Incorporating Auto Accessibility into Statewide Project Prioritization: Feasibility, Behavior, and Equity

A dissertation

presented to

The Department of Engineering Systems and Environment

University of Virginia

In Partial Fulfillment

of the requirements for the Degree

Doctor of Philosophy

By

Richard Atta Boateng

May 2021

APPROVAL SHEET

This dissertation is submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Richard Atta Boateng

The dissertation has been read and approved by the Examining Committee:

John S. Miller (Advisor)
Brian L. Smith (Co-Advisor)
Michael D. Fontaine
B. Brian Park
Andrew S. Mondschein

Accepted for School of Engineering and Applied Science:

Craig H. Benson
Dean, School of Engineering and
Applied Science

May, 2021.

ACKNOWLEDGEMENTS

I am first of all thankful to Almighty God for bestowing in me the opportunity, untiring strength and ability to successfully complete this research study. I would also like to thank Brian L. Smith for his guidance, tolerance and encouragement and also for serving as my co-advisor. His unflinching support in both academic and personal issues was remarkable and will always be remembered. To John S. Miller my advisor, I say a big thank you for providing me freedom in my research and supporting all my research initiative. Over the years that I worked with you at the Virginia Transportation Research Council (VTRC), I found a life changing mentor and pillar of strong support in you. I will forever remember the support, advice, jokes and the walk down to UVA to release stress associated with research and to share personal experiences.

My sincere gratitude goes to Michael D. Fontaine (my dissertation committee member) for his critical role in all aspects of my development. You made yourself readily available to me at all times and provided me opportunities for knowledge enrichment through numerous research collaborations. I will forever be grateful to you. To the rest of my dissertation committee members in the person of B. Brian Park and Andrew S. Mondschein, I greatly appreciate the services you rendered for the successful completion of this dissertation. I would also like to thank Dr. Hyungjun Park for his guidance on research. I also want to thank my good friend Zulqarnain H. Khattak for being a valuable research partner and to my head pastor at the Shepherds House Worship Center, Rev. Dr. Kwame Boateng, I say thank you for your prayers and support.

I would like to say a big thank you to my parents and siblings for your selflessness and patience. Finally, I will like to say a big thank you to the strongest woman in the world in the person of Diana Asamoah, my sweet wife. You are a true champion and a partner. I appreciate every bit of time you have spent on me to ensure that this dream becomes a reality and for taking care of the little ones Jason B. Boateng and Janelle S. Boateng. May God richly bless everyone who has been part of this wonderful journey.

Table of Contents

ABSTRACT

While accessibility, the number of time-decayed jobs available to each zone within a region, has frequently been proposed as an element in transportation project prioritization, widespread adoption of accessibility has been hindered by two obstacles: computational feasibility in a semi-open source manner and longitudinal transparency. Surmounting the former through computational steps such as automation of turn restrictions, error checking for incorrectly formed GIS-based service areas, and accounting for random perturbations in the formation of such areas has been a necessary, but not sufficient, condition to render accessibility usable at a statewide level. This dissertation shows that design choices (e.g., number of centroid connectors or the catchment radius), which historically have not been examined in detail, do not have a single "correct" value. Rather, such design choices implicitly determine which of three paradigms are followed when using accessibility to prioritize projects.

One theory is that accessibility computations should directly respond to stakeholder concerns. In this context, the study finds that alteration of one particular design choice, the catchment radius, affected project rankings in manner opposite of those expected by stakeholders, where project accessibility benefits decreased, rather than increased, as the radius grew, owing to the fact that the marginal increase in accessibility was less than the marginal decrease in population. Proponents of the project in question would have been satisfied, therefore, with a fairly small catchment radius of 5 miles. A second theory is that accessibility computations should maximize the association between forecast and observed traveler behavior: this dissertation finds that accessibility alone has a statistically significant impact on destination choice and that the catchment radius can be selected to maximize this association (where in this particular case, a value of 35 miles yields the strongest association). In this case, accessibility alone explains between 4% and 10% of the variation in destination choice depending on the radius chosen. While such values will seem low to proponents, note that the socioeconomic factors alone (income, housing prices, and location) only add about a percentage point to this variance explained. A third theory concerns a conflict within the planning process: to what extent should accessibility for low-income populations, as opposed to accessibility for total populations, be part of project prioritization. This dissertation shows how to select a catchment radius that reduces this conflict, such that accessibility for both groups are aligned when one sets the catchment radius to about 25 miles.

The fact that the manner in which some project evaluation criterion (accessibility) is computed can affect the ranking of candidate projects will not surprise veteran observers of agency transportation project prioritization processes. However, this dissertation proves that the three theories for how such computations should be done—address stakeholder concerns, maximize association of forecasts with observed traveler behavior, or reduce conflict in the planning process by aligning equity concerns with the population at large—are all implicit decisions that result from the manner in which computations are performed. By making these choices explicit—that is, by showing that a particular catchment radius, once selected, will tend to favor one of these three theories relative to the other two—this dissertation seeks to advance the state of transparency when using accessibility as an element in project prioritization.

RESEARCH CONTRIBUTIONS OF THE THREE PAPERS

While accessibility, the number of time-decayed jobs available to each zone within a region, has frequently been proposed as an element in project prioritization, two challenges have hindered its widespread adoption in a consistent manner: computational feasibility and longitudinal transparency. *Computational steps* are required subroutines that must be performed for accessibility to be "correct" and repeatable on a large scale; examples are importation of legacy networks, automation of turn restrictions, and error checking for incorrectly formed service areas. *Transparency*, however, is the result of immediate design choices which themselves do not have a single "correct" value; examples are the number of centroid connectors, the catchment radius, and the granularity of travel time bins; and critically, these design choices have potential second order impacts in terms of stakeholder concerns, alignment with observed behavior, and equity. By necessity, agencies (and researchers) new to this field tend to tackle the first challenge, and this focus was reflected in both the July 16 proposal and the first paper in this dissertation. However, for accessibility (as quantified herein) to become an accepted element of the project prioritization process, at least some of the broader impacts of its use must become more transparent than is presently the case. Paper 1, the organization supporting this work, and the original July 16 proposal targeted the first challenge—but the committee and some (but not all) reviewers of the first paper were more interested in the second challenge. Both challenges have now been addressed through three papers.

Contributions of Paper 1: Transparency and Feasibility

The first paper, "Increasing Transparency and Feasibility of Auto Accessibility for Project Prioritization" (accepted for publication within the *Journal of the Transportation Research Board* as of April 8, 2021) solves the first challenge and articulates the possibility of the second. *Computational solutions* include developing a semi-automated method to import legacy transportation networks; using an algorithm to check for incorrectly formed service areas that sometimes occur in a random fashion with GIS software; automating turn prohibitions; and creating, on a large scale, realistic centroid connectors. The need for the latter three solutions is not limited to the Virginia dataset used for this study as users elsewhere will still need to consider turn prohibitions and the role of centroid connectors. (Future software enhancements may address the formation of service areas but other users at this point in time would need to consider the quality of service area formation).

Failure to use these approaches gives erroneous results: not solving the problem of incorrectly formed service areas led to the region within 50 miles of a one-mile corridor (where improvements are proposed) having an accessibility almost 40 times higher than the correct value. At large radii, accessibility scores may be underestimated because of random variation in service area creation; the solution is to identify zones with a negative accessibility contribution and to convert this value to zero. (That said, this random variation is not as critical as the above problems: choosing to eliminate negative net accessibility contributions, attributed to geometric approximations in the software, affects forecasts by less 21% at a 35 mile influence area or smaller.) The paper offers values to practitioners by implementing this solution in a GIS environment—logical because GIS is virtually ubiquitous in all 50 state departments of transportation, plus Puerto Rico and the District of Columbia (AASHTO, 2020).

This first paper then comprehensively demonstrates the short-term impact of making different *design choices* on accessibility, which has not been routinely studied elsewhere. Consider, for example, the catchment radius-- the sphere of influence considered for a candidate project being evaluated. Other sources cited therein used a fixed value for this parameter, such as 9, 45, 70, or 300 miles. None of these values of the catchment radius are wrong per se—they are legitimate design decisions the analyst may choose—but they affect accessibility. The paper, shows, for instance, that for one candidate project, the forecast accessibility improvement drops by 80% if this radius is altered from 45 to 15 miles—and altering the radius from 5 to 35 miles changes the relative rankings of five candidate projects. Such design choices have both immediate impacts (e.g., choosing this radius rather than that radius will favor this particular project) and broader impacts (e.g., choosing this radius rather than that radius has a bigger effect on improvements that support work related travel by low-income populations). For that reason, the longitudinal transparency of these design choices matters. (The paper also finds that other design choices matter but to a lesser extent: at a 10 mile radius, the number of centroid connectors affects accessibility by 23% and for most projects, varying the number of centroid connectors caused a score difference of only about 10%. Accordingly, these latter design choices receive less attention in the remaining papers.)

After showing that these design choices materially affect a project's forecast impact on accessibility, both in absolute and relative terms, the first paper then sets the stage for considering three theories for selecting design parameters used in the computations:

- *Theory 1. Parameters should be selected to resolve stakeholder concerns.*
- *Theory 2. Parameters should be selected to confirm user behavior.*
- *Theory 3. Parameters should be selected as a conflict resolution tool.*

The first paper concludes with a nod to the first theory: in Virginia, a key stakeholder concern was that failure to have a very large catchment radius would mean that certain suburban projects would be placed at a disadvantage when candidate projects are evaluated on the merits of accessibility. The first paper refutes this concern, showing that the catchment radius has a different impact than what was expected, owing to the mathematics that are used in Virginia's formula. (For the exurban project in question, a higher accessibility score is attained than comparable projects if a *smaller* catchment radius is used.) This vignette introduces accessibility parameter selection as a process, not selection of a single value, guided by stakeholder concerns (theory 1) rather than the behavioral validation of theory 2. The author does not proclaim that all stakeholders will be pleased with the particular approach used in Virginia, but it does demonstrate how design choices can be made to address qualms noted by a particular stakeholder.

Contributions of Paper 2: Behavior

The second paper, "What is the Association between Auto Accessibility and Traveler Behavior?" (proposed submission to the *Journal of Transportation Planning Education and Research*) examines the second theory: to what extent do differences in accessibility (measured as the distribution of decayed jobs for a given location) explain differences in observed behavior (measured as actual work trips made, based on probe data)? A few studies have examined how accessibility relates to socioeconomic factors such as income, age, household size, auto ownership levels, transit service, and school events (e.g., Bohnet and Gutsche, 2007: Lavieri et al, 2018: and Lasley, 2017), but using accessibility alone to forecast destination choice does not appear to be well explored. In fact, the author is not aware of any studies that have posed the following question: "Given that many factors presumably explain variance in trip destination choice, to what extent, if any, does accessibility alone explain this variance?" This paper initially bridges that gap by quantifying the strength of the association between observed destination choices and accessibility: based on Equation 12, accessibility alone can explain about 5%-10% of this variance in destination choice. Then, the paper considers the role of confounding factors shown in Equation 13 such as whether origin zone i and destination zone j represent disaggregate census block groups or more aggregate census tracts, whether i and j have similar housing costs (and whether there are diverse housing costs near these zones), and the distance of the proposed project from the closest population center.

$$Y_{ij} = \alpha + \beta_1 X_{ij} \qquad \text{(Eq. 12)}$$

$$Y_{ij} = \alpha + \beta_1 X_{ij} + \beta_2 A^p + \beta_3 B_{ij} + \beta_4 C_{ij} + \beta_5 D_j + \beta_6 E_i + \beta_6 F_j \qquad \text{(Eq. 13)}$$

Where;

Y_{ij} = $\frac{\text{Trips from i to j}}{\text{Total trips from i}}$

α = intercept

X_{ij} = $\frac{\text{Decay}_{ij}\text{Employment}_j}{\sum_{j=1}^{n}(\text{Decay}_{ij}\text{Employment}_j}$

A^p = distance between the project and the MPO center for project p

B_{ij} = disparity between housing costs in origin zone i and destination zone j.

C_i = localized diversity of housing costs in zones surrounding origin zone i

D_j = localized diversity of housing costs in zones surrounding destination zone j

E_i = household income for origin zone i

F_j = household income for destination zone j

The paper then varies a key design choice identified previously—notably, the catchment radius—and determines which catchment radius gives the strongest association, between observed behavior (left side of Equation 12 and 13) and accessibility (right side of Equation 12 and 13), finding that these confounding factors, although statistically significant (as was the case with the accessibility term X_{ij} used in Equation 12) add about a percentage point to the real-world (observed) destination variance in destination choice noted in Equation 13. Proponents of accessibility may well be disappointed by the low amount of variation explained by Equation 13—but those same individuals may be comforted by the fact that socioeconomic factors then do not substantially raise this percentage in Equation 13. Certainly others have shown certain factors on the right side of Equation 13 to be statistically significant in terms of impacting behavior—in particular, more disaggregate datasets such as block groups rather than Census tracts are desirable (Richter and Brorsen, 2006; Hartman, 1983)—but others do not appear (at least in studies found by the author) to have specifically stated what portion of variance they explain.

Hence Paper 2 defines validation as replication of traveler behavior for the before period (theory 2), setting the preconditions for determining the strength of the association between auto accessibility and traveler behavior (Equation I1) and then incorporating confounding factors to determine conditions under which this relationship might be strengthened (Equation I2).

Journals have different audiences, and this principle is more apt for Paper 2 than the others: the proposed journal has a pedagogical component, and for this reason an ancillary finding in this work has been highlighted here: accessibility scores tend to decrease as the radius increases because of the way population is used as a normalization component between different types of areas. That is, an accessibility score is the ratio of two values: a project's impact on time-decayed jobs (e.g., for all zones i, the numerator has the sum of $\Delta Population_i Decay_{ij} Employment_j$ which itself is a summation over all zones j) and then the population over which all time-decayed jobs are considered (in the denominator). The catchment radius decreases the accessibility score because in most situations, the marginal increase in population (in the denominator) more than offsets the marginal increase in accessibility (in the numerator) as illustrated in Figures 8 and 9 of that paper. (In fact, the reason the sponsoring agency for this research, the Virginia Department of Transportation, was interested in this analysis was that some stakeholders had wondered if projects in exurban locations, such as Project 5, would fare better if the catchment radius were increased; the paper demonstrates the opposite impact.)

Contributions of Paper 3: Equity

The third paper, "Reducing Conflict: Choosing an Auto Accessibility Sphere of Influence to Explicitly Serve Low-income Populations" (proposed submission to the *Journal of Transport Policy*) also concerns the role of the catchment radius. Hardy and Bell (2019) have explained that the population term in the denominator exists to render urban and rural projects comparable, and in terms of implementation, Sundquist (2017) had mentioned that the sphere of influence (which in that effort was defined as 45 minutes from the project) was "subject to change." The third paper probes the use of this sphere of influence based on theory 3: choose a catchment radius that minimizes the difference between an accessibility score based on total population and an accessibility score based on low-income populations. The author frames this as an "equity" concern taking care to define this term. (In principle, equity can reflect any disparate treatment across different groups, thus one could argue that the urban versus rural consideration that arises in states falls into this category; hence this paper defines equity as equal treatment of benefits for lower income populations.)

Resolution of this conflict between projects benefitting the accessibility of lower income groups in particular versus the total population matters: while Environmental Justice (Executive Order 12898) requires that projects not adversely affect protected groups to a greater extent than all populations, it does not require that benefits to such groups, when evaluating projects, account for a certain percentage of project evaluation. Thus a key design choice—the catchment radius—is examined as a conflict resolution tool to ensure that the use of accessibility scores are not biased against low-income populations. In short, the author examines how to choose the catchment radius (from 5 to 35 miles) such that the net accessibility benefit when considering all populations is the same as the net accessibility benefit when considering low-income populations only.

Five measures of consistency (notably these include the correlation of general population accessibility scores and disadvantaged population accessibility scores and the ranking of candidate projects) are evaluated at each possible catchment radius (5 to 35 miles), and the radius that minimize the conflict between these groups (found to be 25 miles) are chosen. Just as accessibility proponents in Paper 2 may be disappointed by the low percent of variance in trip distribution explained by accessibility alone, equity advocates in Paper 3 may be disheartened by the observation that there is strikingly little conflict (between accessibility benefits for total and disadvantaged populations) based on the measures of consistency, such that equity differences between the best radius (suggested as 25 miles) and the worst radius (10, 15, or 35 miles depending on which consistency measures are chosen) are subtle. Yet, while acknowledging that one of the proposed measures of consistency in fact does not discriminate among the proposed radii, the paper shows that the initial perception of a small amount of conflict is likely due to high correlation at the census tract level between total jobs and low income jobs (or between total population and low-income population). This high correlation does not invalidate the use of the three recommended (out of five examined) measures of consistency, but rather shows that persons concerned with equity need to consider even small differences in these equity measures as meaningful.

The third paper is not unique in seeking to address conflict resolution (Meyer and Miller, 2013), rather, its contribution is to establish design choices to resolve disputes, such as asking "what is the catchment radius that ensures accessibility scores are not biased against low-income populations?" Certainly transit-based accessibility for low-income populations has received attention, but this third paper enables full consideration of auto-oriented accessibility for both low-income and total populations, seeking to choose the radius to eliminate any accessibility disparities between the two groups.

Summary of Research Contributions

Paper 1 firmly demonstrates how to overcome two sets of technical challenges for implementing accessibility. The first set of solutions—such as development of a script to check for errors in service area formation which are endemic to widely used GIS software—are essential for widespread implementation of this accessibility measure. However, persons who are not required to perform this implementation, or who have the ability to procure customized accessibility software, may not place a high priority on these solutions. In response, the first paper offers a second contribution that seemed to have greater acceptance by the initial paper reviewers: crisp documentation of the importance of design choices, such as the distance from the project over which access benefits are tabulated. That first paper simply shows that these design choices matter: legitimate, but different, design decisions may yield different project rankings—but that demonstration is a key contribution.

Paper 2 addresses an area of transportation planning that within the past five years has received substantial attention: to what extent are models validated by behavior? The contribution of paper 2 is threefold: to specify what amount of destination choice variance is explained by accessibility alone, to compare the importance of accessibility to socioeconomic factors believed to influence destination choice (e.g., income, housing costs, and project location relative to an urban center), and then to show how to compute accessibility such that one maximizes the strength of the relationship between observed behavior and accessibility. Certainly others have determined whether accessibility affected travel behavior: Kockelman (1997) found that accessibility (sometimes described as decayed jobs and sometimes described as jobs within a half hour) does affect vehicle miles traveled per household. The author is not aware, however, of studies that have quantified the extent to which accessibility alone forecast behavior, and especially is not aware of studies that defined behavior in the fairly disaggregate and challenging manner of Equation 11: observed trips to a particular zone.

Paper 3 concerns equity—a topic of over 5700 articles, reports, and projects since 1950 and an increasing area of interest (43% of these have been published or initiated since 2010). Just as equity is common, so are papers considering better accessibility for low-income populations, so in neither of those two heavily discussed topical areas does the author claim a contribution. However, the field narrows greatly when one considers auto-oriented accessibility for such populations, with notable exceptions being Carroll et al. (2021) in rural Ireland and Merlin et al. (2018) in San Antonio. Like those two papers, Paper 3 considers auto accessibility for low-income populations. Unlike those two papers, Paper 3 asks how we might choose the catchment radius so that that the relative merits of candidate projects are similar whether we consider all jobs and all people or only low-income jobs and individuals in those jobs. Paper 3's contribution is to show how to select that catchment radius through three somewhat nuanced measures of consistency. Some observers may retort that paper 3 is unnecessary: if equity is so important, simply compute accessibility based solely on low-income populations and use any catchment radius that is desired. While this remains an option, the project prioritization process in general is contentious, and thus having a tool to reduce conflict (in this case, a debate over the extent to which project prioritization processes should explicitly consider low-income populations apart from total populations) has merit.

CONCLUSIONS

This research led to a number of conclusions pertaining to incorporating accessibility into statewide project periodization. The conclusions are presented below;

Feasibility and Transparency

- *When networks are passed from one package to another, care must be exercised to ensure that two-way travel is retained where appropriate.* The solution to this challenge is to alter the direction of the travel attribute such that it follows, rather than contradicts, the digitization direction. This legacy network challenge is not unique to Virginia.

- *It is possible to automate the incorporation of turn prohibitions, thereby saving substantial time (in this case, an estimated 2,000 hours of manual processing).* Rather than digitize turn prohibitions manually, a script from the literature could largely be adapted provided one then performed an iterative additional step: use a custom script to create and populate the turn restriction attribute table because of the FCID varying with each editing session and not being detectable using conventional methods.
- *Service areas can be used provided an automated script, such as that shown herein, corrects inconsistencies.* If service areas are intended to be rings (e.g., employment centers 10 to 11 minutes away from origin A), the service areas should always have the shape of a donut. In random cases, however, some service areas were pie-shaped—showing employment centers within 11 minutes of origin A. This error of mixing service area types (donuts and pies) is hidden from the user. The solution is to modify a sorting subroutine to identify duplicative employment centers automatically and remove them from the accessibility calculation. Manual inspection is infeasible: with 100 origins and 90 bands (1 band for each minute), there are 100 x 90 = 9,000 service areas. The pie-based service areas are not materially wrong, but when service area types are mixed, the accessibility results can be nonsensical.
- *The single most important parameter chosen by the user is the catchment radius*. The accessibility score starts to drop at some relatively small radius, owing to the fact that in most situations, the marginal increase in population (denominator of Equation 5) more than offsets the marginal increase in accessibility (numerator of Equation 5). This radius also affects the impact of other design choices.
- *Other modest network design decisions may materially affect the accessibility score for some projects*. At a catchment radius of 10 miles, for four projects, varying the number of centroid connectors from one to five showed that the highest score was no more than 10% higher than the lowest score. For a fifth project, however, the highest score (with one connector) was almost twice that of the lowest score (with three connectors). A mean absolute deviation analysis can be used to choose the number of centroid connectors that minimizes variability, which in this case was four connectors. As the radius increases, the impact of the number of connectors lessens.
- *At large radii, the accessibility score of projects may be underestimated because of random variation in the creation of service areas where the borders shift slightly.* The solution is to identify block groups yielding a negative contribution to accessibility and then correcting this negative amount to a zero value.

Behavior

- The study indicates that catchment radius 5-35 miles around each project indeed affected accessibility, however at 25, 30 and 35 miles catchment radii, altering the radius did not affect accessibility ranking of the projects.
- The study further indicates that there is significant relationship between observed and forecast behavior and that accessibility alone explains between 4% and 10% of the variation in destination choice.

- The study found accessibility scores to be statistically significant at 95% confidence level with the highest percentage (10.2) of variations in traveler behavior that can be explained by accessibility alone as well as accessibility with other confounding factors occurring at 35 miles catchment radius. While the lowest percent occurred at 5 mile catchment radius (3.7%), the three highest stable percentage of variance that can be explained by accessibility and other confounding factors occurred at 25, 30 and 35 mile catchment radii.
- The study finds three catchment radii that have the potential to provide the best fit between observed and forecast behavior to be at 25, 30 and 35 miles. Among these three, catchment radius at 25 mile will be recommended to be used because it is relatively smaller than the other two, it will require lesser processing and computational time and resources and will be the most cost effective radius to implement.

Equity

- *Consistency of project rankings* is the most intuitive measure: do the rankings remain unchanged when computing accessibility for total populations (ΔA) vs. disadvantaged populations (ΔA')? This showed that of the seven radii considered (0, 5, 10, 15, 20, 25, 30, and 35 miles), consistency was achieved at 15, 25, and 35 miles, as shown in Table 4.
- *Consistency of spatial contribution* allows one to consider whether the geographical benefits of accessibility are similar. This may be measured statistically with the KS test (e.g., if 10% of accessibility benefits comes from zones 2 to 4 miles from the project when considering total populations, is a similar percentage computed when considering low-income populations?). No radius showed perfect consistency, but the greatest consistency was achieved at radii of 5, 10, 15, and 25 miles, as shown in Table 5.
- *Consistency of correlation at the project level* enables detection of a linear association given that these projects were samples, with nominally higher correlation at R = 25 miles.

Future Research Needs and Limitations

All three papers used a similar workflow for the accessibility oriented computations (see Figure 1 in each paper) and natural questions that arise with any data set concern sampling: how would the results be affected if one used projects in a different state, at a different time interval (such that forecast changes in delay were altered), or with different stakeholders (such that other concerns, besides those concerning the Project 5 catchment radius in Paper 1, were raised)? However, even with a much larger dataset that could address these factors, additional research needs and limitations remain fundamental to understanding the use of accessibility in the longer term—that is, over the next decade or so as new projects are constructed:

- For paper 1, what combination of design choices (mostly catchment radii but also centroid connector speeds and possibly the decay values) would enable some basket of well-known projects to reflect accessibility improvements that align with decision-makers' expectations? Paper 2 had suggested using travel behavior to determine the appropriate catchment radius. However, since accessibility scores are ultimately used to rank projects, another way of determining the radius is to identify a sample of well-defined projects, give them to a panel of experts, obtain rankings from the experts, and then determine how to compute accessibility such that accessibility-based project rankings align with those of the experts. Such an expert-based validation is a fundamentally different approach from the behavioral-based validation of Paper 2.

- For paper 2, how does before-after changes in travel behavior relate to before-after changes in accessibility? Paper 2 quantified the association between accessibility at a point in time and destination choice at a point in time. A longer term question would be the association between the changes (in destination choice) and the changes (in accessibility). Crucially, how does latent demand affect this change in accessibility—this should differentiate between urban and rural areas, but it also may be affected by housing costs. (For example, in metropolitan areas where central cities have become desirable as reflected in higher rents and where inner suburbs are less desirable with lower rents, it may be the case that latent demand has a lesser impact on advantaged populations (wealthier individuals living near the CBD and commuting to suburban high tech campuses) than on service workers in the suburbs commuting to the CBD. (This could extend to paper 3 but paper 2 is an appropriate starting point.) While accessibility computation uses auto dataset, the observed trip dataset from streetlight consists of all trips.
- Although paper 3 showed the importance of selecting the appropriate catchment radius as a conflict resolution tool between the total and low-income populations, additional work can be performed to determine the feasibility of a common radius for multiple modes, such as a behavioral analysis: to what extent does the catchment radius affect the alignment of observed origin-destination data with forecast trips? Further, to what extent do stakeholders in the transportation planning process endorse the use of consideration of the radius in this manner? An alternative, for instance, might be to consider simply only low-income populations; this study sought to demonstrate that it is feasible to choose a radius that addresses the needs of all populations. Knowledge of the stakeholder reaction to such practices, as well as the computational details presented in this paper, is essential in conducting the public "vetting" advocated by Sundquist (2017) and Sundquist et al. (2018) to ensure that accessibility is a meaningful metric when candidate transportation projects are evaluated for construction.

PAPER 1: INCREASING TRANSPARENCY AND FEASIBILITY OF AUTO ACCESSIBILITY FOR PROJECT PRIORITIZATION

Submitted to the Transportation Research Board in July 2020

As of April 8, 2021, the paper has been accepted for publication in the Journal of Transportation Research Record.

ABSTRACT

Accessibility, the number of time-decayed jobs available to each zone within a region, can help prioritize candidate transportation investments. This paper demonstrates how to compute auto accessibility using commonly available resources and identifies strategies needed to render calculations feasible and transparent. (The scope excludes transit and pedestrian impacts.)

For the first objective, computational solutions include developing a semi-automated method to import legacy transportation networks; automating turn prohibitions; and using an algorithm to check for inconsistently formed service areas that sometimes occur in a random fashion with GIS software. Failure to exercise quality control using these approaches gives erroneous results: not solving the problem of inconsistently formed service areas led to the region within 50 miles of a 1-mile corridor (where improvements are proposed) having an accessibility almost 40 times higher than the correct value.

For the second objective, the influence area (i.e., catchment radius) matters most: for one project, the forecast accessibility improvement drops by 80% if the area within 45 miles of the project, rather than the area within 15 miles, is the basis of the analysis. Other decisions affect the forecast accessibility improvement by less: the choice of the number of centroid connectors affects forecasts by an average of 23% (with a 10-mile influence area). Choosing to eliminate negative net accessibility contributions, attributed to geometric approximations in the software, affects forecasts by less 21% (35-mile influence area or smaller). Ranking five proposed investments in terms of their forecast accessibility benefit demonstrates the importance of documenting users' computational choices.

Keywords: accessibility, programming, resource allocation, transportation planning, public participation

1. INTRODUCTION

Transportation agencies may use some ability to reach jobs as a metric in the prioritization of candidate transportation projects for construction (*1-8*). Examples are "accessing jobs" in North Carolina *(1)* and the Southeast Iowa Regional Planning Commission *(2)*; supporting "employment" in Delaware *(3)*; providing nonmotorized access to a large employer in Vermont *(4)*; enhancing "economic vitality" for large job concentrations in the Ohio-Kentucky-Indiana Regional Council of Governments *(5)*; or providing "access to business facilities" in Minnesota *(6)*. Accessibility may be a discrete score (e.g., 10 points for direct bicycle access to a large employer [*4*]) or a finer grained calculation (e.g., number of peak hour commutes below thresholds of 20 minutes [*6*] or 50 minutes [*7*]).

Another way of defining accessibility is the sum of time-decayed jobs reachable from any location, which renders nearby jobs more valuable than distant jobs (*8*). Equation 1 shows a simple example: E_j is jobs in zone j; t_{ij} is travel time between zones i and j; $1/t_{ij}$ is the decay function; n is the number of zones in the region; and A_i is the accessibility for zone i.

$$A_i = \sum_{j=1}^{n} \frac{1}{t_{ij}} E_j \qquad (1)$$

Although Equation 1 appears straightforward, Sundquist *(9)* warned that accessibility remains within the domain of "academic or ad hoc studies" until software-based tools for computing accessibility are used in "a vetted and practical way." Earlier work by the same author (*10*) emphasized that accessibility must be "scalable" (enabling comparison of small and large network improvements) and computationally feasible (given many candidate projects). The literature (*11-20*) suggests five guiding principles for determining accessibility if the method is to be suitable for the planning process.

1. The method should be able to evaluate shorter trips (9, 12).
2. The manner in which accessibility is computed should yield the desired characteristic to be measured (13).
3. The method should be automated at the appropriate level of geography (11, 15).
4. The method should be implementable by the agency (16, 17).
5. The method should be transparent (18, 19).
6. The method should be calibrated (20).

Techniques exist to satisfy these principles. For the first (evaluate shorter trips), a detailed network is better than a network of only major facilities; for the third (automation), processes such as incorporation of turn restrictions into the network might be devised. However, because achieving these principles requires effort, it is reasonable that their necessity (or non-necessity) be documented. The researchers are not aware of such documentation.

2. PURPOSE AND SCOPE

This paper reports on ways to implement accessibility-based calculations on a relatively large scale. The motivation was twofold: (1) a desire to understand the impact on accessibility of adhering to the principles from the literature, and (2) a desire to increase the likelihood that calculations are transparent, as suggested elsewhere (21, 22).

The scope of this work was the GIS environment, chosen in part because GIS is virtually ubiquitous in U.S. state departments of transportation, with AASHTO reporting contact information for GIS staff in all 50 states plus Puerto Rico and the District of Columbia (23). The scope was also limited to auto accessibility, excluding transit and nonmotorized impacts.

3. METHODOLOGY

The methodology consisted of four steps:

1. Develop a workflow for computing accessibility.
2. Obtain network datasets, activity datasets, and projects for evaluation.
3. Address data processing challenges.
4. Assess the sensitivity of accessibility scores to computational strategies.

3.1 Workflow for Computing Accessibility

The workflow (Figure 1) entails developing two datasets, one where a candidate transportation project is not built and one where the candidate transportation project is built. The workflow uses ESRI's ArcGIS Network Analyst (ArcMap version 10.3.1), where service areas are generated for each 1-minute travel time interval for each candidate project for the morning peak period. (This period is chosen to reflect, for most jobs, the travel time to work and is used in lieu of other travel times that are available such as the mid-day peak, evening peak, or off-peak travel times.) Equations 2 through 5 compute accessibility scores by intersecting population-based service areas and employment centroids.

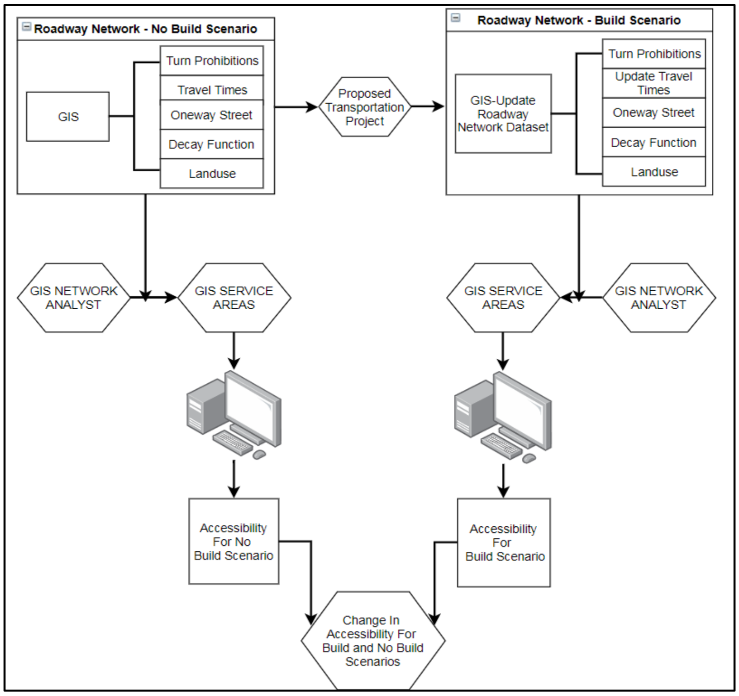


FIGURE 1. Summary of the accessibility computation workflow.

For each census block group i, Equation 2 gives accessibility as employment in zone j multiplied by a nonlinear decay function; these products are summed for all zones j. The decay function is based on travel time from i to j, dropping from 1.00 (4-minute travel time) to 0.01 (90.5-minute travel time). Given decay values (Decay_{ij}) of 0.91 (6 minutes) and 0.86 (7 minutes), a single zone i with 100 jobs (E_j) located 6 minutes away and 1,000 jobs located 7 minutes away has accessibility of (0.91)(100) + (0.86)(1,000) = 951.

$$A_i = \sum_{j=1}^{n} \text{Decay}_{ij}\text{Employment}_j \quad (2)$$

Equation 3 shows that the accessibility for block groups with a large population (Pop_i) is more important than those with a smaller population:

$$A = \sum_{1}^{n}\left(\sum_{j=1}^{n} \text{Decay}_{ij}\text{Employment}_j\right)\text{Pop}_i \quad (3)$$

Virginia's project-driven improvement in accessibility (Equation 4) is normalized by the population within a certain radius (R) of the project (Equation 5), which for auto-oriented projects is 45 miles. This number of block groups within R miles thus gives the n used in Equations 1-5 and enables one to convert the non-normalized change in accessibility (ΔA) to a population-weighted mean change in accessibility (ΔS). This ΔS, also known as the accessibility score, is evaluated for candidate transportation projects in the remainder of this paper.

$$\Delta A = \sum_{i=1}^{n}(A_i^{After} Pop_i) - \sum_{i=1}^{n}(A_i^{Before} Pop_i) \qquad (4)$$

$$\Delta S = \frac{\sum_{i=1}^{n}(A_i^{After} Pop_i) - \sum_{i=1}^{n}(A_i^{Before} Pop_i)}{\sum_{i=1}^{n} Pop_i} \qquad (5)$$

3.2 Network Datasets, Activity Datasets, and Projects for Evaluation

Five projects (Z. Ling, personal communication) were provided by the Virginia Department of Transportation (VDOT) in order to test the workflow, representing diverse areas, facility types, population densities, and improvements (Table 1). Projects 2, 3, and 5, located in the urban areas of Richmond, Hampton Roads, and Northern Virginia, have 3 times the population density (for the area within 35 miles of the project) as project 1 in Charlottesville, which in turn has about 1.5 times the population density of project 4 in Front Royal. Forecast impacts vary: projects 2 and 3 increase expected travel speed by about 6 mph (from 32 to 38 mph for project 2 and from 38 to 44 mph for project 3), whereas projects 1, 4, and 5 increase the forecast travel speed by 4 times that amount (project 1 raises speeds from 13 to 40 mph on a five-lane principal arterial; project 4 raises speeds from 43 to 69 mph in one direction on a rural interstate with heavy trucks; and project 5 raises speeds from 17 to 45 mph on a two-lane minor arterial). Project lengths ranged from 0.9 miles for project 1 to 1.6 miles for project 2.

3.21 Data Quality Challenges

With 3.3 million links, care must be taken to automate certain computations. Although many are specific to the Virginia dataset (e.g., the generation of centroid connectors), three data quality challenges appeared likely to extend to other locations that use network-based accessibility measures in a GIS environment. These challenges were resolved through preprocessing the dataset to solve unexpected problems:

1. Ensure two-way links are represented appropriately.
2. Automate the importation of turn restrictions.
3. Manage inconsistently formed service areas.

TABLE 1. Summary of Input Data Elements

Category	No.	Description
Baseline inputs	1	Highway Links for No-build Scenario: These baseline network data covering the entirety of Virginia include more than 3 million links. The comprehensive road network dataset contains attributes such as distance, speed, travel times during the AM peak, road functional class, travel direction, and digitization direction. Each link has a unique identification number that connects 2 nodes.

Category	No.	Description
	2	Junction Nodes: The dataset contains nearly 1.5 million nodes; each node has a unique code, which was useful when generating centroid connectors.
	3	Block Groups: These zones contain forecast demographic attributes for year 2025 such as population and employment.
	4	Proposed Projects Dataset: For each proposed project, this dataset has links indicating the project's location and, in conjunction with data element 2, enables one to determine how the project will affect link travel times.
	5	Turn Restriction Dataset: The Virginia turn restriction dataset contains codes that correspond perfectly with the junctions of the highway network. Each link in the Virginia highway network dataset also has its unique code. The data were further processed using MySQL to match the nodes that form each link, with identifiers indicating restricted turning movements.
	6	One Minute Bin Decay Values: These reflect the value of a job as a function of travel time. For example, a job that is 5.5 to 6.5 minutes away has a value of roughly 0.96; a job that is 89.5 to 90.5 minutes away has a value of roughly 0.01.
Project inputs	7 [a]	*Project 1: US 250/Route 20 Intersection Improvement (Charlottesville):* Reconstruct the US 250 (Richmond Rd.) and Route 20 (Stony Point Rd.) intersection to improve safety and operations. The project includes additional turn lanes, right of way, medians, and new signals.
	8 [a]	*Project 2: Pole Green Road Widening (Richmond):* Widen Pole Green Rd. (Rt. 627) from 2 to 4 lanes between Bell Creek Rd. and Rural Point Rd. (1.55 miles).
	9 [a]	*Project 3: George Washington Highway Widening (Hampton Roads):* This project will provide improvements to Rt. 17 by expanding the existing 3-lane undivided roadway to a 4-lane divided roadway from Yadkin Rd. to Canal Dr. The project will also include intersection improvements.
	10 [a]	*Project 4: I-81 Exit 300 at I-66E Northbound Widening (Staunton/Front Royal):* Add an additional lane and widen the left shoulder from Mileposts 299.1 to 300.4 Northbound; replace and widen bridge over Water Plant Rd.
	11 [a]	*Project 5: Rt. 2 and Rt. 17 from Lansdowne Rd. Past Shannon Airport (Fredericksburg):* This project improves the intersection at Lansdowne Rd.; widens Rt. 2 past the intersection of Shannon Drive; adds a southbound through lane on Rt. 2 from Bowman Dr. to Shannon Airport Circle; and adds a northbound right-turn lane on Lansdowne and a westbound right-turn lane on Mansfield.

a For each project, the build scenario dataset consists of the same dataset as element 1 with one exception: new speeds and new travel times reflecting the proposed transportation project being evaluated.

3.3 Sensitivity of Accessibility Scores to Design Decisions

Once accessibility could be implemented, the impacts of three key design decisions that the analyst must make were determined. One such decision is the number of centroid connectors (e.g., artificial segments that connect the population centroid of each block group to the roadway network). Thus, for each of the five candidate projects, 10 different networks (5 before and 5 after) were developed. The number of centroid connectors ranged from one to five, and the impact that the number of centroid connectors had on the change in accessibility was determined for each project.

Then, for one project, two additional design decisions were examined: the impact of the catchment radius (i.e., the value of R in Equation 5), and an element of random variation in the service areas (since even when formed as desired, service areas remain an approximation). The former was evaluated by executing the approach for radii from 5 to 85 miles. The latter was evaluated by identifying the relatively few zones i where following construction the A_i^{After} - A_i^{Before} term—i.e., the net accessibility contribution of zone i following project construction—was negative; this difference should be either positive or zero.

4. RESULTS

Following implementation of the workflow in Figure 1, two sets of results were obtained:

1. resolution of data processing challenges
2. assessment of the sensitivity of accessibility scores to computational strategies.

4.1 Resolution of Data Quality Challenges

Three challenges were addressed: import of one-way streets, automation of the importation of turn restrictions for large networks, and management of incorrectly formed service areas.

4.11 Ensure Two-Way Links are Represented Appropriately

Figure 2 (left) shows the expected type of service area where, starting with the blue population centroid, there should be travel in all directions. The problem of not being able to travel south or west from the blue centroid in Figure 2 (right) is an artifact of how legacy networks are imported. A standard link has two pieces of information to determine the permitted travel direction. The first piece is the link's travel direction attribute (Figure 3, left): a value of F means one may travel in the digitized direction, and a value of T means one may travel in the reverse direction. The second piece of information is the original digitization direction (which was set before importation of the links), indicated by the arrows in Figure 3 (right).

For hypothetical two-way Route 999, the two perfectly overlapping links should allow travel in opposite directions. However, as shown in Figure 3 (right), these two links differed in both the digitization direction and the direction of travel attribute, thereby creating two duplicative links, each with travel in the same direction. To be clear, this is not a problem with the software but rather how information within the network is used by the analyst. In this particular case, the same critical information is stored in two places (as the digitization direction and the direction of travel), and the analyst needs to choose just one of them. The simplest solution to this problem is to code the direction of the travel attribute as F for both links.

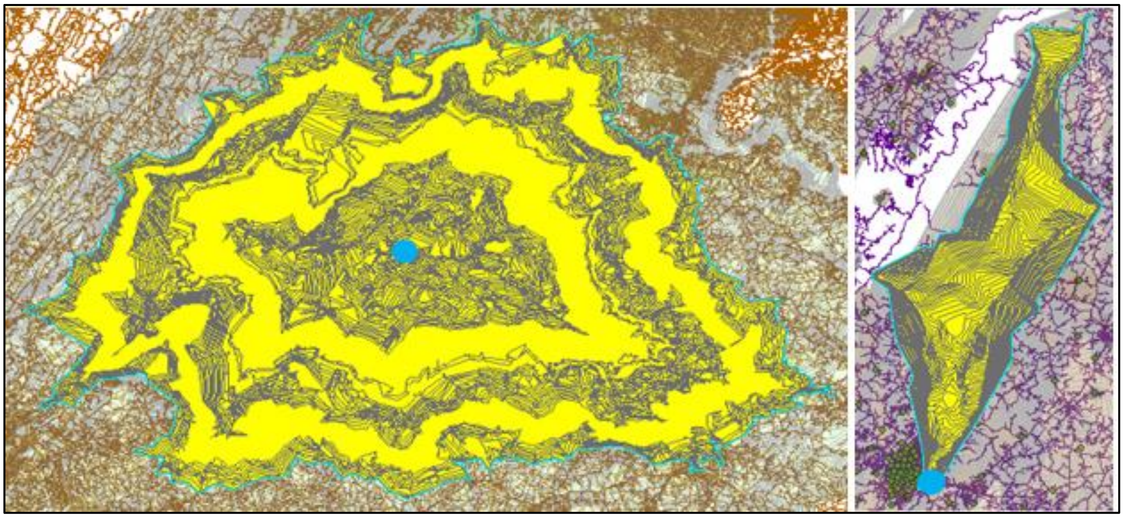


FIGURE 2. Examples of service areas where two-way travel is allowed (left) and where only one-way travel is allowed (right).

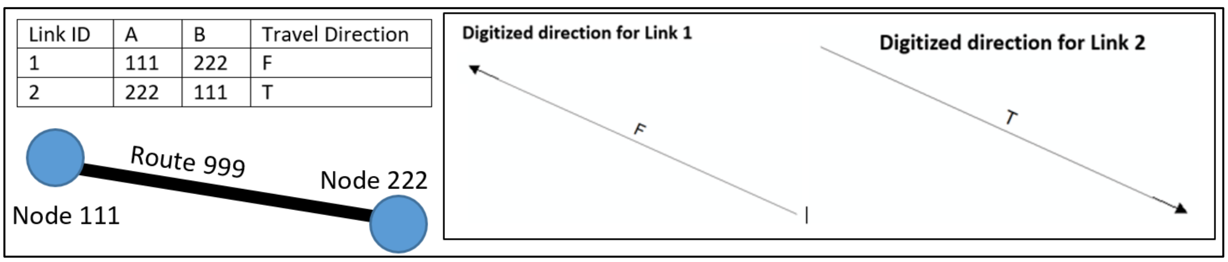


FIGURE 3. A two-way hypothetical route represented as two overlapping links; arrows indicate the digitized direction. F means travel in the digitized direction is permitted; T means travel in the reverse of the digitized direction is permitted.

4.12 Automation of the Importation of Turn Restrictions for Large Networks

Figure 4 illustrates the concept of turn restrictions with an origin (point 1) and a destination (point 2). Without restrictions, travelers follow the shortest path (Figure 4, left). If in reality left-hand turns are prohibited from link 1A to link AB, then travelers might use the path shown in Figure 4 (right). The turn restriction does not deactivate the link entirely; for instance, travelers might still use link AB for other movements. One way to add turn prohibitions is to draw them manually, as shown by the red lines in Figure 4, a process that requires roughly 4 minutes per restriction. Because Virginia has nearly 30,000 locations where turns are prohibited, such a process would require approximately 1 year of staff time.

Thus, an existing eight-step procedure was implemented to create these turn prohibitions automatically *(26, 27)* with one crucial modification: comments (*26*) also identified a problem with the creation of edge data for the turn restriction feature class—a problem also encountered during this effort. Additional literature *(27)* suggests that one can detect the feature class identification number

(FCID) by examining the feature class properties; however, this method was not always successful because the FCID can change throughout editing sessions. The solution is to initiate an editing session with the turn feature class; manually draw a line from one roadway link to another; observe the value of the FCID (which is dynamically assigned by the software); and then insert this value into a script creating these turn restrictions.

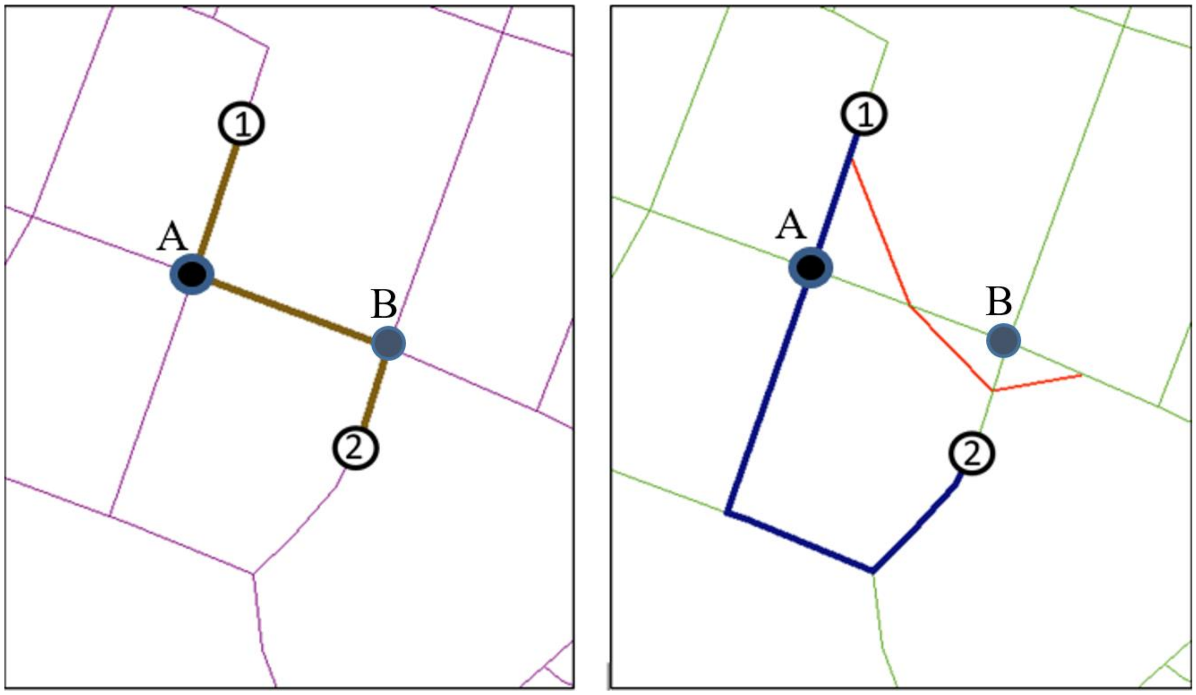


FIGURE 4. Examples of turn restrictions: travel from the origin (point 1) to the destination (point 2) without turn restrictions at point A (left) and with a turn restriction prohibiting a left turn at point A (right).

4.13 Management of Mixed Service Areas

In some cases, with no discernible pattern, service areas were created in a manner different than intended. Figure 5 (left) shows the desired shape of a service area where the yellow donut-shaped band indicates all locations that are 6.5 to 7.5 minutes away from the origin in blue. Figure 5 (right) shows a different service area: although the band should indicate locations that are 68.5 to 69.5 minutes away from the origin, the band instead indicates locations that are 0 to 69.5 minutes away from the origin. The "band" in Figure 5 (right) takes not the form of a donut but rather the form of a pie. In Figure 5 (right), the accessibility for the origin zone will, in Equation 3, incorrectly incorporate the jobs at the aforementioned employment center twice: once for the 6.5 to 7.5-minute band and once for the 68.5 to 69.5-minute band. In short, decayed jobs are tabulated differently for the two shapes shown in Figure 5.

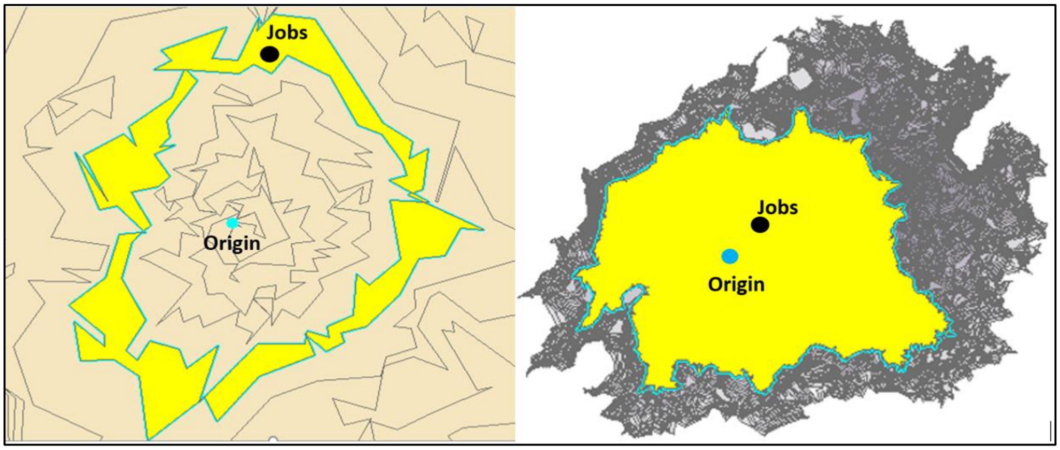


FIGURE 5. Example of a donut-shaped service area (left) and a pie-shaped service area (right).

It is infeasible to check the many thousands of service areas formed for each project. Instead, a script was adapted (*28*) to determine when a service area should be included in the accessibility computations. Table 2 shows the script's algorithm for a single population centroid, X, and two employment centroids, Y and Z. In reality, employment centroid Y (100 jobs) is located 3 minutes from centroid X and employment centroid Z (200 jobs) is located 7 minutes from centroid X. Although the network informs the first four columns, the fifth column is unknown without visual inspection: is the service area formed correctly as a donut or incorrectly as a pie?

The CheckCat attribute separates the population centroid and the employment centroid by the letter A: "XAY" denotes population centroid X and employment centroid Y. If donut-shaped service areas were consistently formed, XAY would appear only once in Table 2. When multiple rows have the same CheckCat value, the algorithm assigns a rank of 1 to the row with the shortest time, as the placement of the employment centroid in that particular service area is correct. For instance, for the 100 jobs in zone Y, the script shows that they truly are 3 minutes (and not 4, 5, 8, or 9 minutes) away from population X.

TABLE 2 Summary of Algorithm to Address Problem of Inconsistently Formed Service Areas

Information From the Network				Unknown	Information From the Script		
Facility ID	Midpoint Service Area Time[a]	JoinFID	JOBS	Service Area	Check Cat	Rank	Explanation (Relative to X)
X	2[a]	-1	Null	Pie	X-1	0	No jobs are 2 minutes away.
X	3	Y	100	Pie	XAY	1	There are 100 jobs located 3 minutes away.
X	4	Y	100	Pie	XAY	0	
X	5	Y	100	Pie	XAY	0	
X	6	-1	Null	Donut	X-1	0	No jobs are 6 minutes away
X	7	Z	200	Donut	XAZ	1	There are 200 jobs located 7 minutes away.
X	8	Y	100	Pie	XAY	0	
X	8	Z	200	Pie	XAZ	0	
X	9	Y	100	Pie	XAY	0	
X	9	Z	200	Pie	XAZ	0	

[a] For example, the first row shows the service area for 1.5 to 2.5 minutes.

Without this algorithm, accessibility differs from the desired value if the algorithm and decay values of Equations 1-5 are to be followed. For instance, for one region (block groups within 50 miles of a corridor where transportation investments were proposed), the present-day accessibility (Equation 5) was 752.4 million. The algorithms dropped this present-day accessibility more than an order of magnitude to 20.2 million. In short, the critical reason for some type of checking as proposed in Table 2 is to avoid mixing donut-shaped and pie-shaped service areas.

4.2 Quantify the Impact of Computational Decisions on Accessibility Scores

4.21 Impact of Centroid Connectors

The accessibility for each of five projects was determined by varying the number of centroid connectors from one to five with three catchment radii R. For each value of R, the maximum travel time considered was 2R + 0.5. For example, for any block group i within R = 5 miles of a project, the reduction in travel time to any employment site j was considered provided the time from i to j was less than 10.5 minutes. (Computationally, this 2R value allows one to have some alignment between the size of the catchment radius and the maximum travel time being considered.) The reason for the "0.5" addition was to catch the times in 1-minute bands with the integer as the midpoint: e.g., a travel time of 10 minutes would be in the 9.5 to 10.5-minute band. Table 3 shows the resultant accessibility scores (e.g., ΔS in Equation 5) and the percent difference in brackets between the score and the median value for each project: project 1 (5 miles, 1 centroid connector) had a score of 68, which was 43% less than the overall median (for all cells associated with project 1) of 120.

As the radius increases, the effect of the number of connectors lessens—an expected effect given that since the maximum travel time allowed for an accessibility improvement is increasing, examined trip lengths are increasing and the portion of the trip on the connector drops. At 10 miles, Table 3, with scores rounded to the nearest integer, shows that the quantity of connectors had a modest impact, on the order of 9% (project 5), 5% (project 3), and 3% (projects 2 and 4). For project 1, the largest accessibility score of 198 (one connector) is 95% larger than the smallest accessibility score of 101 (three centroid connectors). At 5 miles, large impacts were noted for two of the five projects; at 15 miles, large impacts were observed for none of the projects.

One way to choose an appropriate number of connectors is through the mean absolute deviation (Equation 6), under the theory that the centroid connectors themselves are not fundamentally part of the accessibility improvement and thus changes in accessibility that are attributed to changes in the number of connectors are not desired. The mean accessibility score of each block group (x_i) was deducted from the scores of each block group (y_i); the differences were summed and divided by the number of block groups (N). The lowest mean absolute deviation indicates which number of centroid connectors minimizes variability for these scores. For the data in Table 3, four centroid connectors yield the lowest variability (1.363) and hence the greatest stability; this result contrasts with those of other studies (*25, 29*) that recommend one to three centroid connectors be used for accessibility analysis.

$$MAD = \frac{1}{N}\sum_{i=1}^{N}|y_i - x_i| \qquad (6)$$

4.22 Impact of Catchment Radius

The use of project 5 with multiple catchment radii illustrates how these challenges affect project evaluation. Column 5 (Table 4) clearly shows that the accessibility score drops as the catchment radius (R in Equation 5) increases for the case of one centroid connector. For instance, the accessibility score (e.g., ΔS) at R = 45 miles (16.69) is about 15% of the accessibility score at 10 miles (R = 115.89).

TABLE 4 Summary of Results for Project 5

Miles (1)	Base Accessibility (2)	Improved Accessibility (3)	Difference in Accessibility (4)	Difference in Accessibility Negative Values Removed (5)[a]
5	26,712.48	27,240.40	527.93	531.90
10	64,602.95	64,718.84	115.89	119.79
15	69,723.65	69,816.14	92.49	96.60
25	48,145.95	48,194.27	48.32	57.33
35	104,387.14	104,417.03	29.89	36.75
45	270,382.32	270,399.01	16.69	26.35
55	307,023.55	307,028.77	5.22	19.97
65	231,915.39	231,912.16	-3.23	19.80
75	229,195.04	229,193.58	-1.46	20.45
85	180,668.90	180,668.36	-0.54	16.98

[a] Column 5 is the same as column 4 except that any zones with a decrease in accessibility were set to zero. For instance, in the 5-mile case (row 1), there were 51 zones in the project definition, and for 2 zones the accessibility dropped after the project was built. For those 2 zones, the population-weighted accessibility (i.e., the zone population multiplied by decayed jobs) summed

to -488,196. Had this amount been set to zero, the accessibility during the after period would have been 3,346,131,913 rather than 3,345,643,717, which would have increased the after accessibility from 27,240.40 in column 4 to 27,244.38. Thus, the difference in accessibility would have increased from 527.93 (in column 4) to 531.90 (in column 5).

Two reasons appear to explain why the radius R has a large impact on the score.

1. *There may be a higher percentage of affected trips with smaller radii.* For instance, for Project 5 (recognizing that the nature of the origin-destination matrix is such that many interchanges will not be affected by a project), the percentage of origin-destination pairs where the travel time is improved by the project is 1.73% at 25 miles but 0.43% at 45 miles.

2. *The benefit of a reduction in travel time is greatest for shorter trips because the rate of jobs decay follows the Gamma function, decreasing rapidly.* An improvement that reduces a commute from 10 minutes to 8 minutes yields a net improvement of 0.084: the accessibility contribution of the decayed job changes from 0.735 to 0.819. An improvement that reduces a 30-minute commute by 2 minutes yields a smaller net improvement of only 0.028, one-third the value at the larger distance. However, the denominator of this accessibility term—the population—usually increases linearly. As the marginal increase in population (denominator of Equation 5) is generally larger than the marginal increase in accessibility (numerator of Equation 5), it will usually be the case that an increased radius yields a reduced net accessibility.

4.23 Impact of Service Area Creation

At larger values of R, column 5 shows a negative accessibility score at radii of 65 to 85 miles. Examination of the service areas showed that for the 5-mile case, there were rare instances where for a particular origin block group the net accessibility dropped such that A_i^{After} - A_i^{Before} was negative. Such an instance is shown in the rectangular area of Figure 6, where a distance of about 22 feet appears to have caused the shift. The service areas, even when formed in a consistent manner, remain an approximation such that there is a modest element of random variation in their creation and there can be small-scale deviations in the boundary. At smaller radii, negative scores are not perceived as they are dwarfed by the positive scores from most zones i, but as the difference in accessibility score drops—which happens at larger radii—negative scores are possible.

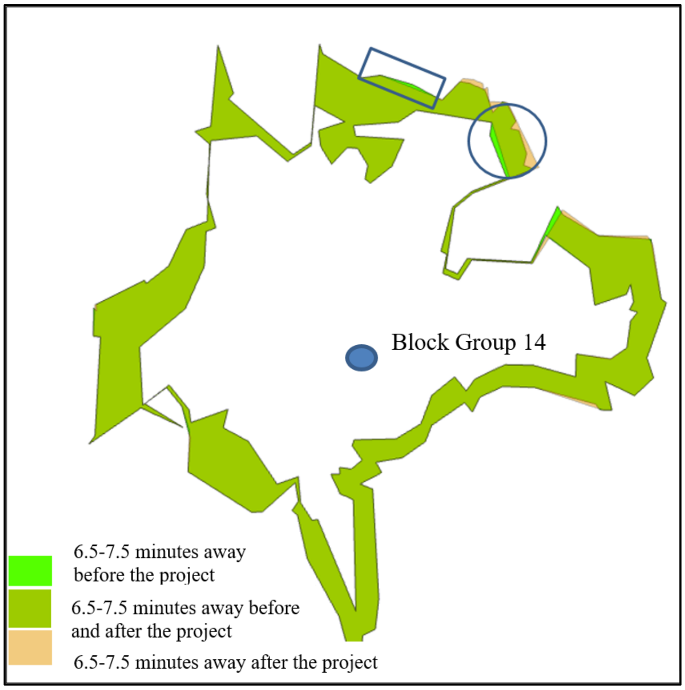


FIGURE 6. Expected Impacts (circle) and Unexpected Impacts (Rectangle) of a Proposed Project on the 7 Minute Service Area.

These service areas are generalized approximate polygons that ESRI (*30*) noted are "fairly accurate, except in the fringes," such that accuracy may be greater for high-density gridded streets in an urban core but may drop in rural areas. Even when a project is situated in an urban location, the odds that these service areas will include rural locations increases as the radius increases. For these two reasons—greater number of fringe areas and greater likelihood of incorporating rural locations—coupled with the observation that accessibility decreases as the catchment radius increases—it is not unexpected that there are some locations where an initially negative accessibility results. In Table 4, this occurred at radii of 65 miles and higher. For a different investment (project 1), additional experiments showed that negative results did not occur until a radius of 75 miles. In short, at large radii where the true delta accessibility approaches zero, this random variation becomes noticeable.

One way to address this variation is to set the decreases in origin zone accessibility to zero, with the view that a project should either have no impact on a zone's accessibility or increase that accessibility. That is, a decision maker might ask: "Presume all such variations (in service area formulation) are adversely affecting how this project affects accessibility. What happens if we eliminate such variations?" Implementing this approach yields the differences in accessibility shown in column 5 of Table 4.

(Note that this discussion considers the case of projects that are all designed to improve auto accessibility. In practice, however, a project might be proposed that would reduce accessibility, such as a speed limit reduction implemented for improving safety. In that instance, the approach would be similar except positive accessibilities are removed. A more challenging case would be a project that had both positive and negative accessibility impacts, such as the addition of street connections which shortened access times for some trips (through less circuitous travel paths) but which increased access times for other trips (through reduced speeds on some routes). In that case, it might be that one would need to evaluate the utility of this error checking at specific radii on a case by case basis.)

A critical lesson therefore is that the quality control of the input data is not completely independent of the design decisions: although the network quality herein was sufficient for radii considered by Virginia (typically 45 miles), one would presumably need a larger degree of quality if it were necessary to consider substantially longer radii. Similarly, one might choose a more precise method of service area creation with a smaller number of projects (where less speed was required) or smaller radii (where greater differentiation over a small space was essential). The one minute increments for time decay might be replaced with larger bins for larger radii, especially for feasibility of computations, or with smaller bins with a finer grained network, which might be the case with pedestrian-oriented improvements.

A critical lesson therefore is that the quality control of the input data is not completely independent of the design decisions: while the network quality herein was sufficient for radii considered by Virginia (typically 45 miles), one would presumably need a larger degree of quality if it were necessary to consider substantially longer radii. Similarly, one might choose a more precise method of service area creation if one had a smaller number of projects (where less speed was required) or smaller radii (where greater differentiation over a small space was essential).

5. DISCUSSION

Solving computational challenges, such as ensuring consistent service area formation, requires resources in the form of staff time and computing power. The same can be said for understanding the impacts of design choices, such as the number of centroid connectors. Accordingly, it is appropriate to understand the relative importance of these endeavors.

For project 5, Table 5 presents the robustness of this accessibility performance measure with regard to computational errors and design decisions. Row 1 shows a base scenario, with no computational errors; three explicit design decisions; and an initial accessibility score of 119.79. The succeeding rows show how a change, whether a computational error or a different design decision, alters the score. Excluding turn restrictions (an error) raises the accessibility score to 163.69—a deviation of roughly 44 units. A different decision (not an error) is a catchment radius of 15 rather than 10 miles. Such a design decision is not incorrect but it matters, altering the accessibility score by 23 units—more than one-half the error of not including turn restrictions.

The rightmost column of Table 5 shows each scenario as either a computational scenario (e.g., an error that should be corrected) or a design scenario (e.g., a parameter that a user could legitimately select). Of interest is the fact that only some of the scenarios are immediately evident when one initiates the accessibility computations. Clearly, most users of transportation networks will know to consider turn restrictions (row 6), and the importance of centroid connectors (rows 2 and 4) has been well defined in the literature (*24*, *28*). However, the need to address inconsistencies in service areas (row 9) is not evident unless one examines many such areas to find the relatively random mistakes that do occur; similarly, the possibility of a negative accessibility score resulting from approximations in service areas (row 3) is not typically realized unless sensitivity testing is performed. Although this latter decision has a minor impact at a 10-mile catchment radius, at a 45-mile catchment radius it can alter the score by about 50%.

TABLE 5 Robustness of the Accessibility Score (Project 5, 10-Mile Radius)

Row	Scenario	Score	Deviation	Scenario Category
1	Compute accessibility score correctly (avoid rows 2-4). Design choices: • 4 centroid connectors • 10-mile catchment radius • Exclude negative accessibility zones	119.79	0	
2	Choose to use 4 centroid connectors	121.66	2	Design
3	Choose not to exclude negative accessibilities resulting from approximations in service areas	115.89	4	Computation
4	Choose to use 2 centroid connectors	132.48	13	Design
5	Choose a different catchment radius of 15 miles	96.60	23	Design
6	Do not include turn restrictions	163.69	44	Computation
7	Choose a different catchment radius of 45 miles	26.35	93	Design
8	Do not fix one-way versus two-way streets	2.20	118	Computation
9	Do not fix service areas shaped like pies rather than donuts	757.57	638	Computation

Other design choices influence accessibility. For instance, the project location may be defined as either the project center or the entire corridor, but this decision ultimately affects the catchment area. Additional experiments showed such variation in location definition had, for the five projects, less than a 5% impact on the accessibility score on average at a 10-mile radius, with the impact dropping as the radius increased. Returning to Table 3, an additional experiment showed that the use of pie-based service areas (rather than donut-shaped service areas) affects the results: for a 10-mile radius with four centroid connectors, the rank of project 3 changed from second to first.

6. CONCLUSIONS

This paper shows ways to implement accessibility-based calculations on a relatively large scale within a GIS environment in a transparent manner. The first set of conclusions concern feasibility—i.e., computational steps needed to implement any accessibility computation in a GIS environment. The second set of conclusions concern transparency—i.e., the implications of design decisions chosen by the user.

6.1 Feasibility

- *When networks are passed from one package to another, care must be exercised to ensure that two-way travel is retained where appropriate.* The solution to this challenge is to alter the direction of the travel attribute such that it follows, rather than contradicts, the digitization direction. This legacy network challenge is not unique to Virginia.
- *It is possible to automate the incorporation of turn prohibitions, thereby saving substantial time (in this case, an estimated 2,000 hours of manual processing).* Rather than digitize turn prohibitions manually, a script from the literature (*26*, *27*) could largely be adapted provided one then performed an iterative additional step: use a custom script to create and populate the turn restriction attribute table because of the FCID varying with each editing session and not being detectable using conventional methods (*27*).
- *Service areas can be used provided an automated script, such as that shown herein, corrects inconsistencies.* If service areas are intended to be rings (e.g., employment centers 10 to 11 minutes away from origin A), the service areas should always have the shape of a donut. In random cases, however, some service areas were pie-shaped—showing employment centers within 11 minutes of origin A. This error of mixing service area types (donuts and pies) is hidden from the user. The solution is to modify a sorting subroutine (*28*) to identify duplicative employment centers automatically and remove them from the accessibility calculation. Manual inspection is infeasible: with 100 origins and 90 bands (1 band for each minute), there are 100 x 90 = 9,000 service areas. The pie-based service areas are not materially wrong, but when service area types are mixed, the accessibility results can be nonsensical.

6.2 Transparency

- *The single most important parameter chosen by the user is the catchment radius*. The accessibility score starts to drop at some relatively small radius, owing to the fact that in most situations, the marginal increase in population (denominator of Equation 5) more than offsets the marginal increase in accessibility (numerator of Equation 5). This radius also affects the impact of other design choices.
- *Other modest network design decisions may materially affect the accessibility score for some projects*. At a catchment radius of 10 miles, for four projects, varying the number of centroid connectors from one to five showed that the highest score was no more than 10% higher than the lowest score. For a fifth project, however, the highest score (with one connector) was almost twice that of the lowest score (with three connectors). A mean absolute deviation analysis can be used to choose the number of

centroid connectors that minimizes variability, which in this case was four connectors. As the radius increases, the impact of the number of connectors lessens.

- *At large radii, the accessibility score of projects may be underestimated because of random variation in the creation of service areas where the borders shift slightly.* The solution is to identify block groups yielding a negative contribution to accessibility and then correcting this negative amount to a zero value.

6.3 Future work and Limitations

Although this paper presents the importance of documenting design decisions, additional work is needed to identify which choices generate public support. One approach would be to use a public participation process where stakeholders would identify best practices through a delphi method (e.g., choose a smaller, rather than larger, catchment radius). A second approach would be to quantify which design choices best align with a priori observations of behavior (e.g., which catchment radii yields the best correlation between a change in accessibility and a change in trip distribution patterns). These approaches are complementary, and probably a combination of the two, in conjunction with transparent computations, is needed to fulfill the “vetting” advocated by Sundquist (*9*, *10*) to bring greater credibility to the use of accessibility in project prioritization.

ACKNOWLEDGMENTS

The study on which this paper was based benefitted from insights provided by the technical review panel guiding the effort (champion Peng Xiao, Ziwen Ling, Amy O’Leary, and Ram Venkatanarayana); IT assistance (Linda Cullop, Eric Hetzer, Michele Mandell, Adam Munro, Alex Ptak, and an anonymous member of the ESRI support team); and editing (Linda Evans). The authors are also grateful for the support for this research by the Virginia Department of Transportation and FHWA. The material in this paper is the responsibility of the authors and does not necessarily represent the viewpoint of these agencies.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: R. Boateng and J. Miller; data collection: R. Boateng; analysis and interpretation of results: R. Boateng and J. Miller; draft manuscript preparation: R. Boateng and J. Miller. Both authors reviewed the results and approved the final version of the manuscript. The authors do not have any conflicts of interest to declare.

REFERENCES

1. ATKINS. North Carolina Statewide Transportation Plan: From Policy to Projects, 2040 Plan. North Carolina Department of Transportation, Raleigh, 2012.

2. National Association of Development Organizations. Transportation Project Prioritization and Performance-Based Planning Efforts in Rural and Small Metropolitan Regions. Washington, D.C., 2011.
3. Delaware Department of Transportation. DelDOT Project Prioritization Criteria. Dover, undated.
4. Vermont Agency of Transportation. Project Prioritization. Montpelier, undated.
5. Ohio-Kentucky-Indiana Regional Council of Governments. 2040 Oki Regional Transportation Plan: Moving the Region Forward: Project Prioritization Process. Cincinnati, 2016.
6. Gunasekera, K., and I. Hirschman. NCHRP 08-36, Task 112: Cross Mode Project Prioritization. 2014. http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36(112)_FR.pdf. Accessed July 17, 2019.
7. Wasatch Front Regional Council. DRAFT 2019-2050 Regional Transportation Plan. 2019. https://wfrc.org/VisionPlans/RegionalTransportationPlan/InProgress2019_2050Plan/DRAFT_FOR_PUBLIC_COMMENT_DRAFT_2019_2050_RTP_Document.pdf. Accessed July 17, 2019.
8. Commonwealth of Virginia. *SMART SCALE Technical Guide.* Richmond, 2018.
9. Sundquist, E. *Project Proposal: Accessibility as a Decision-making Tool for Utah Planning.* Utah Department of Transportation, Salt Lake City, 2018.
10. Sundquist, E., C. McCahill, and L. Dredske. *Accessibility in Practice: A Guide for Transportation and Land Use Decision Making.* Office of Intermodal Planning and Investment, Richmond, Va., 2017.
11. Ford, A. C., S. L. Barr, R. J. Dawson, and P. James. Transport Accessibility Analysis Using GIS: Assessing Sustainable Transport in London. *International Journal of Geo-Information,* 2015. 4: 124-149.
12. Litman, T. A. Evaluating Accessibility for Transport Planning: Measuring People's Ability to Reach Desired Goods and Activities. Victoria Transportation Policy Institute, 2019. https://www.vtpi.org/access.pdf. Accessed Jan. 23, 2020.
13. Kempf, P., V. Gavrilovic, S. Wagg, D. Hurst, and J. Espie. Measuring the Accessibility of Intermodal Centers in Virginia, pp. 21-36. In *Proceedings of 1st National Conference on Intermodal Transportation: Problems, Practices, and Policies.* Eastern Seaboard Intermodal Transportation Applications Center, Hampton, Va., 2012. http://esitac.biz.hamptonu.edu/media/docs/ncit2012_proceedings_rev20121214.pdf. Accessed Jan. 23, 2020.
14. Newmark, G. L., and P. M. Haas. Income, Location Efficiency, and VMT: Affordable Housing as a Climate Strategy. The California Housing Partnership, San Francisco, 2015.
15. Blanchard, S. D., and P. Waddell. UrbanAccess: Generalized Methodology for Measuring Regional Accessibility with an Integrated Pedestrian and Transit Network. *Transportation Research Record: Journal of the Transportation Research Board,* 2017. 2653: 35-44.
16. Williams, K. M., J. Kramer, Y. Keita, L. D. Enomah, and T. Boyd. Integrating Equity Into MPO Project Prioritization. USDOT University Transportation Center, The University of Texas at Arlington, 2019.
17. Lasley, P. *Expanding Your Toolbox of Performance Measures at a Low Cost: A Total Peak-Period Travel Time Using Existing and Available Data.* Transportation Research Board, Washington, D.C., 2015.
18. Proffitt, D. G., K. Bartholomew, R. Ewing, and H. J. Miller. Accessibility Planning in American Metropolitan Areas: Are We There Yet? *Urban Studies,* 2019. 56: 167–192.
19. Wasatch Front Regional Council. Access to Opportunities. 2020. https://wfrc.org/maps-data/access-to-opportunities/#1574886662041-98a852e7-2c2a. Accessed Jan. 24, 2020.
20. Guo, Y., S. Agrawal, S. Peeta, and S. Somenahalli. Impacts of Property Accessibility and Neighborhood Built Environment on Single-Unit and Multiunit Residential Property Values. *Transportation Research Record: Journal of the Transportation Research Board,* 2016. 2568: 103-112.

21. Tsui, J., and M.-A. Bourque. Evaluating Projects in the Regional Municipality of York Transportation Master Plan Using a Quantitative Priority Setting Model. Presented at the 96th Annual Meeting of the Transportation Research Board, Washington, D.C., 2017.
22. Khan, S. I. A Multicriteria Index for Evaluating Alternatives and a Case Study for Transport Investment Decisions. Presented at the 97th Annual Meeting of the Transportation Research Board, Washington, D.C., 2018.
23. AASHTO. GIS for Transportation Symposium: State DOT GIS-T Contact Information. 2020. https://gis-t.transportation.org/who-is-gis-t/state-dot-contacts/. Accessed July 20, 2020.
24. Jafari, E., M. D. Gemar, N. R. Juri, and J. Duthie. An Investigation of Centroid Connector Placement for Advanced Traffic Assignment Models with Added Network Detail. *Transportation Research Record: Journal of the Transportation Research Board,* 2015. 2498: 19-26.
25. LSA Connectics Transportation Group. *WATS Travel Demand Model Improvements. Technical Memorandum 13: Additional Mode Recommendations.* 2008. ftp://ftp.ci.missoula.mt.us/DEV%20ftp%20files/Transportation/MPO/MODEL_ENHANCEMENT/RFP/Proposals/LSA/Reference/Model%20Documentation/Ann%20Arbor%20(Washtenaw%20County)%20Model/TM13_WATSRTM_AdditionalModes_August2008.pdf. Accessed July 23, 2020.
26. Geographic Information Systems Stack Exchange. Using ArcPy to Populate Turn Feature Class for Network Analyst by Reading Edge Identifiers from Separate Table. 2018. https://gis.stackexchange.com/questions/214719/using-arcpy-to-populate-turn-feature-class-for-network-analyst-by-reading-edge-i. Accessed Feb. 8, 2019.
27. ESRI. Copying Source Feature Classes. 2019. http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/network-analyst/copying-source-feature-classes.htm. Accessed Feb. 26, 2019.
28. ArcPy Café. Ranking Field Values. 2013. https://arcpy.wordpress.com/tag/ranks/. Accessed May 2, 2019.
29. Cambridge Systematics, Inc. *KATS Travel Model Update: Technical Documentation.* 2015. https://katsmpo.files.wordpress.com/2015/02/kats-travel-model-update-technical-documentation-06292015.pdf. Accessed July 23, 1010.
30. ESRI. Service Area Analysis. Redlands, California, 2019. https://desktop.arcgis.com/en/arcmap/10.5/extensions/network-analyst/service-area.htm. Accessed October 8, 2020.

APPENDIX

Table 1. Summary of Scripting Steps for Automation of Turn Prohibitions, Adapted From (*26, 30*)

Step (No.)	Example Implementation[a]
Create a turn feature class (1)	MyFeatureDataset=r"E:\State.gdb\VaRoads" *arcpy.CreateTurnFeatureClass_na(MyFeatureDataset,"VaTurnRestrictions","2")* *arcpy.AddField_management("VaTurnRestrictions","TurnPermitted","SHORT")*
Create a line feature class (2)	*arcpy.CreateFeatureclass_management(out_path=MyFeatureDataset, out_name="TurnLines", geometry_type="POLYLINE", template="", has_m="DISABLED", has_z="DISABLED")* *arcpy.AddField_management("TurnLines","TurnPermitted","SHORT")*
Populate a turn prohibition table (3)	TurnNotAllowed=r"E:\Statewide\TurnNotAllowedTable.csv" (Note the table has 3 columns: an identification number, a FROM street ID, and a TO street ID.)
Create a list of prohibited turns (4)	*turnNotAllowedList = [[row[0],row[1]] for row in arcpy.da.SearchCursor("TurnNotAllowed",["FromStreetID","ToStreetID"])]*
Create a data dictionary (5)	*MyLinks= r"E:\State.gdb\VaRoads\VaLink"* *streetDict = {row[1]: row[0].positionAlongLine(0.5,True).firstPoint for row in arcpy.da.SearchCursor(MyLinks, ["SHAPE@","StreetID"])}* *print streetDict*
Create a cursor that will populate the TurnLines feature class (6)	*MyTurnLines=r"E:\State.gdb\VaRoads\TurnLines"* *cursor = arcpy.da.InsertCursor(MyTurnLines, ["SHAPE@","TurnPermitted"])* *for turn in turnNotAllowedList:* *array = arcpy.Array()* *array.add(streetDict[turn[0]])* *array.add(streetDict[turn[1]])* *polyline = arcpy.Polyline(array)* *cursor.insertRow([polyline,-1])* *del cursor*
Obtain the feature class identification number (FCID) (7)	Because the researcher observed that the method in (*4*) did not always provide a reliable FCID, initiate an editing session with the turn feature class. Manually draw a line from one roadway link to another, and observe the value of the FCID.[a]
Create the edges data for the turn restriction file (8)	MyTurnRestrictions=r"E:\State.gdb\VaRoads\VaTurnRestrictions" cursorMy = arcpy.da.InsertCursor(MyTurnRestrictions, ["SHAPE@","Edge1FID","Edge2FID","Edge1End","Edge1FCID","Edge2FCID","Edge1Pos","Edge2Pos","TurnPermitted"]) [b] MyDatabase=r"E:\State.gdb" edit = arcpy.da.Editor(MyDatabase) edit.startEditing(False,False) edit.startOperation() for turn in turnNotAllowedList: array = arcpy.Array() array.add(streetDict[turn[0]]) array.add(streetDict[turn[1]]) polyline = arcpy.Polyline(array) cursorMy.insertRow([polyline,turn[0],turn[1],"Y",**5**[c],**5**[c],0.5,0.5,-1]) del cursorMy edit.stopOperation() edit.stopEditing(True)

[a] Italicized lines are taken directly from Geographic Information Systems Stack Exchange *(1).* The material is italicized, rather than enclosed in quotation marks, as quotation marks are also used in the script.

[b] These attributes were suggested to the researcher by an anonymous member of the ESRI support team

[c] When repeating this procedure, the value of "5" shown in the last row should be replaced with the value of FCID in the previous row.

Table 2. Scripts to Eliminate Employment Based On Incorrectly Formed Service Areas, Adapted From (*28*)

Algorithm Step	**Script**[b]
Adapt from Arcpy Café (*3*) a function can sort two related sets of data. The function will be modified to rank employment centroids for a common population centroid on the basis of distance. Notice that when the rank, shown as row[1], has a value of 1 then the placement of the employment centroid in that particular service area is correct.	*import arcpy* *def addRanks(table, sort_fields, category_field, rank_field='RANK'):* *arcpy.AddField_management(table, rank_field, "SHORT")* *sort_sql = ', '.join(['ORDER BY ' + category_field] + sort_fields)* *query_fields = [category_field, rank_field] + sort_fields* *with arcpy.da.UpdateCursor(table, query_fields,* *sql_clause=(None, sort_sql)) as cur:* *category_field_val = None* *i = 0* *for row in cur:* *if category_field_val == row[0]:* *i += 1* *else:* *category_field_val = row[0]* *i = 1* *row[1] = i*
Modify the function to set the rank for croissant-shaped service areas to zero	if i > 1: row[1]=0 *cur.updateRow(row)*
Create a table that will contain the CheckCat variable	NewTable=r"E:\State.gdb\ServiceAreaCheckTable" arcpy.AddField_management(NewTable,"CheckCat","TEXT",10)
Develop the CheckCat variable based on the population centroid and the employment centroid	arcpy.CalculateField_management (NewTable, "CheckCat", "str(!FacilityID!) + 'A' + str(!JOIN_FID!)", "PYTHON_9.3")
Calculate the Rank variable	addRanks(MyNewTable,['Break'], 'CheckCat', 'rank')
Calculate the decayed jobs and multiply by the rank variable [a]	Partial DecayedEmployment = = max (0, "jobs") * DecayValues ("Break")*"rank"

[a] This is an excerpt of the script. In the full script, a data dictionary is created to store for each population centroid a list of the decayed jobs, such as for zone X, the 100 jobs multiplied by the appropriate decay value for 3 minutes, and the 200 jobs multiplied by the appropriate decay value for 7 minutes.

[b] Italicized lines are taken directly from ArcPy Café (*3*). The material is italicized, rather than enclosed in quotation marks, as quotation marks are also used in the script.

Table 3. Spatial Join Centroids to Service Areas *(30)*

Description	Scripts
Import the file	import arcpy arcpy.env.workspace=r"G:\TRB2021_BEFORE_AFTER_DISK.gdb" arcpy.env.overwrite="True"
Define the location of the new file	NewResults=r"G:\TRB2021_BEFORE_AFTER_DISK.gdb\ResultsSA_P4_Before30Miles"
Spatial Join	arcpy.SpatialJoin_analysis("SA_P4_Before30Miles","P4CC4CENTROIDS30MILES", NewResults, "JOIN_ONE_TO_MANY", "KEEP_ALL", match_option="INTERSECT")

Table 4. Accessibility Scripts *(30)*

Description	Scripts
Import the data	import arcpy arcpy.env.workspace=r"G:\ Project4BeforeAfter.gdb"

Import the Decay factors	def DecayValues(EndingTime):
	AutoDecayList = [[-1,4,1.00000],
	[4,5,0.962622],[5,6,0.912019],[6,7,0.864076],[7,8,0.818653],
	[8,9,0.775618],[9,10,0.734846],[10,11,0.696216],[11,12,0.659618],
	[12,13,0.624943],[13,14,0.592091],[14,15,0.560966],[15,16,0.531477],
	[16,17,0.503539],[17,18,0.477069],[18,19,0.451990],[19,20,0.428230],
	[20,21,0.405719],[21,22,0.384391],[22,23,0.364184],[23,24,0.345040],
	[24,25,0.326902],[25,26,0.309717],[26,27,0.293436],[27,28,0.278011],
	[28,29,0.263396],[29,30,0.249550],[30,31,0.236432],[31,32,0.224003],
	[32,33,0.212228],[33,34,0.201071],[34,35,0.190502],[35,36,0.180487],
	[36,37,0.170999],[37,38,0.162010],[38,39,0.153494],[39,40,0.145425],
	[40,41,0.137780],[41,42,0.130537],[42,43,0.123675],[43,44,0.117174],
	[44,45,0.111014],[45,46,0.105179],[46,47,0.099650],[47,48,0.094411],
	[48,49,0.089448],[49,50,0.084746],[50,51,0.080291],[51,52,0.076070],
	[52,53,0.072072],[53,54,0.068283],[54,55,0.064693],[55,56,0.061293],
	[56,57,0.058071],[57,58,0.055018],[58,59,0.052126],[59,60,0.049386],
	[60,61,0.046790],[61,62,0.044330],[62,63,0.042000],[63,64,0.039792],
	[64,65,0.037700],[65,66,0.035718],[66,67,0.033841],[67,68,0.032062],
	[68,69,0.030376],[69,70,0.028779],[70,71,0.027267],[71,72,0.025833],
	[72,73,0.024475],[73,74,0.023189],[74,75,0.021970],[75,76,0.020815],
	[76,77,0.019721],[77,78,0.018684],[78,79,0.017702],[79,80,0.016771],
	[80,81,0.015890],[81,82,0.015054],[82,83,0.014263],[83,84,0.013513],
	[84,85,0.012803],[85,86,0.012130],[86,87,0.011492],[87,88,0.010888],
	[88,89,0.010316],[89,90,0.009773]]
	AutoDecay=0
	for row in AutoDecayList:
	MidTime=EndingTime-0.5
	if MidTime>row[0] and MidTime<=row[1]:
	return row[2];AutoDecay=row[2]
	if AutoDecay==0:
	print ("Warning: Decay Value set to zero because you are out of range")
	return 0

Create a population data dictionary	PopDataDictionary={}
	PopulationLayer=r" G:\ Project4BeforeAfter.gdb \P4CC4Centroids_30miles"
	TotalPop=0
	with arcpy.da.SearchCursor(PopulationLayer,["OBJECTID","Pop2025"]) as cursor:
	for row in cursor:
	Population=row[1]
	TotalPop=TotalPop+Population
	if row[0] not in PopDataDictionary.keys():
	PopDataDictionary[row[0]]=Population

Associate, with each facility ID, the decayed jobs for each zone	DataDictionary={}
	with
	arcpy.da.SearchCursor("ResultsSA_P4_Before30Miles",["FacilityID","Pop2025","Emp2025","ToBreak","ran
	k"]) as cursor:
	for row in cursor:
	Employment = (max(0,row[2]))*DecayValues(row[3])*row[4]
	if row[0] not in DataDictionary.keys():
	DataDictionary[row[0]]=[Employment]
	else:
	DataDictionary[row[0]].append(Employment)

Compute Accessibility Scores with	import csv FileLocation=r"C:\TRB Results CSV\ ResultsSA_P4_Before30Miles.csv" with open (FileLocation,'wb') as csvfile:

Description	Scripts
Import the data	import arcpy arcpy.env.workspace=r"G:\ Project4BeforeAfter.gdb"

final output as CSV file	csvwriter=csv.writer(csvfile,delimiter=',')
	AlmostGlobalAccessibility=0
	TotalEq3=0
	data=["Population is", TotalPop]
	csvwriter.writerow(data)
	data=["FacilityID","Decayed Employment","Population","Population Decayed Accessibility","Running Total"]
	csvwriter.writerow(data)
	for FacilityID in DataDictionary.keys():
	EmpList=DataDictionary[FacilityID]
	TotalEmp=sum(EmpList)
	#AlmostGlobalAccessibility=AlmostGlobalAccessibility+TotalEmp
	#New part for Equation 3
	Pop=PopDataDictionary[FacilityID]
	Eq3=Pop*TotalEmp
	TotalEq3=TotalEq3+Eq3
	data=[FacilityID,TotalEmp,Pop,Eq3,TotalEq3]
	csvwriter.writerow(data)
	#data=["Almost Total Accessibility is",AlmostGlobalAccessibility]
	#csvwriter.writerow(data)
	data=["New way is ",TotalEq3]
	csvwriter.writerow(data)
	data=["Final Accessibility is", TotalEq3/TotalPop]
	csvwriter.writerow(data)
	#print AlmostGlobalAccessibility
	print TotalEq3/TotalPop

PAPER 2. WHAT IS THE ASSOCIATION BETWEEN AUTO ACCESSIBILITY AND TRAVELER BEHAVIOR?

(Proposed Submission to the Journal of Transportation Planning Education and Research)

ABSTRACT

While accessibility has frequently been proposed as an element in project prioritization, recent work has shown that this formulation is highly sensitive to one particular computational parameter: the sphere of influence considered for the candidate project, formally defined as the catchment radius. The manner for selecting this radius has not been resolved, although multiple approaches are feasible. The Delphi method, for instance, entails convening a panel of experts to select a radius such that the accessibility scores from a series of projects agrees with the apriori ranking of such experts. A public participation approach is to select a radius that addresses stakeholder concerns, such as choosing a large radius in order to account for longer commutes that would benefit from the project. This paper proposes, and then evaluates, another approach: choosing the catchment radius that gives best fit between observed and forecast behavior. The paper also considers the effects of confounding factors on this association, notably: geographical level of aggregation tract; income, project location, disparity between housing costs at the origin and destination, and localized diversity in such housing costs. An ancillary benefit of this work is that the study quantifies the amount of variance in traveler behavior explained by accessibility alone.

The first part of the paper uses accessibility models to compute accessibility and rank the projects for each of the catchment radii. Findings from this part of the study indicate that catchment radius 5-35 miles around each project indeed affected accessibility however at 25, 30 and 35 miles catchment radii, altering the radius did not affect accessibility ranking of the projects. While findings from the second part of the paper indicate that there is significant relationship between observed and forecast behavior, the study further showed that catchment radii 25, 30 and 35 miles recorded the least amount of variations in the percentages for coefficient of correlation and determination ranging from 0.01% to 0.25%. The study found accessibility scores to be statistically significant at 95% confidence level with the highest percentage of variations in traveler behavior that can be explained by accessibility alone occurring at 35 miles catchment radius. While the lowest percent occurred at 5 mile catchment radius (3.7%), catchment radii 25, 30, and 35 recorded the highest percentage of variance representing between 9.4-10.2 percent. Although combining other variables to quantify the amount of variance explained by accessibility and other confounding factors improved the percent of variance, it did not significantly impact the overall percentage and it ranged from 0.4-4.2%.

Intersecting these three analysis, the study finds three catchment radii that have the potential to provide the best fit between observed and forecast behavior to be at 25, 30 and 35 miles. Among these three, catchment radius at 25 mile is recommended to be used because it is relatively smaller than the other two, it will require lesser processing and computational time and resources and will be the most cost effective radius to implement. Although some sources have shown that accessibility statistically significantly affects destination choice or jobs-housing balance, no sources measure the portion of behavior explained by accessibility. Hence this paper will bridge the gap of determining the portion of travel behavior explained by accessibility.

1. Introduction

Project prioritization—that is, the process through which candidate transportation investments are selected for implementation—often makes use of multiple criteria in the areas of safety, operational performance, the physical environment, infrastructure condition, and land development (Commonwealth of Virginia, 2018). One often-considered criterion is "accessibility"—that is, the ease with which the transportation system enables connections between residents and key social functions (Sinha and Labi, 2007) such as employment. Recent work has shown that accessibility score is highly sensitive to one particular computational parameter: the sphere of influence considered for the candidate project, formally defined as the catchment radius, where "catchment" (Hardy and Bell, 2019) reflects the geographical area over which project accessibility benefits are considered.

The manner in which this radius is selected has not been resolved, although multiple approaches are feasible. The Delphi method, for instance, entails convening a panel of experts to select a radius such that the accessibility scores from a series of projects agrees with the apriori ranking of such experts. A public participation approach is to select a radius that addresses stakeholder concerns, such as choosing a large radius in order to account for longer commutes that would benefit from the project. This paper proposes, and then evaluates, another approach: choosing the catchment radius that gives best fit between observed and forecast behavior as well as the effects of other confounding factors on this association, notably: geographical level of aggregation at tract level; difference in rent between Origin and Destination (O-D), difference in house price between O-D, (Both rent and house price are a form of housing cost), localized diversity of rent at O, localized diversity of rent at D, localized diversity of house price at O, localized diversity of house price at D, Income at O, Income at D and Distance to MPO center. An ancillary benefit of this work is that the study quantifies the amount of variance in traveler behavior explained by accessibility alone.

2. Literature Review

Several studies have been conducted to determine if any the relationship between accessibility and travel behavior. The studies show that accessibility has an impact on behavior—but the extent of this impact is not fully known especially with respect to one's choice of destination, where that choice will be a particular census tract or block group. This association is confounded by other factors.

2.1 Accessibility Affects Travel Behavior to Some Degree

Several studies have been conducted to determine the association between accessibility and travel behavior but no study has considered the extent to which accessibility explains variation in destination choice. Kockelman (1997) defined traveler behavior in several ways: as total vehicle miles traveled per household, nonwork vehicle miles traveled per household, whether a trip was made by auto or not (hence a binary variable), and autos per household member (age 5 or older). In each case when such behavior was treated as the dependent variable, accessibility (defined in some instances as jobs within 30 minutes by auto or and in other instance as decayed jobs) was highly significant (p<0.01). However, accessibility was not the only variable in these models: other independent variables reflect the "built environment" (e.g., a dissimilarity index that indicates the extent to which land uses are homogeneous or heterogeneous) or socioeconomic characteristics (such as income)—and often these variables were either significant (*p*<0.10) or highly significant. While the independent variables were significant, these models generally explained a minority of the variance in the dataset: for example, when the dependent variable was VMT per household, the model included six highly significant variables (the constant, household size, auto ownership, income, land use mix, and accessibility to jobs within 30 minutes)—and one significant variable (entropy)—and the model explained 15% of the variation in the dataset. (For nonwork VMT, the model explained 6% of the variation in the dataset.) Thus one would not expect accessibility alone to increase a large proportion of VMT, but the research leaves open the question of the extent to which accessibility can explain any variation in destination choice, as that was not included as a dependent variable.

Merlin (2014) defined regional accessibility as a "measure of accessibility from a specified point to all of the destinations of interest in the metropolitan region", contrasting that with local accessibility as a "measure of accessibility from the same point, but only to destinations located within the boundaries of a particular local geography." The measure of accessibility was the highest of retail and service jobs density in each travel analysis zone. Accessibility share was the ratio of local accessibility to regional accessibility. Merlin (2014) found that such accessibility share variables better explained internal tour capture mixed-use variables; further, accessibility share was found to be the single most influential variable that predicted one's ability to complete a nonwork tour without leaving his/her home, compared to mixed-use variables.

Another study (Merlin, 2015) investigated whether variation in the built environment (measured as residential density, employment density, urban and metro area population, and presence of rail in metro area) can potentially influence households participating in out of home and nonwork activities. The study counted all nonwork activities irrespective of the travel mode of transport used and used the negative binomial model to model to account for travel behavior. The built environment variables have higher influence on household involvement in non-working activities, rising or decreasing rates of activity between 8% and 47%, depending in large part on the level of ownership of a household vehicle. Similarly, Guan and Wang (2019) conducted a study to determine the impacts of built environment (residential and work locations) on married couples' travel behavior at the household level. A structural equation model indicated that household members' travel behaviors are affected by their travel attitudes and the built environment, particularly by the location of their residence and work. Contrary to Merlin (2015) and Guan and Wang (2019), Simma and Axhausen (2003) found that spatial structure has little influence on travel behavior. (Simma and Axhausen [2003] had used structural equation modelling along with the spatial structure as an accessibility measure as well as personal characteristics of travelers.)

2.2 Other Factors Affect Behavior in Addition to Accessibility

While it is plausible that accessibility could explain some portion of trip distribution, the literature (Bohnet and Gutsche, 2007; Zhang et al. 2009; Lavieri et al. 2018; Behara et al. 2018; and Lasley, 2017), suggests that other factors besides accessibility affect the distribution of trips, such as income, age, household size, auto ownership levels, transit service, and school events. Lavieri et al.

2018 found, for instance, that when considering just one mode ("ridesourcing" defined as drivers offering coordinated services based on some type of coordinated system), income played a role, with wealthier motorists using ridesourcing more on weekdays and poorer motorists using ridesourcing on weekends, possibly to compensate for poor transit service at that time. Land use also affects trip distribution: Zhang et al. 2009 found that for the Central Austin MPO, land uses that could be characterized as mixed use had work trips that were 1.9 miles shorter than those that originated from more homogeneous land uses (a reduction of 18%). While it is not surprising that housing prices will influence one's choice of location (or choice of work), Bohnet and Gutsche, 2007 noted that suburban residents make a fundamentally different tradeoff, in terms of "higher costs in time and money" than do more centralized residents, which for this effort would suggest potentially disparate results if one examines suburban versus CBD locations. Behara et al. (2018) found that school holidays could disrupt travel patterns (in the short term); in the longer term, school quality affects the choice of residential location, with key differences being noted between income groups (Lasley, 2017). For example, Lasley (2017) found that, when choosing a residential location where the home will be purchased, while home price was the most important factor for all groups, low-income groups differed from middle and high income groups in that "neighborhood quality" (defined as "reputation, aesthetics, and amenities") mattered less than factors such as school quality. Lin and Yu (2016) investigated the relationship between job accessibility and apartment rents. One of the three questions they wanted to find out was if accessibility to job has any positive influence on apartment rents. Using empirical data from the Taipei Metropolitan Area (Taiwan), linear regression, and quantile regressions, the study concluded that accessibility to job has a positive influence on apartment rents. The study further revealed that fresh policies should target helping housing affordability in cities.

To be clear, these factors exert an influence on trip distribution in different ways: school quality, for instance, may affect the location of a resident, such that once the locations are fixed (e.g., the population of a zone is similar for both observed and modeled trips), one might not expect school quality to materially degrade the performance of a model. However, because school quality may be incorporated into home prices (Lasley, 2017) which in turn will affect trip distribution (e.g., a wealthy resident with a high paying job may choose a different employment location than a poorer resident with a lower paying job), one would expect the location of jobs for a given zone to not be the only factor that affects the distribution of journey-to-work trips. In fact, none of these sources indicate that accessibility alone forecasts commuter behavior, rather, these sources indicate that accessibility is one of several factors that influence behavior. For example, Lavieri et al. (2018) defined transit accessibility as the mean number of buses at an "average" bus stop within the TAZ and found that such accessibility was just one of 13 variables that explained variation in trip generation rates by zone (and the t-statistic of that variable was 1.75). The authors showed that transit accessibility had an impact on trip generation rates after controlling for other variables (e.g., median household income, population density, and other factors such as presence of parks). That said, Lavieri et al. (2018) does not quantify the strength of that association based solely on accessibility as a predictor variable. By contrast, the proposer seeks to initially quantify the extent to which accessibility alone explains trip patterns and then add in other factors (see Equation 1 on page 10) as needed.

2.3 The Role of the Catchment Radius has not been Fully Examined

The impact of the catchment radius on project-driven accessibility has not been examined in detail, although different default catchment radii have been used in previous studies; some studies defined this radius in terms of distance, and other defined this radius in terms of travel time. Kockelman (1997) and Geurs and Wee (2004) both considered jobs that were within 30 minutes by car only. Pyrialakou et al. (2016) also used varying radii depending on the type of opportunity under consideration (all travel times are by auto): 9 miles / 10 minutes (large schools); 18 miles / 20 minutes (museums); 14 miles / 15 minutes (public libraries); 28 miles / 30 minutes (Amtrak stations); and 37 miles /40 minutes (airports). Pokharel and IEDA (2016) used a critical distance of 9 miles, and Conway et al. (2017) used destinations within a median travel time of 45 minutes. Other studies also used larger catchment radii to evaluate candidate projects, but these also varied: Hardy and Bell (2019) used a fixed catchment radius of 45 miles; Merlin et al. (2018) used zone-to-zone accessibilities such that the catchment radius reflected the entire San Antonio Region of about 70 miles, and Texas A&M Transportation Institute and Economic Development Research Group (2014) used a large catchment radius that covers the entire region of roughly 300 miles.

In sum, accessibility has been shown to affect VMT (Kockelman, 1997 and Merlin, 2015), number of trips (Serulle and Cirillo, 2016; Lasley, 2017), but it has not been shown to affect destination choice; further, the extent to which altering the catchment radius (where different values have been used [Kockelman, 1997; Geurs and Wee, 2004; Pyrialakou et al., 2016; Pokharel and Ieda, 2016; Hardy et al, 2019; Merlin et al., 2018; and The Texas A&M Transportation Institute and Economic Development Research Group, Inc., 2014]) affects the relationship between accessibility and destination choice has not been examined. Thus there are two gaps: first, the extent to which altering catchment radius alone can strengthen the accessibility-behavior relationship is not known, and second, the extent to which confounding factors, such as income (Lavieri et al, 2018 and Lasley, 2017) and price (Bohnet and Gutsche, 2007 and Lasley, 2017), affect this relationship has not been fully specified.

3. Purpose and Scope

This study considers the extent to which differences in accessibility scores (measured as the distribution of decayed jobs for a given location) explain differences in observed behavior (measured as trips made for the purposes of traveling to or from work). While accessibility has frequently been proposed as an element in project prioritization, recent work has shown that this formulation is highly sensitive to one particular computational parameter: the sphere of influence considered for the candidate project, formally defined as the catchment radius. The manner for selecting this radius has not been resolved, although multiple approaches are feasible. The Delphi method, for instance, entails convening a panel of experts to select a radius such that the accessibility scores from a series of projects agrees with the a-priori ranking of such experts. A public participation approach is to select a radius that addresses stakeholder concerns, such as choosing a large radius in order to account for longer commutes that would benefit from the project. This paper proposes, and then evaluates, another approach: choosing the catchment radius that gives best fit between observed and forecast behavior.

The paper also considers the effects of other confounding factors on this association at the geographical level of aggregation (block group and tract level); household income, location of the project relative to generation of origin and destination patterns, rent and housing values.

4. Data Used

Virginia-specific data provided by Ling (2019) indicate expected total populations by Census block group for year 2025. Whereas each block group has data for the total populations, there is an equivalent attribute indicating jobs for these populations.

The researcher are aware of three data sets that may be used to obtain some type of trip distribution behavior as suggested by (Ford et al, 2015 and Cambridge Systematics, 2015): location-based systems (LBS) available through StreetLight Insight*,* census journey to work data available through the Census Transportation Planning Package (CTPP), (Table A302103 titled "Means of transportation Workers 16 years and over"), and possibly National Household Travel Survey (NHTS), data available through the Oak Ridge National Laboratory. None of these data sets perfectly fit accessibility data: StreetLight data are available at the block group level but are not available by mode; CTPP data are available by mode but are only available at the larger census tract level rather than at the smaller block group level, and neither data set matches perfectly the timing for the accessibility data set. StreetLight dataset was used for this study because the dataset can be obtained at block group level which is disaggregate and studies have shown the usefulness of disaggregate dataset (Richter and Brorsen, 2006; Hartman, 1983).

While the data and details of the candidate projects used for this study have been summarized in Table 1, the descriptive statistics of the demographic dataset have been summarized in Table 2. These include seven different data sources and five candidate projects located in the Commonwealth of Virginia. These five projects are among the several projects considered for implementation in future by VDOT.

4.1 How the Data Were Acquired

The income, housing value, and rental data in tabular form (U.S. Census Bureau, 2019c, 2019d) were linked to 2018 state-level census tract data in geographic form (U.S. Census Bureau, 2019a) which had been reprojected into Albers Equal Area to minimize spatial distortion. (The data cleansing process showed 1900 geographic tracts and 1907 tabular tracts; the discrepancy appeared to be that there are 7 tabular tracts all with null values for all attributes; all remaining geographic and tabular tracts showed a one-to-one match.)

Diversity measures were tabulated at the county level; for example, the diversity of housing values for each tract in Accomack County was based on the variation in values for the ten tracts in Accomack County. Data from a separate Census tabulation (U.S. Census Bureau, 2019b) shows that in Virginia, within any given year, 6.6% of the population moved to another location within the same county (compared to 9.1% of the population moving to from a different county, state, or county to a given Virginia county). Restricting these figures to persons age 25-64 (in order to focus more on the labor force but avoid college students) still shows similar figures—6.5% of Virginia's population moved during the past year to another location in the same county, compared to 8.8% of Virginia's population moving to a given Virginia county from another county, state, or country. That said, these figure are low compared to historical standards: Holmes (2018) reported that in 2017, only 11% of the U.S. population moved, and that estimate was lower than historical standards.

Because it was not clear if rents or housing values should be used for attribute Bij, two different attributes were tabulated: for tracts i and j, B_Rent_ij is the difference in rents and B_Housing_ij is the difference in home values. Similarly, for variables Ci and Dj which refer to localized diversity of housing costs, two variables were considered: DiversityRent_i is the diversity of rents in the county that encompasses zone i and DiversityHousing_i is the diversity of housing values in the county that encompasses zone I; each variable was computed as the standard deviation divided by the mean for the county using the SummarizeWithin tool in ArcGIS Pro, so that this variable represents a true diversity measure. A new variable Income was also added which is the income for households in zone i.

Table 1 Data and sample projects used for the study

Category	No.	Description
Input variables	1	*Highway Links for No-build Scenario*: These baseline network data covering the entirety of Virginia consist of more than 3 million links. The comprehensive road network dataset contains attributes such as distance; speed; travel times during the AM, peak; road functional class; travel direction, and digitization direction. Each link has a unique identification number that connects 2 nodes.
	2	*Highway Links for Build Scenario*: The build scenario dataset consists of the same dataset as element 1 with one exception: new speeds and new travel times reflecting the proposed transportation project being evaluated.
	3	*Junction Nodes*: The dataset consists of nearly 1.5 million nodes; each node has a unique code, which was useful when generating centroid connectors.
	4	*Block Groups*: These zones contain forecast demographic attributes for year 2025 such as population and employment.
	5	*Proposed Projects Dataset*: For each proposed project, this dataset consists of links indicating the project's location and, in conjunction with data element 2, enables one to determine how the project will affect link travel times.
	6	*Turn Restriction Dataset*: The Virginia turn restriction dataset contains codes that correspond perfectly with the junctions of the highway network. Each link in the Virginia highway network dataset also has its unique code. The data were further processed using MySQL to match the nodes that form each link, with identifiers indicating restricted turning movements.
	7	*One Minute Bin Decay Values:* These reflect the value of a job as a function of travel time. For example, a job that is between 5.5 and 6.5 minutes away has a value of 0.962622; a job that is 89.5-90.5 minutes away has a value of 0.009773.
	8	*Street Light Data:* Location-based systems (LBS) available through StreetLight Insight will be used
Projects	8	*Project 1: US 250/Route 20 Intersection Improvement (Charlottesville):*

Category	No.	Description
		Reconstruct the US 250 (Richmond Rd.) and Route 20 (Stony Point Rd.) intersection to improve safety and operations. Project includes additional turn lanes, right of way, medians, and new signals.
	9	*Project 2: Pole Green Road Widening (Richmond):* Widen Pole Green Rd. (Rt. 627) from 2 to 4 lanes between Bell Creek Rd. and Rural Point Rd. (1.55 miles).
	10	*Project 3: George Washington Highway Widening (Hampton Roads):* This project will provide improvements to Rt. 17 by expanding the existing 3-lane undivided roadway to a 4-lane divided roadway from Yadkin Rd. to Canal Dr. Project will also include intersection improvements.
	11	*Project 4: I-81 Exit 300 at I-66E Northbound Widening (Staunton/Front Royal):* Add an additional lane and widen left shoulder to standard from Milepost 299.1 to 300.4 Northbound; replace and widen bridge over Water Plant Rd.
	12	*Project 5: Rt. 2 and Rt. 17 from Lansdowne Rd. Past Shannon Airport (Fredericksburg):* This project improves the intersection at Lansdowne Rd., widens Rt. 2 past the intersection of Shannon Drive, adds a southbound through lane on Rt. 2 from Bowman Dr. to Shannon Airport Circle, and adds a northbound right-turn lane on Lansdowne and westbound right-turn lane on Mansfield.

Table 2 Descriptive Statistics of the Data Population and Employment dataset using 35 miles catchment radius

Projects	Data Type	Mean	Std. Error	Median	Std. Dev	Variance	Min	Max	Sum	Count
1	*Total Population*	2124	97	1684	1597	2551786	0	9999	571440	269
	Total Employment	1914	93	1539	1531	2344570	0	8582	514792	269
2	*Total Population*	1980	63	1650	1703	2899534	0	27290	1437432	726
	Total Employment	1910	64	1595	1722	2965537	0	27290	1386908	726
3	*Total Population*	1525	29	1332	957	915281	0	10751	1697418	1113
	Total Employment	1464	28	1298	922	849526	0	9712	1629812	1113
4	*Total Population*	1399	80	1251	1316	1731989	0	7250	383341	274
	Total Employment	1244	79	1045	1311	1719919	0	7214	340908	274
5	*Total Population*	2619	92	2095	2168	4701793	0	20299	1466726	560
	Total Employment	2559	92	2047	2177	4739563	0	20297	1433130	560

5. Methodology

The methodology consisted of three steps:

- Develop a workflow for computing accessibility.

- Compute accessibility scores at varying catchment radii
- Conduct statistical analysis to determine the catchment radius that gives best fit between observed and forecast behavior

5.1 Workflow for Computing Accessibility

The workflow (Figure 1) consists of developing two datasets, one where a candidate transportation project is not built and one where the candidate transportation project is built. Data include turn restrictions, travel times, permitted travel direction, factors to decay jobs, and block group attributes for population and employment. The workflow uses ESRI's ArcGIS Network Analyst (ArcMap version 10.3.1), where service areas are generated for each 1-minute travel time interval for each candidate project for the no-build and build scenarios. Equations 1 through 4 compute accessibility scores using the intersection of population-based services areas and employment centroids.

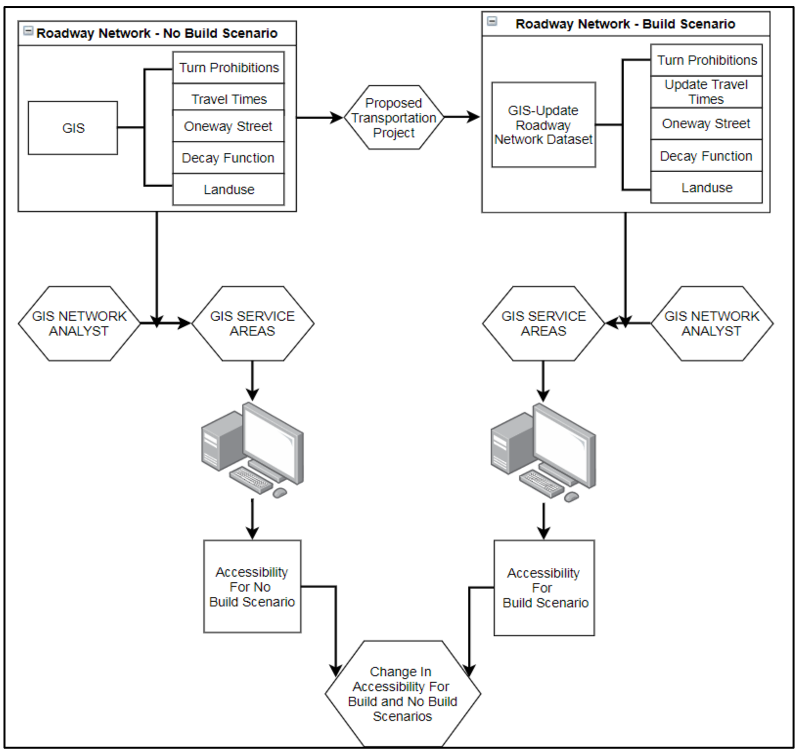


Figure 1. Summary of the accessibility computation workflow.

5.11 Obtain network datasets, activity datasets, and projects for evaluation

The Virginia Department of Transportation (VDOT) provided the 12 data elements detailed in Table 1 (Z. Ling, personal communication). Some data processing obstacles were specific to the Virginia dataset, four computational challenges appeared likely to extend to other locations that might also implement some form of a network-based accessibility measure in a GIS environment. These challenges were resolved through preprocessing the dataset and adapting scripts to solve unexpected problems:

- Convert duplicative one-way streets into unique two-way links.
- Automate the creation of realistic centroid connectors.
- Automate the importation of turn restrictions.
- Manage incorrectly formed service areas.

Once accessibility could be implemented, each of the five candidate projects, 10 different networks datasets (5 before and 5 after) were developed. The networks were identical except that the number of centroid connectors ranged from one to five, and the impact that the number of centroid connectors would have on the change in accessibility score was determined for each project

5.12 Assess the sensitivity of accessibility scores to changing catchment radius.

This study also varied the catchment radius, using values of 5, 10, 15, 20, 25, 30 and 35 miles as shown in Figure 2A for five candidate projects and their corresponding maximum travel times 10.5, 20.5, 30.5, 40.5, 50.5, 60.5 and 70.5 minutes can be seen in Figure 2B. (The reason for these half-minute breakpoints is that the "decay" function, which values jobs that are located further away less than jobs that located nearby, is, given for integer minutes, whereas travel times are continuous. Thus, accessibility is computed for bins of 0-0.5 minutes, 0.5-2.5 minutes, 2.5-4.5 minutes, and so forth with the largest bin being 68.5-70.5 minutes.)

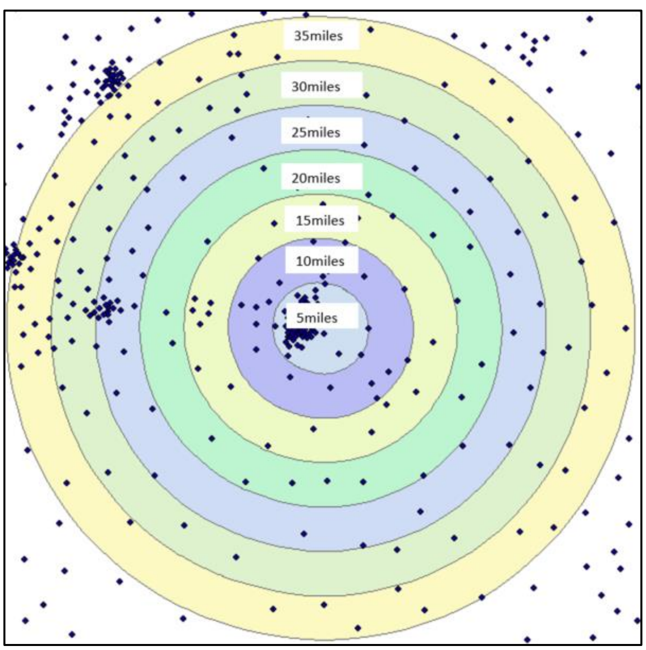


Figure 2A Catchment radius vary between 5, 10, 15 miles for west end of the project, the middle of the project, the east end of the project, and "grouped" the entire segment)

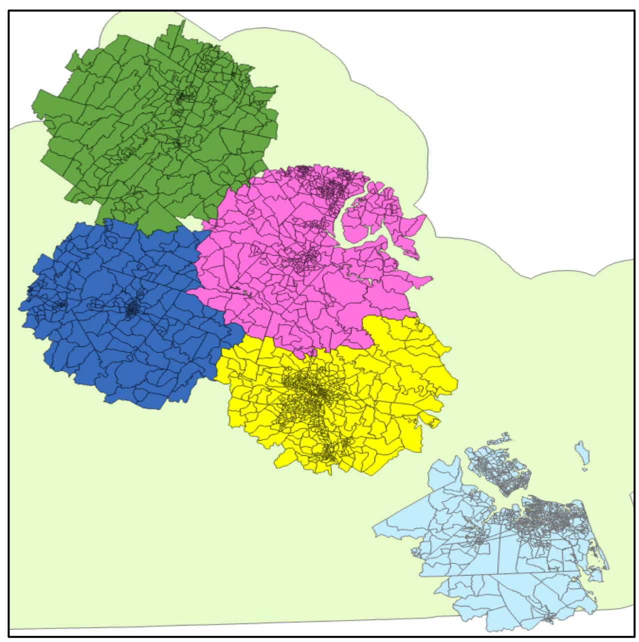


Figure 2B. Examples of Projects with Catchment Radii of 35 Miles Plotted against Area where Observed Data (green) are Available. (Project 1 = dark blue, Project 2 =

yellow, Project 3=light blue, Project 4 = dark green, Project 5 = pink.)

5.2 Compute accessibility scores at varying catchment radii

After establishing a statewide accessibility network, the first part of this study computes project accessibility score as a change-in-accessibility score for each project during the no-built scenario (before) and built scenario (after). Equation 1 is used to compute accessibility where for each census block group i, accessibility is the sum, for all employment zones j, of the employment in zone j multiplied by a step decay function which in turn is based on the travel time from zone i to zone j before the project is built. As travel times increase from 0.5 to 90.5 minutes, the decay function decreases from 1 to 0.01, such that for a travel time of 6 minutes, the decay function has a value of roughly 0.91, which decreases to a value of 0.86 for a travel time of 7 minutes. Thus for block group i, if 100 jobs are located 6 minutes away and 1,000 jobs are located 7 minutes away, then the accessibility is presently (0.91)(100)+(0.86)(1,000) = 951.

$$\mathrm{A_i} = \sum_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{Decay_{ij}Employment_j} \qquad \text{(Eq. 1)}$$

The population term ($\mathrm{Pop_i}$) in Equation 2 weights the accessibility for each block group i by the number of residents. Thus, for block groups with a large population, the accessibility in those block groups is more important than block groups with a smaller population.

$$\mathrm{A_{before}} = \sum_{\mathrm{i}}^{\mathrm{j}} \left(\sum_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{Decay_{ij}Employment_j} \right) \mathrm{Pop_i} \qquad \text{(Eq. 2)}$$

Population-weighted accessibility shown in Equation 2 may be rewritten as Equation 3 and 4, where R, varies between 5, 10, 15, 20, 25, 30 and 35 miles (seven possible values) and C (project definition) is the middle of the project. Equation 4 is this change in accessibility divided by total population within R miles of C. The reason for the "max" term in Equation 3 is that in some cases, slight aberrations in the GIS processing can cause a negative change in accessibility; the maximization term addresses this concern (Boateng and Miller, 2021).

$$\Delta \mathrm{A} = \sum_{\mathrm{i=1}}^{\mathrm{n}} \max(\mathrm{A_i^{After}Pop_i} - \mathrm{A_i^{Before}Pop_i}, 0) \qquad \text{(Eq. 3)}$$

$$\Delta \mathrm{A} = \frac{\sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{A_i^{After}Pop_i}) - \sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{A_i^{Before}Pop_i})}{\text{Population within R miles of C}} \qquad \text{(Eq. 4)}$$

As an example of Equation 4, consider the case of C being the center of a particular link where a project will reduce the time to traverse that link from 3 minutes to 2.5. Let R be 5 miles, and consider only the before accessibility (hence Equation 2 or the right side of Equation 4). While this equation is implemented in practice as a Python script, it may be visualized as a matrix operation where each "row" in Table 3 represents the numerator of Equation 4 for the n= 67 centroids within five miles of the proposed project. For instance, for centroid 1 (column A), the decayed employment for the before condition, that is A_1^{before} , is 3,312 (column B); note this quantity is itself is a summation based on Equation 1. For that first zone, the population-weighted accessibility (based on 1,054 residents shown in column C) is 3,491,495 (A_1^{before} Pop_1). A similar set of calculations are performed for each of the 67 zones within 5 miles of the project.

Table 3 Example of Computing Accessibility for the Before Condition, Project 1, 5 Mile Radius.

A	B	C	D
Population is	104,857		
Facility ID	Decayed Employment	Population	Population Decayed Accessibility
1	3,312	1,054	3,491,495
2	963	2,212	2,130,458
3	339	653	221,160
4	10,453	3,312	34,618,352
5	27,715	1,348	37,360,696
6	16,091	1,149	18,487,212
7	22,130	2,471	54,690,600
8	31,809	5,757	183,111,087
9	24,785	1,007	24,960,989
64	48,717	1,126	54,870,278
65	44,468	4,218	187,585,923
66	56,521	1,227	69,355,225
67	48,273	1,775	85,668,538

After the project is built (Table 4) the accessibilities are computed anew, where the population remains unchanged but generally the decayed employment either remains unchanged or increases because the travel times decrease. In this particular case, the link did not benefit the accessibility for zone 1, but it roughly tripled the accessibility for zone 2 as shown in Table 4.

Table 4 Example of Computing Accessibility for the After Condition, Project 1, 5 Mile Radius

A	B	C	D
Population is	104,857		
Facility ID	Decayed Employment	Population	Population Decayed Accessibility
1	3,312	1,054	3,491,495
2	963	2,212	2,130,458
3	1,114	653	727,913
4	11,062	3,312	36,636,918

A	B	C	D
5	28,444	1,348	38,343,872
6	16,091	1,149	18,487,212
7	22,130	2,471	54,690,600
8	31,809	5,757	183,111,087
9	24,785	1,007	24,960,989
64	48,896	1,126	55,072,487
65	44,474	4,218	187,609,737
66	56,624	1,227	69,480,972
67	48,273	1,775	85,668,538

Equation 4 then takes the difference in these two sets of accessibilities and divides by the population. Thus for Project 3 at a 5 mile radius, Equation 4 is tabulated as 3632806251 (sum of all values in column D of Table 3 including rows not shown) minus 1030019173 (sum of all values in column D of Table 4) divided by 104,856), which is 204. Note also that it is possible to determine the relative contribution of each zone to the accessibility score; for the 13 rows shown in Tables 3 and 4, for example, zones 3 and 4 are contributing substantially do the improved accessibility score, while zone 1 offers virtually no contribution.

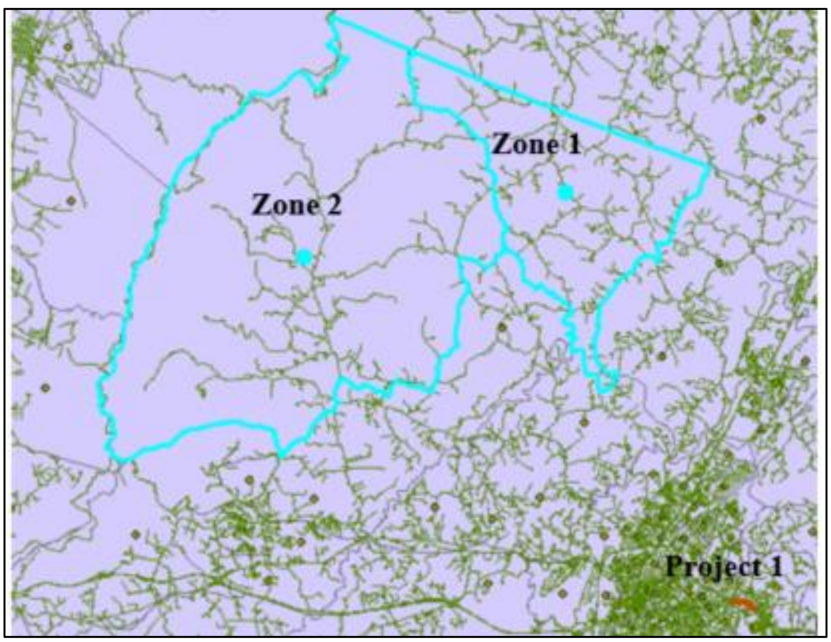


Figure 3. Showing Zones 1 and 2 as well as Project 1 location

5.3 Conduct statistical analysis to determine the catchment radius that gives best fit between observed and forecast behavior

The second part of this study quantifies, as a function of geographical disaggregation, the extent to which differences in accessibility (measured as the distribution of decayed jobs for a given location) explain differences in observed behavior (measured as trips made for the purposes of traveling to or from work). The goal here is to specify the strength of the association between accessibility and destination choices at the zone level, where the "zone" may be the disaggregate census block group or a more aggregate geographical unit such as a census tract. The researchers propose to focus on projects 1, 2, 3, 4 and 5 as a case study. For any particular zone i there are many possible destination zones j. For example; suppose block groups are the unit of analysis. For a catchment radius of 35 miles, there were 269 zones within the vicinity of Project 1, thus j can range from 1 to 269. For Project 2 there will be 726 zones, Project 3 will be 1,113 zones, Project 4 will be 274 zones and Project 5 will be 560 zones. For a catchment radius of 10 miles, however, the number of zones for each project will be smaller than those above. Thus the radius that gives the best strength of association, where the radius can vary from 5 to 35 miles, will be determined. While some aggregation of geographic levels may be desirable, coefficient of correlation by itself may not be useful for evaluating the relationship between accessibility and trip behavior because as the number of zones becomes very small, it will approach 1. Accordingly one can also check the p-value for the coefficient of correlation (Hamburg, 1977), by comparing the t-statistic with the critical t-value based on N-2 degrees of freedom, where N is the number of i-j interchanges and is typically close to n^2.

A sensitivity analysis will be conducted by varying the catchment radius and based on the results at the block group and tract level, the catchment radius that gives the best correlation will be selected. Hence this paper defines validation as replication of traveler behavior for the before period. Although further work is needed to measure behavioral shifts from project-driven accessibility, this paper thus lays the groundwork for a future "after the fact" validation. In the travel demand forecasts—which has been around for decades since the 1950s, but only recently, both in Virginia and in the U.S., have gone back and many years later evaluated longer term forecasts. Thus while longer term accessibility impacts need evaluation, this paper at least sets the preconditions for determining the strength of the association between auto accessibility and traveler behavior.

The computation will be done as follows; for any particular zone i there are many possible destination zones j. For a catchment radius of 10 miles, there were 88 zones within the vicinity of the Project, thus j can range from 1 to 88. The formal definition of accessibility for the entire region at any point in time is given in Equation 5.

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{n} \mathrm{Decay}_{ij}\mathrm{Employment}_{j}\right)\mathrm{Pop}_{i} \qquad \text{(Eq. 5)}$$

Thus for any given zone i, the formal definition of accessibility is Equation 6.

$$\sum_{j=1}^{n}\left(\mathrm{Decay}_{ij}\mathrm{Employment}_{j}\right)\mathrm{Pop}_{i} \qquad \text{(Eq. 6)}$$

However, in terms of observing trips, note that Equation 6 will logically be heavily influenced by the population term for zone i. For instance, there could be a zone i with very high land values and a relatively low population in which case Equation 6 takes on a smaller value than is the case with a large population value, and this might not reflect the term inside parentheses.

Instead, a more promising approach appears to use the relative accessibility for each zone j as it applies to each zone i. That is, for zone i, one would expect there to be some relationship between the relative accessibility contributed by zone j (compared to the total accessibility of zone i) and the relative number of trips attracted by zone j (compared to the total number of trips generated by zone i). Formally one would thus write Equation 7 and colloquially one would write Equation 8.

$$\frac{\mathrm{Decay_{ij}Employment_j}}{\sum_{j=1}^{n}(\mathrm{Decay_{ij}Employment_j}} = \frac{\mathrm{T_{ij}}}{\sum_{j=1}^{n}(\mathrm{T_{ij}})} \qquad \text{(Eq. 7)}$$

$$\frac{\text{Relative accessibility contributed by zone j}}{\text{Total accessibility of Zone i}} = \frac{\text{Trips from i to j}}{\text{Total trips from i}} \qquad \text{(Eq. 8)}$$

To acquire data for the right side of Equations 7 and 8, for project 1, the block groups within the catchment radius of the project was determined within ArcGIS, and the resultant shapefile was used as a zone set within StreetLight Insight ™ . (All zones were set as non-pass through and non-directional in order to use them as origins and destinations.) The data period was 01/01/2018-12/31/2018.

Based on Equations 7 and 8, one may compare, for each origin zone i and destination zone j, the left side of Equation 9 (e.g., accessibility contribution of zone j to zone i relative to the total accessibility of zone i) and the right side of Equation 18 (the trips attracted by zone j from zone i relative to all trips originating from zone i). Thus if there are n zones, there will be n^2 accessibility percentages (the left side of Equation 9) and n^2 trip percentages (the right side of Equation 6). The coefficient of determination, which is the proportion of variation in observed trip percentages explained by the accessibility percentages, is potentially a useful way to determine if there is an association. (Alternatively, one may use the coefficient of correlation, which is the square root of the coefficient of determination). The coefficient of correlation r is computed with Equation 10, where $Y_{observed}$ is the observed trip percentage (the right side of Equation 9) and $Y_{computed}$ is a linear expression based on the accessibility percentage (the left side of Equation 9). For example, for one particular project, $Y_{computed}$ was found to be 0.5022 (Accessibility percentage) + 0.0073.

$$\frac{\mathrm{Decay_{ij}Employment_j}}{\sum_{j=1}^{n}(\mathrm{Decay_{ij}Employment_j}} = \frac{\text{Trips from i to j}}{\text{Total trips from i}} \qquad \text{(Eq. 9)}$$

$$\sqrt{1-\frac{\sum_{z=1}^{n}\left(Y_{observed}-Y_{computed}\right)^2}{\sum_{z=1}^{n}\left(Y_{observed}-\overline{Y}\right)^2}} \quad \text{(Eq. 10)}$$

Some examples of multiple geographic levels of specificity for Project 1, include; at the block group and tract level seen Figure 4. While some aggregation of geographic levels may be desirable, Equation 10 is not useful by itself for evaluating the relationship between accessibility and trip behavior because as the number of zones becomes very small, Equation 10 will approach 1. Accordingly one can also check the p-value for the coefficient of correlation (Hamburg, 1977) by comparing the t-statistic from Equation 11 with the critical t-value based on N-2 degrees of freedom, where N is the number of i-j interchanges and is typically close to n^2.

$$t=\frac{\mathrm{r}}{\sqrt{\frac{(1-\mathrm{r}^2)}{\mathrm{N}-2}}} \quad \text{(Eq. 11)}$$

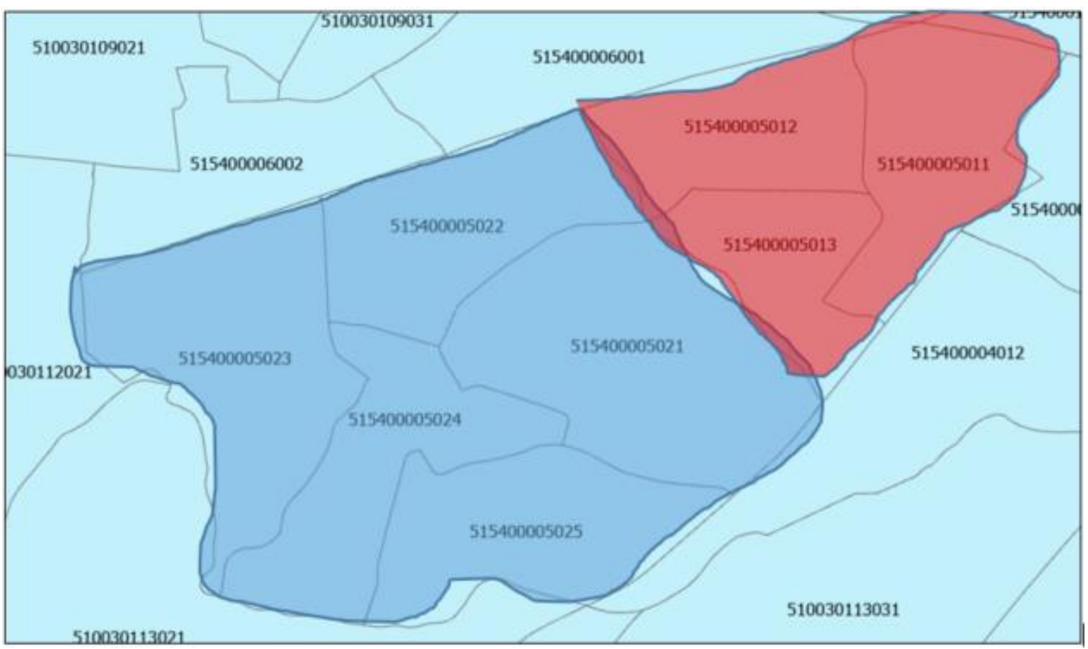


Figure 4 Block Groups, Tracts, and Jurisdictions in Central Virginia. Five block groups are shown in census tract 502 (blue) and the block groups are shown in census tract 501 (red). Those two census tracts are in the City of Charlottesville (where all census tracts begin with the designation 51540). It also shows block groups from Albemarle County (where such census tracts begin with 51003).

5.31 Effects of confounding factors on this association at the geographical level of aggregation

In the third part of the analysis, the study will conduct additional analysis at the census tract level to determine why the association between accessibility and observed behavior is stronger in some locations and weaker in others. Many studies have been done to suggest that other factors besides accessibility affect the distribution of trips, such as income, age, household size, auto ownership levels, transit service, and school events (e.g., Bohnet and Gutsche, 2007: Lavieri et al, 2018: and Lasley, 2017). However, no study has sought to determine the association between accessibility and destination-based traveler behavior after accounting for other confounding factors (geographical level of aggregation, household income, location of the project relative to generation of origin and destination patterns, and housing values). (A few studies have examined how accessibility relates to net amount of travel or auto ownership, such as Kockelman [1997], but using accessibility alone to forecast destination choice does not appear to be well explored.)

This analysis will bridge that gap by determining if there is an association between the percentages of accessibility with confounding factors and observed trip percentages for five transportation projects. Then, because there are other confounding factors of interest that can be investigated when examining this association, the researcher proposes here such that conceptually one can fit a multiple regression model (MRM) of the form shown in Equation 12 and 13. Recall that Equation 9 sought to determine if there is an association between the accessibility percentages (left side) and observed trip percentages (right side) for a single project. There are multiple potential additional factors that can be investigated when examining this association, and the researcher proposes there such that conceptually one is fitting a model of the form shown in Equation 13.

$$Y_{ij} = \alpha + \beta_1 X_{ij} \qquad \text{(Eq. 12)}$$

$$Y_{ij} = \alpha + \beta_1 X_{ij} + \beta_2 A^p + \beta_3 B_{ij} + \beta_4 C_{ij} + \beta_5 D_j + \beta_6 E_i + \beta_6 F_j \qquad \text{(Eq. 13)}$$

Where;

Y_{ij} = $\frac{\text{Trips from i to j}}{\text{Total trips from i}}$

α = intercept

X_{ij} = $\frac{\text{Decay}_{\text{ij}}\text{Employment}_{\text{j}}}{\sum_{j=1}^{n}(\text{Decay}_{\text{ij}}\text{Employment}_{\text{j}}}$

A^p = distance between the project and the MPO center for project p

B_{ij} = disparity between housing costs in origin zone i and destination zone j.

C_i = localized diversity of housing costs in zones surrounding origin zone i

D_j = localized diversity of housing costs in zones surrounding destination zone j

E_i = household income for origin zone i

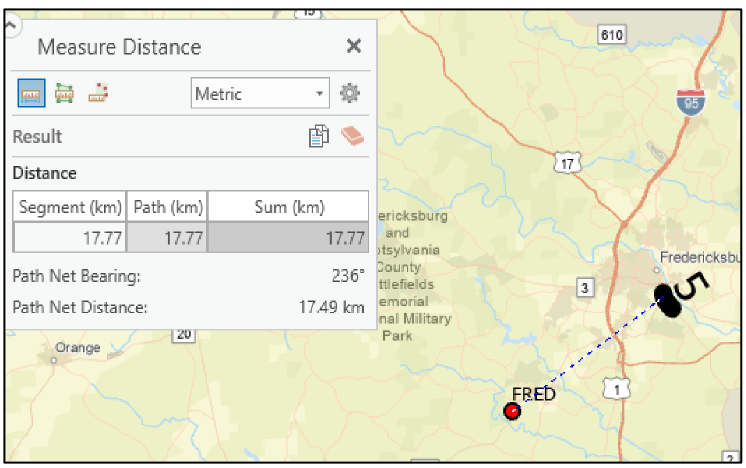
F_j = household income for destination zone j

Note that B_{ij} may be represented with two different variables: monthly rents or median housing values. For this first iteration, both variables have been chosen, but other formulations, such as a combination of the two, may be selected

These factors may be defined as follows:

- Factor A^p. *Distance between the project and the MPO center for project p*, with two projects as a focus, this factor is essentially a binary variable, although there is the flexibility to use it as an interval variable if more than two projects are examined. The MPO center may be defined as the Census block group with the highest sum of population plus employment divided by the square mileage of the Census block group (hence the highest population plus employment density).

  The Virginia Department of Transportation has developed a layer of MPO study areas; this polygon feature class is available through ArcGIS Online. These polygons were downloaded and converted to centroids. The projects were then dissolved such that each project could be represented as a single feature and these were merged into a single feature class. Then, the distance of each project to each MPO Center was tabulated in ArcGIS Pro as seen in Figure 5. For example, for project 5, the closest MPO center is Fredericksburg, about 17,490 meters away when measured by hand but exactly 17,663 meters as shown in Figure 5.



OBJECTID	IN_FID	NEAR_FID	NEAR_DIST	NEAR_RANK
61	5	4	22524.281433	1
62	5	9	87868.132105	2
63	5	11	98335.165845	3
64	5	2	19874.223237	4
65	5	13	144643.36989	5
66	5	15	48226.350436	6
67	5	6	60908.270176	7
68	5	14	73899.618473	8
69	5	5	21266.008654	9
70	5	8	29465.737947	10
71	5	12	313904.86821	11
72	5	3	19239.912169	12
73	5	10	370334.38507	13
74	5	1	61948.165035	14
75	5	7	17663.181731	15

Figure 5. Location of the project relative to generation of origin and destination patterns

- *Factor B_{ij}. Disparity between housing costs in origin zone i and destination zone j*. To some extent, separation of home and work reflects the comparative advantage of the origin zone as a residence. There are several ways in which housing costs can be measured: American Community Survey Table DP04 (80), for example has median rent and the median value of owner occupied units can be used as an indication of housing cost.

- *Factors C_i and D_j. Localized disparity for the area surrounding origin zone i and destination zone j in terms of housing costs*. Figure 6 conceives of two different origin zones: I to the left and I' to the right. Let some destination zone J be a zone with relatively high housing costs. It might be the case that the impact of the disparity between destination zone J and the origin zone depends on whether the origin zone is surrounded by other zones with a similar housing cost (e.g., I in Figure 6, *left*) or is surrounded by zones with substantially different housing costs (e.g., I' in Figure 6, *right*). The third factor is the measure of localized disparity; an initial measure is the *coefficient of variance* (defined as the ratio of the standard deviation to the mean) although other measures of diversity (e.g., the Gini-Simpson diversity index [81] are feasible. Thus the coefficient of variance in Figure 6 if I and I' are origin zones would be C_i = 0.060 (left) and $C_{i'}$ = 0.252 (right). A similar process may be used for destination zones Dj.

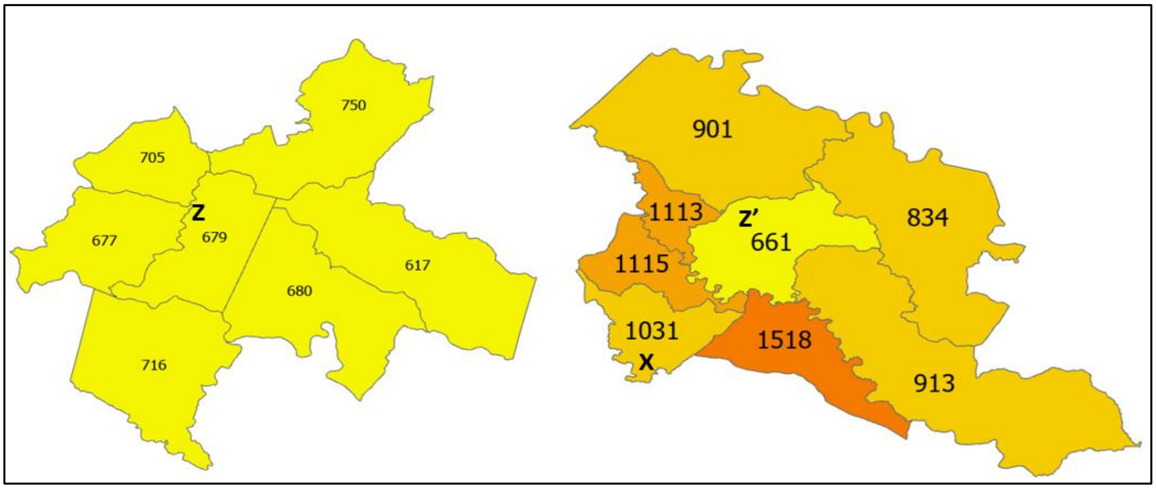


Figure 6. Median Rents by Census Tract for Areas Where Rents are Similar (*Left*) and Dissimilar (*Right*)

Factors E_i and F_j. Household incomes for origin zone i and destination zone j. One rationale for factors Ei and Fj is to assess the role of wealth on proportion of trips between i and i (e.g., if zone j' has higher incomes than other candidate destination zones j, then it might be the case that zone j' attracts a higher number of service-related work trips).

Finally, the researchers propose to perform Equation 13 for projects 1, 2, 3, 4 and 5 with a 5, 10, 15, 20, 25, 30, and 35 mile catchment radius. This analysis enables one to determine the effects of these confounding factors on the association between observed destination choice (Y_{ij}) and destination choice derived from accessibility (X_{ij}).

5.32 How the variable values were computed

For example, consider origin tract 51033030400 and destination tract 51177020309 as shown in Figure 7, the following explanatory variables can be tabulated given that the former is tract i and the latter is tract j (Table 5). These refer one row of data in project 5 with a 35 mile radius. Table 5 shows how these data are used to support Equation 1. Accordingly, one may tabulate six variables based on the values in the upper half of Table 1, letting i = 1 for origin tract 51033030400 in Caroline County and letting j = 2 for destination tract 51177020309 in Spotsylvania County. The difference in rents (B_Rent_{12}) is -$566 and the difference in housing values ($B_Housing_{12}$) is -$136,400, computed as $170,500-

$306,900. If rent is used in Equation 1, then C_Diversity_Rent_1 is 0.223 and D_Diversity_Rent_2 is 0.228; if housing is used instead, then these attributes are 0.116 and 0.154, respectively.

Table 5. Example of Extraction of Data for Equation 13

Data Type		Origin : 51033030400	Destination : 51177020309
Extracted from Census	Zone	i = 1	j = 2
	Tract rent	$851	$1,417
	Mean county rent	$975	$1,244
	Standard deviation of county rent	$217	$284
	Coefficient of variance of county rent	0.223	0.228
	Tract housing value	$170,500	$306,900
	Mean county housing value	$191,929	$299,799
	Standard deviation of county housing value	$22,204	$46,294
	Coefficient of variance of county housing value	0.116	0.154
Applied in Equation 1	Difference in rent (O-D) : B_Rent_{12}	-$566	
	Difference in home values (O-D): $B\_Housing_{12}$	-$136,400	
	Diversity in rent at Origin : C_Rent_1	0.223	
	Diversity in Housing values at origin : $C\_Housing_1$	0.116	
	Diversity in rent at Destination : D_Rent_2	0.228	
	Diversity in housing values at Destination : $D\_Housing_2$	0.154	
	Income at Origin : E_Income1	$54,615	
	Income at Destination : F_Income2	$84,852	
	Distance Between MPO and Project: A_MPO_Distance (meters)	Project1 (54,28), Project2 (17,637), Project3 (21,092), Project 4 (27,897) and Project 5 (22,524)	

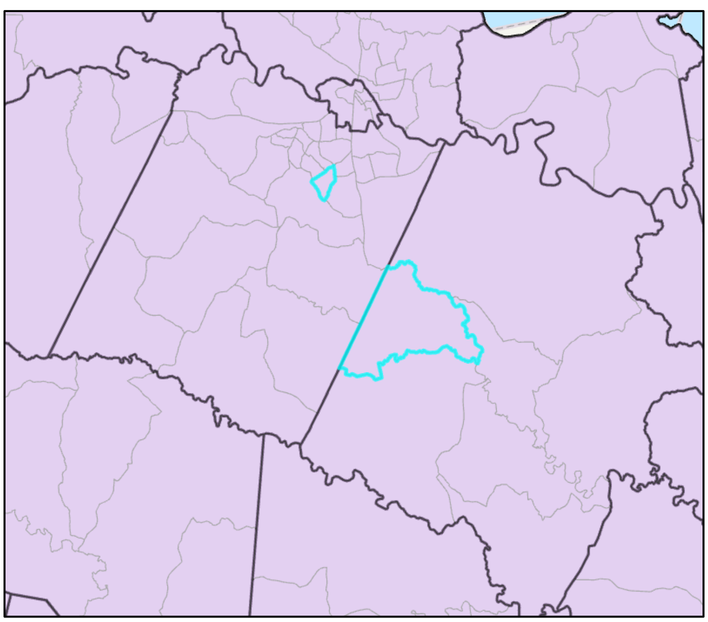


Figure 7. Origin Tract 51033030400 (lower right) and Destination Tract 51177020309 (upper left)

In order to implement multiple regression, the study ensured that the residuals are normally distributed. In Figure 8A, the plotted histogram, P-P plot and residual-by-predicted charts indicate that the data is not normally distributed hence the need to transform the data. Therefore several attempts to transform the dependent variable were made. The successful transformation adapted for this study was Log10 of the dependent variable as shown in Figure 8B. It can clearly be seen that the residuals are normally distributed (Chambers, 2017). The VIF values range from 1.009 to 1.749 indicating that multicollinearity is not an issue in the dataset (Hair et. al, 2010). The models developed here in this study showed a statistically significant F statistics at alpha = 0.05. The rest of the plots can be found in the appendix.

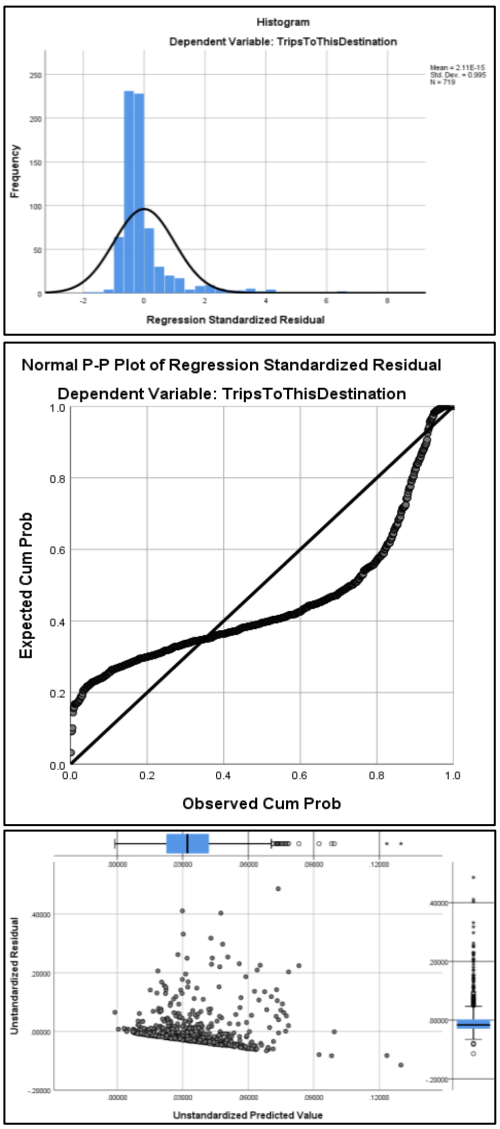


Figure 8A. Residuals are not normally distributed

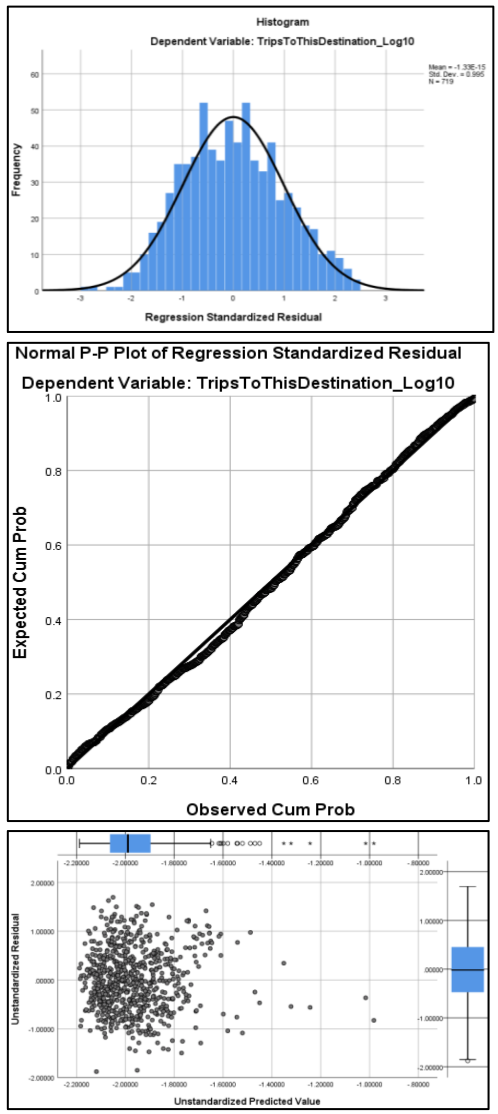


Figure 8B. Residuals are normally distributed

6. Results

Table 6 shows the initial results of Equations 1-4. Consider, for example, the 35 mile radius, shown as Row H. For projects 1, 2, 3, 4, and 5, the accessibility scores are 71, 70, 55, 36, and 40, respectively (based on the total population).

6.1 Impact of altering the radius on accessibility scores

Figure 9 shows that the radius affects the numerical accessibility score. Accessibility drops rapidly as the radius increases from either 5 to 10 miles (for three of the five projects) or from 10 to 15 miles (for the remaining two projects). This drop is better understood when examining the denominator's rate of change in Figure 10: the marginal increase in accessibility (the numerator of Equation 4) is less than the marginal increase in population (the denominator of Equation 4). For that reason, such drops in accessibility become considerably more moderate above about 15 miles. Figure 9 shows that Projects 1, 3, and 5 lose an average of 44% of their maximum accessibility score at 5 miles, but for a radius increase from 15 to 20 miles, those projects lose just, on average, 8% of their maximum accessibility score at 5 miles.

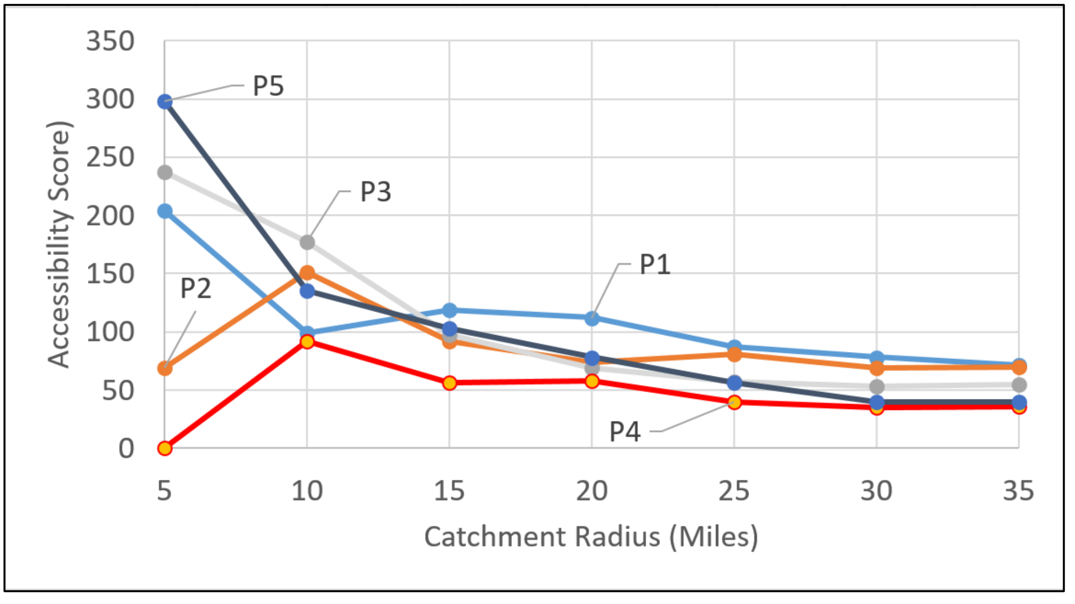


Figure 9. Accessibility Scores by Project and Radius

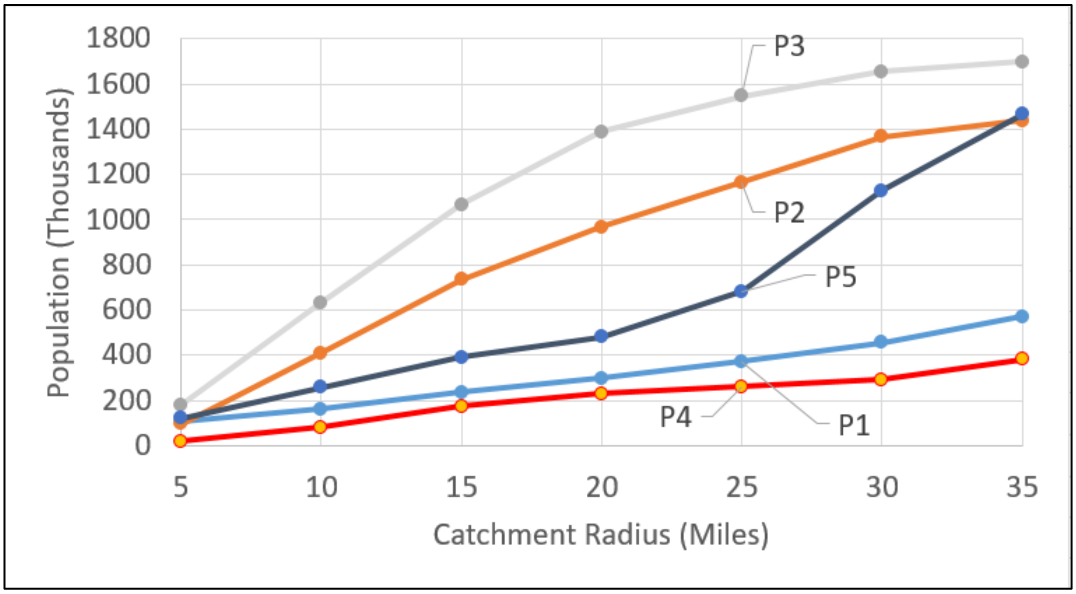


Figure 10. Population by Project and Radius.

Figures 9 and 10 illustrate the type of sensitivity test that can be performed to understand various manifestations of accessibility—not just in the classroom but also in professional practice. For instance, one reason for the denominator term in Equation 4 was to avoid penalizing rural areas (Hardy and Bell, 2019) since a travel time reduction (e.g., the numerator of Equation 4) will tend to favor locations with a greater number of people and jobs. Based on a generalized description of accessibility (e.g., net improvement in population-weighted time-decayed jobs), one might expect that Project 5, located in the exurban Fredericksburg area of the greater Washington, D.C. metropolitan region, would show considerably larger accessibility benefits as the catchment radius increases. While expansion of this radius in fact does increase the number of time-decayed jobs, Figure 10 shows that the population associated with project 5 increases relatively rapidly from 20 to 35 miles, faster than the other projects. This difference is reflected in Figure 9, where not only does the accessibility score for Project 5 fall (e.g., its score at 20 miles is lower than its score at 15 miles), but its standing compared to the other projects drops from 20 to 35 miles (e.g., at 20 miles it is surpassed only by project 1 but thereafter is surpassed by all other projects except Project 4).

Table 6 Accessibility Scores

Project	Catchment Radius (Miles)	Population	Accessibility Score	Project Rank
1	5	104857	204	3
2	5	95718	69	4
3	5	179617	237	2
4	5	19352	0	5
5	5	122819	298	1
1	10	162215	99	4
2	10	410241	151	2
3	10	631703	177	1

Project	Catchment Radius (Miles)	Population	Accessibility Score	Project Rank
4	10	81421	92	5
5	10	256761	135	3
1	15	236460	119	1
2	15	736358	92	4
3	15	1066253	97	3
4	15	174316	56	5
5	15	392093	103	2
1	20	298954	112	1
2	20	968429	74	3
3	20	1388537	69	4
4	20	231452	58	5
5	20	480249	78	2
1	25	371727	87	1
2	25	1165706	81	2
3	25	1545207	57	3
4	25	261152	40	5
5	25	682317	56	4
1	30	455539	78	1
2	30	1366018	69	2
3	30	1655459	53	3
4	30	292710	35	5
5	30	1128209	40	4
1	35	571440	71	1
2	35	1437432	70	2
3	35	1697418	55	3
4	35	383341	36	5
5	35	1466726	40	4

***Not significant at 95% Confidence Level**

6.2 Measuring the Correlation between Accessibility and Observed Behavior

In the second analysis, a comparison of street light data set to accessibility for each of the five projects at the block group and tract level within catchment radius of 5-35 miles was made and the results suggests that as the geographical disaggregation increases, the percent of variation in trip making behavior explained by accessibility on average will generally drop particularly for Projects 1, 3, 4 and 5 (shown in Table 7). For example at 5 miles catchment radius, percentage of variations on average for the five projects drops from 22% at tract level to 11% at the block group level. On average, greater geographical aggregation (from the block group level to the census tract level) increased the variation explained by accessibility by about one percentage point. (For instance, for project 3 at a 35 mile radius, accessibility explained 6% of variation at the block group level but 7% at the tract level.

While differences in geographical aggregation affected the variation explained by 0-2% at catchment radii 10, 15 and 30 miles, the change from the tract to block group level affected variance explained by 1-11% for catchment radii of 5, 20, 25, and 35 miles. These percentages are the square of Equation 9; the coefficient of correlation ranges from 8% to 93%. With the exception of projects 2 and 5 at 5 mile catchment radius and project 4 at 10, 15 and 20 mile catchment radius, the low p-values (<<0.01) for the rest of the projects at 5-35 mile catchment radius suggest that the observed association is not attributable to chance. While catchment radii 20, 25, 30 and 35 provide the least percentage in variance for the coefficient of correlation (0.1-0.5) and coefficient of determination (0.0-0.3), they were also found to be significant at 95% confidence level for the five projects considered for the study except for project 4 at 20 mile radius. Among these 4 selected catchment radius, 25 mile catchment radius provided better percentage of variation in trip making behavior explained by accessibility on average 12% compared with 8% for 30 and 9% for 35 miles radius at the tract level.

Table 7 Percentage of Variation Explained by Accessibility

Projects	Analysis	Block Level							Tract Level						
		Catchment Radius (Miles)							Catchment Radius (Miles)						
		5	10	15	20	25	30	35	5	10	15	20	25	30	35
1	Coefficient of Correlation	15%	16%	15%	15%	24%	25%	26%	24%	25%	26%	28%	41%	28%	29%
	(p-value)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Coefficient of Determination	2%	2%	2%	2%	6%	6%	7%	6%	6%	7%	8%	17%	8%	9%
	Sample Size (N)	2366	3462	4463	5207	6270	7549	9751	364	508	701	735	897	1012	1327
2	Coefficient of Correlation	22%	39%	36%	33%	30%	29%	28%	16%	35%	35%	33%	26%	27%	28%
	(p-value)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.13)*	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Coefficient of Determination	5%	15%	13%	11%	9%	9%	8%	3%	12%	12%	11%	7%	8%	8%
	Sample Size (N)	850	3462	28624	40334	48239	54627	57535	87	1271	2074	2692	3237	3657	3792
3	Coefficient of Correlation	23%	27%	23%	22%	25%	25%	25%	36%	21%	41%	29%	31%	31%	26%
	(p-value)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Coefficient of Determination	5%	7%	5%	5%	6%	6%	6%	13%	4%	17%	8%	10%	10%	7%
	Sample Size (N)	4810	26026	53296	73749	84061	90611	92323	263	1206	2186	2866	3242	3350	3465
4	Coefficient of Correlation	60%	34%	36%	36%	35%	31%	35%	93%	13%	9%	21%	27%	24%	37%
	(p-value)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.55)*	(0.58)*	(0.11)*	(0.01)	(0.01)	(0.00)
	Coefficient of Determination	36%	11%	13%	13%	12%	10%	12%	86%	2%	1%	4%	7%	6%	14%
	Sample Size (N)	76	866	3175	4685	5482	6401	8592	7	25	43	59	84	115	203
5	Coefficient of Correlation	29%	30%	26%	24%	24%	23%	25%	8%	29%	19%	34%	41%	28%	29%
	(p-value)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.44)*	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Coefficient of Determination	8%	9%	7%	6%	6%	5%	6%	1%	9%	4%	12%	17%	8%	8%
	Sample Size (N)	1708	4191	7180	9120	11249	18134	26265	95	296	379	442	600	859	1420
	Mean Coefficient of Correlation	30%	29%	27%	26%	28%	27%	28%	35%	25%	26%	29%	33%	28%	30%
	Mean Coefficient of Determination	11%	9%	8%	7%	8%	7%	8%	22%	7%	8%	9%	12%	8%	9%

It is not surprising that the relationship between accessibility and behavior varies by project: even if two projects had the same impact on travel time, their different locations will mean a different impact on local origin-destination patterns. In practice, one needs to choose a radius for some large number of projects—just five of which are reflected in this study—such that one asks how does the catchment radius influences the relationship between accessibility and behavior without controlling for specific projects? Again, as was the case with Table 7, observed behavior was the dependent variable and accessibility score was the independent variable. The dependent variable was transformed by using Log10 function in SPSS to ensure that the residuals are normally distributed. Table 8 suggests that, as other studies have shown, accessibility statistically influences behavior—although this impact is fairly small, between 4% (at a 5 mile radius) and 10% (at a 35 mile radius). One then wonders if accessibility in combination with additional explanatory variables may increase this percent of variation explained.

Table 8 Trip behavior vs Accessibility scores

Catchment Radius (miles)	Coef. Employment (Accessibility)	Std. Error	t	P>|t|	Coefficient of Variation
5	3.25	0.61	5.31	0.00	3.65%
10	6.55	0.46	14.37	0.00	7.01%
15	8.18	0.45	18.13	0.00	6.97%
20	9.11	0.45	20.42	0.00	7.14%
25	10.77	0.44	24.53	0.00	8.49%
30	9.41	0.39	24.28	0.00	7.65%
35	10.68	0.36	29.49	0.00	9.55%

Dependent Variable: TripsToThisDestination_Log10

6.3 Effects of Confounding Factors on This Association: The Geographical Level Of Aggregation, Household Income, Location Of The Project Relative To Generation Of Origin And Destination Patterns, And Housing Costs.

In the third part of the analysis, the study conducted multiple regression analysis at the census tract level to determine why the association between accessibility and observed behavior varied by different catchment radii (5-35 miles). Accordingly, the study further combined accessibility scores with other explanatory variables to determine the association between accessibility and destination-based traveler behavior after accounting for other confounding factors at census tract level. These confounding factors include; Difference in Rent at origin and destination (O-D), Difference in home values (O-D), Diversity in Rent (O-D), Diversity in Housing values (O-D), Income (O-D), Distance between MPO and Project.

Table 9 indicates that housing price at the origin and rent at the destination has no significant effect on travel behavior. These variables were not found to be significant at all for all the radii investigated. However, accessibility, differences in rent and income and origin and destination variables were generally found to influence travel behavior. At 5 mile catchment radius, an increase in accessibility and income significantly influence travel behavior explaining about 4% of the variance. At 10 and 15 mile catchment radius, variables such as accessibility, differences in rent and income were found to influence travel behavior and it explains about 7.3-7.5% of the variance. At 20 and 35 miles, the study results indicate that variables such as accessibility, differences in rent, housing price and income influence destination choice explaining 7.8-10.2% of the variance. Finally at 25 and 30 mile radius, accessibility, differences in rent, housing prices, income, MPO distances and rent at origin (for 25miles only) accounts for about 9.4-9.7% of the variance that can be explained. It can further be seen that, the highest percentage (10.2) of variance that can be explained by accessibility and other confounding factors occurred at 35 miles catchment radius (See figure 11).

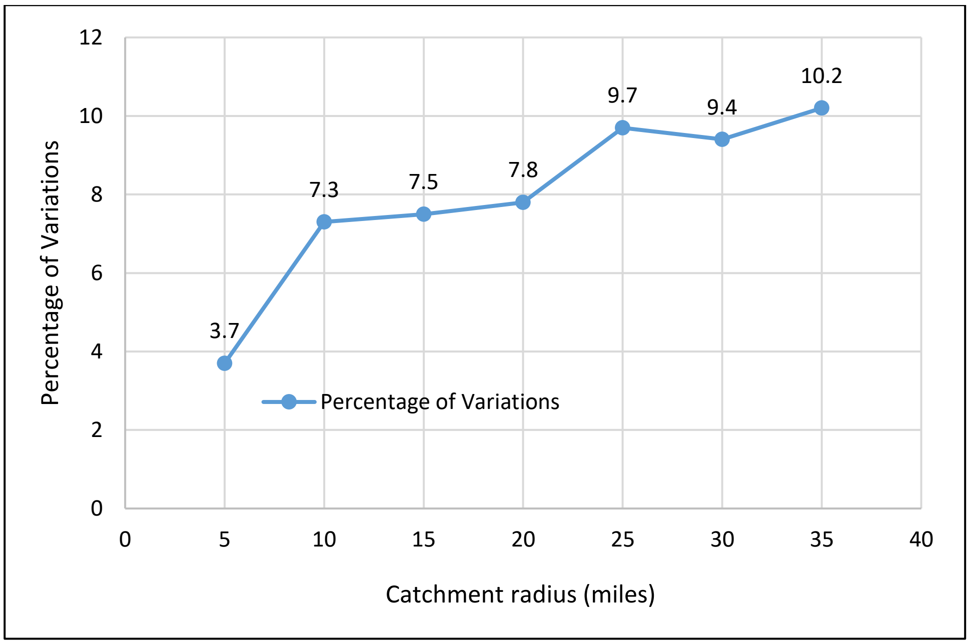


Figure 11. Percentage of variance that explain destination-based traveler behavior

Table 9. Explanatory Variables for Traveler Destination Choice

Radius (miles)	X_ Access	B_ Rent Orig-Dest	B_ Housing Orig-Dest	C_ Rent Orig var	C_ Housing Orig var	D_ Rent Dest var	D_ Housing Dest var	E_ Income Orig	F_ Income Dest	A MPO Distance	Std. Error	P>|z|	Variations %
5	3.05										0.61	0.00	3.7%
								0.00			0.00	0.01	
10	6.55										0.46	0.00	7.3%
		-0.00									0.00	0.00	
								0.00			0.00	0.04	
									-0.00		0.00	0.01	
15	8.20										0.45	0.00	7.5%
		-0.00									0.00	0.00	
								0.00			0.00	0.00	
									-0.00		0.00	0.00	
20	8.88										0.45	0.00	7.8%
		-0.00									0.00	0.00	
					-0.22						0.05	0.00	
									-0.00		0.00	0.00	
25	9.92										0.47	0.00	9.7%
		-0.00									0.00	0.00	
				0.26							0.13	0.05	
							-0.40				0.05	0.00	
									-0.00		0.00	0.00	
										-0.00	0.00	0.00	
30	8.30										0.41	0.00	9.4%
		-0.00									0.00	0.00	
							-0.41				0.05	0.00	
									-0.00		0.00	0.00	
										-0.00	0.00	0.00	
35	10.28										0.37	0.00	10.2%
		-0.00									0.00	0.00	
					-0.28						0.05	0.00	
									-0.00		0.00	0.00	

Orig=Origin; Dest=Destination; Var = Coefficient of variation

7. Discussion on choosing the catchment radius that give the best fit between observed and forecast behavior.

Figures 8-10 show that altering the catchment radius has a substantial impact on accessibility scores and, in some cases, the rankings themselves. Although catchment radii of 25, 30 and 35 miles yielded consistent rankings, this was not the case for other radii. One criterion for choosing a radius is to choose the radius that yields the best agreement between accessibility and what this study calls observed behavior—that is, the destination choices for individuals living in a particular zone j, given that they reside in zone i. Using a probe-based data set serve as the observed choices (from StreetLight InSight), the study examines the extent to which variation in choices is explained by accessibility at different radii of 5-35 miles.

Generally, accessibility alone explains a statistically significant but modest portion of traveler choice (between 3.7% and 9.6% depending on the radius). With additional explanatory variables, as much as 10.2% of the variance can be explained. Despite these nominally small percentages, the distinction between Tables 8 and 9 underscores the relative relevance of accessibility in that the addition of several socioeconomic variables—income, costs of housing (whether as rent or housing value), and variation in these variables relative to adjacent zones—only adds about a few percentage points of explanatory power except for project at 5 mile catchment radius. Without accessibility as a variable in the MRM, the percentage of variations of other influential variables produced 1.3%, 0.4%, 0.5%, 1.2%, 3.7%, 4.2% and 1.7% at catchment radii 5, 10, 15, 20, 25, 30 and 35 miles respectively. In short, while a number of other factors explain destination choice besides those in this study, accessibility alone seems to matter more than the other factors presented here.

The study finally conducted statistical regression modeling to determine if other variables besides accessibility can help forecast travel behavior. Findings from the study confirmed a statistical significant relationship between accessibility and travel behavior with the highest percentage of variations occurring at 35 miles catchment radius (10.2%), followed by 25 miles representing 9.7% and 30 miles representing 9.4%. At 5 mile catchment radius, the study recorded its lowest percentage of variations at 3.7%. It can further be seen that at catchment radius 10, 15 and 20, miles, the variations explained by accessibility alone range from 7.3-7.8% indicating a range in catchment radii to choose from. Intersecting this to the ranking results discussed earlier means that catchment radii 25, 30 and 35 miles can be selected as interim catchment radii to choose the best one from.

Based on the 3 catchment radii, the researchers recommend 25 mile catchment radius because it requires comparatively less amount of computational time and resources than 30 or 35 miles.

8. Conclusions

Accessibility to job has frequently been proposed as an element in project prioritization, recent work has shown that this formulation is highly sensitive to one particular computational parameter: the sphere of influence considered for the candidate project, formally defined as the catchment radius. The manner in which this radius is selected has not been resolved, although multiple approaches are feasible. The Delphi method, for instance, entails convening a panel of experts to select a radius such that the accessibility scores from a series of projects agrees with the a-priori ranking of such experts. A public participation approach is to select a radius that addresses stakeholder concerns, such as choosing a large radius in order to account for longer commutes that would benefit from the project.

- The first part of the paper uses accessibility models to compute accessibility and rank the projects for each of the catchment radii. Findings from this part of the study indicate that catchment radius 5-35 miles around each project indeed affected accessibility however at 25, 30 and 35 miles catchment radii, altering the radius did not affect accessibility ranking of the projects. The second part of this paper proposes, and then evaluates, another approach: choosing the catchment radius that gives best fit between observed and forecast behavior.
- While findings from the second part indicate that there is significant relationship between observed and forecast behavior, accessibility alone explains between 4% and 10% of the variation in destination choice. The study further showed that at catchment radii 25, 30 and 35 miles, the percentage of variation explained by accessibility alone was more than the average percentage of 7.2.
- The third part of the paper also considers the effects of other confounding factors on this association at the geographical level of aggregation (tract level); household income, location of the project relative to generation of origin and destination patterns, rent and housing values. An ancillary benefit of this work is that the study quantifies the amount of variance in traveler behavior explained by accessibility alone. The study found accessibility scores to be statistically significant at 95% confidence level with the highest percentage (10.2) of variations in traveler behavior that can be explained by accessibility alone as well as accessibility with other confounding factors occurring at 35 miles catchment radius. While the lowest percent occurred at 5 mile catchment radius (3.7%), the highest stable percentage of variance that can be explained by accessibility and other confounding factors occurred at 25, 30 and 35 mile catchment radii.
- Intersecting these three analysis, the study finds three catchment radii that has the potential to provide the best fit between observed and forecast behavior to be at 25, 30 and 35 miles. Among these three, catchment radius at 25 mile will be recommended to be used because it is relatively smaller than the other two, it will require lesser processing and computational time and resources and will be the most cost effective radius to implement.

8.1 Future Work and Limitations

There are two categories of future work—one is shorter term and one is longer term. In the shorter term, four different formulations as an alternative to Equation 13 can be examined.

- The use of rent and housing values associated with variables B, C, and D are somewhat, but not entirely, duplicative. There is/is not a strong correlation between housing prices and rents generally. An argument in favor of retaining both variables is that they measure different markets, but it may be feasible to eliminate housing values or develop an index that extracts information from both variables.

- It may be appropriate to compute the disparity between housing costs in origin i versus destination j as an absolute value, since greater disparity is thought to increase travel.
- Because of how Equation 13 is developed, it should be the case that only destination income has an impact, and this was generally the case as shown in Table 9 except at the radius of 5 miles. For exposition purposes, removal of what has been called variable E may be productive.
- Logically, for persons living in origin zone i, higher housing costs in destination zone j', coupled with lower incomes in zone i compared to those in j', would be associated with a greater proportion of trips between i and j' than between i and other destination zones j. Thus a new compute variable G_{ij} computed as follows may be tested:

  G_{ij} = (Housing costs in j – Housing costs in i)(Incomes in j-Incomes in i)

In the longer term, while this study shows the importance of selecting the appropriate catchment radius that has the potential to provide the best fit between observed and forecast behavior, additional work can be performed to determine the feasibility of a common radius for multiple modes rather than auto mode as this paper has used. Stakeholder reaction to such practices, as well as the computational details presented in this paper, are essential to address the public "vetting" advocated by Sundquist (2017, 2018) to ensure that accessibility is a meaningful metric when evaluating candidate transportation projects for construction.

Acknowledgments

The authors are grateful for the assistance of Pankaj Singla with processing the data needed to compute X_{ij} and Y_{ij} in Equations 9 and 12. The deployment of the system shown in Figure 1 was facilitated through several individuals: a panel of experts who reviewed this work (Peng Xiao, Ziwen Ling, Amy O'Leary, and Ram Venkatanarayana); IT assistance (Linda Cullop, Eric Hetzer, Michele Mandell, Adam Munro, Alex Ptak, and an anonymous member of the ESRI support team); and editing (Linda Evans). The authors are also grateful for the support for this research by the Virginia Department of Transportation and FHWA. The material in this paper is the responsibility of the authors and does not necessarily represent the viewpoint of these agencies.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: R. Boateng and J. Miller; data collection: R. Boateng; analysis and interpretation of results: R. Boateng and J. Miller; draft manuscript preparation: R. Boateng and J. Miller. Both authors reviewed the results and approved the final version of the manuscript.

References

Behara, K.N.S., Bhaskar, A., and Chung, E. Classification of Typical Bluetooth OD Matrices Based on Structural Similarity of Travel Patterns- Case Study on Brisbane City. Transportation Research Board 97th Annual Meeting, Washington D.C., 2018. https://trid.trb.org/Results?txtKeywords=school+systems+and+trip+distribution#/View/1495255. Accessed April 3, 2020. abstract only.

Bennie, Jonathan, Karen Anderson, and Andrew Wetherelt. "Measuring biodiversity across spatial scales in a raised bog using a novel paired-sample diversity index." Journal of ecology 99.2 (2011): 482-490. https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2745.2010.01762.x. Accessed June 1, 2020.

Bohnet, M. and Gutsche, J.M. Estimating Land Use Impacts on Transportation - Findings from the Hanover Region, Proceedings of the European Transport Conference 2007 Held 17-19 October 2007, Leiden, The Netherlands, 2007. https://trid.trb.org/View/859861. Accessed March 25, 2020, abstract only.

Cambridge Systematics, Inc., KATS Travel Model Update Technical Documentation, Technical Report, (2015), https://katsmpo.files.wordpress.com/2015/02/kats-travel-model-update-technical-documentation-06292015.pdf

Chambers, J.M. (2017). Graphical Methods for Data Analysis (1st ed.). Chapman and Hall/CRC. https://doi-org.proxy01.its.virginia.edu/10.1201/9781351072304

Conway, M.W., Byrd, A., and van der Linden, M. Evidence-Based Transit and Land Use Sketch Planning, Transportation Research Record: Journal of the Transportation Research Board, No. 2653, 2017, pp. 45–53. http://dx.doi.org/10.3141/2653-06.

Ford, A.C., Barr, S.L., Dawson, R.J., and James, P. Transport Accessibility Analysis Using GIS: Assessing Sustainable Transport in London. International Journal of Geo-Information, 2015, Vol. 4, No. 1, pp. 124-149., https://www.mdpi.com/2220-9964/4/1/124., doi:10.3390/ijgi4010124. Accessed January 23, 2020.

Geurs, Karst T., and Bert Van Wee. "Accessibility evaluation of land-use and transport strategies: review and research directions." Journal of Transport geography 12.2 (2004): 127-140.

Guan, Xiaodong, and Donggen Wang. "Influences of the built environment on travel: A household-based perspective." Transportation Research Part A: Policy and Practice 130 (2019):

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2013). Multivariate data analysis: Advanced diagnostics for multiple regression [Online supplement]. Retrieved from http://www.mvstats.com/Downloads/Supplements/Advanced_Regression_Diagnostics.pdf

Hamburg, M. Statistical Analysis for Decision Making, Second Edition, Harcourt Brace Jovanovich, Inc., 1977

Hardy, D., Bell, A., and McCahill, C. Accessibility Measurement for Project Prioritization in Virginia, Transportation Research Record, Vol. 2673(12) 266–276, Washington, D.C., 2019. DOI: 10.1177/0361198119859319.

Hartman, Raymond S. "The estimation of short-run household electricity demand using pooled aggregate data." Journal of Business & Economic Statistics 1.2 (1983): 127-135.

Holmes, C. The State of the American Mover: Stats and Facts, Move.org, April 23, 2018. https://www.move.org/moving-stats-facts/. Accessed March 2, 2021.

Kockelman, Kara Maria. "Travel behavior as function of accessibility, land use mixing, and land use balance: evidence from San Francisco Bay Area." Transportation research record 1607, no. 1 (1997): 116-125.

Lasley, P. Influence of Transportation on Residential Choice: A Survey of Texas REALTORS® on Factors Affecting Housing Location Choice, Preliminary Report. Texas Transportation Institute, College Station, 2017. https://static.tti.tamu.edu/tti.tamu.edu/documents/PRC-17-33-F.pdf. Accessed April 3, 2020.

Lavieri, P.S., Dias, F.F., Juri, N.R., Kuhr, J., and Bhat, C.R. A Model of Ridesourcing Demand Generation and Distribution, in Transportation Research Record 2018, Vol. 2672(46) 31–40, National Academy of Sciences, TRB, Washington, D.C., 2018, DOI: 10.1177/0361198118756628.

Lin, Jen-Jia, and Yu-Chun Cheng. "Access to jobs and apartment rents." Journal of Transport Geography 55 (2016): 121-128.

Merlin, L. A., J. Levine, and J. Grengs. Accessibility Analysis for Transportation Projects and Plans. Transport Policy, Vol. 69, October, 2018, pp. 35-48.

Merlin, Louis A. "Can the built environment influence nonwork activity participation? An analysis with national data." Transportation 42.2 (2015): 369-387.

Merlin, Louis A. "Measuring community completeness: Jobs—housing balance, accessibility, and convenient local access to nonwork destinations." Environment and Planning B: Planning and Design 41.4 (2014): 736-756.

National Household Travel Survey (NHTS),https://nhts.ornl.gov/Pokharel, R., and Ieda, H. Road Network Evaluation from a Reliability Perspective: An Accessibility and Network Closure Vulnerability Approach. Asian Transport Studies, Volume 4 Issue 1 Pages 37-56, 2016. DOI https://doi.org/10.11175/eastsats.4.37. https://www.jstage.jst.go.jp/article/eastsats/4/1/4_37/_pdf/-char/en. Accessed July 30, 2020.

Pyrialakou, V. Dimitra, Konstantina Gkritza, and Jon D. Fricker. "Accessibility, mobility, and realized travel behavior: Assessing transport disadvantage from a policy perspective." Journal of transport geography 51 (2016): 252-269. 710-724.

Richter, Francisca G-C., and B. Wade Brorsen. "Aggregate versus disaggregate data in measuring school quality." Journal of Productivity Analysis 25.3 (2006): 279-289.

Serulle, Nayel Urena, and Cinzia Cirillo. "Transportation needs of low income population: a policy analysis for the Washington DC metropolitan region." Public Transport 8.1 (2016): 103-123.

Simma, Anja, and Kay W. Axhausen. "Interactions between Travel Behaviour, Accessibility and Personal Characteristics." European Journal of Transport and Infrastructure Research 3.2 (2003).

StreetLight Insight. Travel Between Origins and Destinations. https://insight.streetlightdata.com/#/login. Accessed March 3, 2020.

The Texas A&M Transportation Institute and Economic Development Research Group, Inc. Accessibility Analysis Addendum: Additional Instructions and Clarifications for Use of The Effective Density Market Access Tool, 2014. http://www.tpics.us/tools/documents/C11_ACCESSIBILITY%20ANALYSIS%20ADDENDUM.pdf. Accessed July 30, 2020

U.S. Census Bureau, American Community Survey 2012-2016 Five-year estimates. Special Tabulation: Census Transportation Planning. Table A302103 - Means of transportation (18) (Workers 16 years and over). Washington, D.C., undated. https://ctpp.transportation.org/2012-2016-5-year-ctpp/ Accessed February 26, 2020.

U.S. Census Bureau. Table DP04: Selected Housing Characteristics, 2014-2018 American Community Survey 5-Year Estimates, Washington, D.C. https://data.census.gov/cedsci/. Accessed June 1, 2020

U.S. Census Bureau. Cartographic Boundary Files – Shapefile, Washington, D.C., 2019a. https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.2018.html. Accessed February 28, 2021.

U.S. Census Bureau. Table S0701, Geographic Mobility by Selected Characteristics in the United States, 2014-2018 American Community Survey 5-Year Estimates, Washington, D.C., 2019b, https://data.census.gov/cedsci/. Accessed March 2, 2021.

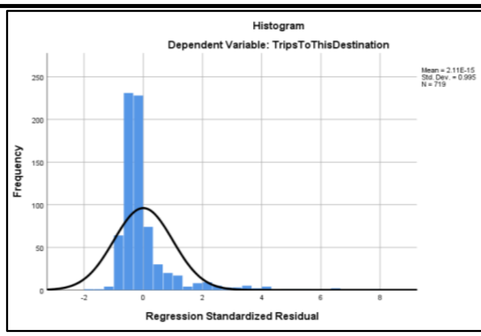
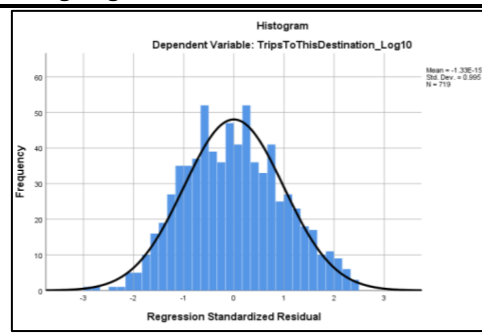
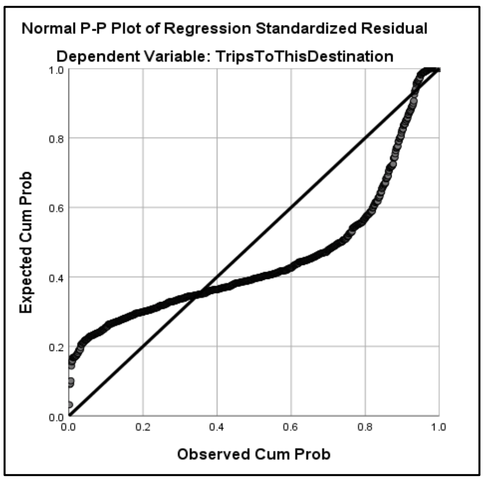
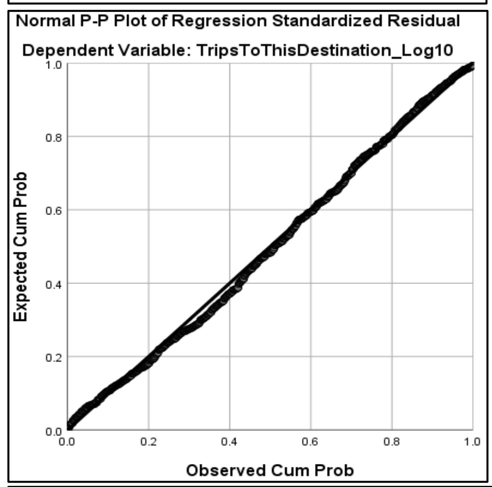
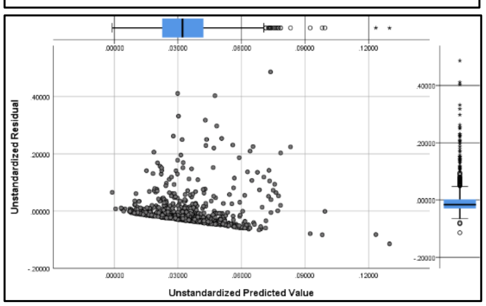
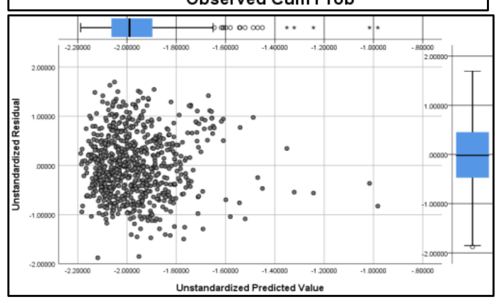
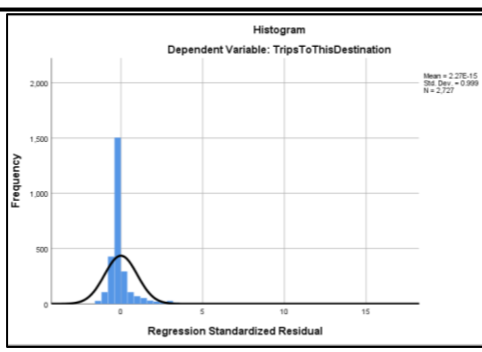
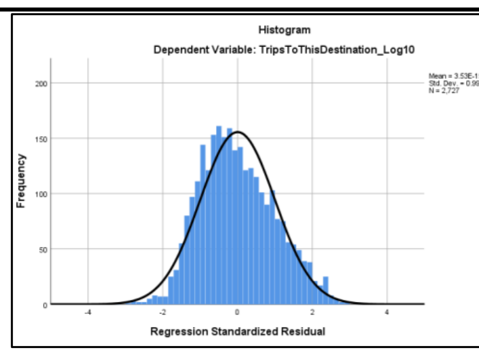
U.S. Census Bureau. Table S1903: Median Income in the Past 12 Months (in 2018 Inflation-Adjusted Dollars), 2014-2018 American Community Survey 5-Year Estimates, Washington, D.C., 2019d. https://data.census.gov/cedsci/. Accessed January 15, 2021.

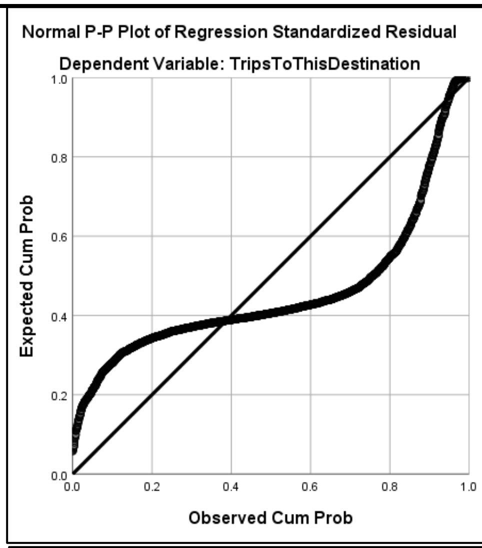
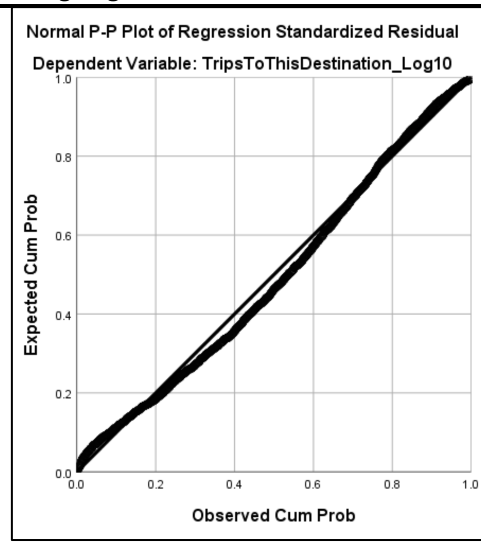
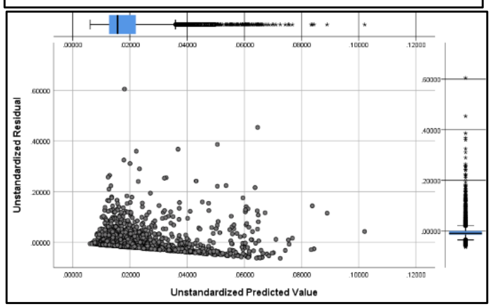
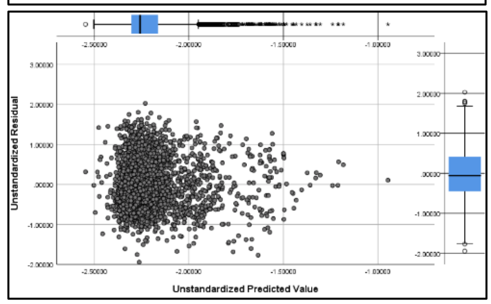
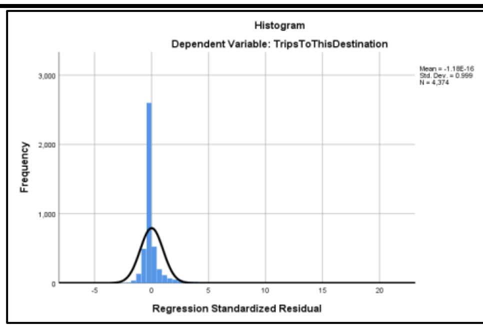
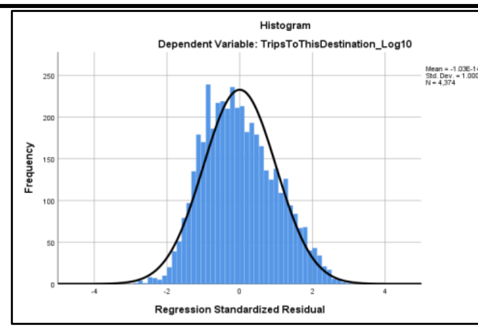
Virginia Department of Transportation. GIS_DATA_TMPD_MPO_STUDY_AREA. Richmond, October 6, 2020

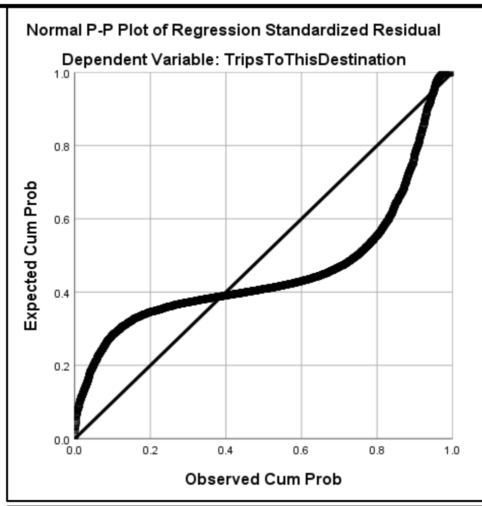
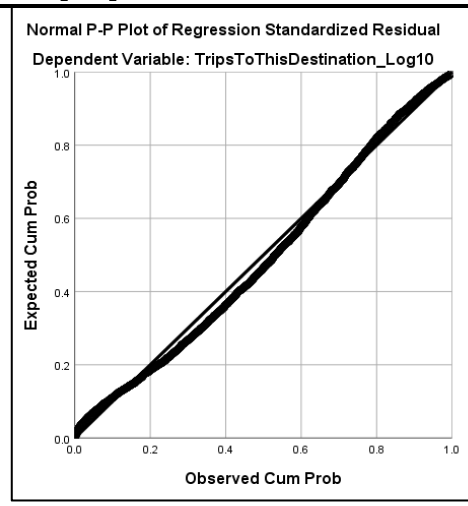
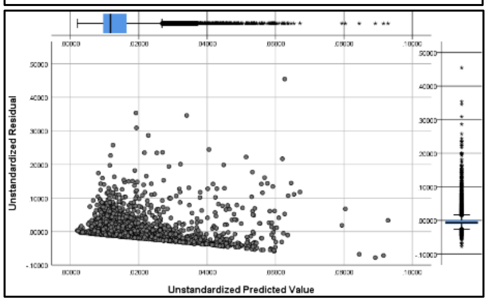
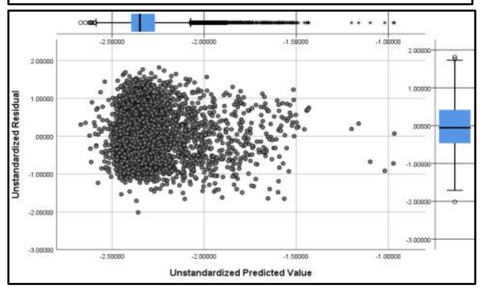
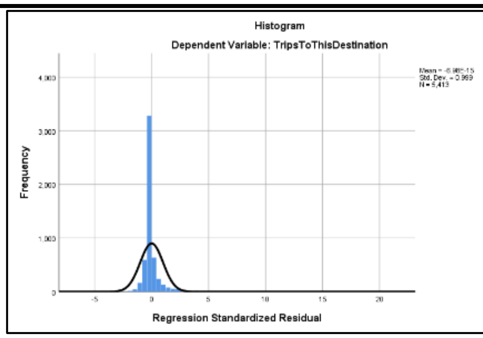
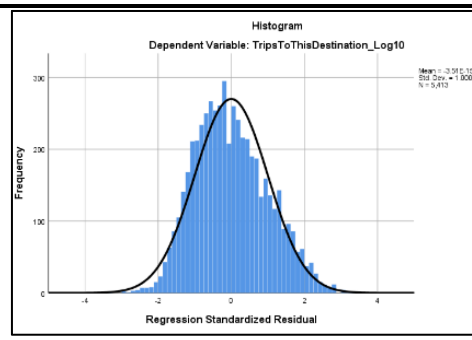
Zhang, M., Kone, A., Tooley, S., and Ramphul, R. Trip Internalization and Mixed-Use Development: A Case Study of Austin Texas, Report No. WUTC/09/169207-1, University of Texas, Austin, 2009. https://static.tti.tamu.edu/swutc.tamu.edu/publications/technicalreports/169207-1.pdf. Accessed March 25, 2020.

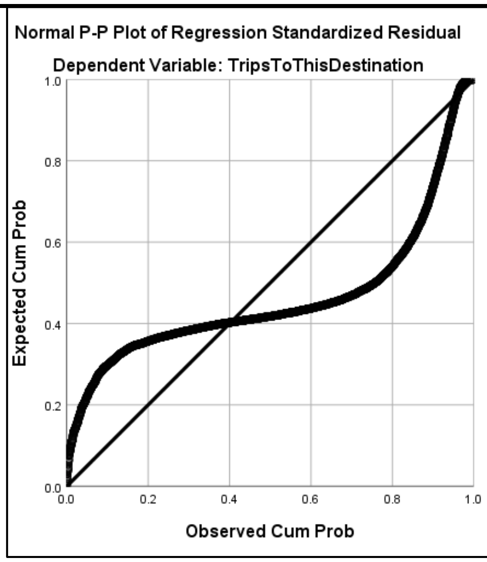
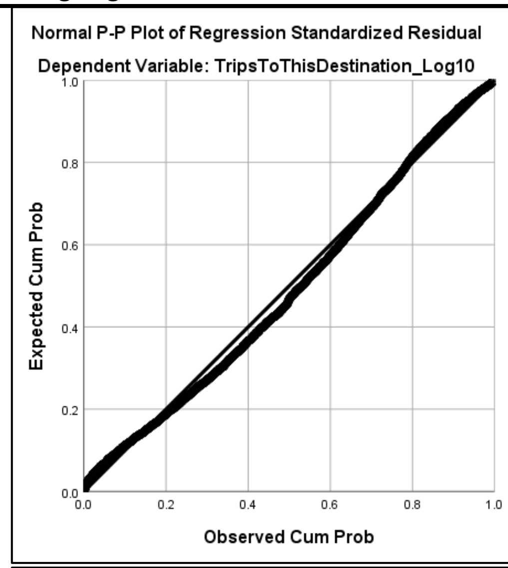
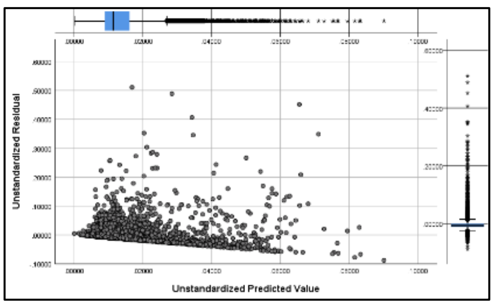
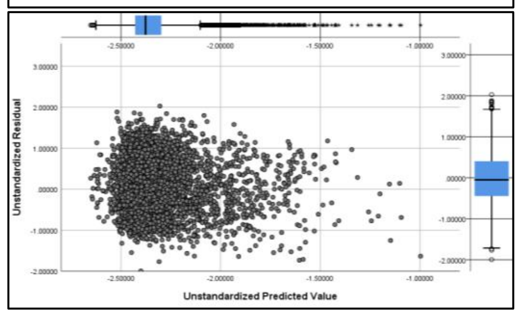
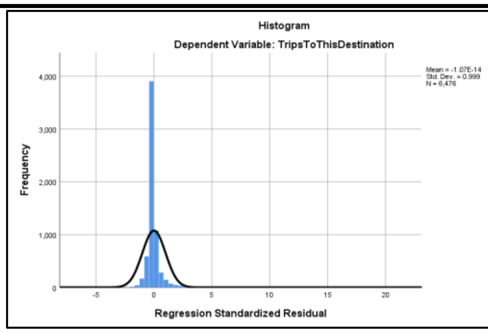
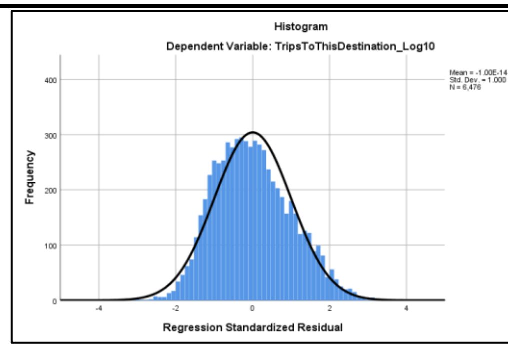
APPENDIX

Table 1. The residuals are normally distributed after Log 10 transformation of the dependent variable

<table>
<tr><th>Radius (Miles)</th><th>Not Normal Dependent Variable</th><th>Transforming the dependent variable using Log10</th></tr>
<tr><td>5</td><td>
</td><td>
</td></tr>
<tr><td>5</td><td>
</td><td>
</td></tr>
<tr><td>5</td><td>
</td><td>
</td></tr>
<tr><td>10</td><td>
</td><td>
</td></tr>
</table>

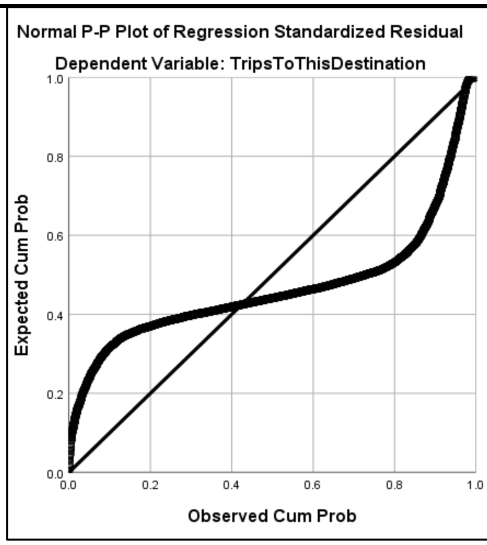
<table>
<tr><th>Radius (Miles)</th><th>Not Normal Dependent Variable</th><th>Transforming the dependent variable using Log10</th></tr>
<tr><td>10</td><td>
</td><td>
</td></tr>
<tr><td>10</td><td>
</td><td>
</td></tr>
<tr><td>15</td><td>
</td><td>
</td></tr>
</table>

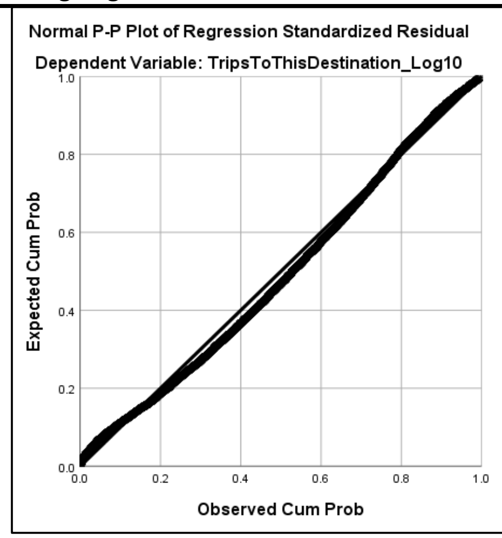
<table>
<tr><th>Radius (Miles)</th><th>Not Normal Dependent Variable</th><th>Transforming the dependent variable using Log10</th></tr>
<tr><td>15</td><td>
</td><td>
</td></tr>
<tr><td>15</td><td>
</td><td>
</td></tr>
<tr><td>20</td><td>
</td><td>
</td></tr>
</table>

<table>
<tr><th>Radius (Miles)</th><th>Not Normal Dependent Variable</th><th>Transforming the dependent variable using Log10</th></tr>
<tr><td>20</td><td>
</td><td>
</td></tr>
<tr><td>20</td><td>
</td><td>
</td></tr>
<tr><td>25</td><td>
</td><td>
</td></tr>
</table>

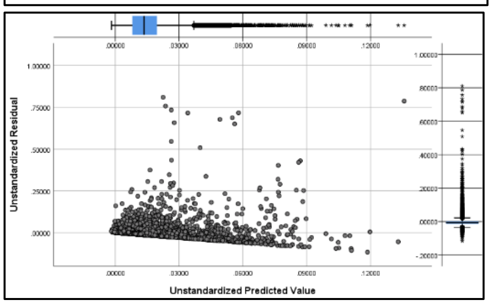
Radius (Miles)	Not Normal Dependent Variable	Transforming the dependent variable using Log10

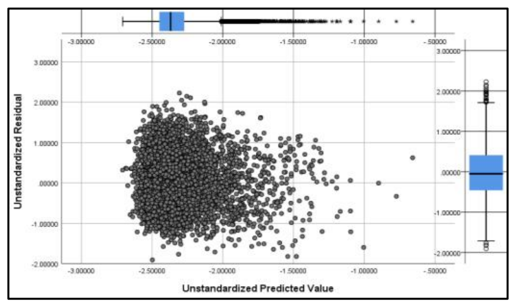
25



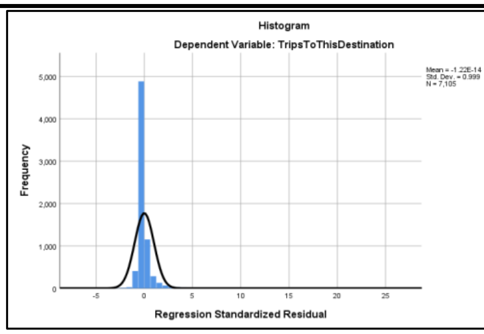


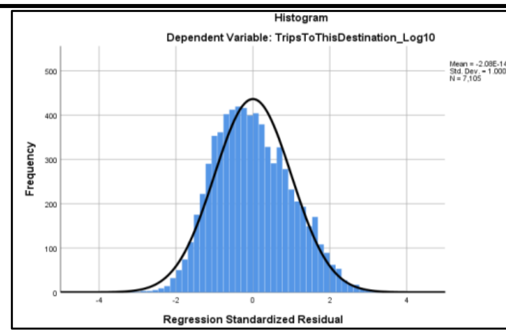
25

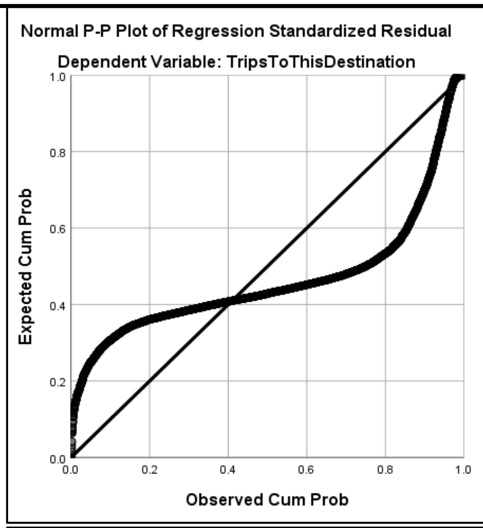
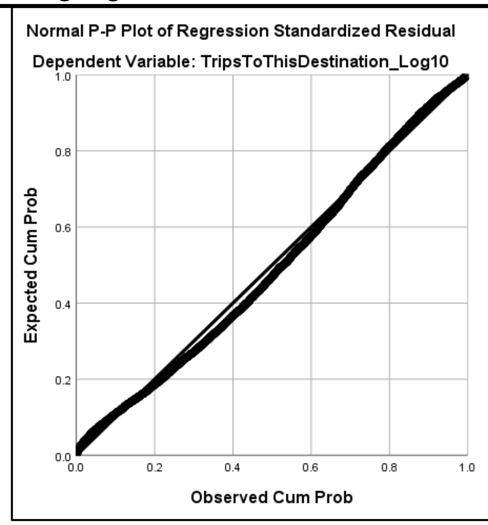
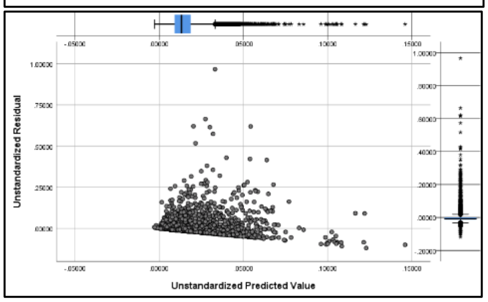
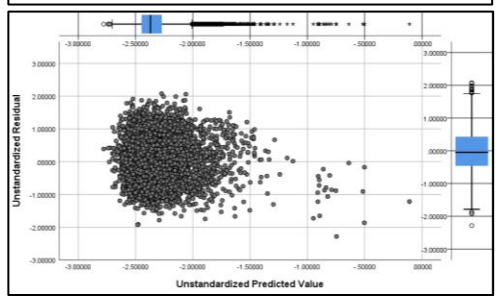
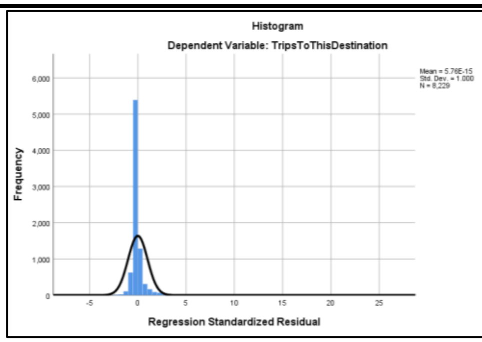
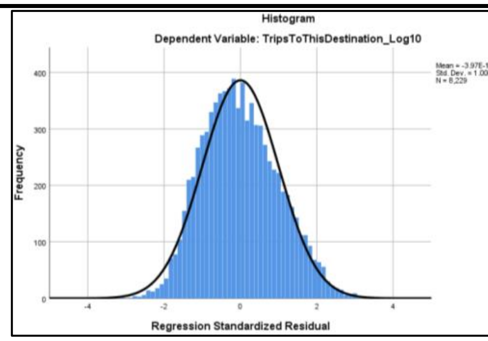


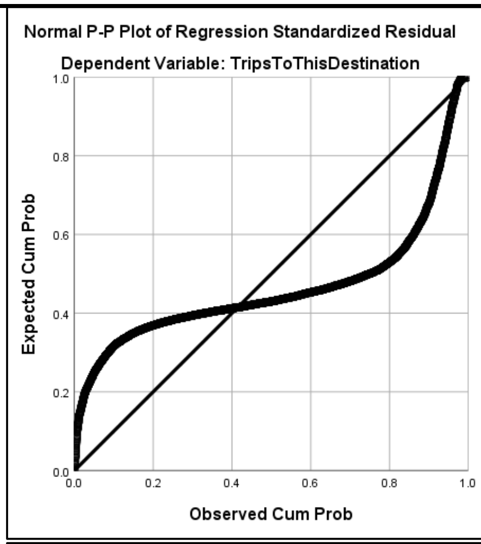
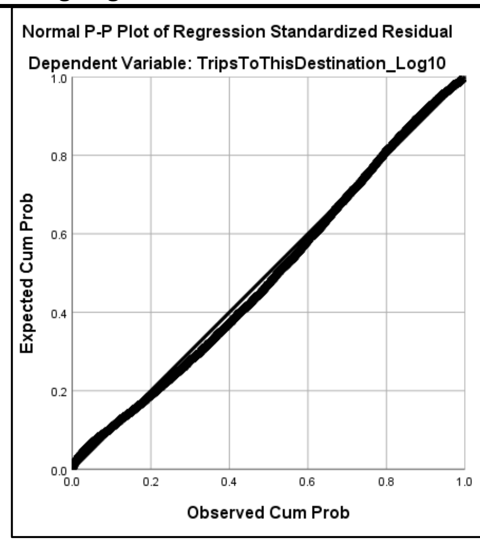
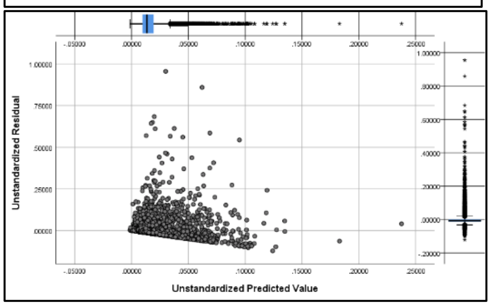
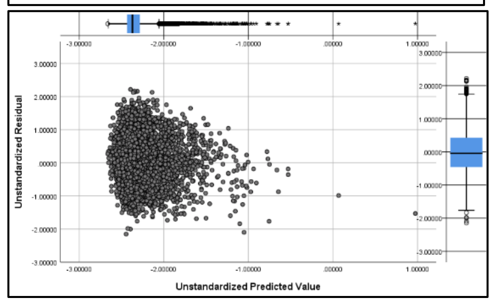


30





Radius (Miles)	Not Normal Dependent Variable	Transforming the dependent variable using Log10
30		
30		
35		

<table>
<tr><th>Radius (Miles)</th><th>Not Normal Dependent Variable</th><th>Transforming the dependent variable using Log10</th></tr>
<tr><td>35</td><td>
</td><td>
</td></tr>
<tr><td>35</td><td>
</td><td>
</td></tr>
</table>

PAPER 3: REDUCING CONFLICT: CHOOSING AN AUTO ACCESSIBILITY SPHERE OF INFLUENCE TO EXPLICITLY SERVE LOW-INCOME POPULATIONS

(Proposed Submission to the Journal of Transport Policy)

Abstract

An implicit choice in the project prioritization process is the extent to which candidate projects should be evaluated with respect to their ability to serve low-income populations. Thus, a conflict can arise regarding the percentage of a project's evaluation that should be earmarked for these groups. As jobs-based accessibility is also a component of project prioritization, this paper puts forth a method for reducing this conflict: judicious selection of the distance from the project for which job access impacts are calculated. Mathematically this distance substantially affects a project's forecast improvement on accessibility. This paper describes how to choose this distance such that the net accessibility benefit of a candidate project when all populations and all jobs are considered is similar to the net accessibility benefit when only low-income populations and associated employment are considered.

Findings from five candidate projects studied indicated that this distance may affect project evaluation when total and low-income populations are considered, based on three proposed measures of consistency. With the first consistency measure, project rankings differed at 5, 10, 20, and 30 miles, with consistent rankings at 15, 25, and 35 miles. With the second consistency measure, the spatial distribution of zonal contributions to accessibility, consistency was lowest at 35 miles; moderate at 20 and 30 miles; and highest at radii of 5, 10, 15, and 25 miles. With the third consistency measure, the correlation between zonal accessibility scores for total and low-income populations, the correlation was relatively strong at all radii except at 10 miles. Cumulatively, these findings suggested that a project distance (drawn from the possibilities of 5 to 35 miles) of 25 miles indicates the least amount of bias against low-income populations when accessibility is used as a measure in project prioritization.

Keywords: accessibility, total population, low-income population, equity

1. Introduction

Project prioritization—that is, the process through which candidate transportation investments are selected for implementation—often makes use of multiple criteria in safety, operational performance, the physical environment, infrastructure condition, and land development (Commonwealth of Virginia, 2018). One often-considered criterion is "accessibility"—that is, the ease with which the transportation system enables connections between residents and key social functions (Sinha and Labi, 2007) such as employment. However, application of this criterion conjures an equity question: to what extent should such an accessibility measure reflect the needs of the population at large vs. the low-income population specifically? Although Executive Order 12898 requires that transportation projects not adversely affect two protected groups—minority and low-income groups—relative to the rest of the population, it does not require that benefits to the two groups account for a particular percentage of project evaluation when projects are evaluated. Certainly, there is no single best way to eliminate this conflict. Stakeholder input is a fundamental component of the transportation planning process, and key decision points for addressing this conflict, such as the establishment of goals and objectives in the long-range plan or public participation in the project development process, are credible ways of resolving such conflicts. However, the difference in project prioritization that results when total populations vs. low-income populations are considered is not fully understood.

2. Literature Review

Novak et al. (2015) distinguished two often-opposing views of equity: alignment of benefits with revenue sources (e.g., toll roads), and the shifting of resources from higher to lower income groups. Regarding the latter, equity analyses may take at least three forms: (1) "procedural" equity (Meyer and Miller, 2020), i.e., whether all individuals have an equal ability to influence investments; (2) "geographic" equity, i.e., whether all locations receive an equal benefit from a given investment; and (3) "social" equity, which concerns the distribution of benefits and disadvantages for "minority and low-income" groups, which is the focus of environmental justice analyses that are required for federally funded projects (U.S. Department of Transportation, 2013). With respect to access in particular, three findings of particular import have been that (1) inequity in accessibility is common; (2) transportation investments have the potential to improve this accessibility, but differences vary by mode; and (3) although accessibility for low-income populations should be explicitly considered, few studies fully explain how to incorporate this consideration into the project prioritization process.

2.1. Inequity in Accessibility Is Common

The literature (e.g., Allen and Farber, 2019; Deboosere et al., 2018) reports that disadvantaged populations tend to have weaker accessibility to jobs, which others (e.g., Preston and Fiana, 2007; Pereira et al., 2017) have stated leads to adverse consequences such as greater transportation costs or an inability to participate in activities. With regard to prioritizing transportation projects, most academic studies have focused on transit or high speed rail, as noted by Merlin et al. (2018); of 16 detailed accessibility studies, only 1 focused on highway-oriented transportation.

Allen and Farber (2019) noted that unequal accessibility by location is an inherent part of the urban transportation system, owing to the fact that there is not uniformity in the spatial distribution of land and transportation networks in cities. The authors stated that one disadvantaged group is low-income households with insufficient public transportation service to their destinations. Although findings for Canadian cities showed that lower income neighborhoods tended to have higher rates of transit accessibility, owing to their tendency to be located in more urban locations, a significant number of low-income urban populations nonetheless experienced low transit accessibility. Deboosere et al. (2018) measured accessibility through public transport to low-income jobs for vulnerable citizens, specifically taking into consideration their travel times, and found that although low-paying jobs were more accessible to these individuals than to others, low-income travelers still suffered from relatively poor accessibility.

Lucas (2012) suggested that poor accessibility, whether it reflects not having a vehicle or suffering from poor transit service, can interact with other social policy concerns such as health inequalities, unemployment, and a poor education level to yield an inability to attain essential goods and services. Other studies (e.g., Preston and Fiana, 2007; Casas, 2007; Pereira et al., 2017) have suggested that this interaction could lead to an increase in the generalized cost of travel, a reduction in activity participation, and social exclusion. Fransen et al. (2019) developed a predictive model to address the limitations of applying aggregate measures to connect persons who are currently unemployed to available jobs. The study used "male, middle-aged (35–54), without migration background, living in a neighborhood with average population density, having a bachelor's degree and preferring a job in the domain of business support, retail and ICT" as a reference category to represent an average population group. The findings indicated that because jobs accessibility has a negative relationship with long-term unemployment, the probability that the most disadvantaged populations (referring to "job seekers with a migration background and with higher age 55 years or older") will remain unemployed in the long term is about two or three times higher than the total population. Fransen et al. (2019) indicated that only the advantaged population (the reference category) appear to have higher accessibility benefits.

2.2 Transportation Investments May Improve Accessibility, But Importance Differs by Mode.

Most literature suggests that the best way to improve accessibility for the disadvantaged is to improve public transportation (Lopez, 2003; Sanchez et al., 2004; Estache et al., 2000; Fan and Chan-Kang, 2005; Serulle and Cirillo, 2016), although others have suggested a more nuanced view by mode (e.g., Acheampong and Silva, 2015; Sun and Zacharias, 2020).

Acheampong and Silva (2015) noted that some individuals may have a greater ability to travel a longer distance than others, owing to differences in auto ownership or income. Sun and Zacharias (2020) examined policies from the perspective of transport equity by assessing the difference in accessibility between public and private vehicles. The study considered time budgets of 30, 45, 60, and 90 minutes, and the results showed that at a 30-minute time budget only 3.7% of the population was able to have access to 70% of the jobs when the private car mode was considered. Crucially, the study revealed a transit limitation: 66.1% of the population was able to access 10% of the jobs by private car, and only 15.1% of the population was able to access 10% of the jobs by public transport). By contrast, Serulle and Cirillo (2016) used the regional travel demand model to analyze the availability of jobs, travel behavior, and trip chaining for low-income populations in the Washington, D.C., area, finding that policies encouraging public transportation investment would deliver higher benefits to low-income populations as opposed to those supporting lower running costs for vehicles.

Carroll et al. (2021) suggested that by considering disadvantaged populations (identified through the deprivation index, which is based on socioeconomic variables such as unemployment, education, change in population, areas with poor public transportation, and locations with above-average auto ownership, one may identify locations where “forced” auto ownership exists and may thus consider interventions such as ridesharing, carsharing, and other transport services that are feasible in low-density areas, especially for increasing jobs-based accessibility. Lopez (2003), Sanchez et al. (2004), Estache et al. (2000), and Fan and Chan-Kang (2005) revealed that there is a direct impact on low-income populations when there is a low transportation connectivity, such that an improvement in mobility can have a positive impact on the job status of persons of low income.

2.3 Few Studies Have Discussed How to Integrate the Expected Impact on Accessibility for Disadvantaged Populations Into the Project Prioritization Process.

Although disaggregate data (e.g., the mode of transport for person x with income y to go to job z) are necessarily a part of studies to illuminate travel behaviors, the project prioritization process generally uses more aggregate information (e.g., how improving a given link will affect travel between multiple combinations of origin and destination zones where the total jobs and population of those zones are known). Certainly, studies have recommended the inclusion of accessibility into the project prioritization process. For example, Deboosere and El-Geneidy (2018) considered transit times for low-income workers and found that such workers have shorter commute times than the total population. Hence, the authors recommended that policy makers take interventions for the most vulnerable populations.

However, studies have not generally shown how to consider this metric given that transportation investments must satisfy a variety of criteria (e.g., reduce pollution, congestion, or crashes) and the data are not necessarily granular. A notable exception is Bocarejo et al. (2012) who developed a method to evaluate public transportation investment; the authors computed accessibility measures to the labor market for subareas in Bogata (Columbia) by using an impedance function that consisted of a travel time budget and percentage of income spent on transport. Although the study found that real accessibility per capita is not entirely dependent on income, inequality of access to job opportunities was observed and in some cases fewer jobs were accessible within a given distance.

To be clear, there are many examples of prioritization processes that consider disadvantaged populations when locating projects: for example, whether a given project is located in an area that serves environmental justice locations is 1 of 10 criteria considered by the Association of Central Oklahoma Governments (2019). However, the authors are not aware of studies that have explicitly examined how to quantify a proposed project’s impact on auto accessibility (for both low-income populations and total populations) where before and after accessibilities are explicitly measured and this information is incorporated as a single input into the multidimensional project prioritization process. Thus, although studies have focused on projects that can improve accessibility for low-income populations, this paper adds to this literature by reporting a study that examined the extent to which considering only low-income populations, as opposed to total population, alters the project’s relative rankings for investment opportunities when focused on the specific criterion of accessibility.

3. Purpose and Scope

This paper reports on a study to determine the extent to which a proposed transportation project's forecast improvement in jobs accessibility—that is, its accessibility score—is altered when the universe is low-income population rather than total population. Because forecast accessibility is sensitive to the distribution of people and jobs relative to the project location, the study examined the access impact of varying the size of the area relative to the project where access is measured. An objective of the study was to identify this radius such that the net accessibility benefit when all populations are considered is the same as the net accessibility benefit when only low-income populations are considered using several proposed measures of consistency and applying them to five candidate transportation projects in urban and rural locations in Virginia.

The scope was limited in two ways. First, the paper focuses on auto access, recognizing that in some locations, auto transport is a critical option for low-income populations. Second, recognizing that there are many computational decisions one must make in applying an accessibility formulation, such as the geographic granularity of the transportation network or the manner in which population centers are connected to this network, the paper focuses on the sensitivity of one particular parameter: the size of the area over which project benefits are tabulated. Previous studies have usually treated this parameter as a fixed value with the notable exception of Pokharel and Ieda (2016) who, after examining distances of 0 to 120 miles from the project, settled on the area within roughly 9 miles, yielding a project sphere of influence of about 250 square miles.

4. Case Study

Five proposed transportation projects were evaluated (see Table 1) where each project's impact on travel time was examined with respect to potential changes in accessibility. Virginia-specific data provided by Ling (2019) indicated expected total populations and total employment by U.S. Census block group for the year 2025. The U.S. Census Bureau (2020a) provided average monthly earnings for jobs via its Longitudinal Employment Dynamics (LED) Extraction Tool, but these averages are available only at the jurisdiction level. However, the "OnTheMap" application of the Longitudinal Employer–Household Dynamics (LEHD) dataset (U.S. Census Bureau, 2020b) provides jobs at the block level (e.g., smaller than a block group) where such jobs are categorized across three annual income levels: \$15,000 or less, more than \$39,996, or between these two amounts. The same application provides the blocks where low-income workers reside. In both cases, block point data were projected to the Universal Transverse Mercator (UTM) Zone 17N, consistent with census block groups used for the project.

A new category of employment and populations, defined as low-income jobs and low-income populations, was determined for each block group by determining the proportion of 2018 jobs and population that were low-income and then multiplying that proportion by 2025 employment and population to estimate 2025 low-income jobs and low-income populations. It was thus possible to determine accessibility using these low-income jobs and the low-income populations.

Table 1 Case Study Jobs and Employment Data

Category	No.	Description
Input variables	1	*Highway Links for No-build Scenario*: These baseline network data covering the entirety of Virginia consist of more than 3 million links. The comprehensive road network dataset contains attributes such as distance; speed; travel times during the AM peak; road functional class; travel direction; and digitization direction. Each link has a unique identification number that connects 2 nodes.
	2	*Highway Links for Build Scenario*: The build scenario dataset consists of the same dataset as element 1 with one exception: new speeds and new travel times reflecting the proposed transportation project being evaluated.
	3	*Junction Nodes*: The dataset consists of nearly 1.5 million nodes; each node has a unique code, which was useful in generating centroid connectors.
	4	*Block Groups*: These zones contain forecast demographic attributes for year 2025 such as population and employment.
	5	*Proposed Projects Dataset*: For each proposed project, this dataset consists of links indicating the project's location and in conjunction with data element 2 enables one to determine how the project will affect link travel times.
	6	*Turn Restriction Dataset*: The Virginia turn restriction dataset contains codes that correspond perfectly to the junctions of the highway network. Each link in the Virginia highway network dataset also has a unique code. The data were further processed using MySQL to match the nodes that form each link, with identifiers indicating restricted turning movements.
	7	*One Minute Bin Decay Values:* These reflect the value of a job as a function of travel time. For example, a job that is 5.5 to 6.5 minutes away has a value of 0.962622; a job that is 89.5 to 90.5 minutes away has a value of 0.009773.
Projects	8	*Project 1: US 250 / Route 20 Intersection Improvement (Charlottesville):* Reconstruct the US 250 (Richmond Rd.) and Route 20 (Stony Point Rd.) intersection to improve safety and operations. Project includes additional turn lanes, right of way, and medians and new signals.
	9	*Project 2: Pole Green Road Widening (Richmond):* Widen Pole Green Rd. (Rt. 627) from 2 to 4 lanes between Bell Creek Rd. and Rural Point Rd. (1.55 miles).
	10	*Project 3: George Washington Highway Widening (Hampton Roads):* This project will provide improvements to Rt. 17 by expanding the existing 3-lane undivided roadway to a 4-lane divided roadway from Yadkin Rd. to Canal Dr. The project will also include intersection improvements.
	11	*Project 4: I-81 Exit 300 at I-66E Northbound Widening (Staunton / Front Royal):* Add an additional lane and widen left shoulder to standard from Milepost 299.1 to 300.4 Northbound; replace and widen bridge over Water Plant Rd.
	12	*Project 5: Rt. 2 and Rt. 17 from Lansdowne Rd. Past Shannon Airport (Fredericksburg):* This project improves the intersection at Lansdowne Rd., widens Rt. 2 past the intersection of Shannon Dr., adds a southbound through lane on Rt. 2 from Bowman Dr. to Shannon Airport Circle, and adds a northbound right-turn lane on Lansdowne and westbound right-turn lane on Mansfield.

Table 2 shows the resultant socioeconomic data associated with each project. For example, the first row of Table 2 shows there are 269 block groups within 35 miles of project 1; some block groups have no residents, the largest has 9,999 residents, and the average population is 2,124. Table 2 shows that low-income employment accounts for between 23% of all jobs (project 2) and 27% of all jobs (project 1). Low-income population ranges from 21% of all people (projects 4 and 5) to 25% of all people (project 3).

Table 2 Descriptive statistics of zones within 35 miles of each project

Projects	Data Type	Mean	Std. Error	Median	Std. Dev	Variance	Min	Max	Sum	Count
1	Total Population	2124	97	1684	1597	2551786	0	9999	571440	269
	Low-Income Population	489	22	392	367	134936	0	2239	131451	269
	Total Employment	772	60	392	989	977200	0	5774	207568	269
	Low-Income Employment	209	17	100	283	79864	0	2205	56224	269
2	Total Population	1980	63	1650	1703	2899534	0	27290	1437432	726
	Low-Income Population	447	13	389	350	122314	0	6023	324783	726
	Total Employment	972	80	300	2150	4621001	0	32573	705529	726
	Low-Income Employment	227	15	84	406	164948	0	3972	164999	726
3	Total Population	1525	29	1332	957	915281	0	10751	1697418	1113
	Low-Income Population	383	7	334	247	61008	0	2995	426615	1113
	Total Employment	970	90	290	3001	9008326	0	67311	1079113	1113
	Low-Income Employment	244	21	88	687	472079	0	17030	271980	1113
4	Total Population	1399	80	1251	1316	1731989	0	7250	383341	274
	Low-Income Population	297	17	278	277	76574	0	1763	81492	274
	Total Employment	667	59	348	970	940595	0	6685	182881	274
	Low-Income Employment	171	16	93	257	66295	0	2434	46793	274
5	Total Population	2619	92	2095	2168	4701793	0	20299	1466726	560
	Low-Income Population	552	20	445	464	215640	0	4178	309135	560
	Total Employment	1029	87	366	2066	4266474	1	27119	576269	560
	Low-Income Employment	266	26	101	626	392008	0	11365	148694	560

5. Methodology

The methodology consisted of four steps:

- Develop a workflow for computing accessibility.
- Compute accessibility scores for the total population and the low-income population.
- Assess the sensitivity of accessibility scores to changing catchment radius.
- Quantify the consistency between accessibility scores for the total and low-income populations.

5.1 Workflow for Computing Accessibility

The workflow (Figure 1) consists of developing two datasets, one where a candidate transportation project is not built and one where the candidate transportation project is built. Data include turn restrictions, travel times, permitted travel direction, factors to decay jobs, and block group attributes for population and employment. The workflow uses ESRI's ArcGIS Network Analyst (ArcMap version 10.3.1), where service areas are generated for each 2-minute travel time interval for each candidate project for the no-build and build scenarios. Equations 1 through 4 compute accessibility scores using the intersection of population-based services areas and employment centroids.

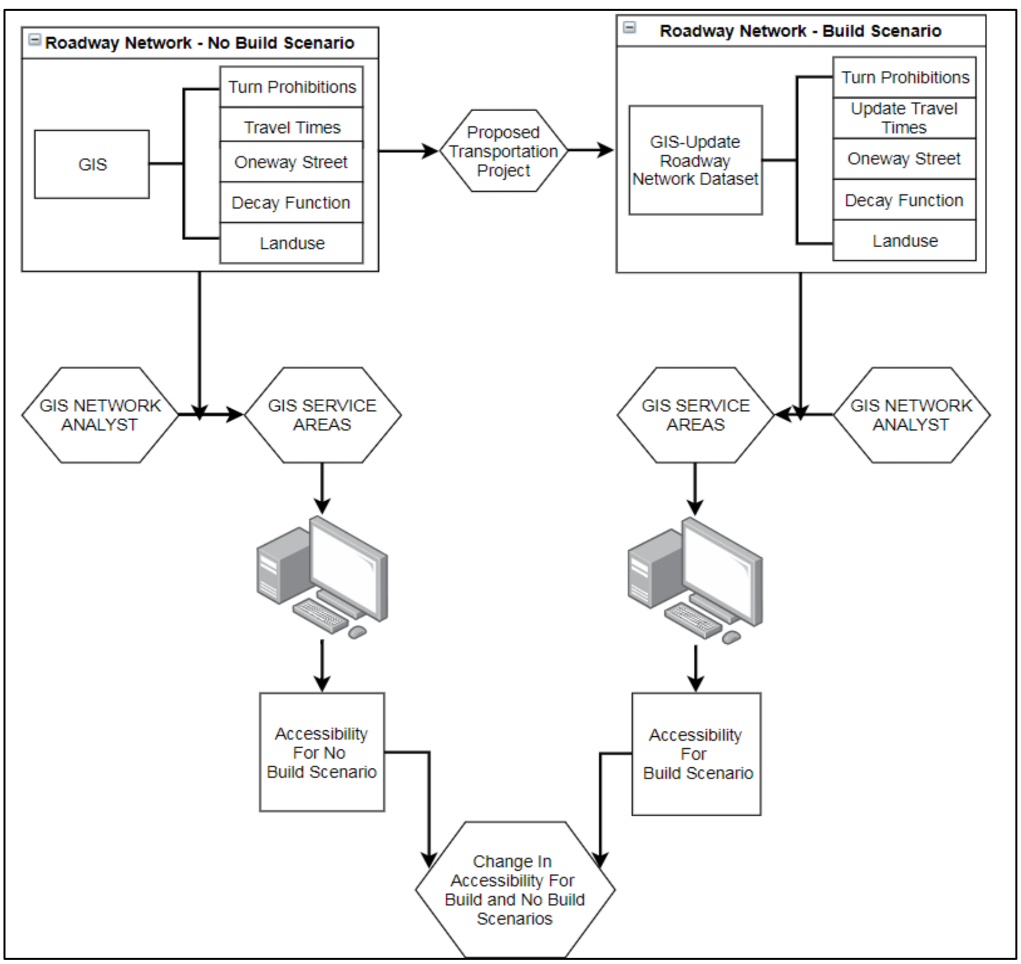


Fig. 2. Summary of the accessibility computation workflow.

The Virginia Department of Transportation (VDOT) provided the 12 data elements detailed in Table 1 (Z. Ling, personal communication). Computational challenges, such as automating the importation of turn restrictions, validating the network creation, and eliminating inconsistencies attributed to the use of GIS-based service areas, were resolved through preprocessing the dataset and adapting scripts to solve unexpected problems (Boateng and Miller, 2020). Once accessibility could be implemented, each of the five candidate projects, 10 different networks datasets (5 before and 5 after), was developed.

5.2 Accessibility Scores for Total Population and Low-Income Population

After a statewide accessibility network was established, the project accessibility score is measured as a change-in-accessibility score that would result if each project were built. Equation 1 is used to compute accessibility where for each block group i, accessibility is the sum, for all employment zones j, of the employment in zone j multiplied by a step decay function that in turn is based on the travel time from zone i to zone j before the project is built. As travel times increase from 0.5 to 90.5 minutes, the decay function decreases from 1 to 0.01, such that for a travel time of 6 minutes, the decay function has a value of roughly 0.91, which decreases to a value of 0.82 for a travel time of 8 minutes. Thus, for block group i, if 100 jobs are located 6 minutes away and 1,000 jobs are located 8 minutes away, the accessibility is presently (0.91)(100) + (0.82)(1,000) = 911.

$$\mathrm{A_i} = \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{Decay_{ij}Employment_j} \qquad \text{(Eq. 1)}$$

The population term ($\mathrm{Pop_i}$) in Equation 2 weights the accessibility for each block group i by the number of residents. Thus, for block groups with a large population, the accessibility in those block groups is more important than in block groups with a smaller population.

$$\mathrm{A_{before}} = \sum_{\mathrm{i}}^{\mathrm{j}} \left(\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{Decay_{ij}Employment_j}\right)\mathrm{Pop_i} \qquad \text{(Eq. 2)}$$

Population-weighted accessibility as shown in Equation 2 may be rewritten as Equations 3 and 4, where R, varies between 5, 10, 15, 20, 25, 30, and 35 miles (seven possible values) and C (project definition) is the middle of the project. Equation 4 is this change in accessibility divided by total population within R miles of C. The reason for the "max" term in Equation 3 is that in some cases, slight aberrations in the GIS processing can cause a negative change in accessibility; the max term addresses this concern (Boateng and Miller, 2021).

$$\Delta\,\mathrm{A} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \max(\mathrm{A_i^{After}Pop_i} - \mathrm{A_i^{Before}Pop_i}, 0) \qquad \text{(Eq. 3)}$$

$$\Delta\,\mathrm{A} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{A_i^{After}Pop_i}) - \sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{A_i^{Before}Pop_i})}{\text{Population within R miles of C}} \qquad \text{(Eq. 4)}$$

As an example of Equation 4, the case of C being the center of a particular link where a project will reduce the time to traverse that link from 3 to 2.5 minutes may be considered. Let R be 5 miles, and consider only the before accessibility (hence, Eq. 2 or the right side of Eq. 4). Although this equation is implemented in practice as a Python script, it may be visualized as a matrix operation where each "row" in Table 3 represents the numerator of Equation 4 for the n = 67 centroids within 5 miles of the proposed project. For instance, for centroid 1 (column A), the decayed employment for the before condition, i.e., A_1^{before}, is 3,312 (column B); this quantity is itself a summation based on Equation 1. For that first zone, the population-weighted accessibility (based on 1,054 residents shown in column C) is 3,491,495 (A_1^{before} Pop_1). A similar set of calculations are performed for each of the 67 zones within 5 miles of the project.

Table 3 Example of Computing Accessibility for the Before and After Condition, Project 1, 5 Mile Radius.

Before				After			
A	B	C	D	E	F	G	H
Population=	104,857			Population=	104,857		
Facility ID	Decayed Employment	Population	Population Decayed Accessibility	Facility ID	Decayed Employment	Population	Population Decayed Accessibility
1	3,312	1,054	3,491,495	1	3,312	1,054	3,491,495
2	963	2,212	2,130,458	2	963	2,212	2,130,458
3	339	653	221,160	3	1,114	653	727,913
4	10,453	3,312	34,618,352	4	11,062	3,312	36,636,918
5	27,715	1,348	37,360,696	5	28,444	1,348	38,343,872
6	16,091	1,149	18,487,212	6	16,091	1,149	18,487,212
7	22,130	2,471	54,690,600	7	22,130	2,471	54,690,600
8	31,809	5,757	183,111,087	8	31,809	5,757	183,111,087
9	24,785	1,007	24,960,989	9	24,785	1,007	24,960,989
64	48,717	1,126	54,870,278	64	48,896	1,126	55,072,487
65	44,468	4,218	187,585,923	65	44,474	4,218	187,609,737
66	56,521	1,227	69,355,225	66	56,624	1,227	69,480,972
67	48,273	1,775	85,668,538	67	48,273	1,775	85,668,538

After the project is built (Table 3 -After), the accessibilities are computed anew, where the population remains unchanged but generally the decayed employment either remains unchanged or increases because the travel times decrease. In this particular case, the link did not benefit the accessibility for zone 1, but it roughly tripled the accessibility for zone 2, as shown in Figure 2.

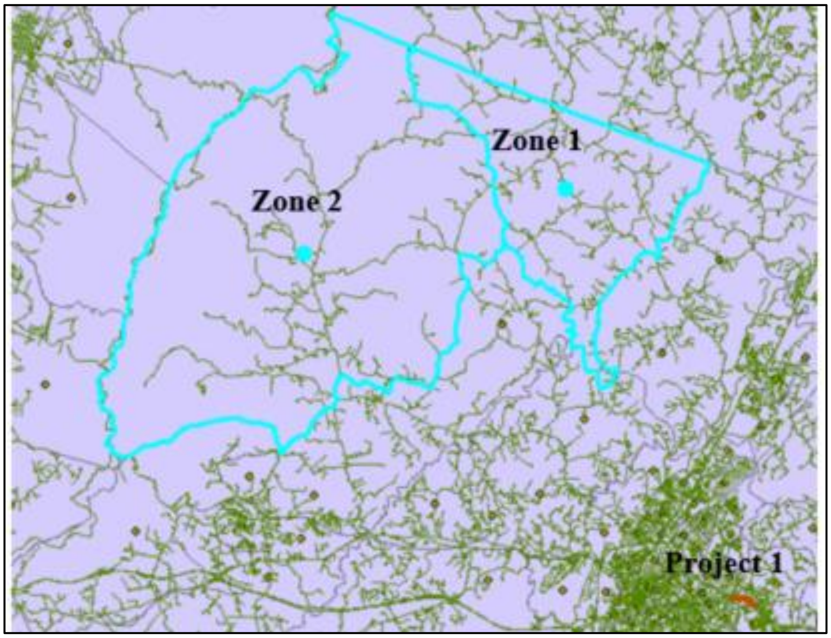


Fig. 2. Zones 1 and 2 and Project 1 location

Equation 4 then takes the difference in these two sets of accessibilities and divides it by the population. Thus, for Project 3 at a 5-mile radius, Equation 4 is tabulated as 3632806251 (sum of all values in column D of Table 3 including rows not shown) minus 1030019173 (sum of all values in column D of Table 4) divided by 104,856), which is equal to 204. It is possible to determine the relative contribution of each zone to the accessibility score; for the 13 rows shown in Tables 3 and 4, for example, zones 3 and 4 are contributing substantially to the improved accessibility score whereas zone 1 offers virtually no contribution.

5.3 Sensitivity of Accessibility Scores to Changing the Project Sphere of Influence

This paper defines this distance from the project for which access impacts are calculated (e.g., 5 miles in Figure 2) as the catchment radius as it ultimately determines the magnitude of the "catchment area" (Hardy and Bell, 2019). The selection of this radius is not a trivial matter as it affects not just the magnitude of the accessibility score but also relative scores: projects that are ranked higher with one radius may be ranked lower with another radius. An interesting application of this property is to consider the catchment radius to address equity concerns. To focus on low-income populations, one can alter the population and jobs terms in Equations 3 and 4. Thus, whereas the numerator of Equation 4 is initially written as Equation 5 in an expanded form, one can, when considering low-income populations only, replace this numerator with Equation 6.

$$A = \frac{\sum_{i=1}^{n} \text{Total population}_i \left(\sum_{j=1}^{n} \text{Decay}_{ij}^{\text{After}} \text{Total employment}_j - \sum_{j=1}^{n} \text{Decay}_{ij}^{\text{Before}} \text{Total employment}_j\right)}{\text{Total population within radius R}} \quad \text{(Eq. 5)}$$

$$A' = \frac{\sum_{i=1}^{n} \text{Low-income population}_i \left(\sum_{j=1}^{n} \text{Decay}_{ij}^{\text{After}} \text{Low-income employment}_j - \sum_{j=1}^{n} \text{Decay}_{ij}^{\text{Before}} \text{Low-income employment}_j\right)}{\text{Low-income population within radius R}} \quad \text{(Eq. 6)}$$

One can then determine how choosing a given radius R affects the correlation between the accessibility scores based on Equations 5 and 6. In this case study, the catchment radius (that is, the denominator of Equation 4) was varied with values of 5, 10, 15, 20, 25, 30, and 35 miles, as shown in Figure 3 with corresponding maximum travel times of 10.5, 20.5, 30.5, 40.5, 50.5, 60.5, and 70.5 minutes. The reason for these half-minute breakpoints is that the "decay" function, which values jobs that are located farther away less than jobs that are located nearby, is given for integer minutes whereas travel times are continuous. Thus, accessibility is computed for bins of 0 to 0.5 minutes, 0.5 to 2.5 minutes, 2.5 to 4.5 minutes, and so forth with the largest bin being 68.5 to 70.5 minutes.

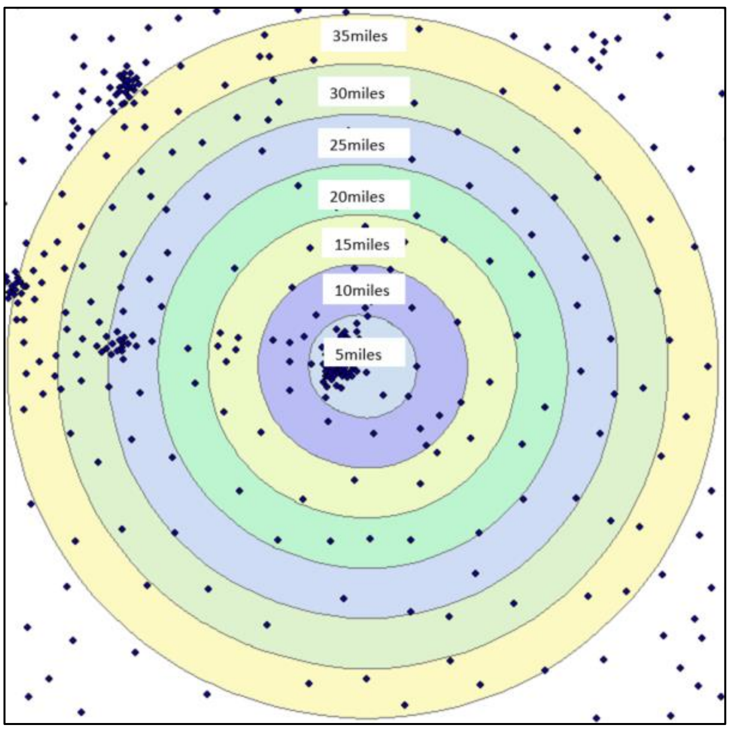


Fig. 3. Approximate Catchment Radii of 5 to 35 Miles for a Project (Center). The points represent centroids of block groups in the vicinity of the project**.**

The final step is to evaluate whether the project rankings are changed (or unchanged) by the use of Equation 5 or 6. Thus, this first test is one of practicality: does the prioritization order for the five projects change?

5.4 Consistency between accessibility scores for total and low-income populations

Several indicators were used to measure the consistency between accessibility scores computed via Equations 5 and 6. The t-test indicates if there is a significant difference at the block group level between the means of the accessibility scores for the total population and the low-income population. The advantage of this test is that it is the most direct; the disadvantage can be that it may simply reflect the fact that the low-income population is substantially less than the total population. For that reason, Equation 7, the first measure, is a starting point but is followed by two additional consistency measures.

$$T = \frac{Mean\ (tot\ pop) - Mean\ (low - income\ pop)}{\frac{S.D.}{\sqrt{(n)}}} \qquad \text{(Eq. 7)}$$

where:

Mean (tot pop) = the average accessibility scores for the total population

Mean (low-income pop) = the average accessibility scores for the low-income population

S.D. = standard deviation of the differences of the accessibility scores for the total and low-income population

n = the sample size

n-1 = the degree of freedom.

The second measure is based on the Kolmogorov-Smirnov (KS) test for discrete distributions, where i reflects the distance from the project in 2-mile bins. At a given radius R, the KS test will help determine if the distributions are the same or different. For example, for a 20-mile radius, i has 10 values: 0 to 2 miles, 2 to 4 miles . . . 18 to 20 miles. The KS test asks whether the contribution of bin i to the total accessibility score for a single project differs between scores based on total and low-income populations. The KS test was executed using Equation 8, where $F_{exp}(x)$ represents the distribution associated with the accessibility scores associated with the total population and $F_{obs}(x)$ represents the distribution associated with the accessibility scores for the low-income population distribution.

$$D_n = \max_{x} \left| F_{exp}\ (x) - F_{obs}\ (x) \right| \qquad \text{(Eq. 8)}$$

The third measure used a correlation analysis to determine the degree of linear association between the accessibility scores of the total and low-income populations. This correlation coefficient (Eq. 9) helps determine whether the relative change in scores altered when total population vs. low-income population was considered. If one wants to analyze how differences in accessibility scores for the total population can be explained by differences in accessibility scores for the low-income population, the coefficient of determination (r^2) can be used instead.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad \text{(Eq. 9)}$$

Where:

r = correlation coefficient

x_i = values of the accessibility scores for total population for each block group

$\bar{x}$ = mean of the values of the accessibility scores for total population

y_i = values of the accessibility scores for low-income population for each block group

$\bar{y}$ = mean of the values of the accessibility scores for low-income population.

In order to determine the statistical significance of the relationship between the accessibility scores of the total and low-income populations, the p-value for the coefficient of correlation (Hamburg, 1977) was computed by comparing the t-statistic (t) from Equation 10 with the critical t-value based on N-2 degrees of freedom, where N is the number of block groups and is typically close to n^2.

$$t = \frac{r}{\sqrt{\frac{(1-r^2)}{N-2}}} \quad \text{(Eq. 10)}$$

In order to evaluate whether the use of low-income populations and low-income jobs in Equation 6 (as opposed to total populations and all jobs in Equation 5) affects the overall accessibility score is to measure the coefficient of correlation at each radius *R*, where the deviance from a perfect correlation of 1.0 indicates the extent to which the method of Equation 6 affects project prioritization. Then, one can use Equation 11, adapted from Wuensch (2019) and verified with Soper (2015) to determine whether these correlations at different radii s and t are significantly different from each other, where p-values below 0.05 typically indicate a significant difference; C_s and C_t are the coefficients of correlation at radii s and t such that $C_s > C_t$, n is the number of projects for which these correlations are performed at each radius; and ϕ is the cumulative normal distribution. Equation 11 presumes two independent samples for computing the p-value. In this analysis, however, the samples are not completely independent because the datasets include common accessibility components for the jobs and populations at radii of 0 to 35 miles. Accordingly, an upper bound on Equation 6 can be developed based on this lack of independence where the pooled variance is less than or equal to the quantity $2*(1/(n-3))^{0.5}$ (Firebug, 2017) such that upper bound for the p-value is Equation 12.

$$p = 2\phi\left[-\frac{0.5\ln\frac{1+C_s}{1-C_s} - 0.5\ln\frac{1+C_t}{1-C_t}}{\sqrt{\frac{2}{n-3}}}\right] \quad \text{(Eq. 11)}$$

$$p = 2\phi\left[-\frac{0.5\ln\frac{1+C_s}{1-C_s} - 0.5\ln\frac{1+C_t}{1-C_t}}{2\sqrt{\frac{1}{n-3}}}\right] \quad \text{(Eq. 12)}$$

The fourth measure is comparable to the third except one may apply this analysis at the project level: to what extent is there correlation between overall project scores? The fifth and final measure is the project rankings: at what radius are the project rankings the most similar?

6. Results

Table 5 shows the initial results of Equations 5 and 6. For example, the 35-mile radius, shown as Row H, may be considered. For projects 1, 2, 3, 4, and 5, the accessibility scores are 71, 70, 55, 36, and 40, respectively (based on the total population) or 18, 15, 13, 10, and 11, respectively (based on the low-income population).

6.1 Impact of Altering the Radius

Clearly the radius affects the numerical score. For instance, an increase in the catchment radius from 5 to 10 miles (considering total population) resulted in a sharp decline in accessibility for projects 1, 3, and 5, and subsequent sharp declines are noted for projects 2 and 4 as one expands beyond a 10-mile radius. The reason for this score dropping beyond these radii is that the marginal increase in accessibility (the numerator of Eq. 4) is less than the marginal increase in population (the denominator of Eq. 4) as the radius grows beyond a relatively low value of 5 to 10 miles.

These sharp declines, however, become more moderate at radii of 15 miles and above. For instance, as the radius increases from 5 to 10 miles, projects 1, 3, and 5 lose an average of 44% of their maximum accessibility score at 5 miles, but for a radius increase from 15 to 20 miles, those projects lose, on average, only 8% of their maximum accessibility score at 5 miles. A similar trend is evident for projects 2 and 4; they peak at 10 miles, lose an average of 39% of their maximum accessibility score as the radius increases to 15 miles, but lose only an additional average of 5% of their maximum accessibility score as the radius increases from 15 to 20 miles. Thus, clearly the radius affects the scores, but the question then becomes how these scores compare across different populations.

6.2 Consistency Measure 1: Differences in Accessibility Scores

Table 4 generally shows that the accessibility scores differ between total and low-income populations. For example, Equation 7 showed that at a catchment radius of 35 miles, there was a significant difference between the means of the scores based on total and low-income populations (row H). The only cases where there was not a significant difference in these accessibility scores were for project 2 (5- and 10-mile catchment radius) and project 4 (10-mile catchment radius).

Table 4 Accessibility Scores for the Total and Low-Income Populations

Row	Projects	Catchment Radius (Miles)	Accessibility Benefit for Total Population	Accessibility Benefit for Low-Income	Ranking for Total Population	Ranking for Low-Income Population	Sample Size	t-Stat	P-Value at 95%
A	1	5	204	50	3	2	67	2.1	0.02
	2	10	151	30	4	4	232	1.4	0.08
	3	5	237	43	2	3	120	1.8	0.03
	4	10	92	25	5	5	43	1.3	0.11
	5	5	298	55	1	1	51	1.9	0.03
B	1	5	204	50	3	2	67	2.1	0.00
	2	5	69	20	4	4	40	1.4	0.08
	3	5	237	43	2	3	120	1.8	0.03
	4	5	0	0	5	5	10	--	--
	5	5	298	55	1	1	51	1.9	0.03
C	1	10	99	23	4	5	88	3.2	0.00
	2	10	151	30	2	3	232	1.4	0.08
	3	10	177	39	1	1	390	3.6	0.00
	4	10	92	25	5	4	43	1.3	0.11
	5	10	135	33	3	2	87	3.1	0.00
D	1	15	119	27	1	1	109	3.6	0.00
	2	15	92	18	4	4	401	2.6	0.01
	3	15	97	21	3	3	666	4.4	0.00
	4	15	56	16	5	5	93	3.6	0.00
	5	15	103	26	2	2	144	3.8	0.00
E	1	20	112	27	1	1	129	5.0	0.00
	2	20	74	16	3	4	511	2.7	0.00
	3	20	69	15	4	5	883	4.1	0.00
	4	20	58	17	5	3	125	3.6	0.00
	5	20	78	19	2	2	183	4.1	0.00
F	1	25	87	21	1	1	172	6.1	0.00
	2	25	81	17	2	2	595	3.2	0.00
	3	25	57	13	3	3	986	5.0	0.00
	4	25	40	11	5	5	152	3.5	0.00
	5	25	56	14	4	4	239	4.3	0.00
G	1	30	78	20	1	1	211	6.1	0.00
	2	30	69	15	2	2	683	3.2	0.00
	3	30	53	12	3	3	1065	5.1	0.00
	4	30	35	11	5	4	193	4.9	0.00
	5	30	40	10	4	5	382	4.4	0.00
H	1	35	71	18	1	1	269 [a]	6.5	0.00
	2	35	70	15	2	2	726	3.5	0.00
	3	35	55	13	3	3	1113	5.3	0.00
	4	35	36	10	5	5	274	4.9	0.00
	5	35	40	11	4	4	560	5.0	0.00

*Not significant at 95% confidence level.

[a] For example, for project 1 there are 269 zones within 35 miles of the project. For zone 37, the quantity ΔA, the numerator of Eq. 5 = 49, which is larger than ΔA', the numerator of Eq. 6 = 4. For the 269 zones, the mean value of ΔA was 150479, which is larger than the mean value of ΔA' = 8631, and this difference was statistically significant ($p < 0.01$).

6.3 Consistency Measure 2: Geographical Contributions to Accessibility

The KS test examines geographical variation in terms of relative distance to the project. For instance, if zones 2 to 4 miles away from the project contribute 75% of the accessibility score based on total populations yet 10% of the accessibility score based on low-income populations and if such differences are observed at other distances from the project, then the KS test would tend to show a difference. The 35-mile radius (far right of Table 5) shows that except for project 1, there is a significant difference in the distribution of the accessibility benefits for the total and low-income populations.

When all radii are considered, the KS test always showed a significant difference in the distribution of the geographic contribution to accessibility scores for the total and low-income populations at catchment radii of 5, 10, 15, 20, 25, 30, and 35 miles. This inequality in geographical contribution would not be evident from examination of Table 2 alone: the percentage of population that is low income within 35 miles of project 3 in Hampton Roads (25%) is slightly higher than the corresponding percentages for the remaining four projects (21%-23%).

Table 5 Using the Kolmogorov-Smirnov (KS) test to determine the distribution of total population and low-income population by varying the catchment radius

Projects	Type of Statistical Test	Catchment Radius (miles)						
		5	10	15	20	25	30	35
1	Test Statistic	0.05	0.09	0.03	0.02	0.01	0.02	0.02 [a]
	D-Critical Value	0.17	0.14	0.13	0.12	0.10	0.09	0.08
2	Test Statistic	0.06	0.07	0.06	0.07*	0.02	0.04	0.94*
	D-Critical Value	0.21	0.09	0.07	0.06	0.06	0.05	0.05
3	Test Statistic	0.16*	0.98*	0.99*	0.98*	0.95*	0.96*	0.99*
	D-Critical Value	0.12	0.07	0.05	0.05	0.04	0.04	0.04
4	Test Statistic	--	0.07	0.05	0.03	0.05	0.03	0.97*
	D-Critical Value	--	0.21	0.14	0.12	0.11	0.10	0.08
5	Test Statistic	0.07	0.03	0.04	0.02	0.03	0.97*	0.97*
	D-Critical Value	0.19	0.15	0.11	0.10	0.09	0.07	0.06

*Distribution is different for the total and low-income populations.

[a] For example, for a catchment radius of 35 miles, the zones that are 6 to 9 miles from project 1 contribute 8.6% of the accessibility score ΔA (numerator of Eq. 5) but 8.4% of the accessibility score for $\Delta A'$ (numerator of Eq. 6). An examination of the cumulative distribution for all 2-mile bins indeed does not show a significant difference for project 1 because the test statistic of 0.02 is smaller than the D-critical value of 0.08.

6.4 Consistency Measure 3: Correlation between zonal accessibility scores

Although it is evident that the accessibility score for a given block group i will tend to be smaller if it is based on the low-income population rather than the total population, a separate metric is the strength of the correlation between these two sets of scores for all block groups. If, for example, the accessibility scores for zones 1, 2, and 3 based on total population were 80, 90, and 100 compared to accessibility scores of 8, 9, and 10 based on low-income population, then this consistency measure would be fairly strong. Table 6 shows the actual correlations: for project 1 at a 10-mile radius, there is a strong relationship between the accessibility score based on the total population and the accessibility score based on the low-income population total with a correlation of 95%. Except for project 3, the correlation is always at least 90%.

Table 6 Correlation of Accessibility Scores for Total and Low-Income Population (Zone Level)

Projects	Analysis	Catchment Radius (Miles)						
		5	10	15	20	25	30	35
1	Correlation (r)	1.00	0.95	0.99	0.99	0.99	0.99	0.98 [a]
	Test statistic	2.00	1.99	1.98	1.98	1.97	1.97	1.97
2	Correlation (r)	0.98	0.99	0.98	1.00	1.00	0.99	0.99
	Test statistic	1.97	2.02	1.97	1.96	1.96	1.96	1.96
3	Correlation (r)	0.95	0.84	0.94	0.91	0.94	0.94	0.95
	Test statistic	1.98	1.97	1.96	1.96	1.96	1.96	1.96
4	Correlation (r)	--	1.00	0.97	0.98	0.95	0.97	0.96
	Test statistic	--	2.02	1.99	1.98	1.98	1.97	1.97
5	Correlation (r)	0.99	0.99	0.98	0.98	0.99	0.99	0.98
	Test statistic	2.01	1.99	1.98	1.97	1.97	1.97	1.96
Mean	Correlation (r)	0.98	0.95	0.97	0.97	0.97	0.98	0.97

[a] For example, for project 1, for zone 50, the quantity ΔA_{50} in the numerator of Equation 5 was 665580 and the quantity $\Delta A'_{50}$ n the numerator of Equation 6 was 32632. Although the latter is only a fraction of the former, there was a strong linear correlation for all 269 zones between ΔA_i and $\Delta A'_i$ of 0.98.

6.5 Consistency Measure 4: Correlation between project level accessibility scores

One can also tabulate a correlation analysis based on total project scores rather than individual zones. Table 7 further shows the comparison of accessibility scores for total population vs. low-income population at radii of 15, 25, and 35 miles. These three catchment radii provided similar ranking results for the two types of population datasets. The correlation comparison showed that a 15-mile catchment radius has a weaker correlation (87%) than a 25-mile (96%) and 35-mile (95%) catchment radius. However, these differences were not significant: Equations 11 and 12 showed that the p-values for the two most disparate radii (15 and 25 miles) were between 0.58 and 0.69. Not surprisingly, the differences between the nominally close correlations of 25 and 35 miles were not significant (p = 0.90-0.93).

Table 7 Correlation of Accessibility Scores for Total and Low-Income Populations (Project Level)

Projects	Accessibility Benefit for Total Population	Accessibility Benefit for Low-Income Population	Accessibility Benefit for Total Population	Accessibility Benefit for Low-Income Population	Accessibility Benefit for Total Population	Accessibility Benefit for Low-Income Population
	15 miles		25 miles		35	
1	119	27	87	21	71	18
2	92	18	81	17	70	15
3	97	21	57	13	55	13
4	56	16	40	11	36	10
5	103	26	56	14	40	11
Correlation (r)	0.88		0.96		0.95	

6.6 Consistency Measure 5: Agreement of rankings

Visual inspection suggests three desirable radii: 15, 25, and 35 miles. That is, Table 4 (rows B, C, E, and G) showed that at 5-, 10-, 20-, and 30-mile catchment radii, the ranking results for total and low-income populations differ. By contrast, these rankings are the same at 15, 25, and 35 miles: for instance, at 35 miles, the rankings were Project 1 (1st), Project 2 (2nd), Project 3 (3rd), Project 5 (4th) and Project 4 (5th).

7. Discussion

One surprise in Tables 4 through 7 was that there was not more discrepancy between the two methods used in Equations 5 and 6. At a 5-mile radius, the exclusive use of low-income population yielded the same top-ranked method as the use of all populations: project 5. It is only if one seeks the second-ranked project that there was a difference, with the low-income population accessibility yielding project 3 and the total population accessibility yielding project 1. One observation with regard to these similarities is that although the accessibility computations were computationally iterative at the zone level, they were not fully disaggregate at the individual level. Thus, because of the wide variation in zone-by-zone populations and jobs (e.g., for project 5 at a 35-mile radius, the standard deviation in zone-by-zone jobs was more than twice the mean value), key determinants of accessibility were naturally the number of people and decayed jobs by zone. Thus, variation in distributions of low-income jobs vs. total jobs (or low-income populations vs. total populations) may be masked by even larger variations in zone size. This does not mean that examination of low-income jobs or low-income population is immaterial, but rather that one should carefully examine differences in accessibility between these two methods *as even small differences (in this aggregate evaluation) may portend larger differences at the level of the individual traveler*.

Tables 4 through 7 may be used to select a catchment radius that therefore minimizes differences between the low-income accessibility and total accessibility, starting with the fact that the marginal differences in the numerator (decayed jobs) and the denominator (population) vary by project. For instance, when only total population is considered (e.g., the left side of Table 4), the best project at radii of 5, 10, and 15 miles were, respectively, projects 5, 3, and 1. Further, as shown in Table 4 row A, the maximum score for three projects was at a 5-mile radius whereas for the other two projects was at a 10-mile radius. Thus, how can these five proposed measures of consistency be used to choose a radius that minimizes conflicts between the two sets of accessibility scores?

- *Measure 1 (differences in scores based on Eq. 7)* showed statistically significant differences between accessibility scores for total and low-income populations among all catchment radii.
- *Measure 2 (differences in geographical contribution to the scores based on Eq. 8*) showed that at the 15- and 25-mile catchment radii, the distribution of the accessibility scores for total and low-income populations were similar for four of the five projects except for the third project. By contrast, at 35 miles, the distribution of the accessibility scores for total and low-income populations differed except for project 1.
- *Measure 3 (correlation between zonal accessibility scores in Eq. 9) s*howed the greatest agreement at R = 5 miles, a fairly strong agreement at R = 15 to 35 miles, and the weakest agreement at R = 10 miles but generally a remarkably strong correlation overall.
- *Measure 4 (correlation between project level accessibility scores)* showed for the five projects in Table 7 that at radii of 15, 25, and 35 miles, the coefficient of correlation (r) between the project scores A and A' from Equations 5 and 6, respectively, were 0.88 at 15 miles, 0.96 at 25 miles, and 0.95 at 35 miles. The correlation coefficient increased from 15 to 25 miles but dropped slightly at 35 miles; the substitution of low-income persons for total persons and low-income jobs for total jobs yielded a strong linear relationship between the two sets of scores at 25 miles.
- *Measure 5 (consistency of rankings)* showed agreement at radii of 15, 25, and 35 miles. Returning to measure 4, although the correlation between A and A' was stronger at 25 miles than at 15 and 35 miles, in neither case did it alter the particular project rankings.

On balance, measures 1 and 3 are not particularly useful: measure 1 will always tend to show a difference between accessibility methods, and measure 3 will tend to show similarities. Measure 5 is, of course, intuitive—do the rankings change—and clearly sensitive to the different radii. Measures 2 and 4 are more subtle but useful given that the projects themselves were samples: based on these latter three measures, the radii of 15 and 25 miles appear most promising. Based on measure 4, the 25-mile radius shows nominally better consistency than the 15-mile radius, although the correlations were not significantly different (p = 0.58-0.69). Accordingly, in this study where equity concerns rendered it desirable to choose a radius that yielded similar rankings, even though the rankings did not change for radii of 15, 25, and 35 miles (measure 5), measure 2 favored the 15- and 25-mile radius and measure 4 showed nominally greater agreement at R = 25, with the caveat that the difference in correlations was not significant.

Overall, the key parameter of interest, which in this case was the catchment radius is selected as a conflict resolution tool to ensure that the use of accessibility scores are not biased against low-income populations. For example, the radius that maximizes the p-value for the t-test, minimizes differences in rankings for the KS test, and increases the correlation is the radius that should minimize conflicts. Formally, this theory is the "goal programming approach" (an element of multicriteria decision making) where one seeks to minimize the difference between scores with total populations and scores with low-income populations (Meyer and Miller, 1984). Certainly, the goal of conflict resolution in transportation project prioritization has received generous attention, notably through explicit identification of multiple alternatives (Liu, 2015), which is the approach used in Virginia's project prioritization process, and through "tradeoffs" (Brody and Margerum, 2009), where some decision makers agree to support one project X (which excels in one particular criterion) in exchange for other decision makers supporting project Y (which excels in a different criterion). This study was not unique in seeking to address conflict resolution; rather, it was unique in that it sought to set parameters in such a way that conflicts between accessibility for the total population and accessibility for the low-income populations were reduced or eliminated. Although agencies could, of course, simply choose to consider only low-income populations when considering accessibility, broader support for projects might result if agencies could choose a radius that considered both populations equally.

8. Conclusions

The case study shows that when accessibility is used as a criterion for the selection of projects, the catchment radius affects the relative ranking; the top project at a radius of 10 miles, for instance, differed from the top project at a radius of 15 miles. Such differences in radii suggest an opportunity to consider the needs of low-income populations: rather than choosing a catchment radius that benefits a particular project (as could be done in Row A of Table 4), a catchment radius could be chosen where project rankings are consistent, regardless of whether total populations (vs, low-income populations) or total jobs (vs. low-income jobs) are used to compute accessibility. To this end, this study considered several measures of consistency, three of which appear appropriately sensitive to determine a suitable catchment radius:

1. *Consistency of project rankings* is the most intuitive measure: do the rankings remain unchanged when computing accessibility for total populations (ΔA) vs. disadvantaged populations (ΔA')? This showed that of the seven radii considered (0, 5, 10, 15, 20, 25, 30, and 35 miles), consistency was achieved at 15, 25, and 35 miles, as shown in Table 4.
2. *Consistency of spatial contribution* allows one to consider whether the geographical benefits of accessibility are similar. This may be measured statistically with the KS test (e.g., if 10% of accessibility benefits comes from zones 2 to 4 miles from the project when considering total populations, is a similar percentage computed when considering low-income populations?). No radius showed perfect consistency, but the greatest consistency was achieved at radii of 5, 10, 15, and 25 miles, as shown in Table 5.
3. *Consistency of correlation at the project level* enables detection of a linear association given that these projects were samples, with nominally higher correlation at R = 25 miles.

For these three measures, note that one may perceive only a small amount of conflict at what seem to be the most inequitable radii. Such small radii are likely due to high correlation at the census tract level between total activity and low-income activity. Such high correlation cannot be eliminated at the tract level, such that unless one has individual level data, the practical implication is this: even small differences in equity at the zone level may well signify larger equity differences at the individual level.

Not all proposed measures are useful: differences in accessibility scores for total and low-income populations tended to show large differences at all radii, and spatial contributions of zones tended to show high correlations.

Although previous studies have focused on projects that can improve accessibility for low-income populations, this study adds to this literature by examining the extent to which considering only low-income populations, as opposed to total population, alters the projects' relative rankings. For this particular case study dataset, a 25-mile catchment radius has the potential to reduce prioritization differences based on considering the accessibility needs of total and low-income populations.

8.1 Implementation of appropriate radius in a project prioritization process

At the metropolitan planning organization (MPO) or state level, the evaluation of candidate transportation projects is based on multiple criteria: examples are 10 for an Oklahoma MPO, 14 for a statewide process in Virginia, 23 for a New York MPO, and 31 for a statewide process in Vermont (Association of Central Oklahoma Governments, 2019; Commonwealth of Virginia, 2021; Meyer and Miller, 1984; Novak et al., 2015). Generally, such processes include fairly detailed documentation of how each of these criteria should be tabulated. Virginia's technical documentation runs to 101 pages and specifies details such as the number of points awarded for hybrid vehicle accommodation; the Minnesota Department of Transportation (2021) provides the real discount rate that should be used for amortizing the benefits of projects (along with a 20-page example of how to calculate a project's return on investment based on safety, operational, and delay savings with additional documentation provided for other criteria.

Other transportation agencies may adapt the approach presented herein by testing for a sample of projects which catchment radius yields the greatest agreement based on the three consistency measures and then incorporating that radius into the guidance for how to compute accessibility. In theory, an agency could compute, of course, the accessibility at multiple radii for all projects and then select the radius having the greatest alignment for total vs. low-income populations. In practice, runs at multiple radii are time-consuming: a run at 45 miles, for instance, takes X times longer than a run at 35 mile, which in turn takes Y times longer than a run at 25 miles. Thus, a sampling approach of geographically dispersed projects, which then informs the radius used for accessibility computation, is one way to accommodate consideration of low-income populations.

8.2 Future work and Limitations

Although this study showed the importance of selecting the appropriate catchment radius as a conflict resolution tool between the total and low-income populations, additional work can be performed to determine the feasibility of a common radius for multiple modes, such as a behavioral analysis: to what extent does the catchment radius affect the alignment of observed origin-destination data with forecast trips? Further, to what extent do stakeholders in the transportation planning process endorse the use of consideration of the radius in this manner? An alternative, for instance, might be to consider simply only low-income populations; this study sought to demonstrate that it is feasible to choose a radius that addresses the needs of all populations. Knowledge of the stakeholder reaction to such practices, as well as the computational details presented in this paper, is essential in conducting the public “vetting” advocated by Sundquist (2017) and Sundquist et al. (2018) to ensure that accessibility is a meaningful metric when candidate transportation projects are evaluated for construction.

Acknowledgments

Figure 1 showed the development of a process for computing accessibility, and this process was rendered feasible by insights from several individuals: a technical review panel guiding the research on which this paper was based (champion Peng Xiao, Ziwen Ling, Amy O’Leary, and Ram Venkatanarayana); IT assistance (Linda Cullop, Eric Hetzer, Michele Mandell, Adam Munro, Alex Ptak, and an anonymous member of the ESRI support team); and editing (Linda Evans). The authors are also grateful for the support for this research by the Virginia Department of Transportation and FHWA. The material in this paper is the responsibility of the authors and does not necessarily represent the viewpoint of these agencies.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: R. Boateng and J. Miller; data collection: R. Boateng; analysis and interpretation of results: R. Boateng and J. Miller; draft manuscript preparation: R. Boateng and J. Miller. Both authors reviewed the results and approved the final version of the manuscript.

References

Acheampong, R.A. and Silva, E.A., 2015. "Land use–transport interaction modeling: A review of the literature and future research directions." Journal of Transport and Land Use 8.3: 11-38.

Allen, J. and Farber, S., 2019. "Sizing up transport poverty: A national scale accounting of low-income households suffering from inaccessibility in Canada, and what to do about it." Transport Policy 74: 214-223.

Bocarejo, S., Pablo, J., and Oviedo D.R.H., 2012. "Transport accessibility and social inequities: a tool for identification of mobility needs and evaluation of transport investments." Journal of Transport Geography 24: 142-154.

Boateng, R.A. and Miller, J.S., 2021. “Incorporating Auto Accessibility Into Statewide Project Prioritization: Transparency & Feasibility.” Annual Transportation Research Board Conference, Paper No. TRBAM-21-00889, TRB, Washington, DC.

Brody, S. and Margerum, R.D., 2009. “Oregon’s ACTs, Cross-Jurisdictional Collaboration, and Improved Transportation Planning.” Report No. FHWA-OR-RD-09-11. Oregon Department of Transportation, Salem.

Carroll, P., Benevenuto, R., and Caulfield, B., 2021. Identifying Hotspots of Transport Disadvantage and Car Dependency in Rural Ireland. Transport Policy, 101, 46–56. https://doi.org/10.1016/j.tranpol.2020.11.004.

Casas, I., 2007. "Social exclusion and the disabled: An accessibility approach." The Professional Geographer 59.4: 463-477.

Commonwealth of Virginia, 2018. “SMART SCALE Technical Guide.” Richmond.

Deboosere, R., El-Geneidy, A.M, and Levinson, D., 2018. "Accessibility-oriented development." Journal of Transport Geography 70: 11-20.

Deboosere, R. and El-Geneidy, A.M., 2018. "Evaluating equity and accessibility to jobs by public transport across Canada." Journal of Transport Geography 73: 54-63.

Estache, A., Gomez-Lobo A., and Leipziger D., 2000. “Utility privatization and the needs of the poor in Latin America.” The World Bank, Washington, DC.

Fan, S. and Chan-Kang, C., 2005. “Road development, economic growth, and poverty reduction in China.” International Food Policy Research Institute, Washington, DC.

Firebug, 2017. How do I determine whether two correlations are significantly different? https://stats.stackexchange.com/q/278808.

Fransen, K., Boussauw, K., Deruyter, G. and De Maeyer, P., 2019. “The relationship between transport disadvantage and employability: Predicting long-term unemployment based on job seekers’ access to suitable job openings in Flanders, Belgium.” Transportation Research Part A: Policy and Practice, 125, pp. 268-279.

Hamburg, M., 1977. “Statistical Analysis for Decision Making, Second Edition.” Harcourt Brace Jovanovich, Inc.

Hardy, D., Bell, A., and McCahill, C. Accessibility Measurement for Project Prioritization in Virginia, Transportation Research Record, Vol. 2673(12) 266–276, Washington, D.C., 2019. DOI: 10.1177/0361198119859319.

Ling, Z. Email to J.S. Miller. February 14, 2019.

Liu, M., 2015. “Urban Transport Project Prioritization Strategy in Developing Countries: A Scenario-Based Multi-Criteria Decision Analysis Perspective.” Ph.D. Dissertation, Columbia University, New York.

Lopez H., 2003. “Macroeconomics and Inequality, Macroeconomic challenges in low income countries.” The World Bank, Washington, DC.

Lucas, K., 2012. "Transport and social exclusion: Where are we now?" Transport Policy 20: 105-113.

Merlin, L.A., Levine, J., and Grengs, J., 2018. Accessibility Analysis for Transportation Projects and Plans. Transport Policy 69, pp. 35–48. https://doi.org/10.1016/j.tranpol.2018.05.014.

Meyer, M.D. and Miller, E.J., 2020. “Transportation Planning: A Decision-Oriented Approach.” MTS Transportation Solutions, Atlanta.

Minnesota Department of Transportation, 2021. How MnDOT scores and selects major capacity expansion projects in the Twin Cities, St. Paul. http://www.dot.state.mn.us/projectselection/major-capacity-method.html.

Novak, D.C., Koliba, C., Zia, A., and Tucker, M., 2015. Evaluating the Outcomes Associated with an Innovative Change in a State-Level Transportation Project Prioritization Process: A Case Study of Vermont. Transport Policy 42, pp. 130–143.

Pereira, R.H.M., Schwanen, T., and Banister, D., 2017. "Distributive justice and equity in transportation." Transport Reviews 37.2: 170-191.

Pokharel, R. and Ieda, H. Road Network Evaluation from a Reliability Perspective: An Accessibility and Network Closure Vulnerability Approach. Asian Transport Studies, Volume 4, Issue 1, pp. 37-56, 2016. DOI https://doi.org/10.11175/eastsats.4.37. https://www.jstage.jst.go.jp/article/eastsats/4/1/4_37/_pdf/-char/en. Accessed July 30, 2020.

Preston, J. and Rajé, F., 2007. "Accessibility, mobility and transport-related social exclusion." Journal of Transport Geography 15.3: 151-160.

Serulle, N.U. and Cirillo, C., 2016. "Transportation needs of low income population: a policy analysis for the Washington DC metropolitan region." Public Transport 8.1: 103-123.

Soper, D., 2021. “Calculator: Significance of the Difference Between Two Correlations.” In Free Statistics Calculators (version 4.0), https://www.danielsoper.com/statcalc/calculator.aspx?id=104.

Sinha, K. and Labi, S., 2020. “Transportation Decision Making: Principles of Project Evaluation and Programming.” John Wiley & Sons, Inc.

Sun, Z., and Zacharias, J., 2020. "Transport equity as relative accessibility in a megacity: Beijing." Transport Policy.

Sundquist, E. Project Proposal: Accessibility as a Decision-making Tool for Utah Planning. Utah Department of Transportation, Salt Lake City, 2018.

Sundquist, E., McCahill, C., and Dredske, L., 2017. Accessibility in Practice: A Guide for Transportation and Land Use Decision Making. Office of Intermodal Planning and Investment, Richmond, VA.

U.S. Census Bureau, Accessed December 7, 2020. “LED Extraction Tool—Quarterly Workforce Indicators.” Washington, DC, 2020a. https://ledextract.ces.census.gov/static/data.html.

U.S. Census Bureau, Accessed December 7, 2020. “Area Profile Analysis in 2017” by Private Primary Jobs, OnTheMap, Washington, DC, 2020b. https://onthemap.ces.census.gov/.

U.S. Department of Transportation, 2013. Equity. Washington, DC. https://www.transportation.gov/mission/health/equity.

Wuensch, K.L., 2019. “Comparing Correlation Coefficients, Slopes, and Intercepts.” http://core.ecu.edu/psyc/wuenschk/docs30/CompareCorrCoeff.pdf.