

# Demand and Level of Service Inflation in Floating Catchment Area (FCA) Methods

--Manuscript Draft--

<b>Manuscript Number:</b>	PONE-D-18-23419R1
<b>Article Type:</b>	Research Article
<b>Full Title:</b>	Demand and Level of Service Inflation in Floating Catchment Area (FCA) Methods
<b>Short Title:</b>	Demand and Level of Service Inflation in Floating Catchment Area (FCA) Methods
<b>Corresponding Author:</b>	Antonio Paez McMaster University Hamilton, Ontario CANADA
<b>Keywords:</b>	accessibility; Health Care; Family Physicians; Floating Catchment Area; Two-Step Floating Catchment Area; 2SFCA; Enhanced 2-Step Floating Catchment Area; E2SFCA; Three-Step Floating Catchment Area; 3SFCA; Demand Inflation; Level of Service Inflation; Disparities; Health Geography; Hamilton CMA; Ontario; Canada
<b>Abstract:</b>	Floating Catchment Area (FCA) methods are a popular tool to investigate accessibility to public facilities, in particular health care services. FCA approaches are attractive because, unlike other accessibility measures, they take into account the potential for congestion of facilities. This is done by 1) considering the population within the catchment area of a facility to calculate a variable that measures level of service, and then 2) aggregating the level of service by population centers subject to catchment area constraints. In this paper we discuss an effect of FCA approaches, an artifact that we term demand and level of service _inflation_. These artifacts are present in previous implementations of FCA methods. We argue that inflation makes interpretation of estimates of accessibility difficult, which has possible deleterious consequences for decision making. Next, we propose a simple and intuitive approach to proportionally allocate demand and level of service in FCA calculations. The approach is based on a standardization of the impedance matrix, similar to approaches popular in the spatial statistics and econometrics literature. The result is a more intuitive measure of accessibility that 1) provides a local version of the provider-to-population ratio; and 2) preserves the level of demand and the level of supply in a system. We illustrate the relevant issues with some examples, and then empirically by means of a case study of accessibility to family physicians in the Hamilton Census Metropolitan Area (CMA), in Ontario, Canada. Results indicate that demand and supply inflation/deflation affect the interpretation of accessibility analysis using existing FCA methods, and that the proposed adjustment can lead to more intuitive results.
<b>Order of Authors:</b>	Antonio Paez  Christopher D. Higgins  Salvatore F. Vivona
<b>Opposed Reviewers:</b>	
<b>Response to Reviewers:</b>	See response to reviewers in the attached documents.
<b>Additional Information:</b>	
<b>Question</b>	<b>Response</b>
<b>Financial Disclosure</b>	AP and CH Social Sciences and Humanities Research Council of Canada Insight Development Grant 430-2017-00627. <a href="http://www.sshrc-crsh.gc.ca/home-accueil-eng.aspx">http://www.sshrc-crsh.gc.ca/home-accueil-eng.aspx</a> The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Enter a financial disclosure statement that describes the sources of funding for the work included in this submission. Review the <a href="#">submission guidelines</a> for detailed requirements. View published research articles from <a href="#">PLOS ONE</a> for specific	

examples.

This statement is required for submission and **will appear in the published article** if the submission is accepted. Please make sure it is accurate.

#### Unfunded studies

Enter: *The author(s) received no specific funding for this work.*

#### Funded studies

Enter a statement with the following details:

- Initials of the authors who received each award
- Grant numbers awarded to each author
- The full name of each funder
- URL of each funder website
- Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript?
- **NO** - Include this sentence at the end of your statement: *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*
- **YES** - Specify the role(s) played.

\* typeset

#### Competing Interests

The authors have declared that no competing interests exist.

Use the instructions below to enter a competing interest statement for this submission. On behalf of all authors,

disclose any [competing interests](#) that could be perceived to bias this work—acknowledging all financial support and any other relevant financial or non-financial competing interests.

This statement **will appear in the published article** if the submission is accepted. Please make sure it is accurate. View published research articles from [PLOS ONE](#) for specific examples.

#### NO authors have competing interests

Enter: *The authors have declared that no competing interests exist.*

#### Authors with competing interests

Enter competing interest details beginning with this statement:

*I have read the journal's policy and the authors of this manuscript have the following competing interests: [insert competing interests here]*

\* typeset

#### Ethics Statement

N/A

Enter an ethics statement for this submission. This statement is required if the study involved:

- Human participants
- Human specimens or tissue
- Vertebrate animals or cephalopods
- Vertebrate embryos or tissues

- Field research

Write "N/A" if the submission does not require an ethics statement.

General guidance is provided below.

Consult the [submission guidelines](#) for detailed instructions. **Make sure that all information entered here is included in the Methods section of the manuscript.**

## **Format for specific study types**

### **Human Subject Research (involving human participants and/or tissue)**

- Give the name of the institutional review board or ethics committee that approved the study
- Include the approval number and/or a statement indicating approval of this research
- Indicate the form of consent obtained (written/oral) or the reason that consent was not obtained (e.g. the data were analyzed anonymously)

### **Animal Research (involving vertebrate animals, embryos or tissues)**

- Provide the name of the Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board that reviewed the study protocol, and indicate whether they approved this research or granted a formal waiver of ethical approval
- Include an approval number if one was obtained
- If the study involved *non-human primates*, add *additional details* about animal welfare and steps taken to ameliorate suffering
- If anesthesia, euthanasia, or any kind of animal sacrifice is part of the study, include briefly which substances and/or methods were applied

## **Field Research**

Include the following details if this study involves the collection of plant, animal, or other materials from a natural setting:

- Field permit number
- Name of the institution or relevant body that granted permission

## **Data Availability**

Authors are required to make all data underlying the findings described fully available, without restriction, and from the time of publication. PLOS allows rare exceptions to address legal and ethical concerns. See the [PLOS Data Policy](#) and [FAQ](#) for detailed information.

Yes - all data are fully available without restriction

<p>A Data Availability Statement describing where the data can be found is required at submission. Your answers to this question constitute the Data Availability Statement and <b>will be published in the article</b>, if accepted.</p> <p><b>Important:</b> Stating 'data available on request from the author' is not sufficient. If your data are only available upon request, select 'No' for the first question and explain your exceptional situation in the text box.</p> <p>Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?</p>	
<p><b>Describe where the data may be found in full sentences. If you are copying our sample text, replace any instances of XXX with the appropriate details.</b></p> <ul style="list-style-type: none"> <li>• If the data are <b>held or will be held in a public repository</b>, include URLs, accession numbers or DOIs. If this information will only be available after acceptance, indicate this by ticking the box below. For example: <i>All XXX files are available from the XXX database (accession number(s) XXX, XXX.).</i></li> <li>• If the data are all contained <b>within the manuscript and/or Supporting Information files</b>, enter the following: <i>All relevant data are within the manuscript and its Supporting Information files.</i></li> <li>• If neither of these applies but you are able to provide <b>details of access elsewhere</b>, with or without limitations, please do so. For example:</li> </ul> <p><i>Data cannot be shared publicly because of [XXX]. Data are available from the XXX Institutional Data Access / Ethics Committee (contact via XXX) for researchers who meet the criteria for access to confidential data.</i></p> <p><i>The data underlying the results presented in the study are available from (include the name of the third party</i></p>	<p>All files for this manuscript are available from:  <a href="https://github.com/paezha/Demand-and-Supply-Inflation-in-Floating-Catchment-Area-FCA-Methods-">https://github.com/paezha/Demand-and-Supply-Inflation-in-Floating-Catchment-Area-FCA-Methods-</a></p>

*(and contact information or URL).*

- This text is appropriate if the data are owned by a third party and authors do not have permission to share the data.

\* typeset

Additional data availability information:

April 5, 2019

To the Editor:

Please find enclosed the revised version of our jointly authored paper “Demand and Level of Service Inflation in Floating Catchment Areas (FCA) Methods” that you kindly invited us to revise and resubmit for publication consideration to PLoS ONE.

We would like to begin by thanking two reviewers for their constructive comments and suggestions. In this letter we describe the actions taken in response to their feedback. You will notice that the paper underwent a major rewrite, as we endeavored to improve the paper in this version.

We thank you for the attention paid to this submission and look forward to hearing back from you in due course.

Sincerely,

Antonio Páez

(on behalf of the coauthors)

# Demand and Level of Service Inflation in Floating Catchment Area (FCA) Methods

Antonio Paez <sup>1</sup> \*, Christopher D. Higgins <sup>2</sup>, Salvatore F. Vivona <sup>3</sup>

**1** School of Geography and Earth Sciences, McMaster University, 1280 Main St W, Hamilton, ON L8S 4K1 Canada

**2** Department of Land Surveying and Geo-Informatics & Department of Building and Real Estate, 11 Yuk Choi Rd, Hung Hom, Hong Kong

**3** Department of Computer Science, University of Toronto, 214 College Street, Toronto, ON, M5T 3A1

\* Corresponding author: paezha@mcmaster.ca

## Abstract

Floating Catchment Area (FCA) methods are a popular tool to investigate accessibility to public facilities, in particular health care services. FCA approaches are attractive because, unlike other accessibility measures, they take into account the potential for congestion of facilities. This is done by 1) considering the population within the catchment area of a facility to calculate a variable that measures level of service, and then 2) aggregating the level of service by population centers subject to catchment area constraints. In this paper we discuss an effect of FCA approaches, an artifact that we term demand and level of service *inflation*. These artifacts are present in previous implementations of FCA methods. We argue that inflation makes interpretation of estimates of accessibility difficult, which has possible deleterious consequences for decision making. Next, we propose a simple and intuitive approach to proportionally allocate demand and level of service in FCA calculations. The approach is based on a standardization of the impedance matrix, similar to approaches popular in the spatial statistics and econometrics literature. The result is a more intiuitive measure of accessibility that 1) provides a local version of the provider-to-population ratio; and 2) preserves the level of demand and the level of supply in a system. We illustrate the relevant issues with some examples, and then empirically by means of a case study of accessibility to family physicians in the Hamilton Census Metropolitan Area (CMA), in Ontario, Canada. Results indicate that demand and supply inflation/deflation affect the interpretation of accessibility analysis using existing FCA methods, and that the proposed adjustment can lead to more intutive results.

## Introduction

An important issue in health geography and health policy is the evaluation of accessibility to healthcare services, with hundreds of research papers published on the topic since the 2000s [1]. However, the concept of accessibility is multi-dimensional, which often presents challenges to its operationalization in empirical research.

According to Joseph and Bantock [2], accessibility can be defined by both aspatial and spatial dimensions. The first dimension considers factors such as the quality of the services and their cost, as well as the income, social class, ethnicity, and mobility profile

of potential users of services. From a geographical perspective, the spatial dimension is key, and considers the distribution of available healthcare services across the landscape, in addition to the cost or friction that potential users incur when trying to reach these services. By taking these geographical factors into account, estimates of accessibility can help researchers, planners, and policy makers identify areas with high or low accessibility to healthcare services. This, in turn, can provide valuable information related to social and spatial inequalities and guidance for health policy and resource allocation.

Spatial accessibility can be estimated in various ways. At a high level, provider-to-population ratios (PPR) offer an indication of the level of service within a community. These measures conceptualize a region as a container of population and services, and therefore are sometimes called container approaches. PPRs are straightforward to interpret as the supply of a service (say number of doctors, beds, etc.) divided by demand (say, number of people who require the service). Despite this convenient and intuitive interpretation, container approaches are limited in the amount of spatial information that they provide, especially if applied to large regions. When applied to smaller regions these approaches present other shortcomings, such as the assumption that the population in the container is captive and does not cross the boundaries of the container in search of services - and that users do not come into the container from other regions to avail themselves of local services.

An alternative to container approaches is provided by gravity measures. Gravity measures offer a more sophisticated approach to measuring spatial accessibility to healthcare [2] that moreover addresses some of the limitations of the container approach. Instead of defining rigid container boundaries, gravity measures consider the mobility characteristics of the public to produce flexible (and often overlapping) catchment areas for both services and population. Accordingly, one of the most popular approaches for estimating healthcare accessibility in the literature is the Two-Step Floating Catchment Area (2SFCA) method proposed by Luo and Wang [3] after research by Radke and Mu [4]. The 2SFCA method is an ensemble of two gravity models with a simplified binary distance function to account for crowding of facilities and allocation of levels of service. Numerous applications of this methods are found in the international literature, including work from Germany [5], South Korea [6], Japan [7], China [8], Australia [9], and Canada [10].

Accessibility to healthcare is estimated in two stages in the 2SFCA: in the first step, a level of service at a given healthcare provider is determined based on the supply (e.g., number of physicians in a clinic) and the estimated demand from the surrounding population within some catchment area. This level of service resembles a local provider-to-population ratio (PPR). In the second step, the level of service of different healthcare providers is aggregated for each population center. By operationalizing accessibility in terms of demand and level of service, the 2SFCA method is appealing for health policy analysis. Still, several improvements have been proposed that seek to address the method's most important perceived shortcomings. The result is a family of Floating Catchment Areas (FCA) methods that include more realistic conceptualizations of the friction of distance by specifying variable catchment area sizes [9] and/or the use of stepped [11], continuous [12], and adaptive [13] distance-decay functions. Other authors have added multi-modal transportation [14], age-adjusted healthcare demand profiles [15], as well as ways to counteract the modifiable areal unit problem [16].

A major focus of FCA research, in addition to the improvements mentioned above, has been the introduction of competition for available opportunities or the allocation of services to the population. More concretely, the original 2SFCA approach has been criticized for over-estimating the levels of demand [17] and/or level of service [18] in the system. This is a consequence of the way catchment areas for facilities and population centers typically overlap in any realistic spatial system - an artifact of FCA methods

that can lead to misleading estimates of accessibility.

In effect, when aggregating the population within the overlapping catchment areas of multiple facilities, the original 2SFCA framework leads to double-counting of the population that tends to inflate the level of demand at supply points in the healthcare system. We call this effect *demand inflation*. Inflated demand, in turn, tends to *deflate* the level of service for populations serviced by the facilities so affected. A similar effect, which we call *level of service inflation*, happens when the levels of service of various service points are aggregated for population centers. Ultimately, accessibility estimates are affected in potentially complex ways, depending on the geography of the problem [18], and their interpretation as PPRs becomes suspect.

Various solutions to the issues of demand and level of service inflation have been proposed, including the addition of selection weights based on a travel impedance function in the Three-Step Floating Catchment Area (3SFCA) method [17]; the use of a Huff model to generate probability-based estimate of the selection weights in the 3SFCA method [19]; and, on the supply side, a modified 2SFCA (M2SFCA) method to address suboptimal spatial configuration of services [18].

In this paper we are interested in the way demand and level of service are calculated in FCA methods. We review how different approaches deal with the issue of inflation, and then propose a simple and intuitive approach to proportionally allocate supply and demand. Our solution consists on adjusting the impedance weights used in the estimation of FCA methods. More concretely, by incorporating methods drawn from the field of spatial statistics and econometrics, proportional allocation has the feature that it preserves the levels of demand and service in the system. To illustrate the key aspects of our proposal, we conduct a case study of access to family physicians in Hamilton, Canada. Our results indicate that the proposed adjustments produce more intuitive measures of accessibility to healthcare measured in terms of local PPRs. Moreover, these outputs can be used to provide estimates of access disparity across a region that are both easily understood and robust to demand and level of service inflation.

## Background: Floating Catchment Area Methods

To motivate the discussion to follow we begin by reviewing some popular FCA methods. In general terms, FCA approaches are implemented as ensembles of two gravity models in two steps, using an impedance function to represent the cost required to overcome distance. Impedance functions implement a distance-decay effect that mimics a commonly observed cost-minimization behavior, namely that people in general prefer to spend less time/money/effort travelling to destinations. In this way, the impedance function defines a *catchment area* for the points of service and population centers alike.

In the first step of FCA methods, the impedance function defines catchment areas for facilities  $j$ , which could be clinics, parks, libraries, etc. A weighted sum of the population within a catchment area is allocated to the corresponding facility or service point to represent demand. In the second step of the algorithm, the catchment areas are “floated” to population centers  $i$ . Accessibility at location  $i$  is calculated as the weighted sum of the level of service at every location  $j$  that includes  $i$  within its catchment area. The following methods are popular in the literature.

### Two-Stage Floating Catchment Areas (2SFCA)

The original 2SFCA implements a binary impedance function  $W$  with a threshold cost  $d_0$  as follows [3]:

$$W(d_{ij} \leq d_0) = \begin{cases} 1 & d_{ij} \leq d_0 \\ 0 & d_{ij} > d_0 \end{cases}$$

This function assumes equal potential within a catchment area (i.e.,  $d_{ij} \leq d_0$ ), and zero beyond ( $d_{ij} > d_0$ ). This implies that 1) travellers are equally likely users of a service point within the catchment area, irrespective of how proximate or distant they are from it; and 2) no users travel to the service point from beyond the threshold cost.

Given the impedance function, the level of demand  $D_j$  is calculated as the weighted sum of the population at  $i$ :

$$D_j = \sum_i D_{ij} = \sum_i P_i W(d_{ij} \leq d_0)$$

The supply  $S$  of the service offered at location  $j$  (say, number of beds/doctors in a clinic) is then divided by the demand to obtain a measure of level of service (e.g., beds/person, sq.m of park space/person, library floor space/person). This gives a level of service  $L_j$  at the service point:

$$L_j = \frac{S_j}{D_j} = \frac{S_j}{\sum_i D_{ij}} = \sum_i \frac{S_j}{D_{ij}} = \sum_i L_{ij}$$

The level of service resembles a PPR. Aggregation of demand creates a congestion effect that depends on the number of potential users from different origins  $i$  that converge at service point  $j$ : at a fixed level of supply, greater demand results in lower levels of service. The different decompositions of  $L_j$  help to understand how different population centers contribute to the level of demand at facility  $j$ .

In the second step of the algorithm, catchment areas are “floated” to population centers  $i$ . A second gravity model is used to calculate the accessibility at  $i$ :

$$A_i = \sum_j L_j W(d_{ij} \leq d_0)$$

Since accessibility is calculated as the weighted sum of the level of service at facilities, it is conventionally interpreted as a PPR.

## Enhanced Two-Stage Floating Catchment Areas (E2SFCA)

A criticism of the binary impedance function of the 2SFCA is that it does not account for the declining probability of using a facility as distance grows. As a result of this criticism, other impedance functions have since been proposed, including the stepwise formulation of the Enhanced Two-Stage Floating Catchment Area method [11]:

$$W(d_{ij}|d_1, d_2, \dots, d_R) = \begin{cases} k_1 & d_{ij} \leq d_1 \\ k_2 & d_1 < d_{ij} \leq d_2 \\ \dots & \\ k_{R-1} & d_{R-1} < d_{ij} \leq d_R \\ 0 & d_{ij} > d_R \end{cases}$$

A stepwise function does not assume identical potential within the catchment area (i.e., the space contained within  $d_{ij} \leq d_R$ ), but rather declining potential with increasing cost of travel. It is worthwhile noting that impedance functions have long been studied in geographical analysis in general [20], and accessibility research in particular [21]. However, it is only relatively recently that alternative impedance functions have been incorporated in FCA approaches, including continuous functions [12] and mixtures of continuous and step functions [22].

Besides the use of a non-binary impedance function, the method remains the same. In the first step, demand is calculated as a weighted sum of the population within the catchment area:

$$D_j = \sum_i D_{ij} = \sum_i P_i W(d_{ij}|d_1, d_2, \dots, d_R)$$

Note that non-binary impedance functions discount the level of demand as a function of cost more rapidly than binary functions. How rapidly this happens depends on the definition of the cutoff values  $d_1, d_2, \dots, d_R$  and weights  $k_1, k_2, \dots, k_{r-1}$  of the function.

In the second step of the algorithm, accessibility at  $i$  is calculated as the weighted sum of the level of service of service points  $j$ :

$$A_i = \sum_j \frac{S_j}{D_j} W(d_{ij}|d_1, d_2, \dots, d_R) = \sum_j L_j W(d_{ij}|d_1, d_2, \dots, d_R)$$

Again, the use of a non-binary impedance function discounts the level of service more rapidly compared to binary functions.

### Three-Stage Floating Catchment Areas (3STCA)

Wan et al. [17] proposed a Three-Stage Floating Catchment Area method (3SFCA) that aims at refining the estimates of level of demand and accessibility by means of the use of *selection weights*. This approach operates by introducing an additional step where selection weights are calculated as follows:

$$G_{ij} = \frac{T(d_{ij})}{\sum_{j \forall d_{ij} \leq d_0} T(d_{ij})}$$

where  $T(d_{ij})$  are Gaussian weights (essentially an impedance function), and the summation in the denominator is for all sites  $j$  that are within a critical threshold  $d_0$ . Notice that a property of the selection weights is that their sum over  $j$  equals one:

$$\sum_j G_{ij} = 1$$

Given a set of selection weights, the level of demand is calculated by this algorithm in the following manner:

$$D_j^* = \sum_i G_{ij} P_i W(d_{ij}|d_1, d_2, \dots, d_R) = \sum_i G_{ij} D_{ij}$$

Notice how demand in this method takes what is essentially the demand in the E2SFCA, and allocates it proportionally to service points  $j$ .

Accessibility, in the final step, becomes (with the subindices of the selection weights reversed, to reflect the displacement of the catchment area to population centers) is calculated in the following manner:

$$A_i^* = \sum_j G_{ji} \frac{S_j}{D_{ij}^*} W(d_{ij}|d_1, d_2, \dots, d_R) = \sum_j G_{ji} L_j^* W(d_{ij}|d_1, d_2, \dots, d_R)$$

### Modified Two-Stage Floating Catchment Areas (M2SFCA)

Delamater [18] discusses the application of FCA methods for systems that are not optimally configured to service the whole population. To address this issue, he proposes a modification to the second step of the 2SFCA algorithm that increases the friction of distance. Demand in this modification is the same as in 2SFCA. However, accessibility is calculated in the following manner:

$$A_i = \sum_j L_j W(d_{ij}|d_1, d_2, \dots, d_R) W(d_{ij}|d_1, d_2, \dots, d_R) = \sum_j L_j (W(d_{ij}|d_1, d_2, \dots, d_R))^2$$

In other words, the level of service is discounted by the square of the impedance function, thus increasing the rate of decay. This is done to reflect the possibility that some population centers may experience increased friction to reach destinations in suboptimally configured systems.

## Inflation Effects in FCA Methods

Having reviewed a selection of FCA approaches, we now proceed to discuss the issue of inflation. Inflation has been identified, among others, by Wan et al. [17] and Delamater [18]. As discussed by these authors, inflation happens when demand or level of service are overestimated. Inflation is a consequence of the way in which  $D_{ij}$  and  $A_i$  are calculated, with some population centers contributing to the level of demand at more than one facility and then the level of service of facilities allocated to multiple population centers. Calculating demand, in particular, generally fails to preserve the population, and therefore lacks the pycnophilactic property discussed by Tobler [23]. In practical terms, this implies that the population used to calculate the demand component of level of service will often exceed (but sometimes fall short of) the actual population in a region, depending on the weighting scheme. We term the consequent effect *demand inflation*.

Let us illustrate this inflation effect by means of a simple example using the conventional 2SFCA approach with a binary impedance function. In this case, the population value at  $i$  is multiplied by zero or one, meaning that the contribution of  $i$  to demand at  $j$  whenever  $d_{ij}$  does not exceed the threshold is:

$$D_{ij} = P_i$$

If we concentrate for a moment on a single population center that enters the catchment areas of several service points (see Fig 1, left panel), we can see that when the demand at each of the service points is calculated, the population in question is added two times, and the levels of service are  $L_1 = L_2 = 1/s00$ .

More generally, when calculating the level of service at  $L_j$ , the population at  $i$  contributes to demand every time that  $d_{ij} \leq d_0$  for any  $j$ . And, since since  $D_{ij} = P_i$ , it follows that the sum of the population to be serviced over all clinics is:

$$\sum_j D_{ij} = K_i P_i$$

where  $K_i$  is the number of service points  $j$  that include  $i$  as part of their catchment areas. Therefore, the system-wide contribution of the population at  $i$  to the level of demand implied by these calculations, vastly exceeds the actual population at  $i$ , since:

$$\sum_j D_{ij} = K_i P_i > P_i$$

Let us consider next what happens when enhanced (i.e., non-binary) impedance weights are used. These functions aim to capture more realistically the rule that most members of the population prefer to travel shorter distances to reach a destination. For the example, assume a set of weights with decay as follows (see Fig 1, right panel) :

$$W(d_{ij}|d_1, d_2, \dots, d_R) = \begin{cases} 0.9 & d_{ij} \leq d_1 \\ 0.8 & d_1 < d_{ij} \leq d_2 \\ 0.4 & d_{R-1} < d_{ij} \leq d_R \\ 0 & d_{ij} > d_R \end{cases}$$

The population center in the example is relatively distant from the service points. Accordingly, its potential demand is reduced by assuming that some people do not travel at all. In this example, the contribution of the population center to demand is only  $0.8P$  to each clinic, and therefore the system-wide demand of this center is  $1.6P$  - less than the all-or-nothing allocation of the binary impedance weights, but still in excess of the actual population.

More generally, when calculating the level of service at  $j$  locations, the population at  $i$  contributes to demand every time that  $d_{ij}$  is within the service area for any  $j$ . The precise contribution depends on the weights in the distance-decay function and the position of the population center with relative to all service points. In a function with faster decay, the total demand attributed to  $i$  (i.e.,  $\sum_i D_{ij}$ ) can be less than the population of  $i$ . In other words, depending on the steepness of decay, the total demand can be greater than, equal to, or less than the population at  $i$ :

$$\sum_j D_{ij} \leq P_i$$

Clearly, only when the full population at  $i$  is allocated exclusively to one service point (i.e., when  $K_i = 1$ ) the implied demand equals the population - something that seldom happens in practical situations.

It is important to acknowledge that demand in accessibility analysis represents the potential for spatial interaction, not realized interaction. That said, the expectation that facilities need to serve multiple times the size of the population in a region can easily lead to misleading conclusions about the need for resources. A logical question, however, is whether the inflation of demand (with the consequence deflation of level of service) is not offset in the second step of the method, when the population at  $i$  has potential access to multiple service points?

Let us consider what happens in the second step of the algorithm in the example, when catchment areas are floated to the population center (see Fig 2). When a binary impedance function is used, the aggregation of the level of service means that, despite the inflation of demand due to double-counting, accessibility matches the level of service as well as the regional PPR of 2/100 (left panel). In the case of the stepwise function, the level of implied demand is less than the population, but the population is also assumed to receive less of the available level of service. In this case, again, the accessibility matches the level of service despite the fact that segments of the population were assumed to not contribute to demand.

Clearly, the example is too simplistic (in fact just a variation of the container approach), and it is unclear what the implications would be for a system with even just a slightly more complex geography. To explore this, consider the addition of two population centers to the landscape (see Fig 3). Notice how the three population centers are in the catchment areas of the two clinics. When the binary impedance function is used, demand at each clinic is calculated as 300, and demand over all clinics is therefore 600, or twice the population of the region. When the stepwise impedance function is used, the demand by each center is:

$$\begin{aligned} D_{1j} &= 0.8 \times 100 + 0.8 \times 100 = 160 \\ D_{2j} &= 0.8 \times 100 + 0.4 \times 100 = 120 \\ D_{3j} &= 0.4 \times 100 + 0.8 \times 100 = 120 \end{aligned}$$

and the total load on the system is therefore 400, still well in excess of the total population of the region.

When demand is used to calculate the level of service, and then accessibility in the second step of the algorithm, the following occurs (see Fig 4). When the binary impedance function is used (left panel), the level of service at each clinic is:

$$\begin{aligned} L_1 &= \frac{10}{300} = 0.033 \\ L_2 &= \frac{10}{300} = 0.033 \end{aligned}$$

The level of service at the clinics is only half of the regional PPR, since each clinic is assumed to serve the entire population of the region. Unfortunately, since demand has

been inflated for each clinic, these levels of service cannot be meaningfully interpreted as local PPRs. The sum over the clinics, on the other hand, is 20/300 - which is consistent with the regional PPR. Interestingly, as seen in the figure, the accessibility of each population center matches the regional PPR - but the sum of accessibility over all population centers exceeds the sum of the level of service over all the clinics as a consequence of allocating the same level of service to several population centers.

Continuing with the stepwise impedance function, we can see (Fig 4, right panel) that the levels of service are calculated as:

$$L_1 = \frac{10}{0.8 \times 100 + 0.8 \times 100 + 0.4 \times 100} = \frac{10}{200} = 0.05$$

$$L_2 = \frac{10}{0.8 \times 100 + 0.4 \times 100 + 0.8 \times 100} = \frac{10}{200} = 0.05$$

Notice how the level of service is higher in this case: this is a consequence of assuming (as the stepwise impedance function does) that some of the population does *not* demand service. Demand, however, is still inflated, and interpretation of the levels of service as local PPRs is still inappropriate. Accessibility is higher for population center 1 but lower for the two peripheral centers. Furthermore, the sum of accessibility over all population centers exceeds the sum of the level of service of all clinics in the region.

At issue is the interpretability of the levels of service, which as the example illustrates do not accurately represent PPRs, and how accessibility, which is a weighted sum of levels of service, cannot be interpreted as the PPR for a population center either.

Two methods reviewed above, namely the Three-Stage Floating Catchment Area method and the Modified Two-Stage Floating Catchment Area method aim to address the overestimation of demand and/or levels of service when calculating accessibility. As discussed previously, they do this by compounding the effect of the impedance function. In the case of 3SFCA, demand is deflated by assuming that demand declines more rapidly with distance. Then, when calculating accessibility, the levels of service are allocated more locally, again, as a consequence of steeper distance-decay. In the case of M2SFCA, demand is not deflated, however, the levels of service are allocated more locally as a consequence of steeper distance-decay. In other words, these methods correct for inflation by assuming that *fewer* people demand health care services, and that the levels of service are allocated to fewer people too.

For comparison, the levels of service and accessibility for the example according to these two methods are shown in Fig 5. Notice how the levels of service in the 3FSCA are considerably higher as a consequence of excluding potential users with a steeper rate of decay. On the other hand, the levels of accessibility are also lower, as a consequence of allocating service more locally. The levels of service in the M2SFCA are identical to the E2SFCA, however, accessibility is lower, again as a result of allocating service more locally.

## A Simulated Example

The examples in the preceding section illustrate the way demand and level of service can be overestimated (and in some cases underestimated) in FCA algorithms. However, they are too simplistic to indicate what would happen in a realistic situation. In particular, it is possible that the consequences depend on the geography of the problem as the examples in Delamater [18] suggest. Based on the way demand and level of service are allocated, we conjecture that the effects are likely more pronounced in areas with higher density of population and service, since inflation is a consequence of overlapping catchment areas. Furthermore, we conjecture that demand inflation will be reduced when stepwise/continuous distance-decay functions are used, since their effect is to reduce the overlap by reducing the contribution of population at different distances,

**Table 1.** Impedance Matrices and Step 1 of the FCA Algorithm

Population Center	Binary Impedance			Stepwise Impedance		
	Clinic 1	Clinic 2	Clinic 3	Clinic 1	Clinic 2	Clinic 3
Population 1	1	0	0	0.242	0.000	0.000
Population 2	1	1	0	0.600	0.242	0.000
Population 3	1	1	1	0.242	0.600	0.242
Population 4	1	1	1	0.600	0.600	0.242
Population 5	0	1	1	0.000	0.600	0.945
Population 6	0	1	1	0.000	0.242	0.600
Population 7	0	1	1	0.000	0.242	0.242
Population 8	0	1	1	0.000	0.600	0.242
Population 9	0	1	1	0.000	0.242	0.242

and to allocate levels of service more locally as well. We explore these issues further by means of a simple but realistic simulated example.

The setup for the simulated example is shown in Fig 6. There are three clinics and nine population centers. Assume that the supply at the three clinics is one physician at clinic 1, three physicians at clinic 2, and two physicians at clinic 3. Further, assume that the population at 1, 2, 8, and 9 is 250; population at 3, 4, and 6 is 250; and population at 5 and 7 is 1000. The total population in the region therefore is 4,500. Under this setup, the level of service across the whole system is 1.33 physicians per thousand people, which we will refer to as the Regional PPR.

For this experiment, we consider binary and stepwise impedance functions. The former is simply the traditional 2SFCA method, whereas the latter is the Enhanced 2SFCA approach. The catchment areas for the first step of the algorithm (demand allocation) are shown in Fig 7 (binary impedance) and Fig 8 (stepwise impedance). Notice that some population centers are inside the catchment areas of more than one clinic. For instance, Population Center 5 is in the catchment areas of Clinics 2 and 3, whereas Population Center 4 is in the catchment areas of all three clinics.

To see how the overlap of catchment areas impacts the calculations in the first step of the algorithm, we define impedance matrices using the same criteria as for the buffers seen in Fig ???. These matrices are shown in Table 1.

The demand for each clinic is calculated as the population of the centers multiplied by the values of the corresponding impedance weight with respect to that clinic, and then aggregated for all population centers. The level of service is the supply divided by the demand, multiplied by 1,000. The last row of the table shows the total population as well as the total demand at each clinic.

**Table 2.** Disaggregated Demand Allocations by Population Center and Clinic, and level of Service by Clinic

Population Center	Population	Binary Impedance			Stepwise Impedance			3SFCA			M2SFCA		
		Clinic 1	Clinic 2	Clinic 3	Clinic 1	Clinic 2	Clinic 3	Clinic 1	Clinic 2	Clinic 3	Clinic 1	Clinic 2	Clinic 3
Population 1	250	250	0	0	60.5	0	0	60.5	0	0	60.5	0	0
Population 2	250	250	250	0	150	60.5	0	106.89	17.388	0	150	60.5	0
Population 3	500	500	500	500	121	300	121	27.013	166.05	27.013	121	300	121
Population 4	500	500	500	500	300	300	121	124.83	124.83	20.307	300	300	121
Population 5	1000	0	1000	1000	0	600	945	0	233.01	578.01	0	600	945
Population 6	500	0	500	500	0	121	300	0	34.777	213.78	0	121	300
Population 7	1000	0	1000	1000	0	242	242	0	121	121	0	242	242
Population 8	250	0	250	250	0	150	60.5	0	106.89	17.388	0	150	60.5
Population 9	250	0	250	250	0	60.5	60.5	0	30.25	30.25	0	60.5	60.5
Total Population/Demand	4500	1500	4250	4000	631.5	1834	1850	319.228	834.191	1007.744	631.5	1834	1850
Supply	NA	1	3	2	1	3	2	1	3	2	1	3	2
Level of Service (per 1,000)	NA	0.667	0.706	0.5	1.584	1.636	1.081	3.133	3.596	1.985	1.584	1.636	1.081

Note:

Darker number colors in a row indicate a greater allocation of demand.

First we discuss the results according to the binary impedance function. As seen in Table 2, the population of Center 3 (which is in the catchment area of three clinics) is assumed to contribute 1,500 patients to the demand across the system, whereas Center 1 (which is in the catchment area of only one clinic) contributes exactly its population of 250. Since the population of several centers is counted multiple times, the apparent demand exceeds the actual population. In effect, when we calculate the total demand (the sum of the demand across clinics), we find that this is 9,750 according to the binary impedance function, which far exceeds the actual population.

Turning now to the stepwise function, we see that Center 3 contributes  $500 \times 0.242 + 500 \times 0.600 + 500 \times 0.242 = 542$  to the demand across the system, but Center 1 contributes only  $250 \times 0.242 = 60.5$ . The total demand now is 4,316, which is less than the total population.

This example illustrates a vexing effect in how FCA methods operate: when multiple service points are within the threshold travel cost of a population center, it is assumed that some (and possibly all) of the same persons crowd more than one service point, resulting in inflated demand and deflated levels of service. On the other hand, when stepwise or continuous functions (e.g., E2SFCA) are used to weigh down the population of distant population centers, the apparent effect is that some segments of the population do *not* demand service, even when clinics are within their threshold travel cost. This effect is even more marked in the case of 3SFCA, which produces considerably higher levels of service, as a consequence of stacking the effects of two impedance functions. In effect, demand is deflated and the level of service is inflated. While the assumption that some members of the population drop out from the total demand pool may be acceptable for discretionary services, it is suspect when it comes to essential services such as many health care services, and particularly primary health care.

Recall as well that the Regional Average PPR in this example is 1.33 physicians per thousand. If the total implied demand according to the binary impedance function is 9,750 the corresponding PPR is 0.615 physicians per thousand, or about half of the regional ratio. The corresponding PPR for the stepwise impedance function (implied demand = 4315.5) is 1.39 physicians per thousand, much closer to the Regional Average PPR. However, this PPR is misleading in that it assumes that some segments of the population are served multiple times, and some are not served at all.

Clearly, the first step of the algorithm can lead to inflation or deflation of the levels of demand. But do these matter? Or do they somehow average out when the levels of service are aggregated in the second step of the algorithm? Again, the situation is not clear-cut when multiple population centers and/or service clinics interact through overlapping catchment areas.

To illustrate this, we proceed to estimate the accessibility for the example using the binary and the stepwise impedance matrices. The results appear in Table 3.

**Table 3.** Disaggregated Level of Service Allocations by Clinic and Population Center, and Accessibility by Population Center

Clinic	Supply	Demand	Level of Service	Center 1	Center 2	Center 3	Center 4	Center 5	Center 6	Center 7	Center 8	Center 9
<b>Binary Impedance</b>												
1	1	1500.000		0.667	0.667	0.667	0.667	0	0	0	0	0
2	3	4250.000		0.706	0	0.706	0.706	0.706	0.706	0.706	0.706	0.706
3	2	4000.000		0.500	0	0	0.5	0.5	0.5	0.5	0.5	0.5
Accessibility	NA	NA		NA	0.667	1.37	1.87	1.87	1.21	1.21	1.21	1.21
<b>Stepwise Impedance</b>												
1	1	631.500		1.584	0.383	0.95	0.383	0.95	0	0	0	0
2	3	1834.000		1.636	0	0.396	0.981	0.981	0.981	0.396	0.396	0.981
3	2	1850.000		1.081	0	0	0.262	0.262	1.02	0.649	0.262	0.262
Accessibility	NA	NA		NA	0.383	1.35	1.63	2.19	2	1.04	0.657	1.24
<b>3SFCA</b>												
1	1	319.228		3.133	0.758	1.88	0.758	1.88	0	0	0	0
2	3	834.191		3.596	0	0.87	2.16	2.16	2.16	0.87	0.87	2.16
3	2	1007.744		1.985	0	0	0.48	0.48	1.88	1.19	0.48	0.48
Accessibility	NA	NA		NA	0.0551	0.367	0.253	0.536	0.525	0.17	0.0514	0.198
<b>M2SFCA</b>												
1	1	631.500		1.584	0.383	0.95	0.383	0.95	0	0	0	0
2	3	1834.000		1.636	0	0.396	0.981	0.981	0.981	0.396	0.396	0.981
3	2	1850.000		1.081	0	0	0.262	0.262	1.02	0.649	0.262	0.262
Accessibility	NA	NA		NA	0.0927	0.666	0.745	1.22	1.55	0.485	0.159	0.652

*Note:*

Darker number colors in a column indicate a greater allocation of level of service or accessibility.

Accessibility in the table is calculated as the level of service of the clinics multiplied by the values of the impedance function with respect to a population center, and then aggregated for all clinics. As seen in the table, the levels of accessibility vary considerably depending on the method. As anticipated, use of non-binary impedance functions reduces the inflation effect, and can even lead to deflation. Consider for instance the case of the binary impedance matrix: the total level of service in the system is the sum of the level of service at the three clinics, or 1.87. The level of service allocated to population centers, on the other hand, is the sum of the accessibility in the system, or 11.8. When using the stepwise impedance function, the total level of service in the system is 4.3, and the level of service allocated to population centers is 11.2. Compare this to the case of 3SFCA, where the total level of service in the system is 8.71, but the level of service allocated to population centers is only 2.21; or the case of M2SFCA, which estimates the total level of service in the system as 4.3 (same as E2SFCA) but allocates 5.74 to population centers.

Clearly, all the methods give qualitatively similar results, with peripheral centers displaying lower accessibility and more central places higher. But there are important differences in how demand and level of service are allocated throughout the system to calculate accessibility. Figure 9 shows how the different methods penalize peripheral centers at different rates. And, since the demand is not consistent with the population and the accessibility is not consistent with the level of service of the clinics, it is difficult to interpret the results in terms PPRs. For instance, when we inspect the results for the binary impedance matrix (2SFCA), we can see in the table that the accessibility of Population Center 1 is simply the level of service of Clinic 1. But, as we saw before, this level of service was deflated by double counting the population of Centers 2, 3, and 4, which contribute to the calculation of demand at multiple clinics. Things become more complex as the number of overlapping catchment areas grows. For example, Population Center 2 contributed to the congestion effect of two clinics. However, demand at one of those clinics was calculated using the population of eight out of nine population centers. What this suggests is that, at the very least, some population centers (likely those in the periphery of regions) will have artificially low accessibility levels as a consequence of demand inflation.

## A Method for Proportional Allocation of Demand and Supply

As the examples in the preceding subsection illustrate, FCA methods can induce quite substantial inflation (or deflation) of demand and level of service. This, in turn, can affect the estimates of accessibility in potentially complex ways. The results, furthermore lack a clear interpretation. In this section, we propose a simple and intuitive adjustment to avoid the inflation artifacts inherent in current implementations of FCA methods.

Refer again to Fig 1. Demand inflation occurs because of the overlap in catchment areas - with the underlying assumption that a service location services the population within its catchment area. More realistically, only a fraction of that population will demand service at the location if other service points are within reach (i.e., inside its “floated” catchment area).

For instance, assuming (as the binary impedance function does), that individuals at Population Center 1 are indifferent between Clinics 1 and 2, then it is reasonable to think that the population will sort itself proportionally to these two clinics - in this example, this means that half of the population will attend one of two different clinics (importantly, this assumes that the services on offer are undifferentiated; one would not

generally consider cancer screening and hair removal clinics competitors). This suggests the following adjustment to the way the level of demand is calculated. Given an impedance function, a set of adjusted weights, say  $W_{ij}^{i*}$ , are precalculated by dividing the original impedance weights by the sum of the weights for population center  $i$  over all service points  $j$ :

$$W_{ij}^i = \frac{W_{ij}}{\sum_j W_{ij}}$$

Please notice that these weights are identical to the selection weights of the 3SFCA method. A key property of the adjusted weights is the following:

$$\sum_j W_{ij}^i = 1$$

This adjustment procedure has the effect that, when the level of demand of  $i$  is summed over all service points  $j$ , the aggregated level of demand due to  $i$  is identical to its population:

$$\sum_j P_i W_{ij}^i = P_i$$

As a result of standardizing the impedance weights, population is allocated proportionally to clinics.

On the supply side, inflation happens because the level of service available at location  $j$  is assumed to be available to every population center  $i$  within its catchment area. To adjust this, another set of weights, say  $W_{ij}^{j*}$ , is pre-calculated by dividing the original impedance weights  $W_{ij}$  by the sum of the weights for service point  $j$  over all population centers  $i$ :

$$W_{ij}^j = \frac{W_{ij}}{\sum_i W_{ij}}$$

Again, the resulting weights have the property that:

$$\sum_i W_{ij}^j = 1$$

As before, the result of this procedure is that, when the level of service of  $j$  is aggregated by population centers, the total level of service for that service point is preserved:

$$\sum_i L_j W_{ij}^j = L_j$$

Note that, since the weights add up to one, they can be interpreted as a *probability* or *frequency* of contact, similar to the Huff model of [19].

In reference to Fig 3 (left panel), we can see that the original binary (unadjusted) weights for Population Centre 1 are  $W_{11} = 1$ ,  $W_{12} = 1$ , the weights of population center 2 are  $W_{21} = 1$ ,  $W_{22} = 1$ , and the weights of population center 3 are  $W_{31} = 1$ ,  $W_{32} = 1$ .

On the demand side, the adjusted weights become for Population Center 1,  $W_{11}^i = 1/2$ ,  $W_{12}^i = 1/2$ , for Population Center 2  $W_{21}^i = 1/2$ ,  $W_{22}^i = 1/2$ , and for Population Center 3  $W_{31}^i = 1/2$ ,  $W_{32}^i = 1/2$ . Using the adjusted weights, it can be seen that the level of demand due to each population center equals its respective population:

$$\begin{aligned}\sum_j D_{1j} &= 1/2P_1 + 1/2P_1 = P_1 \\ \sum_j D_{2j} &= 1/2P_2 + 1/2P_2 = P_2 \\ \sum_j D_{3j} &= 1/2P_3 + 1/2P_3 = P_3\end{aligned}$$

Coming next to the supply side, the adjusted weights for Clinic 1 are  $W_{11}^{j*} = 1/3$ ,  $W_{21}^{j*} = 1/3$ , and  $W_{23}^{j*} = 1/3$ ; for Clinic 2 the adjusted weights are  $W_{12}^{j*} = 1/3$ ,

$W_{22}^{i*} = 1/3$ , and  $W_{32}^{i*} = 1/3$ . It can be seen that the level of service is preserved across clinics, and therefore across the system:

$$\begin{aligned}\sum_i L_{i1} &= L_1/3 + L_1/3 + L_1/3 = L_1 \\ \sum_i L_{i2} &= L_2/3 + L_2/3 + L_2/3 = L_2\end{aligned}$$

The method to adjust the weights used above is identical to a procedure that will be familiar to readers acquainted with the literature in the fields of spatial statistics and econometrics. The same adjustment is widely used there under the names of row- and column-standardization of a weights matrix [24,25].

The proposed adjustment can be easily implemented. We will present next the implementation using a compact matrix notation. Begin by defining the following impedance matrix:

$$\mathbf{W} = \begin{pmatrix} W_{11} & \cdots & W_{1J} \\ \vdots & \ddots & \vdots \\ W_{N1} & \cdots & W_{NJ} \end{pmatrix}$$

where  $W_{ij}$  is an impedance function evaluated at  $d_{ij}$ . Subindex  $i$  is for population centers ( $i = 1, \dots, N$ ) and subindex  $j$  is for service points ( $j = 1, \dots, J$ ). Note that the matrix does not need to be square. A row-standardized set of weights is obtained as follows:

$$\mathbf{W}^i = \begin{pmatrix} \frac{W_{11}}{\sum_j W_{1j}} & \cdots & \frac{W_{1J}}{\sum_j W_{1j}} \\ \vdots & \ddots & \vdots \\ \frac{W_{N1}}{\sum_j W_{Nj}} & \cdots & \frac{W_{NJ}}{\sum_j W_{Nj}} \end{pmatrix}$$

Next, a column-standardized set of weights is calculated as:

$$\mathbf{W}^j = \begin{pmatrix} \frac{W_{11}}{\sum_i W_{i1}} & \cdots & \frac{W_{1J}}{\sum_i W_{iJ}} \\ \vdots & \ddots & \vdots \\ \frac{W_{N1}}{\sum_i W_{i1}} & \cdots & \frac{W_{NJ}}{\sum_i W_{iJ}} \end{pmatrix}$$

In the first example above (see Fig 1), the binary impedance matrix is:

$$\mathbf{W}_{binary} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$$

The row-standardized weights that correspond to this matrix are:

$$\mathbf{W}_{binary}^i = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

and the column-standardized weights are:

$$\mathbf{W}_{binary}^j = \begin{pmatrix} 1/3 & 1/3 \\ 1/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix}$$

The stepwise impedance weights in the example are:

$$\mathbf{W}_{stepwise} = \begin{pmatrix} 0.8 & 0.8 \\ 0.8 & 0.4 \\ 0.4 & 0.8 \end{pmatrix}$$

The row-standardized weights in turn are:

459

$$\mathbf{W}_{stepwise}^i = \begin{pmatrix} 1/2 & 1/2 \\ 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}$$

whereas the column-standardized weights are:

460

$$\mathbf{W}_{stepwise}^j = \begin{pmatrix} 4/10 & 4/10 \\ 4/10 & 2/10 \\ 2/10 & 4/10 \end{pmatrix}$$

Once that the impedance weights have been adjusted, a vector of adjusted level of demand  $\mathbf{D}^*$  can be obtained by multiplying the *transposed* impedance matrix by a vector of population values as follows:

461

462

463

$$\mathbf{D}^* = [\mathbf{W}^i]^T \mathbf{P}$$

where the  $T$  operator is for “transpose”, and  $\mathbf{P}$  is:

464

$$\mathbf{P} = \begin{pmatrix} P_1 \\ \vdots \\ P_N \end{pmatrix}$$

The level of demand for the service points in the binary impedance function example is (in vector form):

465

466

$$\mathbf{D}_{binary}^* = \begin{pmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 100 \\ 100 \\ 100 \end{pmatrix} = \begin{pmatrix} 300/2 \\ 300/2 \end{pmatrix} = \begin{pmatrix} 150 \\ 150 \end{pmatrix}$$

Notice how each clinic is expected to service only 150, and the level of demand over the system is identical to the total population.

467

468

The level of demand for the service points in the stepwise impedance function example is (in vector form):

469

470

$$\mathbf{D}_{sw}^* = \begin{pmatrix} 1/2 & 2/3 & 1/3 \\ 1/2 & 1/3 & 2/3 \end{pmatrix} \begin{pmatrix} 100 \\ 100 \\ 100 \end{pmatrix} = \begin{pmatrix} 50 + 200/3 + 100/3 \\ 50 + 100/3 + 200/3 \end{pmatrix} = \begin{pmatrix} 150 \\ 150 \end{pmatrix}$$

As can be seen, the aggregated level of demand, after the adjustment, equals (as desired) the actual population of the region. In the case of the stepwise function, total demand has been adjusted to the population of the region without the restrictive assumption that some people are excluded from the system. This is achieved by assuming an assortative process that leads to proportional allocation of the demand.

471

472

473

474

475

The levels of demand can then be used to calculate the level of service at the individual clinic locations by performing Hadamard division ( $\oslash$ ) of the vector of supply by the vector of adjusted demand. This is the first step of the 2SFCA (aggregating demand over catchment areas for service points):

476

477

478

479

$$\mathbf{L}^* = \mathbf{S} \oslash \mathbf{D}^*$$

Since Hadamard division is an element-by-element operation, the adjusted levels of service in the first example (using the binary impedance function) are:

480

481

$$\mathbf{L}_b^* = \begin{pmatrix} 10 \\ 10 \end{pmatrix} \oslash \begin{pmatrix} 150 \\ 150 \end{pmatrix} = \begin{pmatrix} 10/150 \\ 10/150 \end{pmatrix} = \begin{pmatrix} 0.067 \\ 0.067 \end{pmatrix}$$

The levels of service in the second example, when using the stepwise impedance function, are also:

$$\mathbf{L}_{sw}^* = \begin{pmatrix} 10 \\ 10 \end{pmatrix} \oslash \begin{pmatrix} 150 \\ 150 \end{pmatrix} = \begin{pmatrix} 10/150 \\ 10/150 \end{pmatrix} = \begin{pmatrix} 0.067 \\ 0.067 \end{pmatrix}$$

Unlike the 2SFCA, E2SFCA, and 3SFCA methods that produce levels of service that resemble PPRs but with values that are inconsistent with total demand given the population, this operation returns values that are genuinely local PPRs that are consistent with the population of the region. As we saw above, the demand equals the population. Here, the supply also equals the number of physicians in the region. Because both demand and supply are not inflated or deflated in this rectified method, these values are easily interpretable relative to the Regional Average PPR of 20/300 or 0.067 physicians per person. In the case of the example, it is clear that both clinics have PPRs that are identical to the Regional Average PPR.

Accessibility, finally, is calculated as the matrix product of the column-standardized weights and the adjusted level of service:

$$\mathbf{A}^* = \mathbf{W}^j \mathbf{L}^*$$

which, continuing with the example, gives the following for the binary impedance function:

$$\mathbf{A}_b^* = \begin{pmatrix} 1/3 & 1/3 \\ 1/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 10/150 \\ 10/150 \end{pmatrix} = \begin{pmatrix} 10/450 + 10/450 \\ 10/450 + 10/450 \\ 10/450 + 10/450 \end{pmatrix} = \begin{pmatrix} 0.044 \\ 0.044 \\ 0.044 \end{pmatrix}$$

Notice how the sum of accessibility over the region is consistent with the total level of service over all clinics (i.e., 0.133). The level of service has been allocated in its totality.

When using the stepwise impedance function, accessibility is calculated as:

$$\mathbf{A}_{sw}^* = \begin{pmatrix} 4/10 & 4/10 \\ 4/10 & 2/10 \\ 2/10 & 4/10 \end{pmatrix} \begin{pmatrix} 10/150 \\ 10/150 \end{pmatrix} = \begin{pmatrix} 4/150 + 4/150 \\ 4/150 + 2/150 \\ 2/150 + 4/150 \end{pmatrix} = \begin{pmatrix} 0.053 \\ 0.040 \\ 0.040 \end{pmatrix}$$

Again, the sum of accessibility is consistent with the level of service available from all clinics in the region. As with the Local PPRs, accessibility is interpreted as population-to-provider ratios for each population center in such a way that all calculations are with total demand and total level of service. In particular, accessibility can be interpreted as the share of level of service that a population center receives from all the clinics that service it.

For the sake of comparison, levels of service and accessibility are reported for the simulated example in Tables 4 and 5.

**Table 4.** Disaggregated Proportional Demand Allocations by Population Center and Clinic, and level of Service by Clinic, using Adjusted Weights

Population Center	Population	Binary Impedance - Adjusted			Stepwise Impedance - Adjusted		
		Clinic 1	Clinic 2	Clinic 3	Clinic 1	Clinic 2	Clinic 3
Population 1	250	250	0	0	250	0	0
Population 2	250	125	125	0	178.15	71.853	0
Population 3	500	166.67	166.67	166.67	111.62	276.75	111.62
Population 4	500	166.67	166.67	166.67	208.04	208.04	83.911
Population 5	1000	0	500	500	0	388.35	611.65
Population 6	500	0	250	250	0	143.71	356.29
Population 7	1000	0	500	500	0	500	500
Population 8	250	0	125	125	0	178.15	71.853
Population 9	250	0	125	125	0	125	125
Total Population/Demand	4500	708.333	1958.333	1833.333	747.815	1891.852	1860.333
Supply	NA	1	3	2	1	3	2
Level of Service (per 1,000)	NA	1.412	1.532	1.091	1.337	1.586	1.075

*Note:*

Darker number colors in a row indicate a greater allocation of demand.

**Table 5.** Disaggregated Proportional Level of Service Allocations by Clinic and Population Center, and Accessibility by Population Center, using Adjusted Weights

Clinic	Supply	Demand	Level of Service	Center 1	Center 2	Center 3	Center 4	Center 5	Center 6	Center 7	Center 8	Center 9
<b>Binary Impedance - Adjusted</b>												
1	1	708.333		1.412	0.353	0.353	0.353	0.353	0	0	0	0
2	3	1958.333		1.532	0	0.191	0.191	0.191	0.191	0.191	0.191	0.191
3	2	1833.333		1.091	0	0	0.156	0.156	0.156	0.156	0.156	0.156
Accessibility	NA	NA		NA	0.353	0.544	0.7	0.7	0.347	0.347	0.347	0.347
<b>Stepwise Impedance - Adjusted</b>												
1	1	747.815		1.337	0.192	0.476	0.192	0.476	0	0	0	0
2	3	1891.852		1.586	0	0.114	0.282	0.282	0.282	0.114	0.114	0.282
3	2	1860.333		1.075	0	0	0.0944	0.0944	0.369	0.234	0.0944	0.0944
Accessibility	NA	NA		NA	0.192	0.59	0.569	0.853	0.651	0.348	0.208	0.377

*Note:*

Darker number colors in a column indicate a greater allocation of level of service or accessibility.

An important point to remark is the following. The use of row- and column-standardized impedance weights assumes that the full population of every population center within the catchment of a clinic will receive service. However, the allocation, although proportional, is different when binary or stepwise impedance weights are standardized. When binary weights are employed, the underlying idea is that potential for use is identical within the catchment area irrespective of distance. When stepwise weights are used, proportionally more of the population is allocated to closer clinics. Depending on the definition of cost of travel, this allows a research to accommodate directional effects as well. For example, use of network travel time would tend to favor movement away from congested locations.

## Suboptimal Systems

The research of Delamater [18] illustrates how accessibility estimates can be misleading when systems are not optimally configured. We understand this to mean that some population centers are located too far away from service points to actually benefit from them. In the modified 2SFCA method (M2SFCA), Delamater addresses this issue by increasing the friction of distance. A slight inconsistency in this approach is that some of the centers that contribute to demand fail to benefit from the service due to the increased friction to which the allocation of the level of service is subjected. Our suggestion in the case of suboptimal systems is to use an impedance function that reflects limiting conditions. For instance, in urban settings a travel time longer than 2 hours might be considered too long to be serviced by any clinic.

## System Efficiency

The approach proposed in this paper allocates population and level of service proportionally and exactly. This assumes that the population sorts itself into clinics in the most efficient way. But what if some members of the population lack full information about the spatial distribution of clinics? Or have some bias towards centric locations? The vagaries of human behavior could create excess demand in some locations, and as a consequence supply surpluses in others. Situations like this can be accommodated in a relatively straightforward way using our approach.

Here, we describe the use of *slack factors*. Demand and level of service are allocated proportionally and exhaustively (i.e., 100%). But the standardization could allow for some slack, by inflating demand and/or supply in a controlled way.

Our proposal to standardize the weights was as follows, for the case of rows and columns respectively:

$$\mathbf{W}^i = \begin{pmatrix} \frac{W_{11}}{\sum_j W_{1j}} & \cdots & \frac{W_{1J}}{\sum_j W_{1j}} \\ \vdots & \ddots & \vdots \\ \frac{W_{N1}}{\sum_j W_{Nj}} & \cdots & \frac{W_{NJ}}{\sum_j W_{Nj}} \end{pmatrix} \text{ and } \mathbf{W}^j = \begin{pmatrix} \frac{W_{11}}{\sum_i W_{i1}} & \cdots & \frac{W_{1J}}{\sum_i W_{iJ}} \\ \vdots & \ddots & \vdots \\ \frac{W_{N1}}{\sum_i W_{i1}} & \cdots & \frac{W_{NJ}}{\sum_i W_{iJ}} \end{pmatrix}$$

A set of slack factors, say  $k_i^i$ , could be introduced in the following manner:

$$\mathbf{W}^i = \begin{pmatrix} \frac{k_1^i W_{11}}{\sum_j W_{1j}} & \cdots & \frac{k_1^i W_{1J}}{\sum_j W_{1j}} \\ \vdots & \ddots & \vdots \\ \frac{k_N^i W_{N1}}{\sum_j W_{Nj}} & \cdots & \frac{k_N^i W_{NJ}}{\sum_j W_{Nj}} \end{pmatrix}$$

A value of  $k_1^i = 1.10$  would inflate the demand of population center  $i = 1$  by 10%, whereas a value of  $k_1^i = 1.20$  would inflate the demand by 20%. In a similar way, a set

of slack factors  $k_i^j$  could be introduced to modulate the allocation of supply:

545

$$\mathbf{W}^j = \begin{pmatrix} \frac{k_1^j W_{11}}{\sum_i W_{i1}} & \dots & \frac{k_J^j W_{1N_J}}{\sum_i W_{iJ}} \\ \vdots & \ddots & \vdots \\ \frac{k_1^j W_{N1}}{\sum_i W_{i1}} & \dots & \frac{k_J^j W_{NJ}}{\sum_i W_{iJ}} \end{pmatrix}$$

A value of  $k_1^j = 0.9$ , for example, would deflate the supply of clinic  $j = 1$  by 10%. The use of slack factors provides an interesting way of modulating demand and level of service allocation in a very precise and controlled way, and presents interesting opportunities as well to introduce expert opinion or other empirical approaches to calibrate slack factors.

546

547

548

549

550

## Empirical Example

551

In the remainder of the paper we present an empirical example to illustrate the application of the methods presented above. Based on the preceding discussion, the adjusted 2SFCA method employed can be summarized as:

552

553

554

$$L_j = \sum_i \frac{S_j}{P_i W_{ij}^i}$$

with row-standardized impedance weights  $W_{ij}^i$  in the first step, and:

555

$$A_i = \sum_j L_j W_{ij}^j$$

with column-standardized impedance weights  $W_{ij}^j$  in the second step. The same approach is used to re-weight the impedance function for the stepwise approach (i.e., E2SFCA).

556

557

558

559

560

561

562

563

The case study is based on accessibility to family physicians in the Hamilton Census Metropolitan Area (CMA), in Ontario, Canada. For this, we use data collected about the distribution of the population and primary health care clinics (i.e., family physicians) in the region. Time use data from Canada's General Social Survey (GSS) was also used to inform the selection of thresholds for the impedance functions. The data collection and preprocessing protocols are described next.

## Family Physicians and Clinic Locations

564

In regards to the supply of clinics, the locations of family physicians were obtained using the College of Physicians and Surgeons of Ontario (CPSO) database for the Province of Ontario. We chose this organization because all physicians practicing in Ontario are required to register with the CPSO, as set out in the Ontario Regulation 865/93: Registration [26].

565

566

567

568

569

Our search of CPSO's database was conducted attending to the following criteria.

570

- 1) Only physicians who are registered as family physicians were selected (this excluded specialists such as pediatric physicians).

571

572

- 2) The spatial extent of the search was determined using forward sortation areas (FSAs), which are the first three initial characters of a postal code. Using a GIS, the regions of interest were selected by choosing FSAs within a 10 kilometer buffer distance from the Hamilton CMA boundary. This involved 72 different FSA regions. Each FSA region code was then searched in the CPSO database in addition to the family physician specification.

573

574

575

576

577

578

The parameters of the search were deliberately conservative, and the search identified a total of 2,224 family physicians practicing in the region, of which, 864 are located in the Hamilton CMA. The resulting dataset was manually verified by the third author to ensure that the information was consistent and suitable for geocoding. Prior to geocoding, locational information was organized in three columns, containing street address, city name, and province name. After family physicians were geocoded, locations were further examined. When family physicians overlapped or were within a 50 meter distance of each other we merged the records to identify 535 unique locations that we term “clinics”. Many of these clinics are not in the Hamilton CMA proper, but provide a buffer to minimize edge effects in the analysis. The distribution of clinics and family physicians is shown in Fig 10 for the Hamilton CMA.

579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589

## Population

590  
591  
592  
593  
594  
595  
596

Population information was obtained from the 2011 Canadian Census. To maximize the spatial resolution, population data were acquired at the Dissemination Area (DA) level of geography for all DAs within the selected FSAs. As a result, this includes DAs not in the Hamilton CMA proper, but that provide a buffer against edge effects. From this, the region contains a population of 2,959,090, of which 720,725 are in the Hamilton CMA. The distribution of population in the Hamilton CMA is shown in Fig 11.

597

## Travel Time Matrix

598  
599  
600  
601  
602  
603  
604

Calculation of impedance weights requires that we evaluate an impedance function at values of  $d_{ij}$ , that is, the cost of travel between DA  $i$  and clinic  $j$ . In this research we used travel time as our cost variable. To this end, we computed a matrix of travel times measured over the road network. To calculate the travel time between population centers and clinics we used the DA centroids and the geocoded locations of clinics. Shortest paths on the network and subsequently travel times were computed using a Geographic Information System.

605

## Impedance Functions

606  
607  
608  
609  
610  
611  
612  
613

For the experiments we use two different impedance functions, corresponding to the 2SFCA and E2SFCA algorithms. We do not implement the 3SFCA or the M2SFCA methods because, as noted above, they are equivalent to using steeper impedances. For the 2SFCA approach, impedance is given by a binary function, whereas for E2SFCA it is given by a stepwise function. The impedance functions require that we define cost (i.e., travel time) thresholds to implement them. To select the thresholds, we retrieved time use data from Canada’s General Social Survey Cycle 24 (see <http://odesi2.scholarsportal.info/webview/>).

614  
615  
616  
617  
618  
619  
620

From the time use files, we filtered all activity episodes corresponding to respondents living in CMAs/CAs (metropolitan regions) in Ontario. Next, we filtered all episodes taking place in a car (as driver) while traveling for personal care activities for household adults (which includes traveling to see a doctor) and traveling for shopping or obtaining services (which includes traveling to go to health clinic or doctor’s office). It is worthwhile noting that travel by car accounts for over 95% of trips for the selected purposes in Ontario CMAs/CAs.

621  
622  
623  
624  
625

Once episodes were filtered by mode of travel and purpose of the trip, their durations (in minutes) were examined by means of quantile analysis, using episode weights to ensure the representativeness of the analysis. From the results, we learned that 50% of all trips by car for the aforementioned purposes are less than 15 minutes long, and we selected this value as the threshold  $d_0$  for the binary function. In other

words, this part of the analysis assumes that any person who has to travel longer than 15 minutes to reach a clinic is outside its catchment area. We deem this value appropriate for the scale, density, and level of congestion of Hamilton CMA.

Quantile analysis of trip durations was also used to calibrate a Gaussian function with standard deviation set at 15 minutes, to match the value selected for the binary impedance above. This produced the following stepwise function, with any trips longer than 45 minutes assumed to be outside of catchment:

$$W(d_{ij}) = \begin{cases} 0.946 & d_{ij} \leq 5 \\ 0.801 & 5 < d_{ij} \leq 10 \\ 0.607 & 10 < d_{ij} \leq 15 \\ 0.411 & 15 < d_{ij} \leq 20 \\ 0.135 & 20 < d_{ij} \leq 30 \\ 0.011 & 30 < d_{ij} \leq 45 \\ 0.00 & 45 < d_{ij} \end{cases}$$

Notice how the stepwise function has weights greater than 0.5 for  $d_{ij} \leq 15\text{min}$  and less than 0.5 for  $d_{ij} > 15\text{min}$ . This means that it will count fewer people than the binary function when  $d_{ij} \leq 15\text{min}$ , but more when  $d_{ij} > 15\text{min}$ .

## Results

We begin our discussion of the results by noting that with a total population of the region of 2,959,090 and 2,222 family physicians, the Regional Average PPR ratio is 0.751 family physicians per 1,000 people. This value is somewhat lower than the value of 1.16 for Ontario reported by CIHI [27] and lower than the 1.20 estimated based on the population and physician data for the Hamilton CMA, which we attribute to our conservative search criteria of family physicians in the rest of the region.

The nominal levels of demand, service, and accessibility are calculated for the 2SFCA and E2SFCA using both the unadjusted and adjusted impedance matrices. Table 6 summarizes the results by each impedance matrix. As seen there, when no adjustment is made, the nominal demand explodes to several times the actual population in the region. However, when the impedance weights are standardized, demand is now only slightly less than the total population for the region, since the system is not optimal in the sense discussed by Delamater [18], and a small proportion of the population turns out to be outside of catchment. The nominal demand under binary impedance is lower due to the stricter catchment area condition (i.e., less than 15 minutes), compared to the stepwise function (i.e., less than 45 minutes). This, in turn, is somewhat lower than the total demand in the Regional Average PPR, which does not impose catchment area constraints within the region.

It is clear that the rectified demand leads to results that are considerably more realistic than the conventional approaches. In addition to the nominal system-wide demand, this is seen as well when calculating the regional provider-to-population ratios for each case (i.e., Family Physicians per 1,000 people). As seen in the table, the mean levels of service for clinics in the region in the case of the adjusted binary and stepwise weights are in line with their corresponding Regional Average PPRs. Since the levels of service in the case of the adjusted weights can be interpreted as local PPRs, this indicates that the average clinic offers approximately the same level of service as the regional system does for the whole population. Furthermore, the mean accessibility of a DA according to the adjusted weights is identical to the mean LOS: this is because the LOS is allocated completely to DAs. The total LOS and accessibility in the region match when the adjusted weights are used. This is not the case when the unadjusted weights are used. Clearly, the use of the unadjusted weights can lead to a substantial

amount of accessibility inflation, by factors as high as five or six times the estimates of  
the proposed proportional allocation approach.

These results demonstrate how inflation of the supply (i.e., the level of service) leads  
to much higher values of accessibility in the case of the conventional 2SFCA and  
E2SFCA methods. The procedure to rectify the population and level of service, on the  
other hand, leads to accessibility outputs that are consistent with the regional  
population and overall supply of health care services. This, in turn, makes  
interpretation of the output more robust and intuitive.

668  
669  
670  
671  
672  
673  
674  
675

**Table 6.** Summary of Results Accessibility Analysis Hamilton CMA

Case	Supply (Doctors)	Nominal Demand	Regional Average PPR	Mean Level of Service per Clinic	Total Level of Service	Mean Accessibility per DA	Total	Accessibility
Binary	2222	176,244,020	0.013	0.025	14.766	5.993	3481.709	
Binary Adjusted	2222	2,861,445	0.777	1.046	607.803	1.046	607.803	
Stepwise	2222	209,666,501	0.011	0.069	40.050	6.023	3499.115	
Stepwise Adjusted	2222	2,957,095	0.751	0.879	510.663	0.879	510.663	

Another important issue is that spatial distribution of inflation of demand and level of service. If inflation happened in a uniform way, the upward bias in the estimates could to some extent be ignored, as long as relative differences by location remained relatively constant. Unfortunately, as seen in Fig 12 and Fig 13, demand inflation is far from uniform. In fact, inflation of demand tends to happen, as per our earlier conjecture, in areas with higher population density. Inflation factors are also substantially higher when the binary impedance function is used. Since this function lacks a gradual distance-decay mechanism, it is more generous in terms of counting population serviced. Notice the magnitude of the inflation factors: since the inflation of demand depends on the number of overlapping catchment areas, a factor of 160, for instance, would suggest that a clinic is expected to *simultaneously* serve approximately that number of DAs in the conventional 2SFCA method, and a proportionally similar number in the conventional E2SFCA method.

The map of accessibility for the implementation of 2SFCA is shown in Fig 14 and with the adjusted weights for proportional allocation in Fig 15. The general patterns observed in the figures are as expected, with higher accessibility in denser, better connected parts of the region. Relatively high accessibility in the north and west of the CMA is due to proximity to other major population centers such as Oakville, Kitchener, and Waterloo. A question, however, is the degree of inflation of accessibility in the original 2SFCA? Fig 16 plots the ratio of the binary and adjusted binary accessibility measures. Here it can be seen that the unadjusted accessibility values are at least three times greater than their adjusted counterparts within the study area. This inflation, moreover, is not uniform across space, with inflation of the binary accessibility values up to 8 times greater than those from the adjusted model at the edges of the city where the 15-minute catchment areas begin to overlap with neighboring municipalities.

Why is this important? As noted by various authors [3,18], in traditional FCA methods, the sum of the population-weighted average of accessibility across all population centers is equal to the regional average provider-to-population ratio [18]. In the present case, the weighted sum of accessibility in the unadjusted binary and stepwise measures is 0.751. However, while this value is indeed identical to the regional average provider-to-population ratio, it is problematic because the share of the population correlates poorly with the pattern of inflation observed (see Fig 17). The key issue here is that accessibility is deflated by the share of the population in a DA  $i$ ; however, the degree of inflation of demand and supply depends not only of the population DA  $i$ , but on the population of every DA  $j$  with which DA  $i$  interacts via overlapping catchment areas. As a consequence, deflating accessibility using population shares in previous FCA methods does not accurately offset demand and supply inflation.

Fig 18 and Fig 19 present the results for the stepwise E2SFCA with and without the rectification. The results are qualitatively similar to the 2FSCA, with the expected differences. The inflation factor is even more substantial, given the larger catchment areas used.

## Disparity Analysis

An advantage of the use of adjusted weights for proportional allocation of demand and level of service is that, after rectifying the inflation artifact, they make it possible to conduct accessibility disparity analysis in a very intuitive way.

For instance, an analyst interested in equity analysis could allocate the total level of service uniformly to every DA. In other words, the total level of service (which equals the sum of accessibility over the system) can be divided by the number of population centers in the system to return the Average Local Population Center PPR. The resulting mean value, call it  $L_i^e$  then would be assigned to the population centers as their “equitable” share of the total level of service in the system. Next, the equititative

distribution of the level of service in each population center is subtracted from the estimated mean accessibility to arrive at a disparity index. When the difference between these two quantities is positive, this would indicate that a DA's accessibility exceeds its equitable share of level of service. On the other hand, when the difference is negative, the DA's accessibility is below its equitable share of the level of service.

This approach is reminiscent of the Spatial Access Ratio (SPAR) proposed by Wan et al. [28], which is calculated as the ratio between a population center's accessibility and the mean accessibility across all population centers. Wan et al. [17] calculate SPAR based on the results of their 3SFCA method, by rescaling the accessibility measures to reflect the percentage difference in each population center's accessibility relative to the mean. This measure is designed to overcome the sensitivity of existing FCA metrics to the impedance function. In contrast, the approach proposed here, enables more intuitive and interpretable results by preserving the system-wide population and level of service. In this way, a disparity index is useful to highlight the absolute difference in accessible provider-to-population ratios across population centers.

Disparity maps for the adjusted binary and stepwise impedance functions are shown in Fig 21 and Fig 22. These figures reveal the spatial distribution in disparity, with levels of access that are lower than the mean in more rural parts of the city (where travel times are longer and the distribution of physicians is more spatially disperse) compared to levels of access that are greater than the mean in the higher-density and more connected urban center.

## Conclusion

Accessibility to healthcare is an issue of continued interest in health geography. One of the most popular approaches to estimating accessibility is the 2SFCA method and its associated family of FCA models due to their simplification of more complex gravity models and their interpretation as proxies for provider-to-population ratios. These properties make FCA approaches particularly appealing for health policy. In this paper, we have argued that the overestimation of demand and level of service in FCA approaches poses a challenge to the interpretation of accessibility and the identification of spatial disparities in access, with potentially deleterious consequences for policy analysis.

The issue of overestimation of demand and level of service has been recognized before, notably by Wan et al. [17] and Delamater [18], and alternative approaches have been proposed that seek to offset or reduce the problem. Nevertheless, the present paper has shown that the inflation of demand is present in all existing FCA methods. Moreover, we also show that in some cases, demand is deflated, and detail the potential for inflation/deflation on the supply side. To overcome these issues, we draw from the fields of spatial statistics and econometrics, to incorporate row-standardized impedance weights in the calculation of demand, and column-standardized impedance weights to adjust the level of service. These adjustments ensure that allocation of demand and level of service are proportional. As a result, both the system-wide population and level of service are preserved in the estimation of accessibility.

The case study in Hamilton CMA reveals the extent of inflation in accessibility inherent in the unadjusted approaches compared to the adjusted binary and stepwise FCA methods. Furthermore, the adjustments result in local provider-to-population ratios which can be easily understood relative to the system-wide equitable level of service through the calculation of a disparity index. The applicability of these values is particularly enhanced by the use of a travel survey to inform the estimated impedance functions. Taken together, these innovations provide estimates of spatial accessibility and disparity that are robust to the regional distribution of supply and demand, as well

as observed travel behaviour. By extension, these properties mean that the adjusted approach employed here can offer more rigorous recommendations for health policy.

Finally, 1) we proposed a set of slack factors to modulate the estimates of demand and/or level of supply to account for system inefficiencies; and 2) demonstrated the use of a disparity index to conduct equity analysis.

In conclusion, the research presented in this paper demonstrates how a relatively simple adjustment of the impedance weights can help to overcome the inflation/deflation issue inherent in previous FCA approaches. By incorporating these methods into the estimation of accessibility to healthcare services, future research can help to ensure that the FCA approach continues to live up to its promise as an intuitive and policy-relevant method for investigating access and disparity.

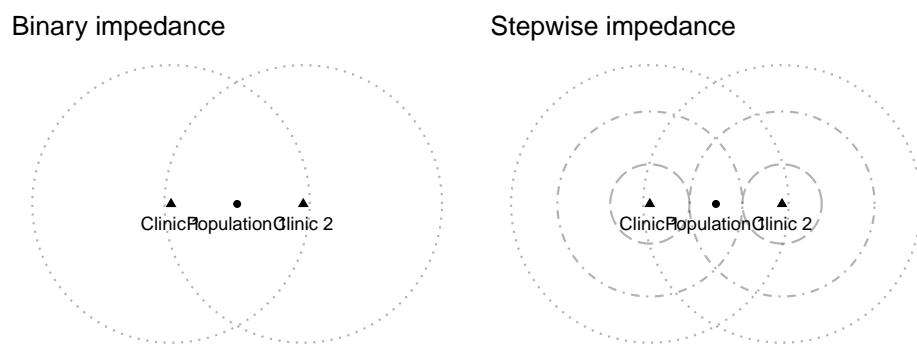
## References

1. Apparicio P, Gelb J, Dube AS, Kingham S, Gauvin L, Robitaille E. The approaches to measuring the potential spatial access to urban health services revisited: Distance types and aggregation-error issues. *International Journal of Health Geographics*. 2017;16. doi:10.1186/s12942-017-0105-9
2. Joseph AE, Bantock PR. Measuring potential physical accessibility to general practitioners in rural areas: A method and case study. *Social science & medicine*. Elsevier; 1982;16: 85–90.
3. Luo W, Wang FH. 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*. 2003;30: 865–884. Available: ISI:000187989500005 C:/Papers/Environment and Planning B/EPB (2003) 30 (6) 865-885.pdf
4. Radke J, Mu L. Spatial decomposition, modeling and mapping service regions to predict access to social programs. *Annals of Geographic Information Sciences*. 2000;6: 105–112.
5. Bauer J, Muller P, Maier W, Groneberg DA. Orthopedic workforce planning in germany - an analysis of orthopedic accessibility. *Plos One*. 2017;12: 15. doi:10.1371/journal.pone.0171747
6. Kim Y, Byon YJ, Yeo H. Enhancing healthcare accessibility measurements using gis: A case study in seoul, korea. *Plos One*. 2018;13: 19. doi:10.1371/journal.pone.0193013
7. Fujita M, Sato Y, Nagashima K, Takahashi S, Hata A. Impact of geographic accessibility on utilization of the annual health check-ups by income level in japan: A multilevel analysis. *Plos One*. 2017;12: 14. doi:10.1371/journal.pone.0177091
8. Song PG, Zhu YJ, Mao X, Li Q, An L. Assessing spatial accessibility to maternity units in shenzhen, china. *Plos One*. 2013;8: 7. doi:10.1371/journal.pone.0070227
9. McGrail MR, Humphreys JS. Measuring spatial accessibility to primary care in rural areas: Improving the effectiveness of the two-step floating catchment area method. *Applied Geography*. 2009;29: 533–541. doi:10.1016/j.apgeog.2008.12.003
10. Shah TI, Bell S, Wilson K. Spatial accessibility to health care services: Identifying under-serviced neighbourhoods in canadian urban areas. *Plos One*. 2016;11. doi:10.1371/journal.pone.0168208
11. Luo W, Qi Y. An enhanced two-step floating catchment area (e2sfca) method for measuring spatial accessibility to primary care physicians. *Health & Place*. 2009;15: 1100–1107. Available: ISI:000270348400023 C:/Papers/Health and Place/Health and Place (2009) 15 (4) 1100-1107.pdf
12. Dai D. Black residential segregation, disparities in spatial access to health care facilities, and late-stage breast cancer diagnosis in metropolitan detroit. *Health & place*.

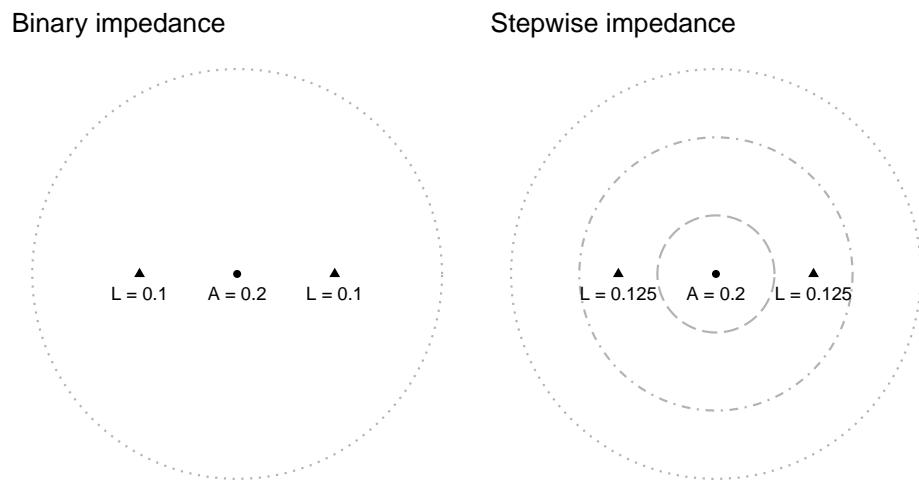
- Elsevier; 2010;16: 1038–1052.
13. Bauer J, Groneberg DA. Measuring spatial accessibility of health care providers - introduction of a variable distance decay function within the floating catchment area (fca) method. *Plos One*. 2016;11. doi:10.1371/journal.pone.0159148
14. Mao L, Nekorchuk D. Measuring spatial accessibility to healthcare for populations with multiple transportation modes. *Health & Place*. 2013;24: 115–122. doi:10.1016/j.healthplace.2013.08.008
15. Ngui A, Apparicio P. Optimizing the two-step floating catchment area method for measuring spatial accessibility to medical clinics in montreal. *Bmc Health Services Research*. 2011;11. Available: <http://www.biomedcentral.com/1472-6963/11/166>
16. Bell S, Wilson K, Bissonnette L, Shah T. Access to primary health care: Does neighborhood of residence matter? *Annals of the Association of American Geographers*. Taylor & Francis; 2013;103: 85–105.
17. Wan N, Zou B, Sternberg T. A three-step floating catchment area method for analyzing spatial access to health services. *International Journal of Geographical Information Science*. 2012;26: 1073–1089. doi:10.1080/13658816.2011.624987
18. Delamater PL. Spatial accessibility in suboptimally configured health care systems: A modified two-step floating catchment area (m2sfca) metric. *Health & Place*. 2013;24: 30–43. doi:10.1016/j.healthplace.2013.07.012
19. Luo J. Integrating the huff model and floating catchment area methods to analyze spatial access to healthcare services. *Transactions in GIS*. Wiley Online Library; 2014;18: 436–448.
20. Taylor PJ. Distance transformation and distance decay functions. *Geographical Analysis*. 1971;3: 221–238. doi:doi:10.1111/j.1538-4632.1971.tb00364.x
21. Kwan MP. Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis*. 1998;30: 191–216. Available: ISI:000074579200001 C:/Papers/Geographical Analysis/Geographical Analysis (1998) 30 (3) 191-216.pdf
22. Schuurman N, Berube M, Crooks VA. Measuring potential spatial access to primary health care physicians using a modified gravity model. *Canadian Geographer-Geographe Canadien*. 2010;54: 29–45. doi:10.1111/j.1541-0064.2009.00301.x
23. Tobler WR. Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*. 1979;74: 519–530. Available: ISI:A1979HS22400001 C:/Papers/Journal of the American Statistical Association/Journal of the American Statistical Association (1979) 74 (367) 519-530.pdf
24. Anselin L. Spatial econometrics: Methods and models. Dordrecht: Kluwer; 1988.
25. Griffith DA. Advanced spatial statistics: Special topics in the exploration of quantitative spatial data series. Dordrecht: Kluwer; 1988.
26. 1991, Medicine Act, S.O. 1991, c. 30.
27. Health Information CI for. Number of physicians by province/territory and specialty. Canadian Institute for Health Information; 2018;
28. Wan N, Zhan FB, Zou B, Chow E. A relative spatial access assessment approach for analyzing potential spatial access to colorectal cancer services in texas. *Applied Geography*. Elsevier; 2012;32: 291–299.

## List of Figures

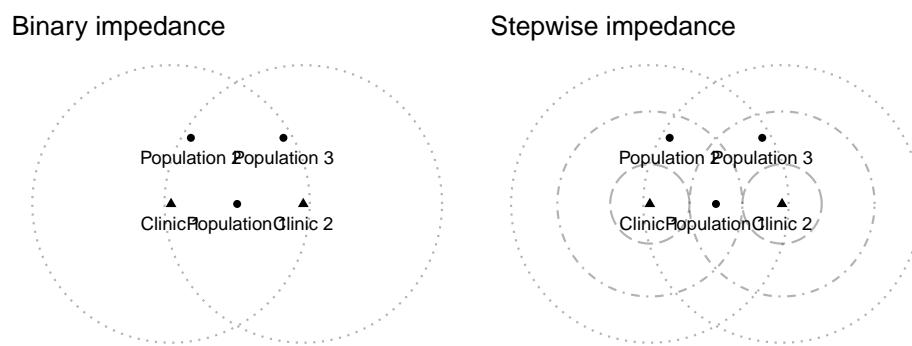
1	A single population center that enters the service areas of two clinics (triangles are clinics, dotted lines are segments of catchment areas) . . . . .	31
2	Levels of service of three clinics and accessibility of one population center (triangles are clinics, dotted lines are segments of catchment areas) . . . . .	32
3	Three population centers and two clinics (triangles are clinics, dotted lines are segments of catchment areas) . . . . .	33
4	Levels of service of two clinics and accessibility of three population centers (triangles are clinics, dotted lines are segments of catchment areas) . . . . .	34
5	Levels of service and accessibility according to 3SFCA and M2SFCA methods (triangles are clinics, dotted lines are segments of catchment areas)	35
6	Setup for the simulation exercise . . . . .	36
7	Catchment areas in step 1, according to binary and stepwise impedance functions (the weights of the stepwise function are 0.945, 0.600, and 0.424)	37
8	Catchment areas in step 1, according to binary and stepwise impedance functions (the weights of the stepwise function are 0.945, 0.600, and 0.424)	38
9	Accessibility by Population Center by Method . . . . .	39
10	Location of clinics and family physicians in the Hamilton CMA and surrounding regions . . . . .	40
11	Population distribution in the Hamilton CMA . . . . .	41
12	Demand inflation, binary impedance function . . . . .	42
13	Demand inflation, stepwise impedance function . . . . .	43
14	Accessibility, binary impedance function . . . . .	44
15	Accessibility, adjusted binary impedance function . . . . .	45
16	Accessibility Inflation factor, binary impedance function . . . . .	46
17	Population share and inflation factors compared . . . . .	47
18	Accessibility, stepwise impedance function . . . . .	48
19	Accessibility, adjusted stepwise impedance function . . . . .	49
20	Accessibility inflation factor, stepwise impedance function . . . . .	50
21	Accessibility disparities, adjusted binary impedance function . . . . .	51
22	Accessibility disparities, adjusted stepwise impedance function . . . . .	52



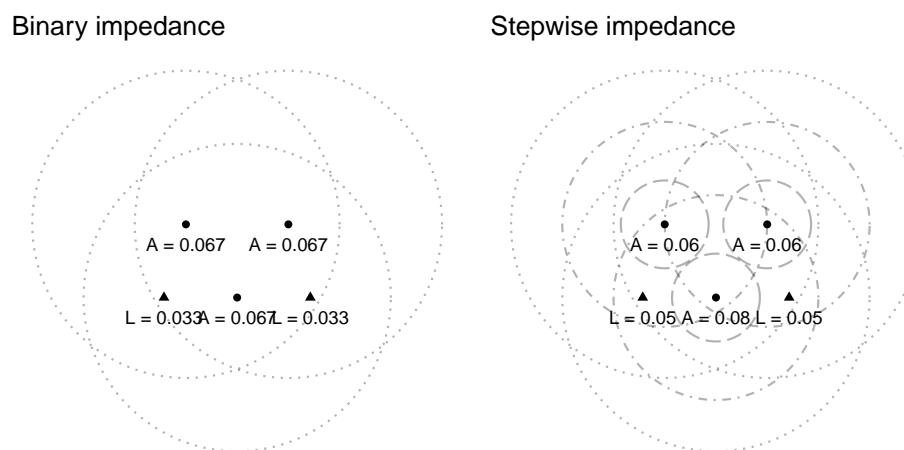
**Fig 1.** A single population center that enters the service areas of two clinics (triangles are clinics, dotted lines are segments of catchment areas)



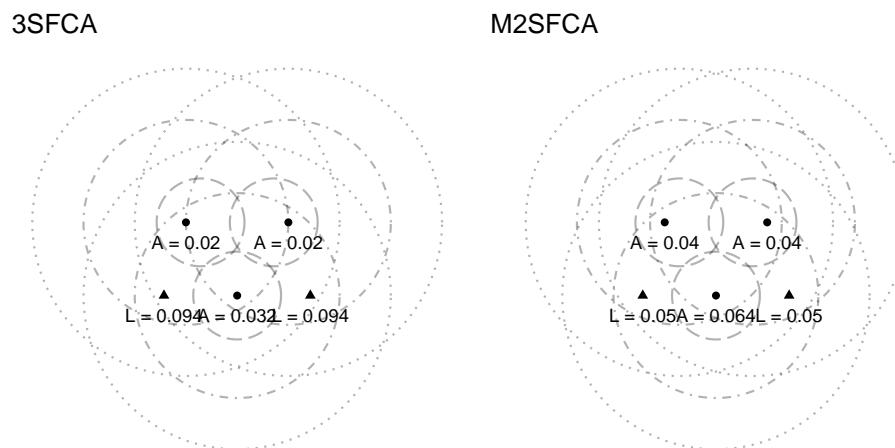
**Fig 2.** Levels of service of three clinics and accessibility of one population center (triangles are clinics, dotted lines are segments of catchment areas)



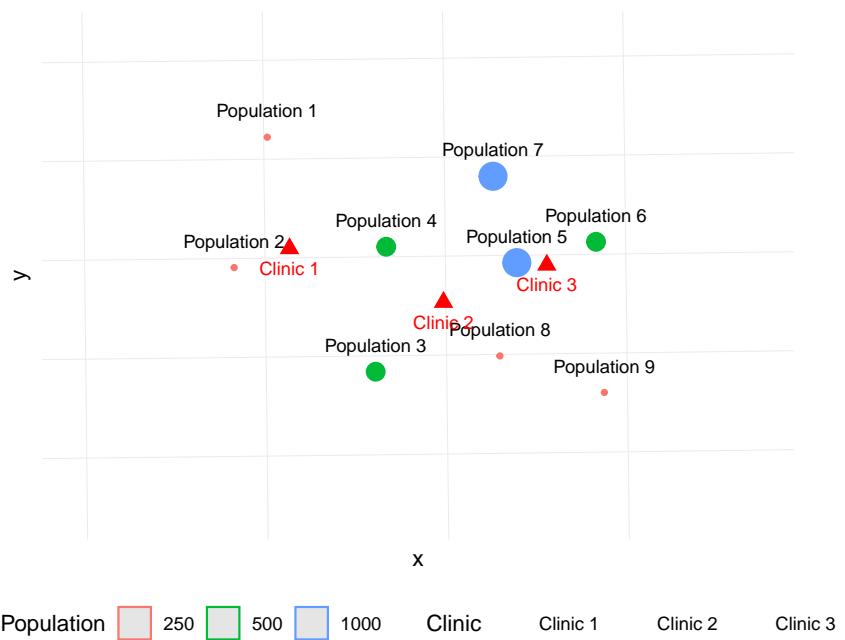
**Fig 3.** Three population centers and two clinics (triangles are clinics, dotted lines are segments of catchment areas)



**Fig 4.** Levels of service of two clinics and accessibility of three population centers (triangles are clinics, dotted lines are segments of catchment areas)

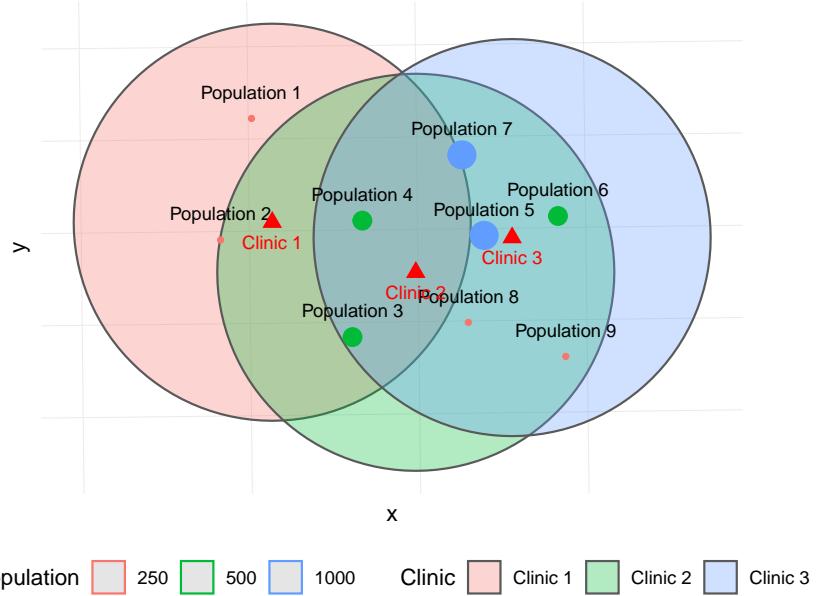


**Fig 5.** Levels of service and accessibility according to 3SFCA and M2SFCA methods (triangles are clinics, dotted lines are segments of catchment areas)



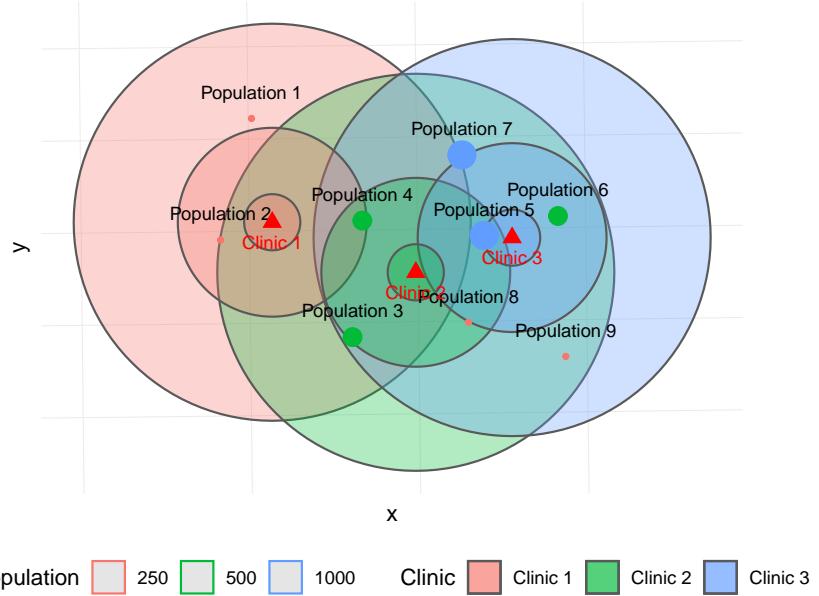
**Fig 6.** Setup for the simulation exercise

Binary impedance: Step one of 2SFCA

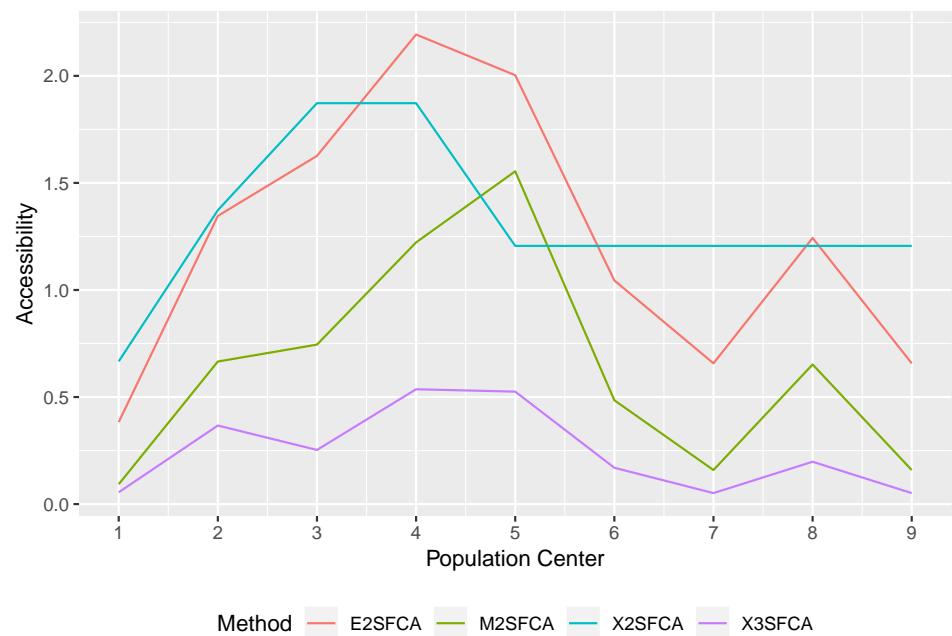


**Fig 7.** Catchment areas in step 1, according to binary and stepwise impedance functions (the weights of the stepwise function are 0.945, 0.600, and 0.424)

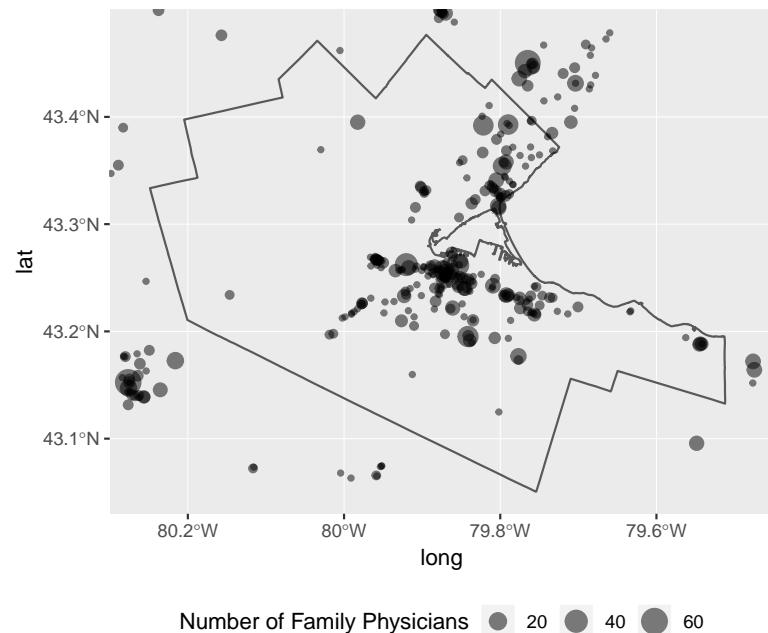
Stepwise impedance: Step one of 2SFCA



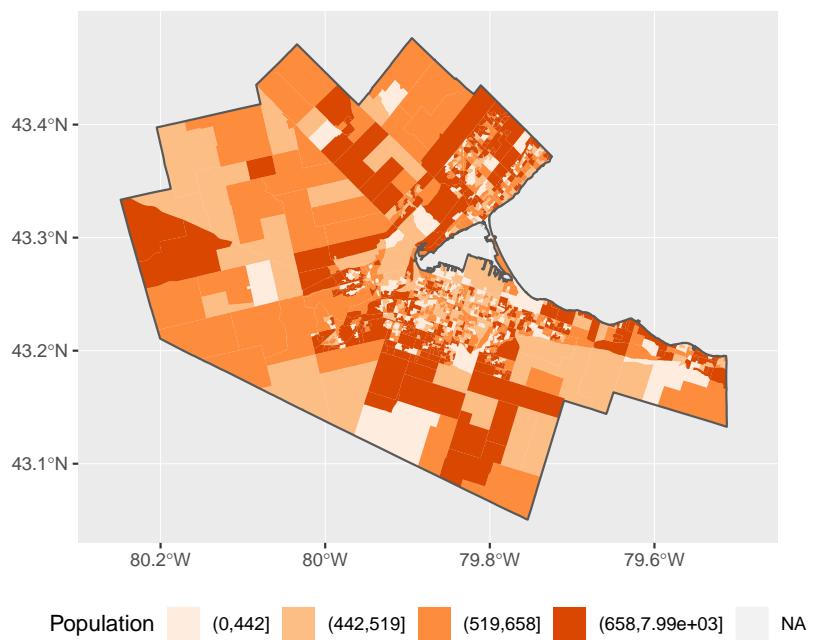
**Fig 8.** Catchment areas in step 1, according to binary and stepwise impedance functions (the weights of the stepwise function are 0.945, 0.600, and 0.424)



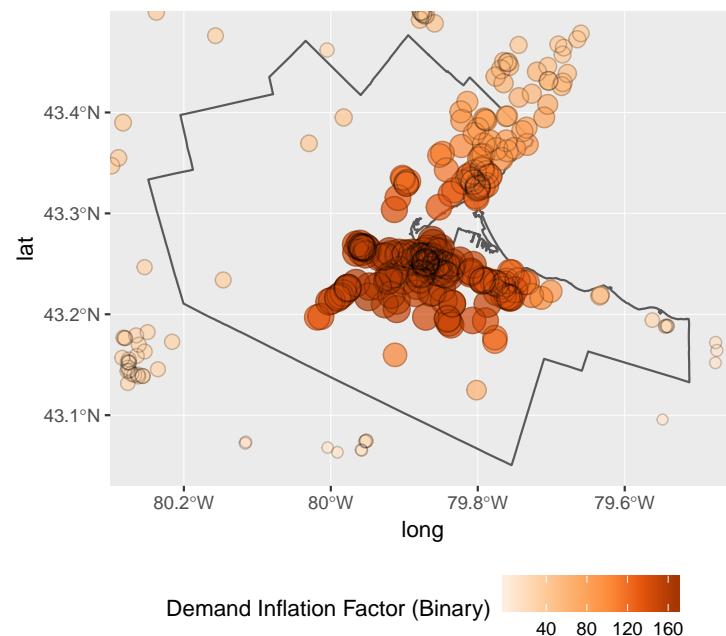
**Fig 9.** Accessibility by Population Center by Method



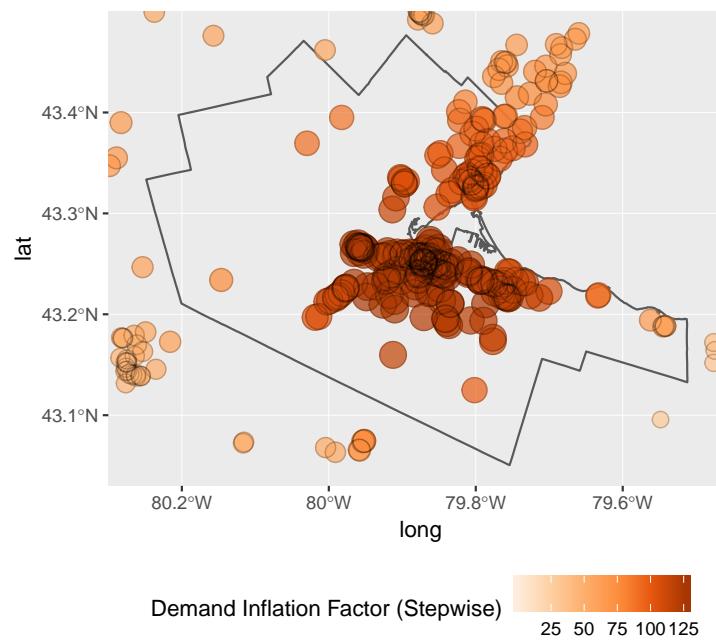
**Fig 10.** Location of clinics and family physicians in the Hamilton CMA and surrounding regions



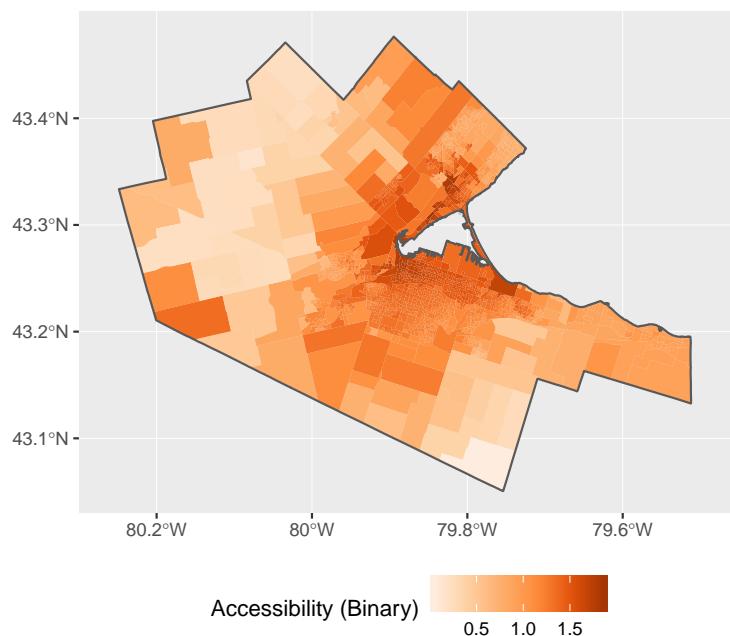
**Fig 11.** Population distribution in the Hamilton CMA



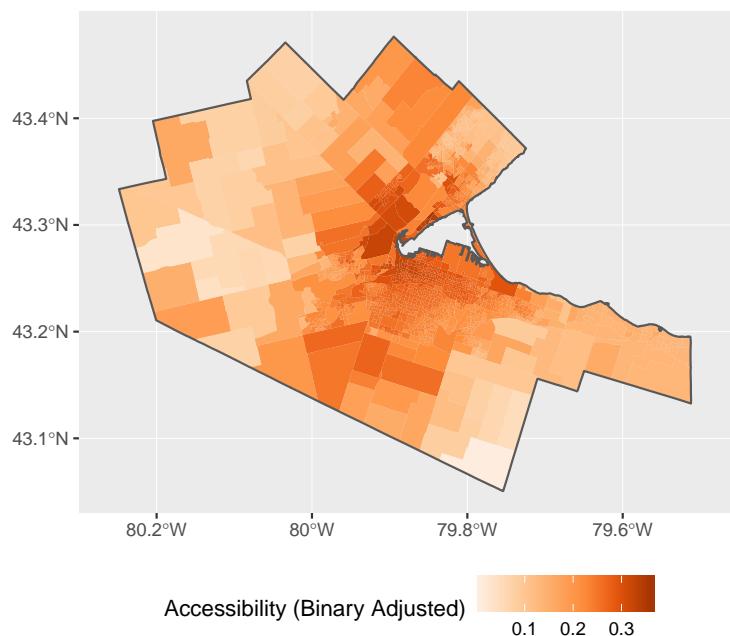
**Fig 12.** Demand inflation, binary impedance function



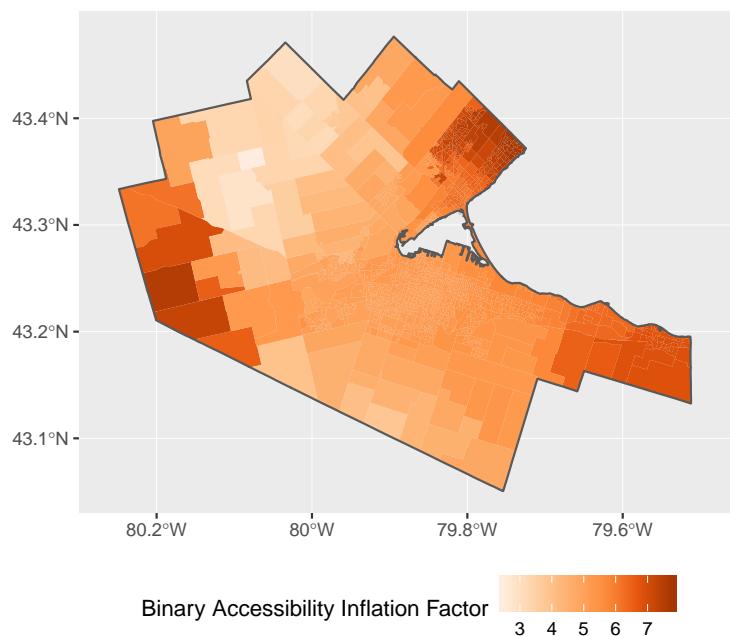
**Fig 13.** Demand inflation, stepwise impedance function



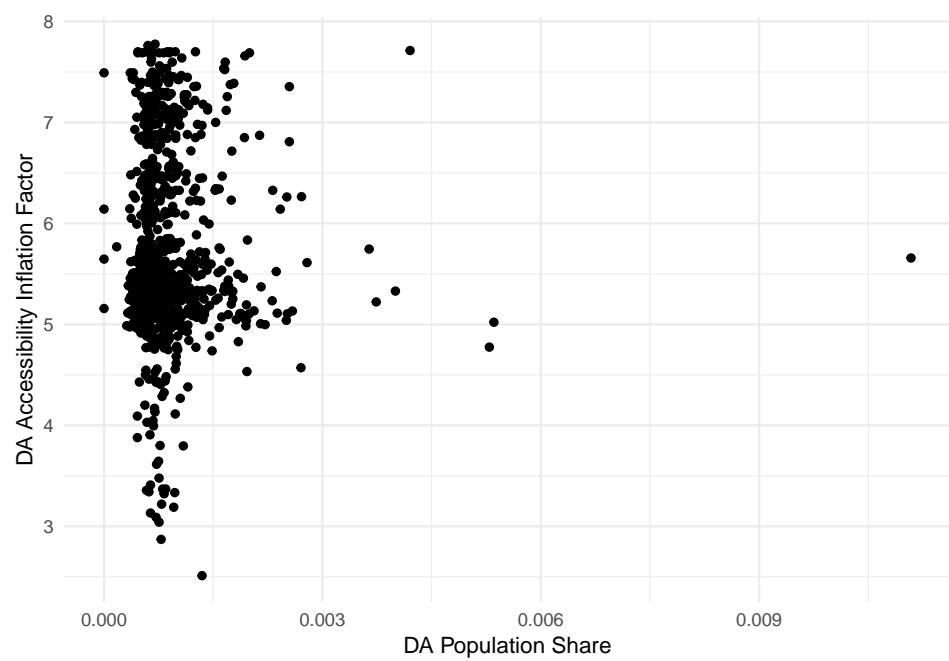
**Fig 14.** Accessibility, binary impedance function



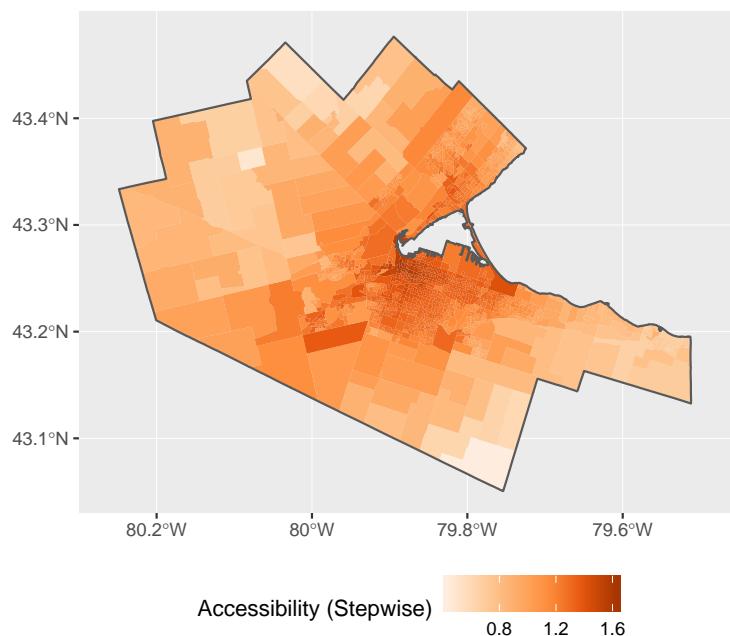
**Fig 15.** Accessibility, adjusted binary impedance function



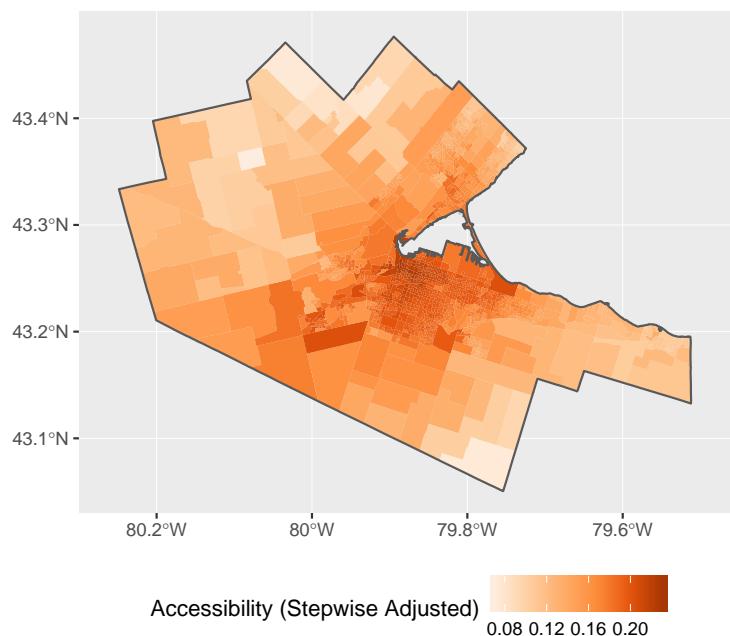
**Fig 16.** Accessibility Inflation factor, binary impedance function



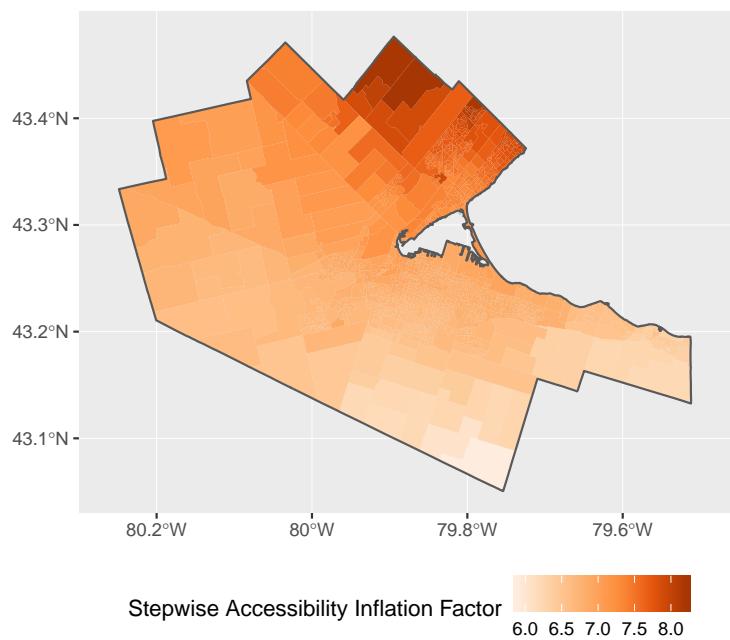
**Fig 17.** Population share and inflation factors compared



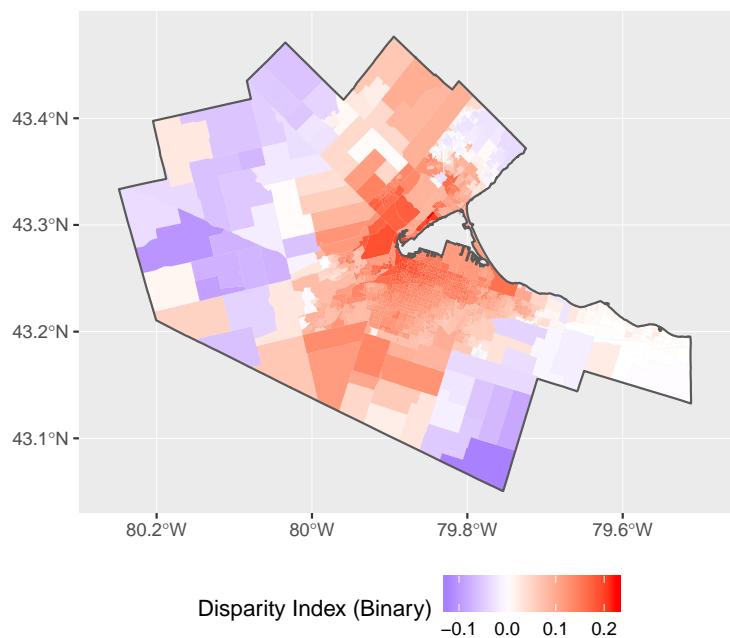
**Fig 18.** Accessibility, stepwise impedance function



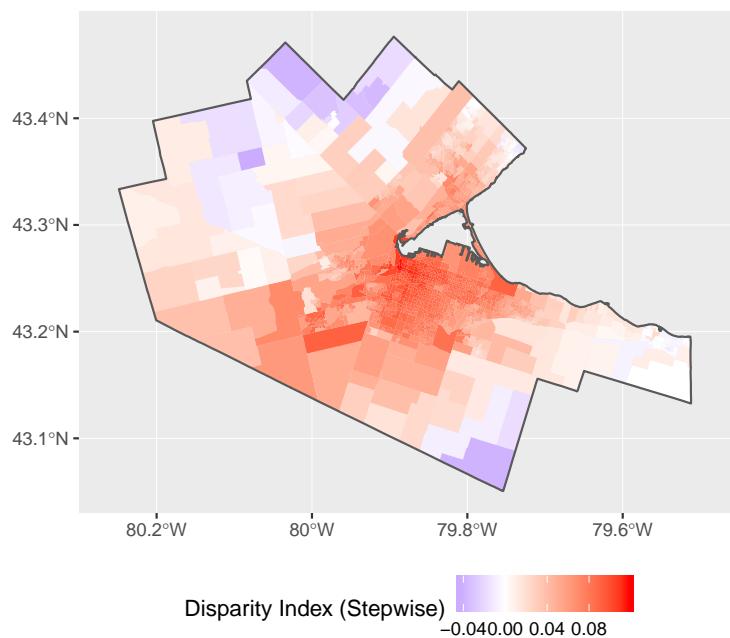
**Fig 19.** Accessibility, adjusted stepwise impedance function



**Fig 20.** Accessibility inflation factor, stepwise impedance function



**Fig 21.** Accessibility disparities, adjusted binary impedance function



**Fig 22.** Accessibility disparities, adjusted stepwise impedance function



Click here to access/download

**Supporting Information - Compressed/ZIP File Archive**  
**figure-latex.zip**

# Delta 01

Antonio Paez

April 5, 2019

## Introduction

~~Evaluating accessibility to healthcare services is an~~An important issue in health geography and health policy ~~is the evaluation of accessibility to healthcare services~~, with hundreds of research papers published on the topic since the 2000s [Apparicio2017]. However, the concept of accessibility is multi-dimensional, which often presents challenges to its operationalization in empirical research. According to Joseph and Bantock [-@Joseph1982], accessibility can be defined by both aspatial and spatial dimensions. The first dimension considers factors such as the ~~availabilityquality~~ of ~~healthcare~~the services ~~(or the supply of services), and~~ their ~~potential or revealed use by the public (demand)~~. Other aspatial factors include the characteristics of the supply, such~~as~~ cost, as the cost of utilizing the service, and the characteristics of the individual, such~~as~~ well as ~~their~~the income, social class, ethnicity, and mobility profile. ~~The~~ of potential users of services. From a geographical perspective, the spatial dimension ~~is key, and~~ considers the geographic distribution of available healthcare services ~~and the costs~~across the landscape, in addition to the cost or friction ~~involved in travelling to them.~~that potential users incur when trying to reach these services. By taking these ~~dimensions~~ geographical factors into account, estimates of accessibility can ~~help researchers, planners, and policy makers~~ identify areas with high or low ~~accessibility~~ to healthcare services ~~and~~. This, in turn, ~~can~~ provide ~~critical~~ valuable information related to social and spatial inequalities and guidance for health policy and resource allocation.

~~At a high level, provider-to-population ratios offer some indication of the level of service within a community. However, these measures lack a true spatial dimension. In contrast, gravity~~Spatial accessibility can be estimated in various ways. At a high level, provider-to-population ratios (PPR) offer an indication of the level of service within a community. These measures conceptualize a region as a container of population and services, and therefore are sometimes called container approaches. PPRs are straightforward to interpret as the supply of a service (say number of doctors, beds, etc.) divided by demand (say, number of people who require the service). Despite this convenient and intuitive interpretation, container approaches are limited in the amount of spatial information that they provide, especially if applied to large regions. When applied to smaller regions these approaches present other shortcomings, such as the assumption that the population in the container is captive and does not cross the boundaries of the container in search of services - and that users do not come into the container from other regions to avail themselves of local services.

An alternative to container approaches is provided by gravity measures. Gravity measures offer a more sophisticated approach to measuring spatial accessibility to healthcare [@Joseph1982]. Nevertheless, these methods have been criticized that moreover addresses some of the limitations of the container approach. Instead of defining rigid container boundaries, gravity measures consider the mobility characteristics of the public to produce flexible (and often overlapping) catchment areas for the difficulty involved in specifying a suitable distance decay function [e.g., @Schuurman2010]. From this both services and population. Accordingly, one of the most popular approaches to estimating healthcare accessibility in the previous literature is the Two-Step Floating Catchment Area (2SFCA) method proposed by Luo and Wang [-@Luo2003; @] after research by Radke and Mu [-@Radke2000], which. The 2SFCA method is based on a-an ensemble of two gravity models with a simplified gravity model with a-binary distance function. to account for crowding of facilities and allocation of levels of service. Numerous applications of this methods are found in the international literature, including work from Germany [@Bauer2017], South Korea [@Kim2018], Japan [@Fujita2017], China [@Song2013], Australia [@McGrail2009], and Canada [@Shah2016].

Within the 2SFCA framework, accessibility to healthcare is estimated across in two stages in the 2SFCA: in the first, the PPR step, a level of service at a given healthcare provider is determined based on the supply (e.g., number of physicians in a clinic) and the estimated demand from the surrounding population within some catchment area. This level of service resembles a local provider-to-population ratio (PPR). In the second stage step, the level of service for of different healthcare providers is summarized aggregated for the each population center. By simplifying gravity measures and operationalizing healthcare accessibility in terms of population-to-provider ratios, the properties of the demand and level of service, the 2SFCA method make it both intuitive and is appealing for health policy.

analysis. Still, several improvements have been made to the 2SFCA method since it was proposed that seek to address the method's most significant important perceived shortcomings. The result is a family of Floating Catchment Areas (FCA) methods that include more realistic conceptualizations of the friction of distance by specifying variable catchment area sizes [@McGrail2009] and/or the incorporation use of stepped [e.g., @Luo2009], continuous [e.g., @Dai2010], and adaptive [@Bauer2016] distance-decay functions. Other authors have added multi-modal transportation [e.g., @Mao2013], age-adjusted healthcare demand profiles [e.g., @Ngui2011], and as well as ways to counteract the modifiable areal unit problem [@Bell2013].

A second major focus area of FCA research, in addition to the improvements mentioned above, has been the addition introduction of competition for available opportunities or the allocation of services to regions. As we discuss further in the next section the population. More concretely, the original 2SFCA approach has been criticized for over-estimating the levels of demand [@Wan2012]. By summarizing] and/or level of service [@Delamater2013] in the system. This is a consequence of the way catchment areas for facilities and population centers typically overlap in any realistic spatial system - an artifact of FCA methods that can lead to misleading estimates of accessibility.

In effect, when aggregating the population within the overlapping catchment areaareas of healthcaremultiple facilities, the original 2SFCA framework often produces leads to double-counting of the population that tends to inflate estimates of the level of demand at supply points in the healthcare system, which. We call this effect demand inflation. Inflated demand, in turn-deflates, tends to deflate the level of service for populations within the study area. Proposed services by the facilities so affected. A similar effect, which we call level of service inflation, happens when the levels of service of various service points are aggregated for population centers. Ultimately, accessibility estimates are affected in potentially complex ways, depending on the geography of the problem [see numerous examples in @Delamater2013], and their interpretation as PPRs becomes suspect.

Various solutions to this problem include the the issues of demand and level of service inflation have been proposed, including the addition of selection weights based on a travel impedance function in what Wan et al. [-@Wan2012] refer to as the Three-Step Floating Catchment Area (3SFCA) method and a modified 3SFCA that uses[@Wan2012]; the use of a Huff model to probabalisticly generate probability-based estimate theseof the selection weights in the 3SFCA method [@Luo2014] based on impedance and the supply of physicians at a given facility. On; and, on the supply side (the allocation of the service to populations), Delamater [-@Delamater2013] proposes a modification to the, a modified 2SFCA (M2SFCA) method to address what he terms a “suboptimal” spatial configuration of services. [-@Delamater2013].

However, althoughIn this paper we are interested in the way demand and level of service are calculated in FCA methods. We review how different approaches deal with the issue of demand overestimation (or inflation) and the alloation of services has been correctly identified in the literature, there are not to our knowledge any existing methods that adequately solve these perceived shortcomings. As we will show, previous approaches inflate or deflate demand and supply to varying degrees. Crucially, this results in potentially misleading estimations of healthcare accessibility. As a consequence, potentially erroneous recommendations for health policy could result, including the imprecise identification of spatial inequalities.

In response, this research proposesinflation, and then propose a simple and intuitive approach to proportionally allocate supply and demand. Our solution consists on adjusting the impedance weights used in the estimation of FCA methods. ByMore concretely, by incorporating methods drawn from the field of spatial statistics and econometrics, we preserve-proportional allocation has the feature that it preserves the levels of demand and supplyservice in the systemand eliminate the inflation and deflation of these parameters in previous FCA approaches. To illustrate the benefitskey aspects of this approachour proposal, we conduct a case study of access to family physicians in Hamilton, Canada. Our results indicate that the proposed adjustments produce more intuitive measures of accessibility to healthcare measured in terms of local provider-to-population ratios.PPRs. Moreover, these outputs can be used to provide estimates of access disparity across a region that are both easily understood and robust to demand and supplylevel of service inflation.

## **Demand Inflation/Accessibility Deflation in FCA Methods**

### **Background: Floating Catchment Area Methods**

To motivate the discussion to follow we begin by quickly reviewing some popular FCA methods.

In general terms, FCA approaches are implemented as ensembles of two gravity models in two steps (2SFCA), using an impedance function to represent the cost required to overcome distance. Impedance functions implement a distance-decay effect that mimics a commonly observed cost-minimization behavior, namely that people in general prefer to spend less time/money/effort travelling to destinations. In this way, the impedance function defines a catchment area for the points of service and population centers alike.

In the first step of FCA methods, the impedance function defines catchment areas are defined for facilities (e.g.,  $j$ , which could be clinics, parks, libraries, etc.) by means of an impedance function. The A weighted sum of the population within a catchment area is allocated to the corresponding facility or service point. to represent demand. In the second step of the algorithm, the catchment areas are “floated” to population centers  $i$ . Accessibility at location  $i$  is calculated as the weighted sum of the level of service at every location  $j$  that includes  $i$  within its catchment area. The following methods are popular in the literature.

### **Two-Stage Floating Catchment Areas (2SFCA)**

The original 2SFCA implements a binary impedance function  $W$  with a threshold cost  $d_0$  as follows [see @Luo2003]:

$$W(d_{ij} \leq d_0) = \begin{cases} 1 & d_{ij} \leq d_0 \\ 0 & d_{ij} > d_0 \end{cases}$$

This creates a demand/congestion effect. The level of function assumes equal potential within a catchment area (i.e.,  $d_{ij} \leq d_0$ ), and zero beyond ( $d_{ij} > d_0$ ). This implies that 1) travellers are equally likely users of a service point within the catchment area, irrespective of how proximate or distant they are from it; and 2) no users travel to the service point from beyond the threshold cost.

Given the impedance function, the level of demand  $D_j$  is calculated as the weighted sum of the population at  $i$ :

$$D_j = \sum_i D_{ij} = \sum_i P_i W(d_{ij} \leq d_0)$$

The supply at the facilityS of the service offered at location  $j$  (say, number of beds/doctors in a clinic) is then divided by the demand to obtain a measure of level of service (e.g., beds/person, sq.m of park space/person, library floor space/person). In the second step, the catchment areas are “floated” to the population centers, and the accessibility at the

~~location is computed as the sum of the level of service of all service points that include the population center as part of their catchment areas. This gives a level of service  $L_j$  at the service point:~~

~~More formally, the level of service at location  $j$  is defined as follows (decomposed in various ways):~~

$$L_j = \frac{S_j}{D_j} = \frac{S_j}{\sum_i D_{ij}} = \sum_i \frac{S_j}{D_{ij}} = \sum_i L_{ij}$$

~~where  $S_j$  is the supply of the service offered at location  $j$  (say, number of beds/doctors in a clinic), whereas  $D_j$  is the level of demand on that service location. It is clear that the congestion effect results from the level of demand, which in turn~~ ~~The level of service resembles a PPR. Aggregation of demand creates a congestion effect that~~ depends on the number of potential users from different origins  $i$  that converge at service point  $j$ : at a fixed level of supply, greater demand results in lower levels of service. The different decompositions of  $L_j$  help to understand how different population centers influencecontribute to the level of demand at facility  $j$ .

~~The level of demand  $D_j$ . In the second step of the algorithm, catchment areas are “floated” to population centers  $i$ . A second gravity model is used to calculate the accessibility at  $i$ :~~

$$A_i = \sum_j L_j W(d_{ij} \leq d_0)$$

~~Since accessibility is calculated using a combination of the population at  $i$  and an impedance function  $W$  that depends on  $d_{ij}$ , an indicator of the cost of travel between  $i$  and  $j$  (e.g., distance, travel time, out-of-pocket expenses, or generalized cost):~~

$$D_j = \sum_i D_{ij} = \sum_i P_i W(d_{ij})$$

~~Accessibility at location  $i$ , in turn, is defined as the weighted sum (via the impedance function) of the level of service at every location  $j$  that includes  $i$  within its catchment area:~~

$$A_i = \sum_j L_j W(d_{ij})$$

~~The impedance function implements the geographical concept of distance decay, which reflects a commonly observed cost minimization behavior, namely that people in general prefer to spend less time than more travelling to destinations. In effect, the impedance function defines the catchment areas for the points of service and population centers alike.~~

~~In early implementations of the 2SFCA approach [see @Luo2003], a binary impedance function was used:~~

$$D_j = \sum_i D_{ij} = \sum_i P_i W(d_{ij}|d_0)$$

with:

$$W(d_{ij} \leq d_R) = \begin{cases} 1 & d_{ij} \leq d_R \\ 0 & d_{ij} > d_R \end{cases}$$

This formulation assumes equal potential for use within a catchment area, and zero beyond. In other words, travellers are assumed to be equally likely users of a service point within the catchment area, irrespective of how proximate or distant they are from facilities, it is conventionally interpreted as a PPR.

## Enhanced Two-Stage Floating Catchment Areas (E2SFCA)

A criticism of the binary impedance function of the 2SFCA is that it does not account for the declining probability of using a facility as distance grows. As a result of this criticism, other impedance functions have since been proposed, including the stepwise formulation of the Enhanced 2-StepTwo-Stage Floating Catchment Area method [E2SFCA; @Luo2009]:

$$D_j = \sum_i D_{ij} = \sum_{i=1}^N \sum_{r=1}^R P_r W(d_{ij}|d_1, d_2, \dots, d_R)$$

where  $W(d_{ij}|d_1, d_2, \dots, d_R)$  takes different values depending on the value of  $d_{ij}$  and cost threshold values  $d_r$ , as follows:

$$\$ \$ W(d_{ij}|d_1, d_2, \dots, d_R) = \left\{ \begin{array}{ll} k_1 & \text{if } d_{ij} \leq d_1 \\ k_2 & \text{if } d_1 < d_{ij} \leq d_2 \\ \dots & \dots \\ k_{R-1} & \text{if } d_{R-1} < d_{ij} \leq d_R \\ 0 & \text{if } d_{ij} > d_R \end{array} \right. \$ \$$$

Clearly, a stepwise function does not assume identical potential within the extent of the catchment area (i.e., the space contained within  $d_{ij} \leq d_R$ ), and better reflects empirical observations but rather declining potential with increasing cost of travel behavior. More recent research has introduced smooth functions to replace the stepwise approach. It is worthwhile noting that impedance functions have long been studied in geographical analysis in general [e.g., @Taylor1971], and accessibility research in particular [e.g., @Kwan1998], but. However, it is only relatively recently that alternative impedance functions have been incorporated in FCA approaches. These include, including continuous functions [e.g., @Dai2010] and mixtures of continuous and step functions [e.g., @Schuurman2010].

Among recent developments in Besides the literature use of a non-binary impedance function, the method remains the same. In the first step, demand is calculated as a weighted sum of the population within the catchment area:

$$D_j = \sum_i D_{ij} = \sum_i P_i W(d_{ij}|d_1, d_2, \dots, d_R)$$

Note that non-binary impedance functions discount the level of demand as a function of cost more rapidly than binary functions. How rapidly this happens depends on FCA

methods, the definition of the cutoff values  $d_1, d_2, \dots, d_R$  and weights  $k_1, k_2, \dots, k_{r-1}$  of the function.

In the second step of the algorithm, accessibility at  $i$  is calculated as the weighted sum of the level of service of service points  $j$ :

$$A_i = \sum_j \frac{S_j}{D_j} W(d_{ij}|d_1, d_2, \dots, d_R) = \sum_j L_j W(d_{ij}|d_1, d_2, \dots, d_R)$$

Again, the use of a non-binary impedance function discounts the level of service more rapidly compared to binary functions.

### Three-Stage Floating Catchment Areas (3STCA)

Wan et al. [-@Wan2012] ~~propose~~proposed a ~~3-Step~~Three-Stage Floating Catchment Area method (3SFCA) that aims at refining the estimates of level of demand and accessibility by ~~incorporating a means of the use of selection weight~~weights. This approach operates by introducing an additional step where selection weights are calculated as follows:

$$G_{ij} = \frac{WT(d_{ij})}{\sum_j \sum_{d_{ij} \leq d_0} T(d_{ij})}$$

These where  $T(d_{ij})$  are Gaussian weights (essentially an impedance function), and the summation in the denominator is for all sites  $j$  that are within a critical threshold  $d_0$ . Notice that a property of the selection weights are used to adjust is that their sum over  $j$  equals one:

$$\sum_j G_{ij} = 1$$

Given a set of selection weights, the level of demand on is calculated by this algorithm in the following manner:

$$D_j^* = \sum_i G_{ij} P_i W(d_{ij}|d_1, d_2, \dots, d_R) = \sum_i G_{ij} D_{ij}$$

Notice how demand in this method takes what is essentially the demand in the E2SFCA, and allocates it proportionally to service points  $j$ .

Accessibility, in the final step, becomes (with the subindices of the selection weights reversed, to reflect the displacement of the catchment area to population centers) is calculated in the following manner:

$$A_i^* = \sum_j G_{ji} \frac{S_j}{D_{ij}^*} W(d_{ij}|d_1, d_2, \dots, d_R) = \sum_j G_{ji} L_j^* W(d_{ij}|d_1, d_2, \dots, d_R)$$

## Modified Two-Stage Floating Catchment Areas (M2SFCA)

Delamater [-@Delamater2013] discusses the application of FCA methods for systems that are not optimally configured to service the whole population. To address this issue, he proposes a modification to the second step of the algorithm in this fashion 2SFCA algorithm that increases the friction of distance. Demand in this modification is the same as in 2SFCA. However, accessibility is calculated in the following manner:

$$D_{ij}^* = G_{ij} P_i W_{ij} = G_{ij} D_{ij}$$

The level of service at  $j$  is calculated in the second step as:

$$L_j = \frac{S_j}{\sum_i D_{ij}^*}$$

Accessibility, in the final step, becomes:

$$A_i = \sum_j G_{ij} L_j W_{ij}$$

Demand and Supply  $A_i = \sum_j L_j W(d_{ij}|d_1, d_2, \dots, d_R) W(d_{ij}|d_1, d_2, \dots, d_R) = \sum_j L_j (W(d_{ij}|d_1, d_2, \dots, d_R))^2$

In other words, the level of service is discounted by the square of the impedance function, thus increasing the rate of decay. This is done to reflect the possibility that some population centers may experience increased friction to reach destinations in suboptimally configured systems.

## Inflation Effects in FCA Methods

An important point in the implementation of FCA methods [see @Having reviewed a selection of FCA approaches, we now proceed to discuss the issue of inflation. Inflation has been identified, among others, by Wan et al. [-@Wan2012], is that demand tends to be] and Delamater [-@Delamater2013]. As discussed by these authors, inflation happens when demand or level of service are overestimated. This inflation is a consequence of the way in which  $D_j$  is calculated, which and  $A_i$  are calculated, with some population centers contributing to the level of demand at more than one facility and then the level of service of facilities allocated to multiple population centers. Calculating demand, in particular, generally fails to preserve the population. In other words, FCA methods lack, and therefore lacks the pycnophilactic property discussed by Tobler [-@Tobler1979]. In practical terms, this implies that the population used to calculate the demand component of accessibility level of service will often exceed (or but sometimes fall short) of the actual population in a region, depending on the weighting scheme. We term the consequent effect demand inflation.

Let us illustrate this inflation effect by means of a simple example using the conventional 2SFCA approach with a binary impedance function. In this case, the population value at  $i$  is

multiplied by zero or one, meaning that the contribution of  $i$  to demand at  $j$  whenever  $d_{ij}$  does not exceed the threshold is:

$$D_{ij} = P_i$$

~~However, when calculating If we concentrate for a moment on a single population center that enters the catchment areas of several service points (see Fig , left panel), we can see that when the demand at each of the service points is calculated, the population in question is added two times, and the levels of service are  $L_1 = L_2 = 1/s00.$~~

~~More generally, when calculating the level of service at  $L_j$  and  $L_k$ , this population is double-counted if  $d_{ij} \leq d_0$  and  $d_{ik} \leq d_0$ . More generally, the population at  $i$  contributes to the demand on multiple service points  $j$ , every time that  $d_{ij} \leq d_0$ . It  $d_0$  for any  $j$ . And, since since  $D_{ij} = P_i$ , it follows, then, that that the sum of the population to be serviced over all clinics is:~~

$$\sum_j D_{ij} = nP_t K_i P_i$$

where  $nK_i$  is the number of service points  $j$  that include  $i$  as part of their catchment areas. ~~Unfortunately, since (as noted above)  $D_{ij} = P_i$ , it turns out that the Therefore, the system-wide contribution of the population at  $i$  to the level of demand implied by the population at these calculations, vastly exceeds the actual population at  $i$  in these calculations, since:~~

$$\sum_j D_{ij} = nP_t K_i P_i > P_i$$

~~In other words, when estimating the level of accessibility with congestion by means of the 2SFCA approach, it appears that the population that needs to be serviced is substantially larger than the actual population. Clearly, this does not make sense, and perhaps worse, it may lead to gross underestimation of the actual level of service. In effect, demand is artificially higher via population inflation.~~

~~This can be more easily seen by means of a simple example.~~

~~Consider the situation shown in Fig (Panel I), with three clinics (labeled  $a$ ,  $b$ , and  $c$ ) and one population center (labeled 1). Let us consider next what happens when enhanced (i.e., non-binary) impedance weights are used. These functions aim to capture more realistically the rule that most members of the population prefer to travel shorter distances to reach a destination. For the example, assume a set of weights with decay as follows (see Fig , right panel) :~~

$$W(d_{ij}|d_1, d_2, \dots, d_R) = \begin{cases} 0.9 & d_{ij} \leq d_1 \\ 0.8 & d_1 < d_{ij} \leq d_2 \\ 0.4 & d_{R-1} < d_{ij} \leq d_R \\ 0 & d_{ij} > d_R \end{cases}$$

The population center in the example is relatively distant from the service points. Accordingly, its potential demand is reduced by assuming that some people do not travel at all. In this example, the contribution of the population center to demand is only  $0.8P$  to each clinic, and therefore the system-wide demand of this center is  $1.6P$  - less than the all-or-nothing allocation of the binary impedance weights, but still in excess of the actual population.

More generally, when calculating the level of service at  $j$  locations, the population at  $i$  contributes to demand every time that  $d_{ij}$  is within the service area for any  $j$ . The precise contribution depends on the weights in the distance-decay function and the position of the population center with relative to all service points. In a function with faster decay, the total demand attributed to  $i$  (i.e.,  $\sum_j D_{ij}$ ) can be less than the population of  $i$ . In other words, depending on the steepness of decay, the total demand can be greater than, equal to, or less than the population at  $i$ :

$$\sum_j D_{ij} \leq P_i$$

Clearly, only when the full population at  $i$  is allocated exclusively to one service point (i.e., when  $K_i = 1$ ) the implied demand equals the population - something that seldom happens in practical situations.

It is important to acknowledge that demand in accessibility analysis represents the potential for spatial interaction, not realized interaction. That said, the expectation that facilities need to serve multiple times the size of the population in a region can easily lead to misleading conclusions about the need for resources. A logical question, however, is whether the inflation of demand (with the consequence deflation of level of service) is not offset in the second step of the method, when the population at  $i$  has potential access to multiple service points?

Let us consider what happens in the second step of the algorithm in the example, when catchment areas are floated to the population center (see Fig ). When a binary impedance function is used, the aggregation of the level of service means that, despite the inflation of demand due to double-counting, accessibility matches the level of service as well as the regional PPR of 2/100 (left panel). In the case of the stepwise function, the level of implied demand is less than the population, but the population is also assumed to receive less of the available level of service. In this case, again, the accessibility matches the level of service despite the fact that segments of the population were assumed to not contribute to demand.

Clearly, the example is too simplistic (in fact just a variation of the container approach), and it is unclear what the implications would be for a system with even just a slightly more complex geography. To explore this, consider the addition of two population centers to the landscape (see Fig ). Notice how the three population centers are in the catchment areas of the two clinics. When the binary impedance function is used, demand at each clinic is calculated as 300, and demand over all clinics is therefore 600, or twice the population of the region. When the stepwise impedance function is used, the demand by each center is:

$$D_{1j} = 0.8 \times 100 + 0.8 \times 100 = 160$$

$$D_{2j} = 0.8 \times 100 + 0.4 \times 100 = 120$$

$$D_{3j} = 0.4 \times 100 + 0.8 \times 100 = 120$$

and the total load on the system is therefore 400, still well in excess of the total population of the region.

When demand is used to calculate the level of service, and then accessibility in the second step of the algorithm, the following occurs (see Fig.). When the binary impedance function is used (left panel), the level of service at each clinic is:

$$L_1 = \frac{10}{300} = 0.033$$

$$L_2 = \frac{10}{300} = 0.033$$

The level of service at the clinics is only half of the regional PPR, since each clinic is assumed to serve the *entire* population of the region. Unfortunately, since demand has been inflated for each clinic, these levels of service cannot be meaningfully interpreted as local PPRs. The sum over the clinics, on the other hand, is 20/300 - which is consistent with the regional PPR. Interestingly, as seen in the figure, the accessibility of each population center matches the regional PPR - but the sum of accessibility over all population centers exceeds the sum of the level of service over all the clinics as a consequence of allocating the same level of service to several population centers.

Continuing with the stepwise impedance function, we can see (Fig., right panel) that the levels of service are calculated as:

$$L_1 = \frac{10}{0.8 \times 100 + 0.8 \times 100 + 0.4 \times 100} = \frac{10}{200} = 0.05$$

$$L_2 = \frac{10}{0.8 \times 100 + 0.4 \times 100 + 0.8 \times 100} = \frac{10}{200} = 0.05$$

Notice how the level of service is higher in this case: this is a consequence of assuming (as the stepwise impedance function does) that some of the population does *not* demand service. Demand, however, is still inflated, and interpretation of the levels of service as local PPRs is still inappropriate. Accessibility is higher for population center 1 but lower for the two peripheral centers. Furthermore, the sum of accessibility over all population centers exceeds the sum of the level of service of all clinics in the region.

At issue is the interpretability of the levels of service, which as the example illustrates do not accurately represent PPRs, and how accessibility, which is a weighted sum of levels of service, cannot be interpreted as the PPR for a population center either.

Two methods reviewed above, namely the Three-Stage Floating Catchment Area method and the Modified Two-Stage Floating Catchment Area method aim to address the overestimation of demand and/or levels of service when calculating accessibility. As discussed previously, they do this by compounding the effect of the impedance function. In the case of 3SFCA, demand is deflated by assuming that demand declines more rapidly with

distance. Then, when calculating accessibility, the levels of service are allocated more locally, again, as a consequence of steeper distance-decay. In the case of M2SFCA, demand is not deflated, however, the levels of service are allocated more locally as a consequence of steeper distance-decay. In other words, these methods correct for inflation by assuming that fewer people demand health care services, and that the levels of service are allocated to fewer people too.

For comparison, the levels of service and accessibility for the example according to these two methods are shown in Fig . Notice how the levels of service in the 3FSCA are considerably higher as a consequence of excluding potential users with a steeper rate of decay. On the other hand, the levels of accessibility are also lower, as a consequence of allocating service more locally. The levels of service in the M2SFCA are identical to the E2SFCA, however, accessibility is lower, again as a result of allocating service more locally.

## A Simulated Example

The examples in the preceding section illustrate the way demand and level of service can be overestimated (and in some cases underestimated) in FCA algorithms. However, they are too simplistic to indicate what would happen in a realistic situation. In particular, it is possible that the consequences depend on the geography of the problem as the examples in Delamater [-@Delamater2013] suggest. Based on the way demand and level of service are allocated, we conjecture that the effects are likely more pronounced in areas with higher density of population and service, since inflation is a consequence of overlapping catchment areas. Furthermore, we conjecture that demand inflation will be reduced when stepwise/continuous distance-decay functions are used, since their effect is to reduce the overlap by reducing the contribution of population at different distances, and to allocate levels of service more locally as well. We explore these issues further by means of a simple but realistic simulated example.

The setup for the simulated example is shown in Fig . There are three clinics and nine population centers. Assume that the supply at the three clinics is identical, say one physician at each clinic 1, three physicians at clinic 2, and two physicians at clinic 3. Further, assume that the population at 1 is 100, 2, 8, and that this is 250; population center at 3, 4, and 6 is part of 250; and population at 5 and 7 is 1000. The total population in the catchment areas of the three service points/region therefore is 4,500. Under this setup, the level of service at each of the clinics is 1 physician per 100 people – even though the service points will not in reality serve one hundred patients each. It is more likely, instead, that each service point will serve only a fraction of all those patients, and that collectively they offer their services to the whole population within their catchment areas. As a consequence of how demand is calculated in the 2SFCA, the level of demand has been overestimated (i.e., demand has been inflated) and the level of service has been underestimated across the whole system is 1.33 physicians per thousand people, which we will refer to as the Regional PPR.

Does this inflation effect matter, though? After all, since accessibility is the sum of the level of service that a population center can reach, the accessibility at 1 is 3 physicians per

~~hundred people, as expected. The situation, however, is less clear-cut when population centers begin to interact with each other via the way they contribute to demand.~~

~~Let us add population center 2 to the previous example, and say that this center has a population of 50 (see Fig , Panel II). Moreover, the population at 2 is part of the catchment area of clinic c only. The level of service at clinics 1 and 2 is still 1 physician per hundred people. The level of demand at clinic 3, on the other hand, is now 150 people, and its level of service (and the accessibility of population center 2) has declined to 1 physician per 150 people—despite the fact that demand on c is likely lower than 150. For the purpose of calculating demand, the population of center 1 has been inflated by a factor of 3.~~

~~Ignoring the demand side for one moment, a similar inflation effect is present on the supply side when the same service point includes multiple population centers as part of its catchment area. In this case, the same level of service is assumed to be available to all those population centers, and therefore supply becomes inflated in the calculation of accessibility. In the example, the algorithm assumes that the level of service available at clinic c to population centers 1 and 2 is  $L_{1c} = L_{2c} = 1/150$ . However, the total level of service available at c for accessibility calculations is  $2/150$ —or twice the actual level of service at the clinic.~~

~~The situation above becomes more For this experiment, we consider binary and stepwise impedance functions. The former is simply the traditional 2SFCA method, whereas the latter is the Enhanced 2SFCA approach. The catchment areas for the first step of the algorithm (demand allocation) are shown in Fig (binary impedance) and Fig Fig (stepwise impedance). Notice that some population centers are inside the catchment areas of more than one clinic. For instance, Population Center 5 is in the catchment areas of Clinics 2 and 3, whereas Population Center 4 is in the catchment areas of all three clinics.~~

~~To see how the overlap of catchment areas impacts the calculations in the first step of the algorithm, we define impedance matrices using the same criteria as for the buffers seen in Fig . These matrices are shown in Table .~~

~~The demand for each clinic is calculated as the population of the centers multiplied by the values of the corresponding impedance weight with respect to that clinic, and then aggregated for all population centers. The level of service is the supply divided by the demand, multiplied by 1,000. The last row of the table shows the total population as well as the total demand at each clinic.~~

~~First we discuss the results according to the binary impedance function. As seen in Table , the population of Center 3 (which is in the catchment area of three clinics) is assumed to contribute 1,500 patients to the demand across the system, whereas Center 1 (which is in the catchment area of only one clinic) contributes exactly its population of 250. Since the population of several centers is counted multiple times, the apparent demand exceeds the actual population. In effect, when we calculate the total demand (the sum of the demand across clinics), we find that this is 9,750 according to the binary impedance function, which far exceeds the actual population.~~

Turning now to the stepwise function, we see that Center 3 contributes  $500 \times 0.242 + 500 \times 0.600 + 500 \times 0.242 = 542$  to the demand across the system, but Center 1 contributes only  $250 \times 0.242 = 60.5$ . The total demand now is 4,316, which is less than the total population.

This example illustrates a vexing when using enhanced impedance functions (i.e., non-binary). Such is the case of the enhanced 2SFCA approach (E2SFCA), whereby the contribution of the population at  $i$  to demand at service location  $j$  is:

$$D_{ij} = P_i W(d_{ij})$$

An example of an impedance function is a stepwise function as follows [see the slow decay function in @Luo2009]:

$$W(d_{ij}) = \begin{cases} 1 & d_{ij} \leq d_1 \\ 0.68 & d_1 < d_{ij} \leq d_2 \\ 0.22 & d_2 < d_{ij} \leq d_3 \\ 0 & \text{otherwise} \end{cases}$$

Let us revisit example (see Fig , Panel I) now using a stepwise impedance function. The population at 1 is still within the catchment area of the three clinics ( $a$ ,  $b$ , and  $c$ ), but now the impedance weights are  $W(d_{1a}) = 1$ ,  $W(d_{1b}) = 0.68$ , and  $W(d_{1c}) = 0.68$ . If, as before, the population at 1 is 100, the implied demand is inflated as follows:

$$\sum_j D_{1j} = 100 + 68 + 68 = 236 > P_1 = 100$$

Although inflation is lower in the case of the stepwise function compared to the binary function, the same location  $i$  can still potentially contribute multiple times its population value to different service points  $j$ , with similar consequences.

The implied demand of population center 2, on the other hand, is:

$$\sum_j D_{2j} = 0 + 0 + 34 = 34 < P_2 = 50$$

and therefore demand has been deflated.

This example demonstrates an internal contradiction effect in how FCA methods operate: when multiple service centers are within the threshold travel cost, they assume of a population center, it is assumed that some (maybe and possibly all) of the same persons crowd more than one of those centers service point, resulting in inflated demand and deflated levels of service. On the other hand, when only one service center is available, the assumption is that some individuals may stepwise or continuous functions (e.g., E2SFCA) are used to weigh down the population of distant population centers, the apparent effect is that some segments of the population do not demand service, even when the center is clinics are within their threshold travel cost. While this assumption This effect is even more marked in the case of 3SFCA, which produces considerably higher levels of service, as

a consequence of stacking the effects of two impedance functions. In effect, demand is deflated and the level of service is inflated. While the assumption that some members of the population drop out from the total demand pool may be acceptable for discretionary services, it is suspect when it comes to essential services such as many health care services, and particularly primary health care.

~~Continuing with this example, population center 2 (see Fig., Panel II) has accessibility of 1 physician per 118 people – higher than when the binary impedance function is used, but still likely biased for two reasons: 1) demand from center 1 is higher than the actual population; and 2) demand from center 2 is lower than the actual population.~~

~~Although it purports to fix this, the issue of demand and supply inflation is not resolved by the 3SFCA method. In this case, the selection weights are  $G_{1a} = 1/2.36$ ,  $G_{1b} = 0.68/2.36$ ,  $G_{1c} = 0.68/2.36$  for location 1, and  $G_{2a} = 0$ ,  $G_{2b} = 0$ ,  $G_{2c} = 0.68$  for location 2. The adjusted levels of demand are then:~~

$$\begin{aligned} D_{1a}^* &= G_{1a} D_{1a} = (1/2.36)(100) = 42.37 & D_{2a}^* &= G_{2a} D_{2a} = (0)(0) = 0 \\ D_{1b}^* &= G_{1b} D_{1a} = (0.68/2.36)(68) = 19.59 \text{ and } & D_{2b}^* &= G_{2b} D_{2b} = (0)(0) = 0 \\ D_{1c}^* &= G_{1c} D_{1a} = (0.68/2.36)(68) = 19.59 & D_{2c}^* &= G_{2c} D_{2c} = (1)(34) = 34 \end{aligned}$$

~~It follows that the demand aggregated by population center is:~~

$$\sum_j D_{1j}^* = D_{1a}^* + D_{1b}^* + D_{1c}^* = 42.37 + 19.59 + 19.59 = 81.55 < P_1 = 100$$

~~for center 1, and:~~

$$\sum_j D_{2j}^* = D_{2a}^* + D_{2b}^* + D_{2c}^* = 0 + 0 + 34 = 34 < P_2 = 50$$

~~for center 2.~~

~~It appears from this example that the 3SFCA method has deflated demand somewhat overzealously. This is perhaps not surprising, when we realize that the 3SFCA method operates, essentially, by stacking the effects of two related impedance functions (Delamater's [ @Delamater2013] Modified 2SFCA approach does the same but on the supply side). Recall that the selection weights are calculated using the impedance weights  $W(d_{ij})$ . These selection weights then multiply the impedance when computing the level of service (since  $D_{ij} = P_i W(d_{ij})$ ). The net effect is to make impedance steeper by a factor that depends on the impedance [the same mechanism is present in the Huff-enhanced model of @Luo2014, even if different impedance functions are used to compute the selection weights and the level of service/accessibility].~~

~~In addition to illustrating the demand/supply inflation (or deflation) effect, the examples in this section also suggest that the magnitude and even direction of the effect depend on the geography of the problem (i.e., the locations of population centers and clinics), as well as the kind of impedance function used. If the inflation effect were even (i.e., all accessibility estimates are inflated/deflated at the same rate) then a simple solution would be to apply a~~

~~constant inflation factor. Alas, it is more likely that the effect will not be even across space (e.g., we can anticipate that demand will be more inflated in denser parts of a region). For this reason, a more general approach to offset the inflation of supply and demand in a systematic way seems to be called for. We discuss this next.~~

Recall as well that the Regional Average PPR in this example is 1.33 physicians per thousand. If the total implied demand according to the binary impedance function is 9,750 the corresponding PPR is 0.615 physicians per thousand, or about half of the regional ratio. The corresponding PPR for the stepwise impedance function (implied demand = 4315.5) is 1.39 physicians per thousand, much closer to the Regional Average PPR. However, this PPR is misleading in that it assumes that some segments of the population are served multiple times, and some are not served at all.

Clearly, the first step of the algorithm can lead to inflation or deflation of the levels of demand. But do these matter? Or do they somehow average out when the levels of service are aggregated in the second step of the algorithm? Again, the situation is not clear-cut when multiple population centers and/or service clinics interact through overlapping catchment areas.

To illustrate this, we proceed to estimate the accessibility for the example using the binary and the stepwise impedance matrices. The results appear in Table .

Accessibility in the table is calculated as the level of service of the clinics multiplied by the values of the impedance function with respect to a population center, and then aggregated for all clinics. As seen in the table, the levels of accessibility vary considerably depending on the method. As anticipated, use of non-binary impedance functions reduces the inflation effect, and can even lead to deflation. Consider for instance the case of the binary impedance matrix: the total level of service in the system is the sum of the level of service at the three clinics, or 1.87. The level of service allocated to population centers, on the other hand, is the sum of the accessibility in the system, or 11.8. When using the stepwise impedance function, the total level of service in the system is 4.3, and the level of service allocated to population centers is 11.2. Compare this to the case of 3SFCA, where the total level of service in the system is 8.71, but the level of service allocated to population centers is only 2.21; or the case of M2SFCA, which estimates the total level of service in the system as 4.3 (same as E2SFCA) but allocates 5.74 to population centers.

Clearly, all the methods give qualitatively similar results, with peripheral centers displaying lower accessibility and more central places higher. But there are important differences in how demand and level of service are allocated throughout the system to calculate accessibility. Figure shows how the different methods penalize peripheral centers at different rates. And, since the demand is not consistent with the population and the accessibility is not consistent with the level of service of the clinics, it is difficult to interpret the results in terms PPRs. For instance, when we inspect the results for the binary impedance matrix (2SFCA), we can see in the table that the accessibility of Population Center 1 is simply the level of service of Clinic 1. But, as we saw before, this level of service was deflated by double counting the population of Centers 2, 3, and 4, which contribute to the calculation of demand at multiple clinics. Things become more complex as the number

of overlapping catchment areas grows. For example, Population Center 2 contributed to the congestion effect of two clinics. However, demand at one of those clinics was calculated using the population of eight out of nine population centers. What this suggests is that, at the very least, some population centers (likely those in the periphery of regions) will have artificially low accessibility levels as a consequence of demand inflation.

## A Method to Rectify for Proportional Allocation of Demand and Supply Levels to System-wide Totals

As the examples in the preceding subsection illustrate, FCA methods can induce quite substantial inflation of supply and demand (and even, in some cases, negative inflation, (or deflation),) of demand and level of service. This, in turn, can lead to artificially lower or higher affect the estimates of accessibility, as the case may be in potentially complex ways. The results, furthermore lack a clear interpretation. In this section, we propose a simple and intuitive adjustment to avoid the inflation artifacts inherent in current implementations of FCA methods.

Refer again to Fig . Demand inflation occurs because of the overlap in catchment areas - with the underlying assumption is that all service location services the population within its catchment area will be serviced by its corresponding service location. More realistically, only a fraction of that population will demand service at any given the location if other service points are within reach (i.e., inside its “floated” catchment area).

For instance, assuming (as the binary impedance function does), that individuals at location Population Center 1 are indifferent between clinics a, b, Clinics 1 and e2, then it is reasonable to think that the population will sort itself proportionally to the these two clinics - in this example, this means that each third half of the population will attend one of three two different clinics (importantly, this assumes that the service points services on offer are undifferentiated services; one would not generally consider cancer screening and hair removal clinics competitors). This suggests the following adjustment to the way the level of demand is calculated. Given an impedance function, a set of adjusted weights, say  $W_{ij}^{i*}$ , are precalculated by dividing the original impedance weights by the sum of the weights for population center  $i$  over all service points  $j$ :

$$W_{ij}^i = \frac{W_{ij}}{\sum_j W_{ij}}$$

Please notice that these weights are identical to the selection weights of the 3SFCA method. A key property of the adjusted weights is the following:

$$\sum_j W_{ij}^i = 1$$

This adjustment procedure has the effect that, when the level of demand of  $i$  is summed over all service points  $j$ , the aggregated level of demand due to  $i$  is identical to its population:

$$\sum_j P_i W_{ij}^i = P_i$$

As a result of standardizing the impedance weights, population is allocated *proportionally* to clinics.

On the supply side, inflation happens because the level of service available at location  $j$  is assumed to be available to every population center  $i$  within its catchment area. To adjust this, another set of weights, say  $W_{ij}^{j*}$ , is pre-calculated by dividing the original impedance weights  $W_{ij}$  by the sum of the weights for service point  $j$  over all population centers  $i$ :

$$W_{ij}^j = \frac{W_{ij}}{\sum_i W_{ij}}$$

Again, the resulting weights have the property that:

$$\sum_i W_{ij}^j = 1$$

As before, the result of this procedure is that, when the level of service of  $j$  is aggregated by population centers, the total level of service for that service point is preserved:

$$\sum_i L_j W_{ij}^j = L_j$$

Note that, since the weights add up to one, they can be interpreted as a *probability* or *frequency* of contact, similar to the Huff model of [@Luo2014].

In reference to Fig , (left panel), we can see that the original binary (unadjusted) weights for population centre Population Centre 1 are  $W_{11} = 1$ ,  $W_{12} = 1$ , and  $W_{13} = 1$ , whereas the weights of population center 2 are  $W_{21} = 0$ ,  $W_{22} = 0$ , and  $W_{23} = 1$ ,  $W_{31} = 1$ , and the weights of population center 3 are  $W_{31} = 1$ ,  $W_{32} = 1$ .

On the demand side, the adjusted weights become for population center Population Center 1,  $W_{11}^i = 1/3$ ,  $W_{12}^i = 1/3$ , and  $W_{13}^i = 1/3$ , and for population center 2,  $W_{21}^i = 1/2$ ,  $W_{22}^i = 0$ ,  $W_{23}^i = 1/2$ , and for Population Center 3  $W_{31}^i = 0$ , and  $W_{32}^i = 1/2$ . Using the adjusted weights, it can be seen that the level of demand due to each population center equals its respective population:

$$\begin{aligned} \sum_j D_{1j} &= 1/3P_1 + 1/3P_1 + 1/3P_1 = P_1 \\ \sum_j D_{2j} &= 0 + 0 + P_2 = P_2 \end{aligned}$$

$$\begin{aligned}\sum_j D_1 j &= 1/2P_1 + 1/2P_1 = P_1 \\ \sum_j D_2 j &= 1/2P_2 + 1/2P_2 = P_2 \\ \sum_j D_3 j &= 1/2P_3 + 1/2P_3 = P_3\end{aligned}$$

Coming next to the supply side, the adjusted weights for ~~service point  $a$~~ Clinic 1 are  ~~$W_{1a}^{j*} = 1$~~ ,  ~~$W_{11}^{j*} = 1/3$~~ ,  ~~$W_{21}^{j*} = 1/3$~~ , and  ~~$W_{2a}^{j*} = 0$~~ ;  ~~$W_{31}^{j*} = 1/3$~~ ; for ~~service point  $b$~~ Clinic 2 the adjusted weights are  ~~$W_{1b}^{j*} = 1$~~  and  ~~$W_{2b}^{j*} = 0$~~ , and for service point  $c$  are  ~~$W_{12}^{j*} = 1/3$~~ ,  ~~$W_{1b}^{j*} W_{22}^{j*} = 1/23$~~ , and  ~~$W_{2b}^{j*} W_{32}^{j*} = 1/23$~~ . It can be seen that the level of service is preserved across clinics, and therefore across the system:

$$\begin{aligned}\sum_i L_{1a} &= L_{1a} + 0 = L_a \\ \sum_i L_{1b} &= L_{1b} + 0 = L_b \\ \sum_i L_{1c} &= L_{1c}/2 + L_{2c}/2 = L_c \\ \sum_i L_{i1} &= L_1/3 + L_1/3 + L_1/3 = L_1 \\ \sum_i L_{i2} &= L_2/3 + L_2/3 + L_2/3 = L_2\end{aligned}$$

The method to adjust the weights used above is identical to a procedure that will be familiar to readers acquainted with the literature in the fields of spatial statistics and econometrics. The same adjustment is widely used there under the names of row- and column-standardization of a weights matrix [see, inter alia, @Anselin1988; @Griffith1988].

The proposed adjustment can be easily implemented. We will present next the implementation using a compact matrix notation. Begin by defining the following impedance matrix:

$$\mathbf{W} = \begin{pmatrix} W_{11} & \cdots & W_{1J} \\ \vdots & \ddots & \vdots \\ W_{N1} & \cdots & W_{NJ} \end{pmatrix}$$

where  $W_{ij}$  is an impedance function evaluated at  $d_{ij}$ . Subindex  $i$  is for population centers ( $i = 1, \dots, N$ ) and subindex  $j$  is for service points ( $j = 1, \dots, J$ ). Note that the matrix does not need to be square. A row-standardized set of weights is obtained as follows:

$$\mathbf{W}^i = \begin{pmatrix} \frac{W_{\overline{11}}}{\sum_j D_{\overline{1j}}} & \frac{W_{11}}{\sum_j W_{1j}} & \cdots & \frac{W_{\overline{1J}}}{\sum_j D_{\overline{1j}}} & \frac{W_{1J}}{\sum_j W_{1j}} \\ \vdots & \ddots & & \vdots & \vdots \\ \frac{W_{\overline{N1}}}{\sum_j D_{\overline{Nj}}} & \frac{W_{N1}}{\sum_j W_{Nj}} & \cdots & \frac{W_{\overline{NJ}}}{\sum_j D_{\overline{Nj}}} & \frac{W_{NJ}}{\sum_j W_{NJ}} \end{pmatrix}$$

Next, a column-standardized set of weights is calculated as:

$$\mathbf{W}^j = \begin{pmatrix} \frac{W_{\overline{11}}}{\sum_i D_{\overline{i1}}} & \cdots & \frac{W_{\overline{1J}}}{\sum_i D_{\overline{iJ}}} \\ \vdots & \ddots & \vdots \\ \frac{W_{\overline{N1}}}{\sum_i D_{\overline{i1}}} & \cdots & \frac{W_{\overline{NJ}}}{\sum_i D_{\overline{iJ}}} \end{pmatrix}$$

$$\mathbf{W}^j = \begin{pmatrix} \frac{W_{11}}{\sum_i W_{i1}} & \cdots & \frac{W_{1J}}{\sum_i W_{iJ}} \\ \vdots & \ddots & \vdots \\ \frac{W_{N1}}{\sum_i W_{i1}} & \cdots & \frac{W_{NJ}}{\sum_i W_{iJ}} \end{pmatrix}$$

In the first example above (see Fig ), the binary impedance matrix is:

$$\mathbf{W}_b = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{W}_{binary} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$$

The row-standardized weights that correspond to this matrix are:

$$\mathbf{W}_b^i = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{W}_{binary}^i = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

and the column-standardized weights are:

$$\mathbf{W}_b^j = \begin{pmatrix} 1 & 1 & 1/2 \\ 0 & 0 & 1/2 \end{pmatrix}$$

~~In the second example (see Fig ), the~~  $\mathbf{W}_{binary}^j = \begin{pmatrix} 1/3 & 1/3 \\ 1/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix}$

The stepwise impedance weights in the example are:

$$\mathbf{W}_s = \begin{pmatrix} 1 & 0.68 & 0.68 \\ 0 & 0 & 0.68 \end{pmatrix}$$

$$\mathbf{W}_{stepwise} = \begin{pmatrix} 0.8 & 0.8 \\ 0.8 & 0.4 \\ 0.4 & 0.8 \end{pmatrix}$$

The row-standardized weights in turn are (with some rounding):

$$\mathbf{W}_s^i = \begin{pmatrix} 0.424 & 0.288 & 0.288 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{W}_{stepwise}^i = \begin{pmatrix} 1/2 & 1/2 \\ 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}$$

whereas the column-standardized weights are:

$$\mathbf{W}_s^j = \begin{pmatrix} 1 & 1 & 1/2 \\ 0 & 0 & 1/2 \end{pmatrix}$$

$$\mathbf{W}_{stepwise}^j = \begin{pmatrix} 4/10 & 4/10 \\ 4/10 & 2/10 \\ 2/10 & 4/10 \end{pmatrix}$$

Once that the impedance weights have been adjusted, a vector of adjusted level of demand  $\mathbf{D}^*$  can be obtained by multiplying the *transposed* impedance matrix by a vector of population values as follows:

$$\mathbf{D}^* = [\mathbf{W}^i]^T \mathbf{P}$$

where the  $^T$  operator is for “transpose”, and  $\mathbf{P}$  is:

$$\mathbf{P} = \begin{pmatrix} P_1 \\ \vdots \\ P_N \end{pmatrix}$$

The level of demand for the service points in the binary impedance function example is (in vector form):

$$\mathbf{D}_b^* = \begin{pmatrix} 1/3 & 0 \\ 1/3 & 0 \\ 1/3 & 1 \end{pmatrix} \begin{pmatrix} 100 \\ 50 \end{pmatrix} = \begin{pmatrix} 100/3 \\ 100/3 \\ 250/3 \end{pmatrix}$$

$$\mathbf{D}_{binary}^* = \begin{pmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 100 \\ 100 \\ 100 \end{pmatrix} = \begin{pmatrix} 300/2 \\ 300/2 \\ 300/2 \end{pmatrix} = \begin{pmatrix} 150 \\ 150 \\ 150 \end{pmatrix}$$

Notice how each clinic is expected to service only 150, and the level of demand over the system is identical to the total population.

The level of demand for the service points in the stepwise impedance function example is (in vector form):

$$\mathbf{D}_s^* = \begin{pmatrix} 0.424 & 0 \\ 0.288 & 0 \\ 0.288 & 4 \end{pmatrix} \begin{pmatrix} 100 \\ 50 \\ 100 \end{pmatrix} = \begin{pmatrix} 42.4 \\ 28.8 \\ 78.8 \end{pmatrix}$$

$$\mathbf{D}_{sw}^* = \begin{pmatrix} 1/2 & 2/3 & 1/3 \\ 1/2 & 1/3 & 2/3 \\ 100 & 100 & 100 \end{pmatrix} = \begin{pmatrix} 50 + 200/3 + 100/3 \\ 50 + 100/3 + 200/3 \\ 150 \end{pmatrix} = \begin{pmatrix} 150 \\ 150 \\ 150 \end{pmatrix}$$

As can be seen, the aggregated level of demand, after the adjustment, equals (as desired) the total actual population of the region. In the case of the stepwise function, total demand has been adjusted to the population of the region without the restrictive assumption that some people are excluded from the system. This is achieved by assuming an assortative process that leads to proportional allocation of the demand.

The levels of demand can then be introduced in used to calculate the next stage level of service at the accessibility calculations individual clinic locations by performing Hadamard division ( $\oslash$ ) of the vector of supply by the vector of adjusted demand. This is the first step of the 2SFCA (aggregating demand over catchment areas for service points):

$$\mathbf{L}^* = \mathbf{S} \oslash \mathbf{D}^*$$

Notice that levels of service are given in terms of units of supply per units of demand, say physicians per person. This can be interpreted as the provider-to-population ratio from the perspective of each service point. These values are consistent with the regional provider-to-population ratio (RPRR), since:

$$RPRR = \frac{\sum_j S_j}{\sum_t D_t^*} = \frac{\sum_j S_j}{\sum_t P_t}$$

The average of  $L^*$ , on the other hand, can be interpreted as the average clinic level of service, considering the spatial distribution of supply and demand.

Since Hadamard division is an element-by-element operation, the adjusted levels of service in the first example (using the binary impedance function) are:

$$\mathbf{L}_b^* = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \oslash \begin{pmatrix} 100/3 \\ 100/3 \\ 250/3 \end{pmatrix} = \begin{pmatrix} 3/100 \\ 3/100 \\ 3/250 \end{pmatrix} = \begin{pmatrix} 0.03 \\ 0.03 \\ 0.012 \end{pmatrix} \begin{pmatrix} 10 \\ 10 \\ 150 \end{pmatrix} \oslash \begin{pmatrix} 150 \\ 150 \\ 150 \end{pmatrix} = \begin{pmatrix} 10/150 \\ 10/150 \\ 10/150 \end{pmatrix} = \begin{pmatrix} 0.067 \\ 0.067 \\ 0.067 \end{pmatrix}$$

The levels of service in the second example, when using the stepwise impedance function, are also:

$$\mathbf{L}_s^* = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \oslash \begin{pmatrix} 42.4 \\ 28.8 \\ 78.8 \end{pmatrix} = \begin{pmatrix} 0.024 \\ 0.035 \\ 0.013 \end{pmatrix}$$

$$\text{Note } \mathbf{L}_{sw}^* = \begin{pmatrix} 10 \\ 10 \\ 150 \end{pmatrix} \oslash \begin{pmatrix} 150 \\ 150 \\ 150 \end{pmatrix} = \begin{pmatrix} 10/150 \\ 10/150 \\ 10/150 \end{pmatrix} = \begin{pmatrix} 0.067 \\ 0.067 \\ 0.067 \end{pmatrix}$$

Unlike the 2SFCA, E2SFCA, and 3SFCA methods that the values of level produce levels of service that resemble PPRs but with values that are inconsistent with total demand given

the population, this operation returns values that are genuinely local PPRs that are consistent with the demand and supply population of the region. As we saw above, the demand equals the population. Here, the supply also equals the number of physicians in the region. Indeed Because both demand and supply are not inflated or deflated in this rectified method, these values are consistent with easily interpretable relative to the regional provider-to-population ratio Regional Average PPR of  $3/150/20/300$  or  $0.02067$  physicians per person. With this in mind In the case of the example, it is clear that locations a and b offer better provider both clinics have PPRs that are identical to population ratios than the regional average, whereas location c offers a lower provider to population than the regional average the Regional Average PPR.

Accessibility, finally, is calculated as the matrix product of the column-standardized weights and the adjusted level of service:

$$\mathbf{A}^* = \mathbf{W}^j \mathbf{L}^*$$

which, continuing with the example, gives the following for the binary impedance function:

$$\begin{aligned} \mathbf{A}_b^* &= \begin{pmatrix} 1 & 1 & 1/2 \\ 0 & 0 & 1/2 \\ 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0.03 \\ 0.03 \\ 0.012 \end{pmatrix} = \begin{pmatrix} 0.03 + 0.03 + 0.006 \\ 0 + 0 + 0.006 \\ 0.006 \end{pmatrix} = \begin{pmatrix} 0.066 \\ 0.006 \\ 1/3 \end{pmatrix} \begin{pmatrix} 1/3 & 1/3 \\ 1/3 & 1/3 \\ 10/150 & 10/150 \end{pmatrix} \\ &= \begin{pmatrix} 10/450 + 10/450 \\ 10/450 + 10/450 \\ 10/450 + 10/450 \end{pmatrix} = \begin{pmatrix} 0.044 \\ 0.044 \\ 0.044 \end{pmatrix} \end{aligned}$$

and for Notice how the sum of accessibility over the region is consistent with the total level of service over all clinics (i.e., 0.133). The level of service has been allocated in its totality.

When using the stepwise impedance function, accessibility is calculated as:

$$\begin{aligned} \mathbf{A}_b^* &= \begin{pmatrix} 1 & 1 & 1/2 \\ 0 & 0 & 1/2 \\ 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0.024 \\ 0.035 \\ 0.013 \end{pmatrix} = \begin{pmatrix} 0.024 + 0.035 + 0.0065 \\ 0 + 0 + 0.0065 \\ 0.0065 \end{pmatrix} = \begin{pmatrix} 0.0655 \\ 0.0065 \\ 1/3 \end{pmatrix} \begin{pmatrix} 4/10 & 4/10 \\ 4/10 & 2/10 \\ 2/10 & 4/10 \end{pmatrix} \begin{pmatrix} 10/150 \\ 10/150 \end{pmatrix} = \\ &= \begin{pmatrix} 4/150 + 4/150 \\ 4/150 + 2/150 \\ 2/150 + 4/150 \end{pmatrix} = \begin{pmatrix} 0.053 \\ 0.040 \\ 0.040 \end{pmatrix} \end{aligned}$$

Again, the sum of accessibility is consistent with the level of service available from all clinics in the region. As with the Local PPRs, accessibility is interpreted as population-to-provider ratios as perceived from the perspective of for each population center, in such a way that all calculations are with total demand and total level of service. In particular, accessibility can be interpreted as the share of level of service that a population center receives from all the clinics that service it.

~~For the sake of comparison, levels of service and therefore accessibility are local versions of the provider reported for the simulated example in Tables and .~~

~~An important point to population ratio. Furthermore, the total accessibility in the system remark is the following. The use of row- and column-standardized impedance weights assumes that the full population of every population center within the catchment of a clinic will receive service. However, the allocation, although proportional, is different when binary or stepwise impedance weights are standardized. When binary weights are employed, the underlying idea is that potential for use is identical within the catchment area irrespective of distance. When stepwise weights are used, proportionally more of the population is allocated to the system-wide level of service, closer clinics. Depending on the definition of cost of travel, this allows a research to accommodate directional effects as well. For example, use of network travel time would tend to favor movement away from congested locations.~~

## **Suboptimal Systems**

~~The research of Delamater [-@Delamater2013] illustrates how accessibility estimates can be misleading when systems are not optimally configured. We understand this to mean that some population centers are located too far away from service points to actually benefit from them. In the modified 2SFCA method (M2SFCA), Delamater addresses this issue by increasing the friction of distance. A slight inconsistency in this approach is that some of the centers that contribute to demand fail to benefit from the service due to the increased friction to which in turn is consistent with the supply and demand values of the system. To summarize, then, the adjustment the allocation of the level of service is subjected. Our suggestion in the case of suboptimal systems is to use an impedance function that reflects limiting conditions. For instance, in urban settings a travel time longer than 2 hours might be considered too long to be serviced by any clinic.~~

## **System Efficiency**

~~The approach proposed succeeds at preserving the level of demand to the total population, and the level of service to the total resources available in the system. In this way, it in this paper allocates population and level of service proportionally and exactly. This assumes that the population sorts itself into clinics in the most efficient way. But what if some members of the population lack full information about the spatial distribution of clinics? Or have some bias towards centric locations? The vagaries of human behavior could create excess demand in some locations, and as a consequence supply surpluses in others. Situations like this can be accommodated in a relatively straightforward way using our approach.~~

~~Here, we describe the use of *slack factors*. Demand and level of service are allocated proportionally and exhaustively (i.e., 100%). But the standardization could allow for some slack, by inflating demand and/or supply in a controlled way.~~

~~Our proposal to standardize the weights was as follows, for the case of rows and columns respectively:~~

$$\mathbf{W}^i = \begin{pmatrix} \frac{W_{11}}{\sum_j W_{1j}} & \dots & \frac{W_{1J}}{\sum_j W_{1j}} \\ \vdots & \ddots & \vdots \\ \frac{W_{N1}}{\sum_j W_{Nj}} & \dots & \frac{W_{NJ}}{\sum_j W_{Nj}} \end{pmatrix} \text{ and } \mathbf{W}^j = \begin{pmatrix} \frac{W_{11}}{\sum_i W_{i1}} & \dots & \frac{W_{1J}}{\sum_i W_{iJ}} \\ \vdots & \ddots & \vdots \\ \frac{W_{N1}}{\sum_i W_{i1}} & \dots & \frac{W_{NJ}}{\sum_i W_{iJ}} \end{pmatrix}$$

A set of slack factors, say  $k_i^i$ , could be introduced in the following manner:

$$\mathbf{W}^i = \begin{pmatrix} \frac{k_1^i W_{11}}{\sum_j W_{1j}} & \dots & \frac{k_1^i W_{1J}}{\sum_j W_{1j}} \\ \vdots & \ddots & \vdots \\ \frac{k_N^i W_{N1}}{\sum_j W_{Nj}} & \dots & \frac{k_N^i W_{NJ}}{\sum_j W_{Nj}} \end{pmatrix}$$

A value of  $k_1^i = 1.10$  would inflate the demand of population center  $i = 1$  by 10%, whereas a value of  $k_1^i = 1.20$  would inflate the demand by 20%. In a similar way, a set of slack factors  $k_i^j$  could be introduced to modulate the allocation of supply:

$$\mathbf{W}^j = \begin{pmatrix} \frac{k_1^j W_{11}}{\sum_i W_{i1}} & \dots & \frac{k_J^j W_{1Nj}}{\sum_i W_{iJ}} \\ \vdots & \ddots & \vdots \\ \frac{k_1^j W_{N1}}{\sum_i W_{i1}} & \dots & \frac{k_J^j W_{NJ}}{\sum_i W_{iJ}} \end{pmatrix}$$

A value of  $k_1^j = 0.9$ , for example, would deflate the supply of clinic  $j = 1$  by 10%. The use of slack factors provides a more intuitive interpretation of accessibility an interesting way of modulating demand and level of service allocation in a very precise and controlled way, and presents interesting opportunities as well to introduce expert opinion or other empirical approaches to calibrate slack factors.

## Empirical Example

In this and the following sections reminder of the paper we present an empirical example to illustrate the application of the methods presented above. Based on the preceding discussion above, the adjusted 2SFCA method employed in this research can be summarized as:

$$L_j = \sum_i \frac{S_j}{P_i W_{ij}^i}$$

with the incorporation of the row-standardized impedance weights  $W_{ij}^i$  in the first step, and:

$$A_i = \sum_j L_j W_{ij}^j$$

with the column-standardized impedance weights  $W_{ij}^j$  incorporated into the second step. The same approach is used to re-weight the impedance function for the stepwise approach (i.e., E2SFCA).

The case study is based on accessibility to family physicians in the Hamilton [Census Metropolitan Area \(CMA\)](#), in Ontario, Canada ~~using. For this, we use~~ data collected about the distribution of the population and primary health care clinics (i.e., family physicians) in the region. Time use data from Canada's General Social Survey ([GSS](#)) was also used to inform the selection of thresholds for the impedance functions. The data collection and preprocessing protocols are described next.

## Family Physicians and Clinic Locations

In regards to the supply of clinics, the locations of family physicians were obtained using the College of Physicians and Surgeons of Ontario (CPSO) database for the Province of Ontario. We chose this organization because all physicians practicing in Ontario are required to register with the CPSO, as set out in the Ontario Regulation 865/93: Registration [[@Ontario1991](#)].

Our search of CPSO's database was conducted attending to the following criteria.

- 1) Only physicians who are registered as family physicians were selected (this excluded specialists such as pediatric physicians).
- 2) The spatial extent of the search was determined using forward sortation areas (FSAs), which are the first three initial characters of a postal code. Using a GIS, the regions of interest were selected by choosing FSAs within a 10 kilometer buffer distance from the Hamilton CMA boundary. This involved 72 different FSA regions. Each FSA region code was then searched in the CPSO database in addition to the family physician specification.

The parameters of the search were deliberately conservative, and the search identified a total of 2,224 family physicians practicing in the region, of which, 864 are located in the Hamilton CMA. The resulting dataset was manually verified by the third author to ensure that the information was consistent and suitable for geocoding. Prior to geocoding, locational information was organized in three columns, containing street address, city name, and province name. After family physicians were geocoded, locations were further examined. When family physicians overlapped or were within a 50 meter distance of each other we merged the records to identify 535 unique locations that we term "clinics". Many of these clinics are not in the Hamilton CMA proper, but provide a buffer to minimize edge effects in the analysis. The distribution of clinics and family physicians is shown in Fig for the Hamilton CMA.

## Population

Population information was obtained from the 2011 Canadian Census. To maximize the spatial resolution, population data were acquired at the Dissemination Area (DA) level of geography for all DAs within the selected FSAs. As a result, this includes DAs not in the Hamilton CMA proper, but that provide a buffer against edge effects. From this, the region contains a population of 2,959,090, of which 720,725 are in the Hamilton CMA. The distribution of population in the Hamilton CMA is shown in Fig .

## Travel Time Matrix

Calculation of impedance weights requires that we evaluate an impedance function at values of  $d_{ij}$ , that is, the cost of travel between DA  $i$  and clinic  $j$ . In this research we used travel time as our cost variable. To this end, we computed a matrix of travel times measured over the road network. To calculate the travel time between population centers and clinics we used the DA centroids and the geocoded locations of clinics. Shortest paths on the network and subsequently travel times were computed using a Geographic Information System.

## Impedance Functions

For the experiments we ~~used~~use two different impedance functions, corresponding to the 2SFCA and E2SFCA algorithms. We do not implement the 3SFCA or the M2SFCA methods because, as noted above, they are equivalent to using steeper impedances. For the 2SFCA approach, impedance is given by a binary function, whereas for E2SFCA it is given by a stepwise function. The impedance functions require that we define cost (i.e., travel time) thresholds to implement them. To select the thresholds, we retrieved time use data from Canada's General Social Survey Cycle 24 (see <http://odesi2.scholarsportal.info/webview/>).

From the time use files, we filtered all activity episodes corresponding to respondents living in CMAs/CAs (metropolitan regions) in Ontario. Next, we filtered all episodes taking place in a car (as driver) while traveling for personal care activities for household adults (which includes traveling to see a doctor) and traveling for shopping or obtaining services (which includes traveling to go to health clinic or doctor's office). It is worthwhile noting that travel by car accounts for over 95% of trips for the selected purposes in Ontario CMAs/CAs.

Once episodes were filtered by mode of travel and purpose of the trip, their durations (in minutes) were examined by means of quantile analysis, using episode weights to ensure the representativeness of the analysis. From the results, we learned that 50% of all trips by car for the aforementioned purposes are less than 15 minutes long, and we selected this value as the threshold  $d_0$  for the binary function. In other words, this part of the analysis assumes that any person who has to travel longer than 15 minutes to reach a clinic is outside ~~of~~its catchment area. We deem this value appropriate for the scale, density, and level of congestion of Hamilton CMA.

Quantile analysis of trip durations was also used to calibrate a Gaussian function with standard deviation set at 15 minutes, to match the value selected for the binary impedance above. This produced the following stepwise function, with any trips longer than 45 minutes assumed to be outside of catchment:

$$W(d_{ij}) = \begin{cases} 0.946 & d_{ij} \leq 5 \\ 0.801 & 5 < d_{ij} \leq 10 \\ 0.607 & 10 < d_{ij} \leq 15 \\ 0.411 & 15 < d_{ij} \leq 20 \\ 0.135 & 20 < d_{ij} \leq 30 \\ 0.011 & 30 < d_{ij} \leq 45 \\ 0.00 & 45 < d_{ij} \end{cases}$$

Notice how the stepwise function has weights greater than 0.5 for  $d_{ij} \leq 15\text{min}$  and less than 0.5 for  $d_{ij} > 15\text{min}$ . This means that it will count fewer people than the binary function when  $d_{ij} \leq 15\text{min}$ , but more when  $d_{ij} > 15\text{min}$ .

## Results

We begin our discussion of the results by noting that with a total population of the region of 2,959,090 and 2,222 family physicians, the regional provider-to-population Regional Average PPR ratio is 0.751 family physicians per 1,000 people. This value is somewhat lower than the value of 1.16 for Ontario reported by CIHI [-@CIHI2018] and lower than the 1.20 estimated based on the population and physician data for the Hamilton CMA, which we attribute to our conservative search criteria of family physicians in the rest of the region.

The levelnominal levels of demand is, service, and accessibility are calculated for the 2SFCA and E2SFCA using both the unadjusted and adjusted impedance matrices. Table reports summarizes the total level of demand calculated results by each impedance matrix. As seen there, when no adjustment is made, the nominal demand explodes to several times the actual population in the region. However, when the impedance weights are standardized, demand is now only slightly less than the total population for the region, since the system is not optimal in the sense discussed by Delamater [-@Delamater2013], and a small proportion of the population turns out to be outside of catchment areas. The totalnominal demand under binary impedance is lower due to the more strictstricter catchment area condition (i.e., less than 15 minutes), compared to the stepwise function (i.e., less than 45 minutes), which. This, in turn, is somewhat lower than the total demand in the regional provider-to-population ratioRegional Average PPR, which does not impose catchment area constraints within the region.

It is clear that the rectified demand leads to results that are considerably more realistic. This is also than the conventional approaches. In addition to the nominal system-wide demand, this is seen as well when calculating the regional provider-to-population ratios for each case (i.e., Family Physicians per 1,000 people). As seen in the table, the mean levels of

service for clinics in the region in the case of the adjusted binary and stepwise weights are in line with their corresponding Regional Average PPRs. Since the levels of service in the case of the adjusted weights can be interpreted as local PPRs, this indicates that the average clinic offers approximately the same level of service as the regional system does for the whole population. Furthermore, the mean accessibility of a DA according to the adjusted weights is identical to the mean LOS: this is because the LOS is allocated completely to DAs. The total LOS and accessibility in the region match when the adjusted weights are used. This is not the case when the unadjusted weights are used. Clearly, the use of the unadjusted weights can lead to a substantial amount of accessibility inflation, by factors as high as five or six times the estimates of the proposed proportional allocation approach.

These results demonstrate how inflation of the supply (i.e., the level of service) leads to much higher values of accessibility in the case of the conventional 2SFCA and E2SFCA methods. The procedure to rectify the population and level of service, on the other hand, leads to accessibility outputs that are consistent with the regional population and overall supply of health care services. This, in turn, makes interpretation of the output more robust and intuitive.

Another important issue is that spatial distribution of inflation of demand and level of service. If inflation happened in a uniform way, the upward bias in the estimates could to some extent be ignored, as long as relative differences by location remained relatively constant. Unfortunately, as seen in Fig and Fig , demand inflation is far from uniform. In fact, inflation of demand tends to happen, as per our earlier conjecture, in areas with higher population density. Inflation factors are also substantially higher when the binary impedance function is used. Since this function lacks a gradual distance-decay mechanism, it is more generous in terms of counting population serviced. Notice the magnitude of the inflation factors: since the inflation of demand depends on the number of overlapping catchment areas, a factor of 160, for instance, would suggest that a clinic is expected to simultaneously serve approximately that number of DAs in the conventional 2SFCA method, and a proportionally similar number in the conventional E2SFCA method.

The next step in the algorithm is to calculate the levels of service, that is, the number of physicians (supply) by level of demand. Since in the case of the adjusted impedance weights the demand is rectified to the population, the level of service is likely going to be higher than when the unrectified demand is used as in the conventional 2FSCA and 3SFCA implementations. Higher levels of service, however, do not necessarily translate in the second step of the algorithm into higher accessibility, since levels of service are also rectified so that total supply is not exceeded.

Accessibility maps for the implementation of 2SFCA are shown in Fig and Fig . The map of accessibility for the implementation of 2SFCA is shown in Fig and with the adjusted weights for proportional allocation in Fig . The general patterns observed in the figures are as expected, with higher accessibility in denser, better connected parts of the region. Relatively high accessibility in the north and west of the CMA is due to proximity to other major population centers such as Oakville, Kitchener, and Waterloo.

~~The results demonstrate how inflation of the supply (i.e., the level of service) leads to much higher values of accessibility in the case of the binary 2SFCA method. The procedure to rectify the population and level of service, on the other hand, leads to accessibility output that is in line with the regional system-wide provider-to-population ratio. A question, however, is the degree of inflation of accessibility in the original 2SFCA? This, in turn, makes interpretation of the output more robust and intuitive.~~

~~How much has access been inflated in the original 2SFCA?~~ Fig plots the ratio of the binary and adjusted binary accessibility measures. Here it can be seen that the unadjusted accessibility values are at least three times greater than their adjusted counterparts within the study area. This inflation, moreover, is not uniform across space, with inflation of the binary accessibility values up to 8 times greater than those from the adjusted model at the edges of the city where the 15-minute catchment areas begin to overlap with neighboring municipalities.

Why is this important? As noted by various authors [e.g., @Delamater2013; @Luo2003], in traditional FCA methods, the sum of the population-weighted average of accessibility across all population centers is equal to the regional average provider-to-population ratio [see the examples in @Delamater2013]. In the present case, the weighted sum of accessibility in the unadjusted binary and stepwise measures is 0.751. However, while this value is indeed identical to the regional average provider-to-population ratio, it is problematic because the share of the population correlates poorly with the pattern of inflation observed (see Fig ). The key issue here is that accessibility is deflated by the share of the population in a DA  $i$ ; however, the degree of inflation of demand and supply ~~depend~~ depends not only of the population DA  $i$ , but on the population of every DA  $j$  with which DA  $i$  interacts via ~~coincident overlapping~~ catchment areas. As a consequence, deflating accessibility using population shares in previous FCA methods does not accurately offset demand and supply inflation.

Fig and Fig present the results for the stepwise E2SFCA with and without the rectification. The results are qualitatively similar to the 2FSCA, with the expected differences. The inflation factor is even more substantial, given the larger catchment areas used.

## Disparity Analysis

~~Since neither supply or demand are inflated as in existing FCA methods, An advantage of the use of adjusted weights for proportional allocation of demand and level of service is that, after rectifying the inflation artifact, they make it is possible to conduct accessibility disparity analysis in a very intuitive way.~~

For instance, an analyst interested in equity analysis could allocate the total level of service uniformly to every DA. In other words, the total level of service (which equals the sum of accessibility over the system) can be divided by the number of population centers in the system. ~~to return the Average Local Population Center PPR.~~ The resulting mean value, call it  $L_i^e$  then would be assigned to the population centers as their “equitable” share of the total level of service in the system. Next, the equitative distribution of the level of service in each population center is substracted from the estimated mean accessibility to arrive at a

disparity index. When the difference between these two quantities is positive, this would indicate that a DA's accessibility exceeds its equitable share of level of service. On the other hand, when the difference is negative, the DA's accessibility is below its equitable share of the level of service.

This approach is reminiscent of the Spatial Access Ratio (SPAR) proposed by Wan et al. [-@Wan2012SPAR], which is calculated as the ratio between a population center's accessibility and the mean accessibility across all population centers. While Wan et al. [-@Wan2012] calculate SPAR based on the results of their 3SFCA method, ~~its use here with the adjusted demand and supply parameters would enable more intuitive results.~~ Nevertheless, SPAR rescales by rescaling the accessibility measures to reflect the percentage difference in each population center's accessibility relative to the mean ~~and was~~. This measure is designed to overcome the sensitivity of existing FCA metrics to the impedance function. In contrast, the ~~preservation of approach proposed here, enables more intuitive and interpretable results by preserving~~ the system-wide population and level of service ~~in the adjusted approach enables the~~. In this way, a disparity index ~~is useful~~ to highlight the absolute difference in accessible provider-to-population ratios across ~~the~~ population centers.

From this, the disparity Disparity maps for the ~~adjusted~~ binary and stepwise impedance functions are shown in Fig and Fig . These figures reveal the spatial distribution in disparity, with levels of access that are lower than the mean in more rural parts of the city (where travel times are longer and the distribution of physicians is more spatially disperse) compared to levels of access that are greater than the mean in the higher-density and more connected urban center.

## Conclusions

### AccessConclusion

Accessibility to healthcare ~~remains a critical is an~~ issue ~~of continued interest~~ in health geography. One of the most popular approaches to estimating accessibility is the 2SFCA method and its associated family of FCA models due to their simplification of more complex gravity models and their interpretation ~~in terms of as proxies for~~ provider-to-population ratios. These ~~intuitive~~ properties make FCA approaches particularly appealing for health policy. However, In this paper, we have argued that the overestimation of demand ~~and level of service~~ in FCA approaches poses a ~~serious~~ challenge to the interpretation of accessibility and the identification of spatial disparities in access, with potentially deleterious consequences for ~~health~~ policy analysis.

Recognizing this, The issue of overestimation of demand and level of service has been recognized before, notably by Wan et al. [-@Wan2012] and Delamater [-@Delamater2013], and alternative approaches have been proposed that seek to offset or ~~minimize~~ reduce the ~~demand overestimation~~ problem. Nevertheless, the present paper has shown that the inflation of demand is present in all existing FCA methods. Moreover, we also show that in

some cases, demand is deflated, and detail the potential for inflation/deflation on the supply side. To overcome these issues, we draw from the fields of spatial statistics and econometrics, to incorporate row-standardized impedance weights in the estimation of an adjusted population calculation of demand parameter, and column-standardized impedance weights to adjust the supplylevel of physiciansservice. These adjustments ensure that allocation of demand and level of service are proportional. As a result, both the system-wide population and levelslevel of service are preserved in the estimation of accessibility, thereby rectifying the inflation/deflation issue.

The case study application in Hamilton revealedCMA reveals the extent of inflation in accessibility inherent in the unadjusted approaches compared to the adjusted binary and stepwise FCA methods. Furthermore, the adjustments result in local provider-to-population ratios which can be easily understood relative to the system-wide equitable level of service through the calculation of a disparity index. The applicability of these values is particularly enhanced by the use of a travel survey to inform the estimated impedance functions. Taken together, these innovations provide estimates of spatial accessibility and disparity that are robust to the regional distribution of supply and demand, as well as observed travel behaviour. By extension, these properties mean that the adjusted approach employed here can offer more rigorous recommendations for health policy.

Our work is not without its limitations. For example, although we use a travel survey to define their parameters, the use of binary and stepwise impedance functions assumes a constant level of access within the travel-time isochrones of a catchment area. While we employed these functions for illustration purposes, the adjustment procedure outlined here is also suitable for re-weighting continuous impedance functions. Second, our adjusted measures of supply and demand account for potential interactions between all physicians and population in the system. This is in contrast to Delamater's [ @Delamater2013] M2SFCA, where some supply may not be allocated due to the spatial configuration of opportunities relative to population centers. Beyond overcoming the inflation issue inherent in Delamater's modification, the assumption of full allocation in the present case seems reasonable in the context of a single-payer healthcare system with undifferentiated service. In essence, this approach measures the potential for spatial interaction; whether or not members of the population opt to actually use a given service is assumed to be an outcome of the aspatial dimension of accessibility.

Nevertheless, the proposed adjustments Finally, 1) we proposed a set of slack factors to modulate the estimates of demand and/or level of supply to account for system inefficiencies; and 2) demonstrated the use of a disparity index to conduct equity analysis.

In conclusion, the research presented in this paper demonstrates how a relatively simple adjustment of the impedance weights can help to overcome the inflation/deflation issue inherent in previous FCA approaches. By incorporating these methods into the estimation of accessibility to healthcare services, future research can help to ensure that the FCA approach livescontinues to live up to its promise as an intuitive and policy-relevant method for investigating access and disparity.

## References

**Editorial Office**

Figure copyrights.

[\*\*Please note that all figures are original.\*\*](#)

### **Reviewer #1**

The authors propose a very interesting adjustment to the 2SFCA method. I'm still unsure that these adjustments are the panacea as suggested by the authors (i.e. inflation issues are 'eliminated') – I'd like to do my own testing (in the future); however, they are a worthy addition to the literature for further academic debate.

First of all, many thanks for your thoughtful response to our submission. If you are interested in conducting testing of your own, we would like to remark that the data and code are placed in a public GitHub repository as follows:

<https://github.com/paezha/Demand-and-Supply-Inflation-in-Floating-Catchment-Area-FCA-Methods->

In what follows we respond to your comments and indicate the actions that we have taken in response.

One fundamental problem throughout the paper is their referral to step 1 as if is measuring the actual or real demand on that supply point. Clearly that is not the case because, in most scenarios, populations will have multiple supply points that are accessible and thus their demand will be 'averaged' across these.

Thank you for this comment. This is something that we have clarified. "Demand" in a model of accessibility represents potential demand – since we do not really know how many members of the population actually have a family doctor. Even in a universal care system such as Canada, there are members of the public who do not have a regular health care provider – approximately 15.8% of the population in 2016, for example (Statistics Canada, 2017). Assuming, however, that all members of the population *do* demand health care (universal coverage as the design parameter, so to speak), one thing that becomes clear from our approach is that the "averaging" that intuitively takes place across the multiple supply points tends to be biased in earlier FCA approaches.

Typically the 2SFCA is measuring 'potential demand', but the authors effectively are presenting it as 'actual demand' (when in reality this is far from true). This very important distinction needs clarification upfront. The authors present an adjustment method that I distinguish as recalculating the 'likely demand on supply points'. Similarly, accessibility (Step 2) is 'averaged' across multiple supply points. The effect of demand inflation and supply inflation to the 2sfca method really should be looking at the combined effect – i.e. at the  $A_i$  (accessibility) scores.

Many thanks for this comment. Indeed, in addition to clarifying the distinction between potential access (accessibility) and actual access, we note that our approach assumes that the system operates in an efficient way, by fully allocating both demand and level of service proportionally. To allow for system inefficiencies, we suggest the use of "slack factors", similar to your suggestion of "likely demand on supply points". Furthermore, we show with the example how the combined effect in the accessibility scores is more intuitive when demand and level of service are rectified to their global levels by means of our standardized weights.

Their simple examples in Figure 1 and 2 are highly extreme (wholly un-real) scenarios, which are very effective at demonstrating how the 2sfca method can fail. However, their extremeness also makes it hard for the reader to evaluate whether, in a more 'normal' scenario, the same deficiencies still exist or

are a significant concern. I believe these examples need to introduced as such or similar: "These examples have been purposefully designed to demonstrate key weaknesses of the 2sfca method; we acknowledge these scenarios are not representative of typical demand and supply landscapes"

In this version of the paper, we recognize the artificial simplicity of the original example, and introduce a small but realistic simulation to complement our presentation.

Overall, I like the concept of the demand vector summing to equal the original population (demand) size – even after the application of impedance weights. In theory, this enables the generation of accessibility scores that more truly are interpretable ratios (e.g. health providers per population). Having said that, their adjusted weights – which divide the original weight by the sum of all weights for (potential) service points – still assumes that all service points are equally 'attractive'. I strongly believe this is not the case at urban fringe locations where larger urban populations are unlikely to be attracted to nearby rural services, but the reverse scenario is very different.

In fact, points are equally attractive only when binary weights are used as the basis for obtaining a standardized set of weights. When the basis weights are from an enhanced (i.e., non-binary) impedance function, the attractiveness of more distant points declines, and more demand is allocated proportionally to closer points (same with the allocation of level of service). We hope that the new examples will help to clarify this point.

Similarly the concept of saturation of intervening opportunities and variable catchment sizes seem pertinent here (i.e. why would a population travel further if sufficient opportunities are available at nearer locations?). Some recent 2sfca methods papers have introduced variable catchment size modelling, which I believe are critical for larger geographic scale applications (e.g. for national modelling).

Our toy example and small simulation indeed assume a heterogeneous and isotropic landscape, and as a consequence all catchment areas are of fixed sizes. The empirical application, in contrast, is based on network distance, and therefore lead to adaptive catchment areas.

As seen in Figures 5 and 6, the demand inflation factor is mostly a factor of urbanized populations only. I didn't understand the Accessibility scores in Figures 7 and 8 or 11 and 12 – how do you get an Accessibility ratio of 1 (i.e. equal supply/demand ratio)?

In the revised version of the paper, we clarify that accessibility is interpreted as the share of provider-to-population ratio allocated to a population center. Therefore, it would mean supply (i.e., doctors) divided by demand (number of patients).

Again, the values in Figure 9 and 13 make no sense to me. Are the authors truly claiming that 'current' accessibility scores (e.g. 2sfca, e2sfca) are inflated 4 to 9 times higher than they should be? In all of the published work on floating catchments, I've never seen any evidence that even vaguely suggests they are out by this factor level.

We have added information to explain this. At first glance, a factor of 4 to 9 is startling. However, when we compare the accessibility estimates obtained from current FCA methods, it turns out that they indeed tend to exceed the regional PPR by a very large factor (see Table 6 in new version of paper).

Figures 14 and 15 are very interesting in that they suggest a very systematic bias between metropolitan and fringe/rural locations (strongest reds in the most metropolitan, strongest blues in the outer edges) – I suggest this is a key finding that needs further discussion.

We have revised the discussion to respond to this.

Again, we thank you for your constructive and thoughtful comments, and hope that you will find our responses and actions in revising the paper satisfactory.

## Reviewer #2

This is a very interesting paper that presents an issue that I've grappled with in the past. Per PLOS ONE's reviewer policy, I'm going to unblind myself in an effort to promote a constructive dialog (and because the content of my review would likely do this anyways).

Many thanks for your comments and suggestions for revisions. This work was in good measure inspired by Delamater (2013), and so we are particularly pleased to receive your input.

Major concerns/issues/comments

L89: This equation is somewhat confusing as presented. Are these all from different metrics?

These are the same metric, just decomposed in different ways. For instance,  $D_j = \sum_i D_{ij}$ .

L104-105: Could this also be conceptualized as the cost required to overcome distance?... e.g., rather than making it about preferences, it seems like this passage should acknowledge that overcoming distance/time is a "cost" of using the service, rather than just framing it as preference-based.

We revised this as per your suggestion.

L347-349: I feel like I'm missing something here. The overall ratio of the system does not appear to be preserved, given my understanding of what the numbers represent. I think that they are (for 2SFCA): Prov/Pop ratio for A = 0.066 and for B = 0.006. In that scenario, the overall "system" ratio would be 0.46 (not 0.02) in this calculation. (If I'm interpreting the output incorrectly, I apologize. However, if that is the case, I think it would be extremely helpful to extend the example calculation to clearly demonstrate the output values and how the overall system ratio is preserved in the resulting values)

We have completely reworked the examples, and we hope that they will be clearer now.

Using the row standardized weight matrix to "allocate" the population to the various facilities in the first step appears to be very similar to the example I provided in Delamater, 2013 (citation 20 in the ms) on pages 33 and 34 and illustrated in System's #4 and #5. In this case, I was concerned about the overall effect of splitting populations into discrete "pieces" and it appears that this approach might have similar issues. If I'm understanding the authors' approach correctly, I think that the 2SFCA calculation for the Figure 1(I) system would result in an accessibility value of  $(0.03 + 0.03 + 0.03) = 0.09$  facilities per person. This output doesn't really make sense to me, purely from a logical perspective – and it would concern me that the approach does not perform well in the most simple of systems.

We have expanded the discussion and added summaries of the results to illustrate the way different implementations of FCA operate. Hopefully you will be persuaded that our approach to proportionally allocate supply and level of service provides a more accurate representation of provider-to-population ratios, and therefore are easier to interpret.

My overall thought is that the problem of supply and demand overestimation might be "problem" when each is considered in a vacuum, but not in the output/results when the two steps are integrated. I would suggest that the Demand and Supply Inflation section be expanded to show more simple examples – and most importantly how inflation actually effects the final sp accessibility values (A) in these systems, as a "proof of concept". As of right now, I can clearly see the issue in the two separate steps, but I'm having a hard time understanding what the demonstrated effects of this are on the final

outcome [e.g., is A (sp accessibility) under/over estimated for Pop A?... Pop B?]. This was one of my concerns when I critiqued the Wan et al 3SFCA paper in Delamater, 2013... I was not sure that competition really “needed” to be accounted for to adequately capture potential sp accessibility. In this case, I'm not sure if the inflation happening in the two steps is having a detrimental effect on the results. I think that clearly demonstrating issues with the output calculations of the E2SFCA and 2SFCA (in more than a single system) would provide a much better justification for this approach (than is presented currently).

Thank you for your suggestion. We have reworked the paper to include examples of 3SFCA and M2SFCA. As you correctly point out, it is important to think whether the final outcome (the accessibility) is affected by the potential inflation, and if so how. Currently, our understanding is that 3SFCA deflates demand by stacking a second impedance function that is standardized to allocate proportionally the results of E2SFCA. This is done again when allocating the level of service in the accessibility calculations. As our examples show, this deflates both demand and level of service, but in an ad-hoc way that is not consistent with the global demand or level of service. M2SFCA, on the other hand, deflates the level of service by squaring the impedance function – in other words, by shrinking the catchment areas. The standardized weights that we provide do not affect the catchment areas, but rather ensure that both demand and level of service are allocated proportionally within them.

We recognize the possibility of suboptimal and/or inefficient systems, following in part the discussion in Delamater (2013). To this end, we discuss the selection of the base impedance weights and the use of what we call “slack factors” that can modulate the allocation of demand and/or level of service in a simple and intuitive way.

I did not do a deep dive into the Results yet, given my concerns about the approach. However, I do wonder if using a ratio to describe inflation is a good approach, especially given that the output has true units (people and supply/people), e.g., from 0.002 to 0.004 and 2 to 4 would have a similar inflation factor... but would have very different absolute differences.

In case that you would like to do a deep dive on the analysis, the manuscript, code, and data files to reproduce the paper are publicly available at:

<https://github.com/paezha/Demand-and-Supply-Inflation-in-Floating-Catchment-Area-FCA-Methods->

We thank you again for your thoughtful and constructive comments and look forward to hearing back from you with your impressions regarding this new version of the paper.

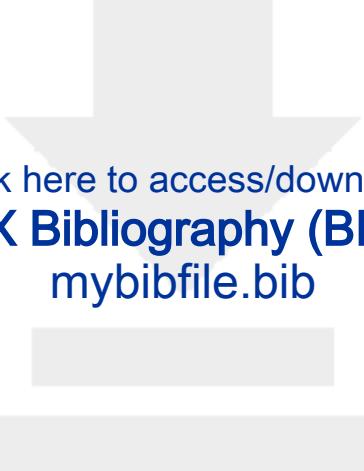


Click here to access/download

**LaTeX Source File (TEX file)**

Supply\_and\_Demand\_Inflation\_in\_FCA\_Methods\_v2.0.te

x



Click here to access/download  
**LaTeX Bibliography (BIB file)**  
mybibfile.bib

Data Review URL



Click here to download Data Review URL

<http://github.com/paezha/Demand-and-Supply-Inflation-in-Floating->