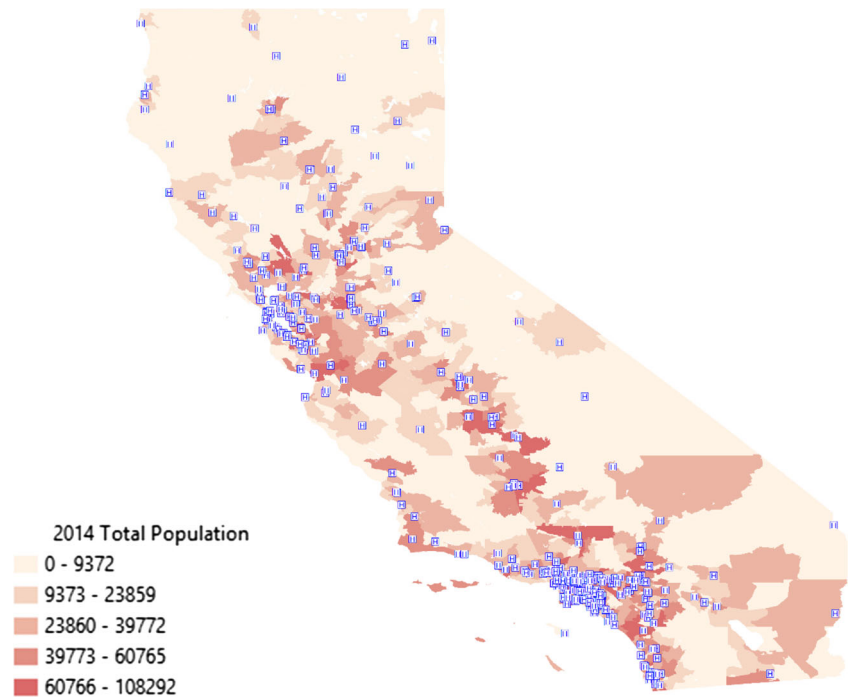
To calculate the supply provider-to-population ratio R_j , we use the following formula:

$$R_j = \frac{S_j}{\sum_{k \in \{Distance(k,j) \leq d_0\}} P_k}$$

Where R_j is the provider-to-population ratio at each provider j , S_j is a medical capacity of that location. P_k is the population at site k , and d_0 is the travel threshold of the provider j . In this step, demand is brought in as a data input.

To emphasize the need for accessible and high quality healthcare, we incorporate hospital's quality St_j into the

Fig. 2 California Population and Hospitals



formula. We obtain the star rating of each site j through HCAHPS as a proxy for hospital’s quality. With hospitals without star ratings, we impute the median score of 3 for the site. The exclusion of the sites is not feasible for our study, since a removal of a hospital site jeopardizes the accuracy of the overall accessibility index. As a result, an imputation is performed to maintain the necessary data points.

The formation of the travel costs are established through utilizing a curve fitting function. The travel cost is an inverse function: the farther a location is from one another, the higher the cost it would occur. The cost function also depicts the inverse relationship between the travel cost and the supply index. The curve fitting function is performed for each location and it follows a general formula:

$$TC_j = \left(\frac{1}{k} \sum_1^k E(C_{j,k}|c_{j,k}) \right)^{-1}$$

Where the expected value of the cost function is:

$$E(C_{j,k}|c_{j,k}) = A_j * c_j^2 + B_j * c + \varepsilon_j$$

Where A and B are the coefficients of the cost c at site j , ε is the residual of the function. We utilize quadratic curve fitting because prior research has demonstrated its accuracy and suitability for such type of analysis (Robinson and Clegg 2005; Soliman et al. 1988).

For supply catchments, we utilize the 30 catchments to depict a minute incremental of drive-time from 1 to 30 min. This approach provides an intricate catchment determination to illustrate the gradual change in travel behavior. The results

of the catchment creation inform the curve fitting mentioned above.

In the *Demand Step*, we derive the quality healthcare spatial accessibility index utilizing this formula:

$$A_i = \beta_1 * D_i + \beta_2 * TC_i$$

The Healthcare Accessibility Index at a site i (A_i) is calculated using two determinants: D_i , demand determinant, and TC_i , travel cost determinant. The weight, β , depicts the relationships between the demand determinants. Similar to the Supply Index calculation, these weights equal each other and add up to 1.

For the demand determinant D_i , the calculation is as follows:

$$D_i = \sum_{j \in \{Distance(i,j) \leq d_0\}} SI_j$$

Where D_i is the demand determinant of each population center i ; SI_j is the supply index at each provider j ; d_0 is the travel boundary of each catchment; $Distance(i, j)$ is the travel time between location i and j .

For the travel cost determinant, TC_i , the calculation to derive at the cost value is as follows:

$$TC_i = \left(\frac{1}{j} \sum_1^j E(C_{i,j}|c_{i,j}) \right)^{-1}$$

Where the expected value of the cost function is:

$$E(C_{i,j}|c_{i,j}) = A_i * c_i^2 + B_i * c + \varepsilon_i$$

Where A and B are the coefficients of the cost c at site i , ε is the residual of the function.

For demand catchments, we replicate the procedure mentioned above in the Supply Step. The results are the 30 concentric catchments originating from the population centers.

4.3 Curve fitting validation

To validate curve fitting as a preferred travel cost function, we perform several steps. First, we utilize the Service Area option within the Network Analyst toolbox in ESRI's ArcGIS 10.4 to calculate 30 concentric one-minute catchments. In addition, to preserve the differences in travel time between the polygons, we perform the Polygon to Raster Function to integrate all polygons into one layer for each location. The raster layer transformation produces a cost number that could be used for curve fitting. Finding the travel cost from each hospital to population centroids, $c_{j,m}$, we construct an Origin-Destination (OD) Cost Matrix. The OD Cost Matrix is a collection of travel time from the origin to the destination (ESRI 2016a). From the travel time, we calculate the appropriate costs pairwise and then we derive the hospital's travel cost. Results of the curve fitting for supply are presented in Table 2 and Fig. 3 below.

In Fig. 3, there are nine supply points that do not adhere to the general curve fitting function. As a result, we impute travel costs for these outliers in order to include those supply points in the subsequent analysis. We replace their original curve fitting with the hospital's curve fitting with the lowest R^2 but still within the three standard deviation below the mean (i.e. $R^2 = 0.58$). This method ensures the preservation of the hospitals' locations while providing a good enough estimate for their travel costs.

We perform similar curve fitting for the demand to estimate travel cost for the population. The results are shown in Table 3 and Figs. 4 and 5 below.

The coefficients of determinations of the curve fitting function for population centroids are comparable to that of the hospitals. We also impute 29 outlier population centroids by using the lowest coefficient of determination's population centroid that is within three standard deviations from the mean (i.e. $R^2 = 0.6$). The process ensures the preservation of the population centroids as well as their travel cost estimation. Overall, the high coefficients of determinations from supply and demand justify our choice in choosing quadratic curve fitting function to model travel cost.

5 Results

Using the algorithm design above, we calculate the quality healthcare accessibility index. This index takes into

consideration quality of care, travel cost, supply, and demand. These variables demonstrate the novelty of the current work and the originality of the proposed methodology. In addition, we rescaled the computed score from 0 to 100 to conform to existing and widely accepted methodologies such as the Walkability Score⁵ or the Sun Number Score.⁶ As opposed to the previous accessibility index, our accessibility index has a larger range. This is in part due to the additional determinants.

At first glance, Californians as a whole have low access to good quality healthcare. As we normalized the quality healthcare accessibility index to represent a score that ranges from 0 to 100, we reveal a startling result of a low access to good healthcare. The disparity of healthcare access in California is glaring; visually, we can detect vast areas that in the lowest quintile of quality healthcare accessibility index. The areas which are in the top quintile are sparse and located within the three big metropolitan areas: Los Angeles, San Francisco, and San Diego. The histogram reveals this disparity (Fig. 6).

The histogram shows that accessibility is skewed to the right, with more than 500 ZIP Codes that have less than accessibility score of 1. Specifically, there are 472 ZIP Codes that have the minimum accessibility scoring (0.000907), which indicates no accessibility to quality healthcare. There are 3.1 million Californians, or 6763 Californians per ZIP Code, that reside in these ZIP Codes. Such low access to quality healthcare undermines the effort of improving community health.

The data is statistically highly volatile as depicted in the descriptive statistics on Table 4. The Normality Test confirmed that the data is not normally distributed (430.92, $p < 0.001$) (R. d'Agostino and Pearson 1973; R. B. d'Agostino 1971). Given the standard deviations and the mean, there are 18 ZIP Codes that are identified as outliers. This demonstrates the variance of quality healthcare across the state of California.

As the data suggests, the interplay between supply and demand is at work: populous metropolitan areas with many hospitals do not necessarily translate to great access, but when access is combined with the hospital's quality of care, the population has a tendency to experience greater access. This can be seen in the high access in the high density area: Los Angeles, San Francisco, and San Diego metropolitan areas. Other areas that are less populous also have adequate access to quality care. It is also interesting to observe that there are isolated clusters throughout California that have adequate access.

It is also interesting to note that there are ZIP Codes that have high access surrounded by very low access ZIP Codes.

⁵ <https://www.walkscore.com/>, accessed on August 18, 2016

⁶ <http://www.sunnumber.com/>, accessed on August 18, 2016

Table 2 Supply Curve Fitting Results

	mean	standard deviation	min	25%	50%	75%	max
Coefficient of determination R^2	0.941086	0.126937	0.065752	0.951162	0.974657	0.984535	1

As soon as we move away from these ZIP Codes, quality healthcare accessibility greatly diminishes. This is partially due to population sparseness. The greater factor that affects the accessibility of the surrounding ZIP Codes is the proximity to hospitals that are at over-capacity.

To further explore the spatial relationship phenomena between quality of care and accessibility among the ZIP Codes, we conducted an optimized hot spot analysis. Optimized hot spot analysis is a spatial statistical process that reveals the clusters of high values and low values within a region. High values and low values are hot spots and cold spots, respectively. Optimized hot spot analysis provides a method to visualize and statistically check the data for relevance and significance. Optimized hot spot analysis produces the Getis-Ord G_i^* statistic, which is then used for hot spots/cold spots identification (ESRI 2016b). The resultant map is a bi-color progression one: red depicts hot spots with high significant while blue signifies cold spots with high significance. Figure 7 displays the optimized hot spot analysis map of quality healthcare accessibility in California.

The optimized hot spot analysis shows that there are significant hot spots with 99% statistical confidence in only the two biggest metropolitan areas and there are no other hotspots with lower degree of confidence. The hot spot clusters between San Francisco in the north and Los Angeles, in the south vary significantly in size. The hot spot clusters in San Francisco extend eastward to Oakland, southward to Silicon Valley, and northward to Sausalito. These depict the highly populated area with high access to quality care. The hot spot clusters in Los Angeles encompass Los Angeles County, Orange County, and parts of the Riverside and San

Bernardino Counties. These are known as an extended Los Angeles metropolitan area, where residents have access to quality healthcare. It is also interesting to see that San Diego, another large metropolitan area, is not identified as a hot spot.

There are several statistically significant spatial clusters of low values depicted as cold spots on the map. It is intriguing to see that quality healthcare accessibility has changed drastically when inching away from the metropolitan areas. The polarization in spatial clustering could be seen in San Francisco metropolitan area and the surrounding areas. Furthermore, there are several stretches of cold spots in California, where low quality healthcare accessibility scores are prevalent.

The analyses demonstrate that the framework we propose is validated and successfully implemented using large data sets on healthcare in California. We follow best practices and guidelines in DSR and positively evaluate the artifact through an instantiation, an algorithm, to apply it in a real world setting. Our work showcases the need to incorporate relevant variables into healthcare accessibility research to be more evenly distributed in order to improve individuals' access and align with the nationwide efforts of providing better quality of care and making a positive impact on healthcare outcomes. The findings of this research provide a better understanding of the challenges associated with measuring quality healthcare accessibility. This study also offers researchers and practitioners a validated framework to improve how quality healthcare accessibility is measured.

6 Discussion

6.1 Research questions revisited

The analyses have provided grounds for answering all three research questions adequately. First, our work demonstrates that there is sufficient evidence to generalize the quality healthcare accessibility index measurements and to provide a comprehensive methodology to evaluate not just the travel time, supply, and demand but also to take into consideration the quality of care and travel cost. These other variables add more depth to the analysis and demonstrate the novelty of our work.

With respect to the second research question, our framework and its instantiation in the form of an algorithm present compelling evidence that current FCA methodologies can benefit significantly from considering new and relevant

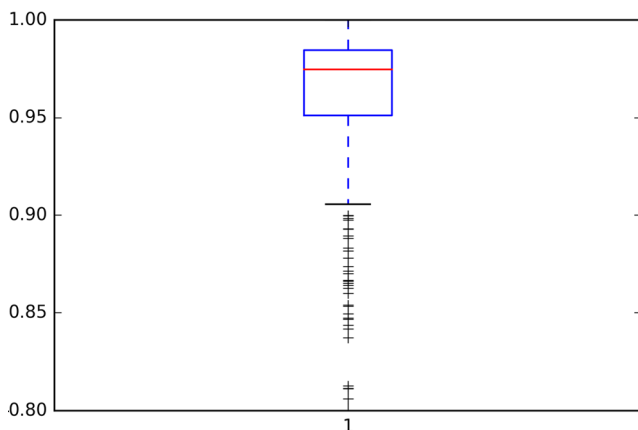


Fig. 3 Supply Curve Fitting Boxplot Analysis

Table 3 Demand Curve Fitting Results

	mean	standard deviation	min	25%	50%	75%	max
Coefficient of determination R^2	0.93654	0.116631	0	0.939075	0.966642	0.980893	1

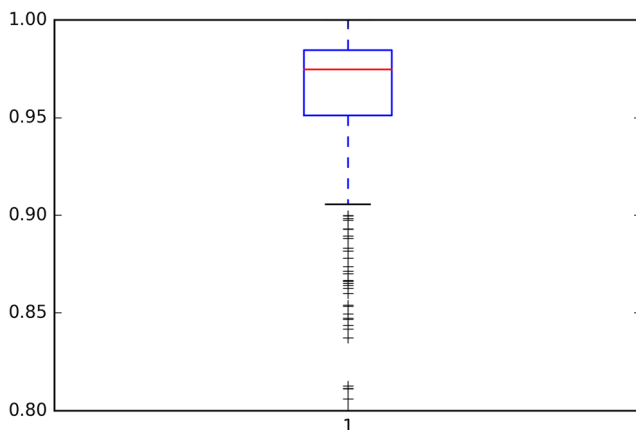
variables in the computation process. Prior studies have not accounted for the quality of care or travel costs, which are crucial for creating a comprehensive quality healthcare accessibility index.

And lastly, we were able to validate the proposed framework and to offer a scalable approach to compute quality healthcare accessibility. While our study focuses on a single state in the US and uses a small set of variables, it reveals the potential of this methodology to utilize Big Data Analytics to extend study sites to other states and countries and to include all pertinent variables in future research.

6.2 Theoretical and research implications

The current study makes several important contributions to Spatial Big Data Analytics. First, it presents a novel framework for conceptualizing the FCA methods. It employs principles from gravity-based models to incorporate quality of care, travel cost, supply, demand, and distance in the characterization of the spatial accessibility of quality healthcare resources. As an artifact, this framework is generalizable and can be utilized in a number of fields beyond healthcare to address the problem of measuring accessibility. We further evaluate the framework using healthcare in California as an instantiation. The evaluation revealed that quality healthcare accessibility in California is sparse. Therefore, the assessment elucidate a call for action: California needs to improve its healthcare accessibility in both quality and reach.

Second, the study summarizes the abundance of existing knowledge on measuring spatial accessibility. This research

**Fig. 4** Demand Curve Fitting Boxplot Analysis

organizes previous studies in a comprehensive manner and provides a brief discussion on each of them. The summarization allows both researchers and practitioners to understand the current state of healthcare spatial accessibility in particular and spatial accessibility in general.

Third, the healthcare accessibility case study in California serves two purposes: instantiation and evaluation. It is an assessment of the artifact Floating Catchment Method General Framework to showcase the framework's utility and applicability. We ground our artifact in DSR principles and we incorporate the notions of relevance and rigor to provide a theoretical solution to a practical problem in a real environment. We follow DSR best practices and guidelines and conduct multiple iterations of the framework. The tool offers a scalable approach to measuring quality healthcare accessibility. We demonstrate that Big Data and analytics can significantly benefit from taking a DSR approach.

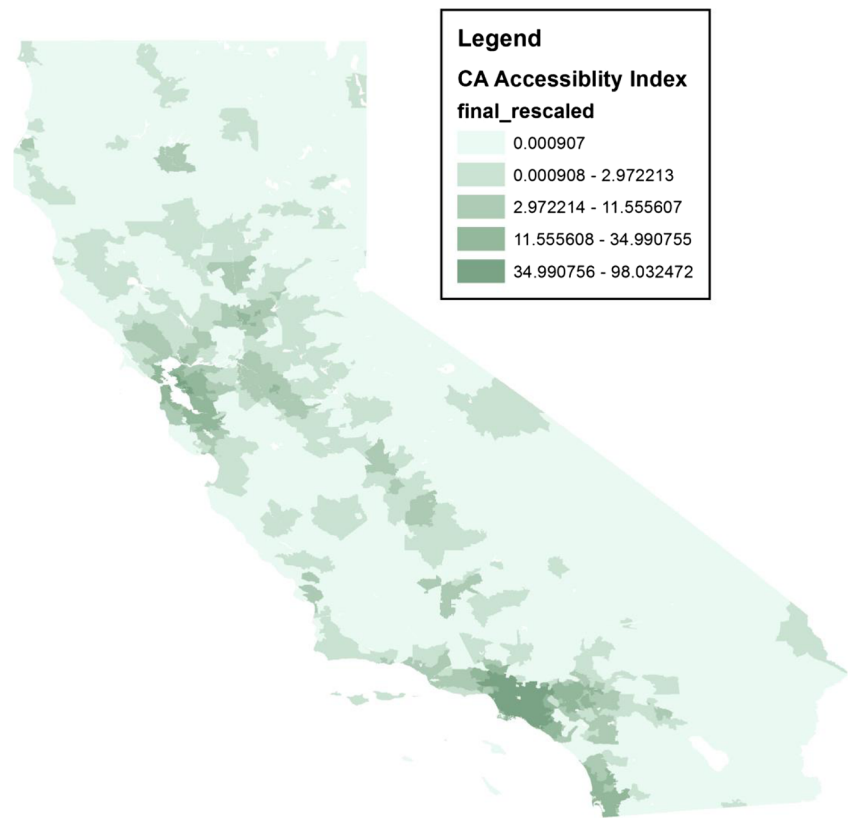
Finally, the case study serves as a prototype for Big Data Applications and Analytics. The amount of spatial and non-spatial data, the computing processes, and the generated results in this case study are replicable. Therefore, it presents opportunities to scale the analysis vertically and horizontally. In essence, in order for researchers to perform high impact studies to make the world better, Big Data Analytics utilizing overarching framework like the one proposed in our study should be seriously considered.

6.3 Practical implications

The current paper does not only contribute to research, but also it provides important implications for practice. First, the study offers decision makers a comprehensive summary of existing methods for measuring healthcare accessibility. This information can be of critical significance when assessing physician shortage areas (Wang and Luo 2005). These areas would have higher priority for decision-makers when considering building new facilities or offering incentives for relocating. In addition, the current study organizes existing knowledge on healthcare accessibility by providing a corpus of the FCA method family.

Second, we offer a conceptualized methodology evaluated with healthcare data from California, encouraging replication and improvement of the proposed artifact. We expand on previous research on the FCA family by taking into consideration travel cost as well as quality of care and creating an easy to

Fig. 5 California Quality Healthcare Accessibility Index



interpret accessibility score from 0 to 100. The presented methodology contains detailed instructions and explanations, which are accessible to a much broader audience compared to prior studies. Our study can be used as a guideline by both researchers and practitioners who are interested in pursuing further research.

And finally, by providing a framework to further analyze the challenges of healthcare and evaluating the capabilities of hospitals, we raise awareness of some important issues. Facilities that demonstrate quality and access to healthcare should be exhibited in smaller urban and rural communities and not just in big metropolitan areas. These objectives align closely with the goals of the “Meaningful Use” program and

support the concept of improving health outcomes by making healthcare more patient-centered.

6.4 Limitations

The current study does not come without limitations. First, we used only healthcare data obtained through the California Office of Statewide Health Planning and Development. Including additional hospital data from other sources in the analysis would be more informative and may enlighten other aspects for consideration in regards to quality healthcare accessibility. Underutilized departments within one hospital may compensate for the expense of others that may be underutilized. The purpose of this analysis is to demonstrate the utility of the proposed framework, thus we encourage additional research to be conducted using larger and more diverse sample sizes using the artifact we created and following the processes we outlined in this study.

The second limitation of this study is the generalizability of the results. We used only data from a single state – California. However, California is the most populous and diverse in the US according to the 2010 US Census⁷ with a population of about 38 million estimated for 2014. Thus, it can be considered

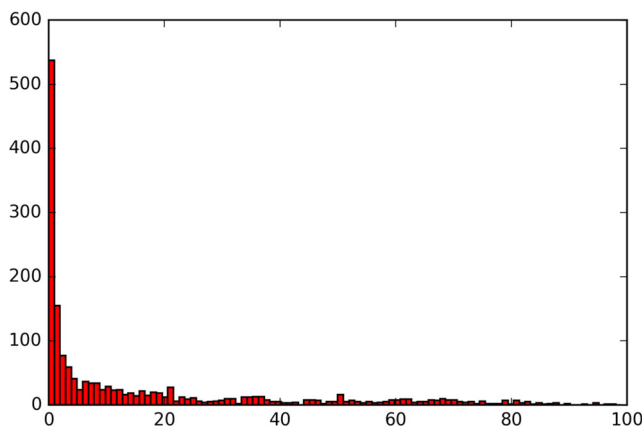


Fig. 6 California Quality Healthcare Accessibility Index Histogram

⁷ <http://quickfacts.census.gov/qfd/states/06000.html>, accessed on August 18, 2016

Table 4 Descriptive Statistics of Quality Healthcare Accessibility Index of California

Mean	Standard Error	Median	Mode	Standard Deviation	Sample Variance	Kurtosis	Skewness
15.86004	0.557535	3.964309	0.000907	22.81809	520.6654	1.433503	1.579
Range	Minimum	Maximum	Sum	Count	25 percentile	50 percentile	75 percentile
98.03157	0	98.032472	26,565.56	1675	0.000907	3.964309	21.459004

representative of the total population of the US and the results can be generalized to other states and medical facilities.

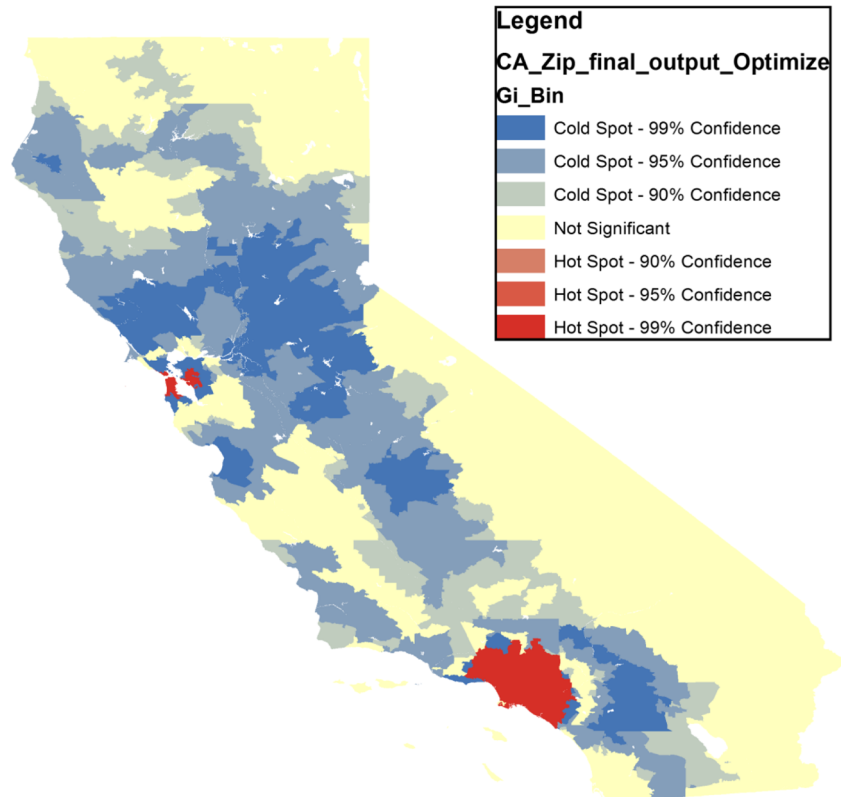
6.5 Future work

Although the proposed framework focuses on quality healthcare, the guidelines provided for measuring accessibility may also be extended to other domains. We hope that others would consider our artifact and employ it in a variety of disciplines such as education, finance, or public administration. Measuring accessibility in those fields is still scarce and further research is essential to support site analysis and distribution of facilities in needed areas. In addition, we suggest to include more variables and assign appropriate weights for the analysis. That can be facilitated through a Big Data approach incorporating much larger and more complex data sets in various fields.

7 Conclusion

In conclusion, measuring quality healthcare accessibility is of growing importance to the US government and it is vital for researchers and practitioners to use reliable data sets and be consistent in their analyses. Thus, the current study provides valuable aids for them to extend this work and measure quality healthcare accessibility in other counties and states. This would also have a positive effect on healthcare services offered and would provide more useful resources for managers when making decisions regarding new facilities. As a result of the Affordable Care Act, there is now Big Data on healthcare and it is important for researchers and practitioners to have a validated tool such as the proposed framework to process this data and further utilize it in various areas.

Fig. 7 Quality Healthcare Accessibility Index Hot Spot Analysis



References

- Aday, L. A., & Andersen, R. (1974). A framework for the study of access to medical care. *Health Services Research*, 9(3), 208.
- Barrett, M. A., Humblet, O., Hiatt, R. A., & Adler, N. E. (2013). Big data and disease prevention: from quantified self to quantified communities. *Big Data*, 1(3), 168–175.
- Canlas, R. (2009). *Data mining in healthcare: Current applications and issues*. School of Information Systems & Management: Carnegie Mellon University, Australia.
- d'Agostino, R. (1971). An omnibus test of normality for moderate and large size samples. *Biometrika*, 58(2), 341–348.
- d'Agostino, R., & Pearson, E. S. (1973). Tests for departure from normality. Empirical results for the distributions of b^2 and $\sqrt{b^1}$. *Biometrika*, 60(3), 613–622.
- Delamater, P. L. (2013). Spatial accessibility in suboptimally configured health care systems: a modified two-step floating catchment area (M2SFCA) metric. *Health & Place*, 24, 30–43.
- Dony, C. C., Delmelle, E. M., & Delmelle, E. C. (2015). Re-conceptualizing accessibility to parks in multi-modal cities: a variable-width floating catchment area (VFCA) method. *Landscape and Urban Planning*, 143, 90–99.
- ESRI (2016a). OD cost matrix analysis. <http://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/od-cost-matrix.htm>. Accessed 18 Aug 2016.
- ESRI (2016b). Optimized hot spot analysis. <http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/optimized-hot-spot-analysis.htm>. Accessed 20 Aug 2016.
- Fransen, K., Neutens, T., De Maeyer, P., & Deruyter, G. (2015). A commuter-based two-step floating catchment area method for measuring spatial accessibility of daycare centers. *Health & Place*, 32, 65–73.
- Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3), 255–261.
- Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The 'big data' revolution in healthcare. <http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care>. Accessed 16 Nov 2016.
- Guagliardo, M. F. (2004). Spatial accessibility of primary care: concepts, methods and challenges. *International Journal of Health Geographics*, 3(1), 3.
- Hevner, A., & Chatterjee, S. (2010). *Design research in information systems: theory and practice* (Vol. 22): Springer.
- Hevner, A., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- Higgs, G. (2004). A literature review of the use of GIS-based measures of access to health care services. *Health Services & Outcomes Research Methodology*, 5(2), 119–139.
- Jee, K., & Kim, G.-H. (2013). Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. *Healthcare informatics research*, 19(2), 79–85.
- Khan, A. A. (1992). An integrated approach to measuring potential spatial access to health care services. *Socio-Economic Planning Sciences*, 26(4), 275–287.
- Langford, M., Higgs, G., & Fry, R. (2016). Multi-modal two-step floating catchment area analysis of primary health care accessibility. *Health & Place*, 38, 70–81.
- Luo, J. (2014). Integrating the huff model and floating catchment area methods to analyze spatial access to healthcare services. *Transactions in GIS*, 18(3), 436–448.
- Luo, W., & Qi, Y. (2009). An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health & Place*, 15(4), 1100–1107.
- Luo, W., & Wang, F. (2003). Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. *Environment and Planning. B, Planning & Design*, 30, 865–884.
- Luo, W., & Whippo, T. (2012). Variable catchment sizes for the two-step floating catchment area (2SFCA) method. *Health & Place*, 18(4), 789–795.
- Mao, L., & Nekorchuk, D. (2013). Measuring spatial accessibility to healthcare for populations with multiple transportation modes. *Health & Place*, 24, 115–122.
- McGrail, M. R., & Humphreys, J. S. (2014). Measuring spatial accessibility to primary health care services: Utilising dynamic catchment sizes. *Applied Geography*, 54, 182–188.
- Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA*, 309(13), 1351–1352.
- Musa, G. J., Chiang, P.-H., Sylk, T., Bavley, R., Keating, W., Lakew, B., et al. (2013). Use of GIS mapping as a public health tool—from cholera to cancer. *Health Services Insights*, 6, 111.
- Ngui, A. N., & Apparicio, P. (2011). Optimizing the two-step floating catchment area method for measuring spatial accessibility to medical clinics in Montreal. *BMC Health Services Research*, 11(1), 166.
- Polzin, P., Borges, J., & Coelho, A. (2014). An extended kernel density two-step floating catchment area method to analyze access to health care. *Environment and Planning. B, Planning & Design*, 41(4), 717–735.
- Radke, J., & Mu, L. (2000). Spatial decompositions, modeling and mapping service regions to predict access to social programs. *Geographic Information Sciences*, 6(2), 105–112.
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3.
- Robinson, M., & Clegg, J. (2005). Improved determination of Q-factor and resonant frequency by a quadratic curve-fitting method. *IEEE Transactions on Electromagnetic Compatibility*, 47(2), 399–402.
- Shah, N., & Pathak, J. (2014). Why health care may finally Be ready for big data. *Harvard Business Review*.
- Siegel, M., Koller, D., Vogt, V., & Sundmacher, L. (2016). Developing a composite index of spatial accessibility across different health care sectors: a German example. *Health Policy*, 120(2), 205–212.
- Soliman, S., Christensen, G., & Rouhi, A. (1988). A new technique for curve fitting based on minimum absolute deviations. *Computational Statistics and Data Analysis*, 6(4), 341–351.
- Vo, A., Plachkinova, M., & Bhaskar, R. (2015). Assessing healthcare accessibility algorithms: A comprehensive investigation of two-step floating catchment methodologies family.
- Wan, N., Zou, B., & Sternberg, T. (2012). A three-step floating catchment area method for analyzing spatial access to health services. *International Journal of Geographical Information Science*, 26(6), 1073–1089.
- Wang, F., & Luo, W. (2005). Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. *Health & Place*, 11(2), 131–146.

Miloslava Plachkinova is an Assistant Professor of Cybersecurity in the Sykes College of Business at the University of Tampa, FL. She holds a PhD in Information Systems and Technology from Claremont Graduate University, CA. She is a Certified Information Security Manager (CISM) and a Project Management Professional (PMP). Dr. Plachkinova's research focuses on information security and healthcare. She investigates how human behavior leads to data breaches and her work in the healthcare field investigates security and privacy issues in mobile health (mHealth) and electronic health records (EHR) on the cloud. Dr. Plachkinova also has extensive industry experience working for both the private and the public sectors.

Au Vo is a PhD Candidate in Information Systems and Technology at the Claremont Graduate University. His current research spans across healthcare systems, spatial analytics, and data analytics. He is also an Associate Director for the Center for Information Technology and Business Analytics at California State University, Fullerton, assisting the Center with maintaining and expanding its relationships with the organizations in the Southern California regions. He is currently an adjunct lecturer at the University of La Verne and California State University, Fullerton. He also maintains his experience and relationships with the industry through various consulting projects.

Dr. Rahul Bhaskar is a Professor of Information Systems and Decision Sciences at the Steven G. Mihaylo College of Business and Economics, California State University, Fullerton. He holds a PhD in Management Information Systems from University of Madison, Wisconsin. Dr. Bhaskar's experience spans across diverse industries. He is experienced in revenue management, big data, data warehouse, and business intelligence. He has been a recipient of grants from the Department of

Homeland Security and other private institutions. His expertise is in data analytic, business intelligence, cyber terrorism, marketing information systems, and health care information systems.

Brian N. Hilton received a Ph.D. in Management Information Systems from Claremont Graduate University, School of Information Science, Claremont, CA, USA. Dr. Hilton's research and teaching focuses on advanced geographic information system (GIS) analysis, the research and development of advanced GIS solutions, and the use of emerging technologies in information system development. Dr. Hilton is the Director of the Advanced GIS Lab which stresses a transdisciplinary approach to knowledge, research, and problem-solving using core ideas, methods, and concepts from several disciplines to critically examine a broad range of real-world problems. Currently, the Lab and its associated students and faculty are examining the implementation of GIS technologies to improve community health, better understand road transportation safety, support humanitarian efforts, and examine ecosystem services.