Analysis Plan

Project Name: Simulated analysis of equity in federal broadband programs
Project code: 2115B
Date finalized: 2/9/2024

In this project, we use simulated and real data to propose a framework for evaluating equity performance of federal broadband programs. This project is designed to identify the ways in which policymakers' decisions in defining program eligibility have important implications for equity in program implementation and delivery.

In designing this project, we construct a hypothetical but realistic broadband program that demonstrates the impact policymakers' decisions can have on service delivery as they design and evaluate federal infrastructure programs. This simulation study is intended to assess methodological and policy tradeoffs. This analysis plan explains our analytical decisions, but does not attempt to confirm a specific set of empirical hypotheses about any real, existing broadband program. This plan has been written and posted after researchers had access to outcome data, prior to any analysis.

Project description

We propose to first investigate how different indicators of program prioritization – who should be considered eligible for a program – affect the equity of broadband program allocation. We then explore how "equity performance" – our determination of how equitable the allocation was – depends on which underserved groups are examined in the equity evaluation and which equity benchmarks are set. We expect that whether or not a given indicator of program prioritization results in a more equitable distribution is contingent on the definition of equity employed. Our core hypothesis is that what appears to be equitable under one prioritization indicator may not be equitable under another indicator. This analysis plan outlines a methodological approach to explore these issues using a mix of computer-simulated infrastructure program data and real Census socio-demographic data.

Since the COVID-19 pandemic, the federal government has undertaken an unprecedented investment in broadband infrastructure, spread across twelve programs and totalling roughly $90 billion.[1] So far, eligibility for broadband infrastructure has been defined using speed-based indicators of prioritization and program eligibility; communities with broadband connections that are not capable of high speeds are prioritized for infrastructure investment. The alternative to this definition would consider internet adoption - whether or not an individual has an internet connection at all. We make the simplifying assumption in this project that anyone lacking Internet (in terms of speed or adoption) could be classified as in need of Internet, and therefore either speed or adoption are valid prioritization criteria.

[1] Recent federal broadband programs emerged from two key pieces of legislation: the American Rescue Plan Act of 2021 (ARP) and the Infrastructure Investment and Jobs Act of 2021 (IIJA, also known as the Bipartisan Infrastructure Deal).

We explore how defining infrastructure prioritization using either speed or measures of self-reported Internet adoption leads to more or less equitable allocations of broadband programs. Each prioritization indicator may target different populations, as each indicator captures a distinct barrier to accessing broadband services. We expect these barriers to access to be distributed differently across space. Spatially-differentiated barriers to access likely have particularly pronounced implications for infrastructure programs, which are targeted at specific, "eligible" geographies, rather than individuals who can receive programs irrespective of their neighbors' eligibility.

This project has two primary research questions and objectives:

1. RQ1: What metrics can we use to evaluate equity in infrastructure programs?:
    a. Objective: Develop a methodology for evaluating equity performance for federal infrastructure programs.
2. RQ2: How do different ways of prioritizing infrastructure allocation affect program equity?
    a. Objective: Apply this methodology to assessing the equity performance of simulated federal broadband program allocations.

In addition to demonstrating the implications of different program prioritizations and equity goals for equity performance evaluations, this approach can also serve as a model that agencies can use to assess how different eligibility guidelines or program prioritizations may result in different equity performance evaluations ex-ante to program implementation. This can help policymakers make data-driven and evidence-based decisions about how to equitably implement programs in early stages of policy design.

Selection of prioritization indicators and equity definitions

How we're thinking about prioritization indicators

The two biggest broadband programs that emerged from recent legislation (see Table 1) both issued guidance that limit funding eligibility to places of high broadband need. The Capital Projects Fund (CPF) emphasized that priority should be given to communities that "currently lack access to the **affordable, reliable, high-quality broadband internet** that is necessary for full participation in school, healthcare, employment, social services, government programs, and civic life."[2] Recipients of broadband infrastructure must have a critical need for **better access, affordability, reliability, and/or consistency.** For CPF, many states defined critical need as lacking access to upload/download speeds greater than 100/20 mbps (defined as underserved) and 25/3 (defined as unserved). The Broadband Equity Access and Deployment (BEAD) funding was distributed based on the number of unserved locations in each grant area, defined as **lacking access to reliable broadband service at speeds greater than 25/3** or latency levels that can support real-time and interactive applications.[3] (Emphasis ours, throughout).

In these cases, program administrators prioritized communities based on proximity to broadband infrastructure that is capable of a given speed, but referenced a range of other need

[2] https://home.treasury.gov/system/files/136/Capital-Projects-Fund-Guidance-States-Territories-and-Freely-Associated-States.pdf
[3] https://broadbandusa.ntia.doc.gov/sites/default/files/2022-05/BEAD%20NOFO.pdf, p.2

conceptualizations, including availability, affordability, and reliability. The concept of broadband need is multifaceted, particularly when policies seek to address or measure digital equity. Need includes not only connectivity (can you access the internet at a reasonable speed for a fair price) but also digital literacy and comfort with a range of digital applications like remote work, online training opportunities, or telehealth.

For this project, we limit ourselves to prioritization indicators that capture challenges with connectivity: that is, availability, adoption, affordability, reliability, and speed of connection to the internet. We leave concerns about digital literacy and online learning, work, or health opportunities to other evaluations, as connectivity was explicitly mentioned in the federal funding guidance we received. However, within those possible indicators, we are limited to those that are measurable with publicly available data that is holistic and comparable across space. Neither affordability nor reliability is available publicly. Availability is a binary indicator that records speeds of 0 or greater than 0, and is therefore a more coarse measure than speed.

In this project, we use two prioritization indicators: speed (download/upload speeds) or adoption (having an Internet subscription at home).[4] We have selected these two indicators because they identify distinct yet overlapping necessary conditions for Internet access.

- **Internet speed:**[5] Internet at a usable speed is a necessary but insufficient condition of broadband access and use. Speed is determined by the physical infrastructure that facilitates fast speed– the cables, wires, servers, etc. that provide households with the capability of broadband connections.[6] It may be especially expensive to connect rural areas to broadband infrastructure, where homes are more spread out and difficult to reach, particularly for last mile connectivity that brings central connections to individual households.[7] Speed can be defined as actual (observed) speeds within a household, or the speed capability (advertised) of the physical connections. Speed is measured along two dimensions, using both upload speed and download speed, as both are required to access the Internet and stream videos. Actual speed may diverge from speed capabilities[8] for many reasons, including: household willingness to pay for full Internet speed, distance from the physical infrastructure, and congestion or throttling based on network use by those around you. We can consider advertised speed to be the upper bound of Internet use in a community. If an area lacks physical infrastructure that is even capable of a given speed, then other connectivity metrics are less relevant.
- **Adoption:** Physical infrastructure is a necessary but insufficient condition for adoption. However, overlooking Internet adoption rates can lead to an incomplete picture of Internet access, digital equity, and the digital divide across the U.S. We measure adoption as having an Internet subscription. Only a subset of people who live near the physical

---

[4] There are other dimensions of Internet need that we could include in our definitions: disruption in service, for example. We focus on these two for parsimony based on data availability (which is, ultimately, a constraint faced by policymakers deciding on the allocation of real-world broadband projects).
[5] Note that within speed, we are including the extreme example of "slow internet" as no speed. So, having no Internet at all would be considered a download and upload speed of 0.
[6] https://www.pewtrusts.org/en/research-and-analysis/fact-sheets/2022/07/how-broadband-infrastructure-gets-built
[7] https://www.pewtrusts.org/en/research-and-analysis/articles/2022/03/29/states-considering-range-of-options-to-bring-broadband-to-rural-america
[8] https://www.fcc.gov/reports-research/reports/measuring-broadband-america/measuring-fixed-broadband-tenth-report

infrastructure actually connect to that infrastructure. Decisions to not subscribe to the Internet can emerge from cost considerations, to comfort and familiarity with the Internet, to beliefs about the quality of the connection to which you are subscribing.

Throughout this project, we make the simplifying assumption that everyone needs an Internet connection, and that moving from having no Internet connection at home to having one is welfare-enhancing for everyone. Therefore, anyone lacking Internet (in terms of speed or adoption) can be prioritized.

How we're thinking about equity

This project focuses on three underserved, equity-relevant groups (ERGs): people of color (POC); people living with disabilities (PWD); people living in rural areas (RURAL). There are many other groups who have faced structural disadvantages that we could have focused on and who are listed in various executive orders and in legislation. We emphasize that it is out of scope of this project to recommend the prioritization of particular groups, and that our choice of equity-relevant groups for our analysis is descriptive, not normative.

Specifically, we focus on these three groups because we expect that they will most clearly illustrate the tradeoffs involved in the program design decisions around broadband allocation. Specifically, their geographic distribution is different in ways that we expect will matter when choosing a definition of need. On the one hand, POC and Rural populations are both clustered in specific areas of the country, but these tend to not be the same areas. This means that better equity outcomes for one group might inadvertently lead to deprioritization of the other. On the other hand, PWD tend not to be geographically concentrated, potentially making it difficult to target such populations with spatially-concentrated broadband programs.

Intersectionality of these groups is also important to consider: for example, just because the program is equitably distributed to POC and rural populations, does not mean that it is equitably distributed to POC in rural areas. To demonstrate the implication of membership in multiple underserved groups, we look at two multi-group definitions of equity: the intersection, and the union of all three of our selected underserved groups. Later in this analysis plan, we explicitly outline how we define and measure the size of each underserved group, the intersection of all three groups, and the union of all three groups.

The second component in our definition of equity we refer to as the equitable distribution goal. Definitions of equity can consider only current (and future) treatment, or can also consider past treatment. For example, EO 13985 states programs should "redress inequities" and "not [perpetuate] systemic barriers to opportunities and benefits." This suggests a much broader measure of equity: one that (somehow) weighs historical experiences as well as current program treatment.[9]

We focus on a specific distributional benchmark that a broadband program can reach. Identifying this benchmark – which is based on population distributions of underserved groups – allows us to compare the share of all underserved individuals who live in a Census tract that received a

---

[9] This distinction is commonly thought of as the difference between equality and equity. In OES's Equity Evaluation Series, they write that "Equality refers to uniformity in the type of input (everyone receives the same treatment), while equity refers to uniformity in outcomes, especially between groups (everyone receives the amount of treatment needed to obtain a desired outcome).

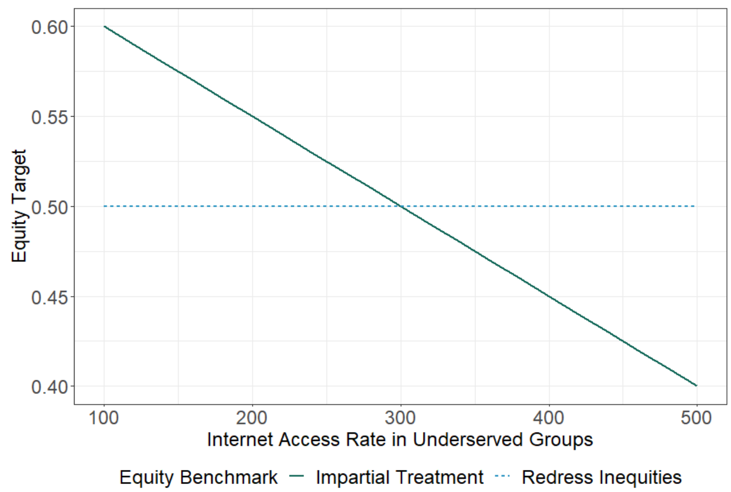
simulated broadband program to a benchmark of what equitable allocation would look like. To create these benchmarks, we consider two specific definitions of equity that are explicitly described in the EO 13985.

1. **Goal 1, impartial treatment:**[10] This equity benchmark holds that a broadband program is equitable to a given ERG if the share of benefiting households who belong to that ERG is at least as large as the share of that ERG in the population as a whole. For example, if 20% of households in the population are rural, then a broadband program is considered equitable to rural households under the proportional distribution benchmark if at least 20% of the households who get a program in their Census tract are rural. This equity benchmark is based only on the size of the underserved group in the population, and stipulates that new connections are proportional to the population distribution of groups.
2. **Goal 2, redress inequities:**[11] This equity benchmark holds that a broadband program is equitable to a given ERG if the share of benefiting households who belong to that ERG is at least as large as the minimally sufficient share that would close a pre-existing gap in internet availability between ERG members and non-members in the population. For example, suppose that the proportion of rural households living in a Census tract with inadequate speed was 10 percentage points higher than the proportion of non-rural households living in tracts with inadequate speed. Then there is a 10 percentage point gap between rural and non-rural households. If it were necessary to provide at least 50% of all programs to rural Census tracts and the remainder of programs to non-rural Census tracts in order to close the 10 percentage point gap, then the gap-closing benchmark would be 50%. This equity benchmark is based on both the size of the underserved group in the population and any disparities in access or uptake that existed before the program was implemented. This benchmark stipulates that the sum of new and pre-existing connections is proportional to the population distribution of groups.

Later in the analysis plan, we derive the formula to calculate the benchmark for redressing inequities. In brief, the benchmark for redressing inequities is derived from making the ratio of new and existing connections for underserved groups to that of non-underserved groups equal to the population ratio of underserved groups to non-underserved groups. Figure 1 visualizes the relationship between the two benchmarks, demonstrating that the “redressing inequities” benchmark changes as a function of existing internet access in underserved groups, while the “impartial treatment” benchmark remains constant as pre-existing access changes.

[10] Equity EO 13985, Section 2(a)
[11] Equity EO 13985, Section 1, paragraph 2

Figure 1. Relationship between two equity benchmarks as the pre-existing access rates change



Project definitions and terminology

This analysis presents a conceptual and measurement framework to evaluate equity performance in simulated broadband program allocations Throughout this analysis, we will employ the following concepts to refer to different aspects of the equity evaluation framework and methodological approach.

- Prioritization indicator: Which measure is used to determine eligibility for broadband program allocation. This determines which tracts are eligible and prioritized for upgraded connections. In this analysis, the two indicators are **speed** and **adoption**.
- Allocation scheme: A particular allocation of broadband programs, based on our computer simulations. In this project, the options for allocation schemes are **speed-based** and **adoption-based**.
    - Infrastructure upgrades are allocated at the level of the Census tract. Each household (and therefore individual) in the tracts selected for an upgrade within a given allocation scheme receives a new and/or upgraded **broadband connection**.
- Underserved/equity-relevant groups (ERGs): The priority groups for evaluating equity in this project, taken from the list of underserved communities in the Executive Order on Advancing Racial Equity and Support for Underserved Communities. The three equity-relevant groups that we study in this project are listed above.
- Equity distribution goals: What does equity actually look like? This is the equity goal that an allocation scheme hopes to accomplish. This could include distributing broadband programs so that each equity-relevant group is not underrepresented in the share of recipients relative to their population, or closing a pre-existing gap in access or priority outcomes between an equity-relevant group and the population as a whole. The two equity distribution goals that we study in this project are **based on the goal of impartial retreatment** or **redressing inequities.**

- Equity benchmark: This is a number that represents the share of the allocation scheme each equity-relevant group would need to receive to meet the scheme's equity goals. To evaluate equity performance, we will compare simulated allocation schemes to equity benchmarks. Each benchmark is based on the equity distribution goal.
- Allocation scheme outcomes: This measures how the simulated program based on a given prioritization indicator was distributed among the population with respect to the underserved groups.
- Equity performance: This is an evaluation on a scale from -1 to 1 regarding how well a given program allocation met an equity benchmark.[12] In this project, equity performance can either be based on a specific underserved group, or can be summed over all relevant underserved groups for a given allocation (see details below). There will be separate evaluations of equity performance for each equity distribution goal.

We summarize the selected categories of these definitions in the table below.

Definition	Selected categories
Prioritization indicator	Speed Adoption
Program allocation	Speed Adoption
Underserved groups (ERGs)	People of color People living in rural areas People with disabilities Union of all groups Intersection of all groups
Equity benchmark	Impartial treatment Redressing inequities
Equity performance	ERG-specific: Assessment of equity performance against an equity benchmark of an allocation scheme *for a given ERG*

Project analytical approach

Our analysis will consist of the following steps:

1. Select prioritization indicators based on two salient barriers to access: availability of Internet with usable speeds and/or lack of adoption of internet at home.
2. Measure prioritization empirically using publicly available, granular data at the Census tract level on broadband speed capabilities, and rate of adoption of broadband services.

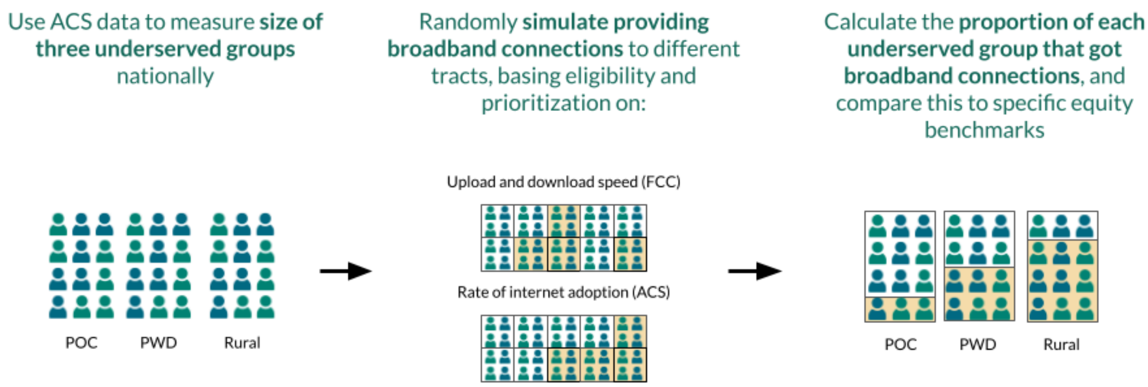
[12] We think of equity as a continuum, where the continuum is the maximum size of the difference in two proportions. A negative value indicates that the program was inequitable with respect to a given group and given equity benchmark.

- Describe similarities and dissimilarities of which areas would be prioritized based on each of the indicators.

3. Simulate broadband program allocation using random computer simulations designed to mimic allocations of broadband programs based on prioritization indicators.
4. Define and measure equity, by specifying two components of equity: equity-relevant groups (ERGs) and equity distribution goals.
5. Estimate an equity benchmark that aligns with each definition empirically using publicly available data on population shares and current levels of broadband access.
6. Measure allocation scheme outcomes relative to the equity benchmark associated with each definition of equity.
7. Evaluate the equity performance of the allocation scheme and compare relative equity performance across allocation schemes and groups, given different definitions of equity.

Figure 2 summarizes our approach.

Figure 2. Summary of simulation procedure and analysis steps



This approach will allow us to make the following comparisons:

1. Compare observed allocation for an ERG to a given benchmark, providing us with an indicator of equity performance.
2. Compare equity performance between simulated allocations based on two different prioritization indicators.
3. Examine the trade-off between equity for one ERG and equity for another ERG.

Preregistration details

This Analysis Plan will be posted on the OES website at oes.gsa.gov.

Hypotheses

Our hypotheses pertain to differences in equity performance generated by the two competing equity benchmarks.

We hypothesize that:

1. **[H1: Impartial Treatment Equity Performance]** The share of simulated program allocations awarded to an ERG will differ from the impartial treatment benchmark.
2. **[H2: Redressing Inequities Equity Performance]** The share of simulated program allocations awarded to an ERG will differ from the redressing inequities benchmark.
3. **[H3: Comparing ERG-Specific Equity Performance across Allocation Schemes]** For a given benchmark, the equity performance for a given ERG will differ between the speed-based simulated program allocations and the adoption-based simulated program allocations.

Data and data structure

This section describes the data sources we plan to use to construct variables to be analyzed, as well as changes to be made to the raw data with respect to data structure and variables.

Data source(s):

We will use data on current broadband infrastructure and connectivity, and local demographics, combining three publicly-available datasets to create our final dataset, which will be at the level of Census tract. All data for this project are publicly available. Form 477 data on broadband speed will be downloaded from the FCC Form 477 website, using the data dated December 31, 2021. Decennial Census and American Community Survey data will be downloaded from the Census API via the tidycensus package in R. The tidycensus package allows users to pull estimates at specified geographic levels (e.g., the Census tract mean, median, or total, with associated margin of error – MOE). We will download all data at the level of Census tracts.[13]

We use Form 477 data from 2021 and ACS data from 2015-2019. We use these data because they predate large federal investment in broadband infrastructure. It is important to capture broadband access levels before the current round of federal investment because these federal programs had an equity focus, so the relationship between broadband access and geographic concentration of equity-relevant groups may differ during this time period. Using this historical time period also allows us to mimic the protocol that agencies may take in identifying areas to prioritize for infrastructure programs.

Our datasets are:

[13] For privacy reasons, the ACS limits the individual-level data available through their Public Use Microdata Sample (PUMS). These data are geocoded to the level of "Public Use Microdata Areas (PUMAs)", which are non-overlapping areas that partition the state into areas of 100,000 residents. This limitation means that we cannot rely on PUMS as our data source.

1. FCC Form 477 ("Form 477"): All Internet Service Providers (ISPs) are required to complete a Form 477 twice a year, at the level of 2010 Census blocks. ISPs must fill out this form for each place where they offer Internet service access (defined as offering access faster than 200 kbps in at least one direction). Speeds lower than 200 kbps are not considered Internet access, and providers do not have to fill out the Form 477 for those areas. Within the areas where the provider does offer services, they are required to report the fastest advertised downstream and upstream speed offered to any house in the area,[14] in addition to these other fields. The FCC releases the Form 477 data every six months at the block level. We will use the Form 477 data dated from December 31, 2019.
2. American Community Survey 5-year estimates ("ACS"): We will use the 5-year tract level estimates spanning 2015 - 2019. This will be at the level of 2010 Census tracts.
3. Rural-Urban Commuting Area ("RUCA")[15] data. To identify urban versus rural areas, we will use the Primary RUCA Code 2010[16] variable in the Rural-Urban Commuting Area (RUCA) dataset provided by the U.S. Department of Agriculture. These data are provided at the Census tract level. RUCA code values of 1, 2, 4, and 5 will be coded as RURAL=0 and RUCA code values of 3, 6, 7, 8, 9, and 10 will be coded as RURAL=1.

Outcomes to be analyzed:

The outcomes central to this analysis are: (1) the share of connections allocated to ERGs; (2a) the difference between the share of connections allocated to ERGs and the equity benchmark; and (2b) the aggregate equity performance.

Variable name	Variable	Description
$erg\_share_{ERG, allocation\_scheme}$	Share of total program allocation allocated to each ERG	Continuous variable that ranges from 0 to 1 and represents the proportion of all simulated broadband connections that were allocated to ERGs. There will be a separate variable for each ERG and program allocation scheme.
$group\_perform_{ERG, allocation\_scheme,equity\_goal}$	Equity performance at the ERG level	Continuous variable that ranges from -1 to 1 and represents the difference between the share of total program allocations to each ERG and the equity benchmark for that ERG and equity goal.

[14] The FCC caveats this data by writing: "A provider that reports deployment of a particular technology and bandwidth in a particular census block may not necessarily offer that particular service everywhere in the census block. Accordingly, a list of providers deployed in a census block does not necessarily reflect the number of choices available to any particular household or business location in that block, and the number of such providers in the census block does not purport to measure competition."
[15] See this resource for more on characterizing geographies as urban/rural.
[16] The 2020 updated RUCA data are not yet available. While there will be measurement error in using the 2010 data, we have no reason to believe this error would systematically bias our findings.

Additional variables and transformations:

We outline additional variables and associated transformations in Appendix A1.

The raw Form 477 data is at the provider-census block level. To merge the datasets, we first need to transform the raw Form 477 data into a dataset with only one observation per block by aggregating over providers . When we first download the data, each census block will contain a row for all the ISPs offering Internet within its borders. We then merge the Form 477 data with a master list of Census blocks in the US census, downloaded via tidycensus. For blocks that were missing from the Form 477 data but appear in the master list of Census blocks, we code their speed as 0 missingness is indicative of having no ISPs. For most blocks, the number of ISPs is greater than one, yielding multiple observations per block. We will transform this data and merge it with the ACS 5-year estimates through the following steps:

- Collapse the Form 477 data to have one row per tract and retain / create the following variables as described below. We exclude entries with a TechCode value of "60" (denoting a Satellite connection).[17]
- We then merge the "ACS" dataset with the "Form 477" dataset. We do this with a left_join on Tract. We do not expect there to be any Census tracts that appear in the Form 477 data but not the ACS data. This will produce the "ACS & 477" dataset.
- Finally, we merge the "RUCA" dataset with the "ACS & 477" dataset. We do this with a full_join on Tract. We do not expect there to be any Census tracts that appear in the ACS & 477 data but not the RUCA data. This will produce the "Merged" dataset.
- We set a threshold for being prioritized for a broadband program that identifies tracts with speeds under 100/20 or adoption rates under 65%. Each tract with a speed under this threshold will get a 1, with 0 otherwise. Each tract with an adoption rate under this threshold will get a 1, with 0 otherwise.

The **Merged** dataset will include all Census tracts in the 50 US states, the District of Columbia and Puerto Rico. The dataset will be at the Census tracts level and each Census tracts will only be observed once, yielding a dataset of an expected 86,020 rows (Census tracts = or equivalents in Tribal areas, Puerto Rico areas).[18]

Data exclusion:

In the case of missing prioritization indicators or equity variables across territories, we will only include territories that have complete data. Because we are looking to compare equity evaluations across different measurement strategies, we will restrict our final sample in the **Merged** dataset to include only complete observations (i.e., Census tracts with no missing data on the variables included in our analysis), so that we do not conflate differences in outcomes with sample bias.

We also exclude tracts with 0 population in the American Community Survey.

[17] See the FCC description of technology codes for fixed broadband deployment data here.

[18] The Census reports the following tallies for 2020: 84,414 Census tracts in the United States, 981 in Puerto Rico, 133 in the Island Areas, and 492 Tribal Census tracts.

Treatment of missing data:

Prior to processing of the ACS data by US agencies, we expect both item nonresponse and unit nonresponse. Item nonresponse may come in the form of individuals who do not fill out all the information requested by the ACS, or ISPs who do not return Form 477 in some areas where they operate. Unit nonresponse may come in the form of individuals who do not respond to the ACS questionnaire at all, or ISPs (likely smaller ISPs) who do not complete the Form 477 data at all. The ACS takes a number of nonresponse adjustment procedures (including raking) to account for item- and unit-level missingness.

In the processing of the ACS data and aggregating it to the tract level, certain variable values will be omitted due to deductive disclosure avoidance (i.e., a particular combination of traits can be de-anonymizing when that combination is rare). As this is more common in sparsely populated and/or homogenous areas, this could result in systematic missing data in the **Merged** dataset (and constituent datasets). We will follow the approach described in OES 2305 to predict the missing values from the non-missing values using a Random Forest approach.

Descriptive statistics, tables, & graphs

The first set of descriptive statistics in the project abstract will summarize the spatial concentration of broadband need, the number and proportion of ERGs targeted under each program allocation scheme. We will report the following quantities, along with their standard errors (calculated using the procedure outlined below):

- We will assign each tract a vector of prioritization variables, taking a 1 if a tract was prioritized under a prioritization indicator and a 0 otherwise. We will report the total number and proportion of tracts, and number and proportion of households that:
    - Are not prioritized under either indicator (receive a 0 for each prioritization variable)
    - Are prioritized under speed only (receive a 1 under speed and a 0 under adoption)
    - Are prioritized under adoption only (receive a 0 under speed and a 1 under adoption)
    - Are prioritized under both (receive a 1 under speed and a 1 under adoption)
- For each program allocation, the average (across simulations, as described below) of the total number and proportion of tracts and total number of households who receive connections

The second set of descriptive statistics in the program abstract will summarize the total number of people who receive upgraded connections under the speed-based program allocation and under the adoption-based program allocation in each of the following ERG groups:

- People of color
- People living in rural communities
- People living with disabilities

The third set of descriptive statistics in the project abstract will report the equity benchmarks and equity performance for each program allocation and definition of equity.

The fourth set of descriptive statistics in the project abstract will display the results of statistical tests that test our hypotheses regarding how equity performance differs across equity-relevant groups and program allocation schemes. The goal of this project is to illustrate the implications for different prioritization indicators and equity definitions in evaluating equity performance.

Statistical models & hypothesis tests

This section describes the statistical models and hypothesis tests that will make up the analysis.

Statistical analysis:

Simulating program spending

After cleaning the data to identify tracts eligible to receive upgraded broadband connections under the speed-based and adoption-based allocation schemes, we then allocate upgraded broadband connections to Census tracts via a simulation. In writing this simulation, we want to capture, as accurately as possible, the policy environment in which program offices make their allocation decisions. To this end, we outline three possible allocation policies.[19] **Note that all allocation policies still only allocate funding to program-eligible tracts**: those under the 100/20 speed and 65% adoption threshold we define above. In other words, **ineligible tracts receive no connections** (as such, individuals in ineligible tracts receive no upgraded Internet infrastructure). The key difference between the allocation policies is whether and how they prioritize *among eligible Census tracts* on the basis of the two prioritization indicators. These are simplified versions of a wide array of different approaches that we think capture common program design decisions.

1. **Prioritize high-need**: This is an allocation policy where programs decide to allocate projects to the neediest (in terms of broadband) tracts. Specifically, among eligible tracts, those with above-median need receive two-thirds of all connections, and those below-median need receive one-third of all connections. Within these above- and below-median strata, the program is provided to tracts at random.
2. **Need agnostic**: This is an allocation policy where programs do not allocate based on prioritization indicators. Eligible tracts have the same probability of receiving connections, irrespective of whether they are above- or below-median prioritization indicators.
3. **Low-hanging fruit**: This is an allocation policy where programs decide to first allocate connections to the "low-hanging fruit", which we conceptualize as providing two-thirds of the connections to eligible but below-median need tracts, and one-third of connections to the above-median need eligible tracts. We conceptualize this as the "low-hanging fruit" approach because we assume that places that are relatively better-off (within the "need" dichotomy) will be easier to connect.

[19] These allocation policies are necessary simplifications. In the broadband programs we seek to emulate, ISPs play an important role in compiling project areas and applying for funding from the state broadband offices. Cost, ease of building infrastructure, and other factors like proximity to existing infrastructure will influence which areas are even included in grant applications in the first place. However, we lack the data to properly mirror this allocation strategy. Moreover, we focus on stages of the allocation process where policymakers could utilize our evidence. The three allocation policies we propose could be built into program guidelines or adopted by state granting offices in the absence of clear program guidance on how to prioritize application between eligible project areas.

For this project, we treat the "prioritize high-need" allocation policy as our workhorse model. We discuss analyses of simulated program funds based on the other two allocation policies in the section on exploratory analyses. We discuss development of the simulation approaches associated with each allocation policy in Appendix A5.

To mimic the allocation policy environment, we also want to constrain the program budget. To do this, we set a total number of households that will be funded.[20] We hold this number constant for the allocated funding for both need bases.

1. Classify eligible tracts as high-need or low-need based on whether they fall above or below the median value of the indicator among the eligible. We do this by assigning each tract to a need strata:
2. Set the budget allocation for the given allocation policy (in this case, high-need prioritization) for each strata. We set a budget allocation of reaching 1 million households. Budget allocation is a scalar variable indicating the number of households that will be connected, and prioritization is a vector that indicates whether a tract is prioritized under the given prioritization scheme.
3. Assign program receipt probabilistically on the basis of need strata, prioritization level, and budget constraints. We do this by first randomly ordering tracts within a need strata. We then award a program to tracts in each strata until we have run out of programs to award based on the number of households in that tract. Note that the randomness in this process is achieved by randomly assigning each tract a number drawn from a uniform distribution and then ordering the tracts within a strata by row.[21]
4. Repeat the simulation for each of the two prioritization indicators, generating two versions of the binary variable that indicates if a tract received a project.

We run this simulation M = 5,000 times. This gives 5,000 different answers to questions such as: "What proportion of those programs went to rural Census tracts?" As we describe below, we need to take this variation in the simulations into account when deriving the point estimates and their standard errors. We discuss how we are aggregating across simulations in the section on Measuring Program Allocation Outcomes.

Key assumptions in broadband program allocation simulation

This approach relies on several assumptions that simplify our analysis and facilitate generating clear conclusions that will be applicable to a variety of infrastructure programs with different allocation methodologies. We outline these assumptions below. Where relevant, we describe possible alternative approaches and justify why we prefer the method outlined in this analysis plan.

Ignoring within-tract variation in need and program targeting: In our simulation, allocation correlates with the prioritization indicators at the tract level and cannot account for within-tract

[20] We conceptualize budget in terms of the number of upgraded connections. To make our analysis more tractable, we assume that each upgraded connection costs the same amount of money.

[21] Randomness appears in two stages of this process: The randomness of assigning each tract a number drawn from the uniform distribution, *and* randomness of repeating the allocation *M* times (discussed at the end of this section).

variation in need. This does limit the analysis. If we think about two communities each with ten households, for example, we could have two very different distributions of need with similar average speeds: One tract where two households have no Internet connection at all (0 out of 10) and eight households that have very fast Internet (10/10). This tract would have the same average speed (8) as a community where four households have a speed of 5/10 and six households have speed of 10/10. While we would think about the need levels in these two tracts very differently, our approach cannot distinguish between them, as we lack the individual-level data needed to identify these individual-level distributions. However, we believe that the same or a similar constraint is faced by many broadband program designers, so in that respect our approach does incorporate a real limitation on program design.

Relatedly, our simulation strategy then assumes that the simulated program allocation would benefit each household in the Census tract equally. This simplification is necessary for our analysis as individual level data (such as the IPUMS ACS data) does not identify individuals' tracts.

Funding is conditional on eligibility but correlated with prioritization indicators: We allocate programs to tracts – among eligible tracts – at a higher probability in tracts with higher values of the prioritization indicator. Available program documentation for federal broadband programs does not prescribe that among eligible tracts, tracts in greater need should be prioritized in program allocation. The logic of how to prioritize eligible tracts for funding appears to vary between states and programs, and many of the considerations in prioritizing some program areas over others will not be available in our data.[22]

Calculating equity benchmarks

See Appendix A4 for a summary of the estimators we plan on using for these calculations.

Our approach to assessing equity relies on the use of equity benchmarks - target levels of allocation that imply meeting an equity distribution goal. Specifically, we calculate equity benchmarks that represent the share of upgraded Internet connections each equity-relevant group would need to receive to meet an equity distribution goal.

Calculating the "Impartial Treatment" benchmark

Our impartial treatment benchmark is based on the national share of individuals belonging to the ERG. For the impartial treatment benchmark that considers the union of ERGs, we use tract-level ACS data to estimate the proportion of the population that falls into the POC and PWD categories at the tract level before examining whether the tract is rural or non-rural, and subtract the intersection. To estimate the impartial treatment benchmark for the rural population, $\widehat{ergshare}_{rural}$, we sum the total number of individuals in all tracts that are coded as rural in the RUCA data, and divide this by the total number of individuals in the United States. To estimate the impartial treatment benchmark for the POC and PWD, $\widehat{ergshare}_{POC}$ and $\widehat{ergshare}_{PWD}$, we sum

[22] These include project feasibility and difficulty (which is, in part, determined by geographic features), project potential return-on-investment and cost effectiveness, among other factors. In previous scoping work, the OES team had conversations with some states that prioritized connecting the low-hanging fruit with early grant funding and more difficult-to-reach and unserved locations with subsequent funding. Other states prioritized the amount of funding that ISPs were able to provide themselves in the applications.

the tract-level variables of total population of POC (total_nonwhite) and PWD (total_disability) across all tracts and divide this by the total number of people in the United States in each case.

To estimate the impartial treatment benchmark when considering the intersection of ERGs, we will use the ACS data to sum the tract-level population of POC who are also PWD (total_nonwhite_disability) living in tracts coded as rural in the RUCA data, and divide this by the total number of individuals in the United States.

To calculate the standard errors for these point estimates, we rely on the ACS guidance for approximating standard errors. We do this as follows:

1. Obtain standard error on tract-level total number of households: divide Census-produced margin of error (MOE) by 1.645 for all tracts.
2. Obtain standard error on total number of households and number of households in ERG by pooling all standard errors created in step 1: specifically take the square root of the sum of the tract-level squared standard errors.
3. Obtain standard error on the share of the population who in the ERG, indexed by e:

$$\widehat{SE}(\widehat{ergshare}_e) = \frac{1}{\widehat{total\_pop}} \sqrt{([SE(\widehat{total\_erg}_e)]^2 - \frac{\widehat{total\_erg}_e^2}{\widehat{total\_pop}^2}[SE(\widehat{total\_pop})]^2}$$

Calculating the "Redressing Inequities" benchmark

To calculate the redressing inequities benchmark, we focus on pre-existing inequities based on broadband adoption only, and use this as a benchmark for all analyses. Specifically, we use IPUMS data to look at whether a household has a broadband connection at home. We also record whether that person lives in a rural area, is a person of color, or is a person with disabilities. When an allocation scheme results in a broadband program being implemented in a given tract, we assume that everyone in that tract now has Internet. This is of course a simplifying assumption: in practice, broadband programs may not improve speed for everyone in a tract, or result in more affordable internet. However, this simplifying assumption allows us to clearly identify gaps on a coarse level and how broadband programs close them.

To calculate the redressing inequities benchmarks, we focus on pre-existing inequities based on adequate speed only, and use this as the benchmark across simulations and ERGs. Specifically, we consider any household in a tract that has average upload speed of less than 20 mbps and download speed of less than 100 mbps as living in a tract without access to adequate internet, and any household who lives in a tract with speeds above this as living in a tract with adequate access to the internet. When an allocation scheme results in a broadband program being implemented in a given tract, we assume that everyone in that tract now has access to adequate speed. This is of course a simplifying assumption: in practice, broadband programs may not improve speed for everyone in a tract, or result in more affordable internet. However, this simplifying assumption allows us to clearly identify gaps on a coarse level and how broadband programs close them.

We need to estimate many more quantities for this goal, so use the following notation:

- a = the number of ERG households who already live in a tract with adequate speed. We estimate this by summing the tract-level total number of ERG households across all tracts that have adequate speed.
- A = the total number of ERG households. We calculate this summing tract-level counts as above for the proportional goal.
- b = the number of non-ERG households who already live in a tract with adequate speed. We estimate this as above, but focusing on the non-ERG households.
- B = the total number of non-ERG households. Estimated as above.
- q_A = the number of new connections for ERG households. This is one of the things we are trying to calculate using the formulas below, a variable defined by the other measured quantities.
- q_B = the number of new connections for non-ERG households. Similarly, this is a variable we are trying to calculate, defined by the other measured quantities.
- Q = q_A + q_B = The total number of new connections. This is a budget constraint given in the simulations: 1,000,000 households.

We can then define the pre-existing gap we want to close in order to redress inequities:

- b/B - a/A

Specifically, we close this gap by finding the number of ERG households who would need to be living in tracts that got a broadband upgrade, q*_A, so as to ensure the ex-post rates of access (sum of new and existing connections) would be proportional to their population share. Closing the gap means:

- (a + q*_A) / A = (b + q*_B) / B

Using some algebra not included here and the constraint that Q = q_A + q_B, which implies q_B = Q - q*_A, we can define q*_A in terms of fixed quantities as follows:

- q*_A = (Ab - Ba + AQ) / (A + B)

which also gives q*_B:

- q*_B = Q - q*_A

Once we have q*_A, we can calculate the target share of the funding that the ERG should receive in order to close the gap: $\widehat{erggap_e} = q_A^{*}/Q = ((Ab - Ba + AQ)/(A + B))/Q$.

Estimating the standard error for this quantity, $\widehat{SE}(\widehat{erggap_e})$ proceeds similarly to the above, but requires some substitution. For example, we use the product rules in the ACS guidance to derive separate point estimates and SEs for Ab, Ba, and AQ. We then substitute those into the formula for the SE of a sum, to get the SEs for Ab-Ba+AQ and A + B. We substitute those into the formula for a ratio to get a standard error for (Ab - Ba + AQ)/(A+B). And finally, we use the ratio formula once more to get the standard error for $\widehat{SE}(\widehat{erggap}_e)$. We devised this approach based on guidance to approximating standard errors in the American Community Survey.

Note that this derivation also implies that the ratio of ex-post broadband connections of ERGs to that of non-ERGs should be equal to the ratio of ERGs in the population to non-ERGs.

Measuring program allocation outcomes

We now define a procedure for estimating M answers to the question, “How did equity-relevant groups fare under a given simulation of funding?” where M is the number of simulations. We also define a procedure for summarizing over the M simulations to arrive at a single answer to the question, “How did equity-relevant groups fare under a given allocation scheme?” Finally, we define a procedure for estimating the variance in our answers that arises both from the simulations and from the sampling variance in the Census data sources, to answer the question, “How confident are we in these answers?”

The ACS variables total_nonwhite, total_disability, and rural estimate, with some margin of error, the number of people of color in the Census tract, the number of people with disabilities in the Census tract, and whether or not the Census tract is rural. To calculate the number of rural people in the Census tract, the rural binary indicator will be multiplied by total_pop to generate the variable total_pop_rural.

For each simulation

In each of the simulations, which tracts get simulated broadband upgrades will differ. For each of the simulations, we estimate the number and share of individuals who are members of ERGs and live in tracts that received infrastructure upgrades, as well as the standard errors for these numbers and proportions that derive from the sampling variability in the ACS. For each of the M simulations, we will calculate the variables outlined in the Variables table and repeat the calculations outlined in Appendix A4.

Aggregating across simulations

The steps above produce, for each ERG and allocation method, M point estimates and standard errors for each ERG and allocation method. We follow Rubin’s Rules for accounting for multiple sources of variation when conducting imputation:

To get a single point estimate, we take an equally-weighted average of the M point estimates:

$$\overline{prop\_\widehat{ERI\_funded}_{se}} = \frac{1}{M}\sum_{m=1}^{M} prop\_\widehat{ERI\_funded}_{sem}$$

To calculate a single standard error, we first calculate the within-simulation variance, which is generated by ACS sampling variability:

$$W_{se} = \frac{1}{M}\sum_{m=1}^{M} \widehat{SE}_{sem}(prop\_\widehat{ERI\_funded}_{sem})^2$$

We then calculate the between-simulation variance, which is generated by the randomness in the simulation procedure:

$$B_{se} = \frac{\sum_{m=1}^{M} (\widehat{prop\_ERI\_funded}_{sem} - \overline{\widehat{prop\_ERI\_funded}_{se}})}{M-1}$$

This gives us the pooled standard error, which takes account of both the simulation and sampling variation:

$$\widehat{SE}(\overline{\widehat{prop\_ERI\_funded}_{se}}) = \sqrt{W_{se} + B_{se} + \frac{B_{se}}{M}}$$

Evaluating equity performance of the allocation scheme

From here, we can calculate equity performance by comparing the percentage of broadband investment that went to each ERG to the equity benchmark for that ERG, across schemes, and across benchmarks. To do this, we take a simple difference between the allocation scheme outcomes and the equity benchmark. For the aggregate equity performance, we sum up the equity performances across the individual ERGs.[23]

Confirmatory analyses:

This suggests several confirmatory tests. The following table summarizes the research questions and associated confirmatory hypothesis tests that we intend to run. The project abstract will highlight a subset of the findings from these tests. The project team will select tests to feature in the abstract to convey valuable insights regarding the implications of the different definitional and measurement trade-offs that we discuss in this analysis plan.

[23] Note: For the aggregate equity measure, we are just calculating equity performance across the three ERGs in isolation, not the intersection and union groups.

Research question	Outcome	Possible hypothesis test	# of tests
Comparing observed allocation for an ERG to a given benchmark			
How equitable are speed-based and adoption-based allocation schemes for each ERG?	Difference between the equity benchmark and the average allocation outcome across *M* simulations for each ERG under each allocation scheme	Two-tailed t-test using the standard errors calculated as above to take account of both sampling and simulation variability	20 - 2 benchmarks - 5 groups - 2 allocations
Comparing equity performance between speed- and adoption-based allocations			
Is speed-based allocation more or less equitable than adoption-based allocation for each ERG?	Difference in equity performances across allocation schemes. Percentage point difference ranging from -2 to 2. Calculated by subtracting the equity performance of speed-based allocation (for a given group and benchmark) from that of adoption-based allocation (for the same group and benchmark).	Two-tailed t-test of the difference in differences, using the standard errors of the differences calculated in the first test.	10 tests - 5 groups - 2 benchmarks
Examining the trade-off between equity for one ERG and equity for another ERG			
Is equity performance for people of color statistically distinguishable from equity performance for rural populations?	Correlation between equity performances between POC and rural populations using the share of POC and rural populations receiving projects in each of *M* simulations as the dataset	1) Regress the proportion of connections allocated to POC on the proportion of allocations allocated to rural using observations from *M* simulations 2) K-S test comparing the distribution of proportions of new connections allocated to people-of-color across M simulations to the distribution of proportions of new connections allocated to rural across M simulations	2 - 2 allocations

Exploratory analysis:

<u>Changing the allocation strategy</u>

In this exploratory analysis, we extend our approach by simulating program allocation under the "need agnostic" and "low-hanging fruit" allocation policies. We repeat all the steps outlined above for each of these two allocation policies for a subset of equity relevant groups, equity benchmarks, and allocation schemes (e.g., look at equity under a speed-based allocation strategy and proportional distribution goal for the three ERGs alone), which are defined elsewhere in the analysis plan.

Slippage parameter

To complement our equity target measures, we incorporate a spillage parameter for each combination of simulated allocation and ERG. This slippage parameter allows for uncertainty in the probability that a given ERG e under spending s missed out on the broadband program in their tract. We would calculate this as follows:

As elsewhere in this document, s indexes spending allocation, e indexes equity-relevant group and j indexes tract.

This slippage parameter would allow us to show how equity outcomes would be affected if there was inequitable targeting within a census tract. This addresses one of the project's limitations that we are – in some ways – assuming a "best case scenario" of equitable allocation at the tract level by allowing all residents within the tract to benefit from the program if they are in a tract that is selected.

Inference criteria, including any adjustments for multiple comparisons:

We will conduct tests of differences in means or distributions between the program allocations and equity-relevant groups, as described above. We will reject the null hypothesis of no difference between the quantities of interest if $p < 0.05$.

We will adjust for multiple hypothesis testing within each family of comparisons. We characterize a family as being within the same research question in the table on confirmatory analyses. We will use simulation to find the test-wise alpha we would need to reach the family-wise alpha of 0.05.

Limitations:

We wish to highlight one key limitation of our approach; other limitations are discussed throughout the analysis plan: it assumes that the lack of Internet connectivity can be addressed by providing new and/or upgraded physical broadband connections. However, broadband connectivity is driven only partly by physical infrastructure, and factors such as affordability, perceptions of quality and reliability, and digital literacy all play important roles in expanding broadband access. Appropriate behavioral interventions to increase uptake of broadband, like the Affordable Connectivity Program (ACP) or programs to improve digital literacy, are all important parts of a strategy to increase broadband connectivity equitably. This project is unable to address the extent to which upgrading a physical broadband connection could successfully improve broadband usage and increase in-home broadband subscriptions.

Link to an analysis code/Script:

N/A

Appendix

A1. Additional variables and transformations

Variable name	Description	Transformations
American Community Survey variables[24]		
tract	Identifier for Census tract that is used to match across datasets	N/A
total_hh	Total number of households in the Census tract	N/A
total_pop	Total number of individuals in the Census tract	N/A
prop_hh_subscription	Continuous variable that measures the proportion of households in the of Census tract with an Internet Subscription[25]	Takes the number of households that respond "**Yes**" to the question about whether anyone in the household has a subscription to the Internet and divides by the total number of households in the Census tract.
adoption_need	Binary variable that indicates whether a Census tract proportion of households with an Internet subscription falls below the need threshold of 65% for adoption[26]	Transformed by taking the proportion of households with an Internet subscription in a tract and coding them as 1 if the average adoption rate is 65% or lower. Census tracts with 0 population are coded as NA.
total_nonwhite	Total number of individuals in the Census tract who report their race as something other than White only	Calculated from ACS variables B01001_001 and B18101H_001
total_disability	Total number of individuals in the Census tract who report that they live with a disability[27]	Calculated from ACS variables B18101_004, B18101_007, B18101_010, B18101_013, B18101_016, B18101_019, B18101_023, B18101_026, B18101_029, B18101_032, B18101_035, and B18101_038

[24] For each of these variables, we will also access the associated margins of error for the population estimates

[25] ACS asks respondents several questions about access to the Internet, including whether the household owns a device that can access the Internet (e.g., desktop or laptop, smartphone, or tablet), if they have access to the Internet, and how they access the Internet. Specifically, we will use the variable category: "Presence and Types of Internet Subscription in the Household"# and use the specific variable: "Estimated Total with an Internet Subscription" as our primary measure of broadband adoption.

[26] See Appendix A2 for a discussion of determining need thresholds.

[27] To identify those living with a disability, ACS asks a series of questions about each respondent's difficulty with critical actions and functions, specifically: difficulty hearing; difficulty seeing even when wearing glasses; concentrating, remembering, or making decisions; walking or climbing stairs; dressing or bathing; difficulty doing errands alone such as visiting a doctor's office or shopping. B18101 codes anyone facing at least one of these difficulties as living with a disability.

total_nonwhite_disability	Total number of individuals in the Census tract who report that they are nonwhite *and* live with a disability	Calculated from total_disability and ACS variables B18101H_003, B18101H_006, and B18101H_009
Variable name	**Description**	**Transformations**
FCC 477 Variables		
tract	Identifier for Census tract that is used to match across datasets	Transformed from the Block variable by preserving only the first 11 characters of the Census GEOID
max_adv_down	Maximum consumer-advertised Internet download speed in the tract	Transformed by taking the maximum download consumer-advertised Internet speed in any block in the tract
max_adv_up	Maximum consumer-advertised Internet upload speed in the tract	Transformed by taking the maximum upload consumer-advertised Internet speed in any block in the tract
mean_adv_down	Mean consumer-advertised Internet download speed in the tract	Transformed by taking the mean download consumer-advertised Internet speed across the blocks after imputing zeros for missing data
mean_adv_up	Mean consumer-advertised Internet upload speed in the tract	Transformed by taking the mean upload consumer-advertised Internet speed across the blocks after imputing zeros for missing data
mean_speed	Continuous variable that takes the average of mean_adv_down and mean_adv_up to have a single measure of average speed	Transformed by averaging mean_adv_down and mean_adv_up
mean_speed_reverse	Continuous variable that reverse codes mean_speed so that higher values indicate more need	Transformed by taking the maximum value of mean_speed across the US, adding 1 and substracting the tract-specific mean_speed (i.e., reverse coding the variable)
speed_need	Binary variable that indicates whether a Census tract average upload and download speeds fall below the speed threshold of 100 down and 20 up[28]	Transformed by taking the mean upload and download means (mean_adv_down and mean_adv_up) and coding them as 1 if the speed is less than the need cut-off
Variable name	**Description**	**Transformations**

[28] See Appendix A2 for a discussion of determining need thresholds.

Combined speed and adoption data		
need_type	Categorical variable that takes the values 0, 1, 2, 3 according to the type of need in a Census tract	Transformed from speed_need and adoption_need: • 0: Tract is not classified as in-need under either definition • 1: Tract is classified as in-need under the speed definition only • 2: Tract is classified as in-need under the adoption definition only • 3: Tract is classified as in-need under both definitions • NA: Tract has 0 population

Variable name	Description	Transformations
Rural-urban commuting area		
rural	Identifier for whether a Census tract is rural	Coded as 1 if RUCA value is 4-10 Coded as 0 if RUCA value is 1-3[29]
total_pop_rural	Total sum of rural residents	Equals total_pop if rural = 1 0 otherwise

Variable name	Description	Transformations
Simulation and simulated variables		
strata_speed	Identifier for whether a Census tract is in the top or bottom half of average speed	Transformed from mean_speed: • 1: Census tract has a mean_speed at or below the median of mean_speed • 0: Census tract has a mean_speed above the median of mean_speed • NA: mean_speed is missing
strata_adoption	Identifier for whether a Census tract was in the top or bottom half of adoption rates	Transformed from prop_hh_subscription • 1: Census tract has a prop_hh_subscription rate at or below the median of prop_hh_subscription • 0: Census tract has a prop_hh_subscription above the median of prop_hh_subscription • NA: prop_hh_subscription is missing
budget	Continuous variable for the total number of households that will be funded in the simulated program allocation	Total number represents the number of households, held constant between each allocation scenario
strata_allocation	Continuous variable that represents how many household connections will reach each strata	Transformed by taking the total number of households funded (budget) and multiplying it by the prioritization scheme

[29] Guidance from this Census guide on using the ACS to analyze rural/urban areas.

$funded\_adoption_{ERG}$	Continuous variable that represents the total number of tracts that received infrastructure upgrades under the adoption allocation for each ERG	Transformed by taking the total number of tracts that received an infrastructure upgrade under the simulated adoption allocation for each ERG.
$funded\_speed_{ERG}$	Continuous variable that represents the total number of tracts that received infrastructure upgrades under the speed allocation	Transformed by taking the total number of tracts that received an infrastructure upgrade under the simulated adoption allocation for each ERG.

A2. Summary of speed and adoption data

We plot the distribution of our Internet access variables – average maximum advertised Internet speed in a tract (both download and upload speeds) and the proportion of the tract who has an Internet subscription at home. In Table A1.1, we display summary statistics for each of these measures, as well as for our single composite measure of speed that averages download and upload speeds. For these scoping analyses, we used 2017-2021 ACS data and 2021 FCC data, but will use 2015-2019 ACS data and 2019 FCC data for the project analysis.

As expected, the speed variable does not contain any missingness, while the adoption variable contains a small amount of missingness (~ 1.3% of tracts), which accounts for the US census tracts with 0 population.

Table A1.1. Summary of speed and adoption data at the tract level (N = 85,395)

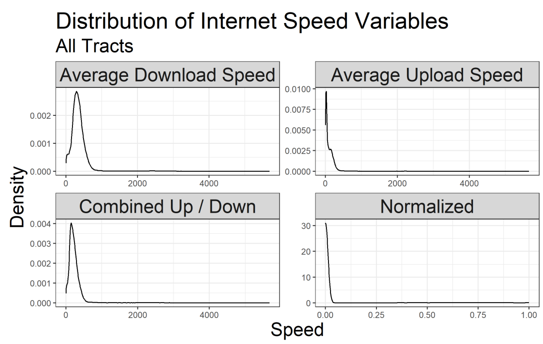
Variable	Minimum	Median	Maximum	Missing
Average, Download	0.0	322.5	5690.4	0
Average, Upload	0.0	69.94	5658.97	0
Average Combined	0.0	199.4	5674.7	0
% with Internet subscription	0.0	0.88	1.00	1123

We also show the distribution of the speed and adoption variables in Figure A1.1. Both Internet speed and adoption among all tracts is heavily right-skewed, particularly Internet speed, which has registered average maximum advertised download speeds exceeding 5,000 mbps. The variable of tract-level adoption rates is also skewed, as most tracts have a low proportion of residents without Internet subscriptions. However, when we look at the distribution of speed variables only among speed-eligible tracts, the degree of skewness is diminished – though still present.

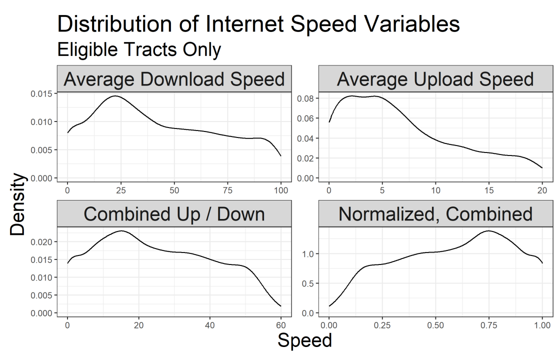
As we discuss in Appendix A2, our data generating process for simulating program allocation – which is correlated with need, needs to consider the different distributions of the two need variables, accounting for both the skewness and the differing variable ranges (0 to ~5700 for speed, and 0 to 1 for broadband adoption).

Figure A1.1. Distribution of Internet need variables, all tracts by speed (A) and speed in tracts, conditional on a 100/20 eligibility cut-off (B) and in all tracts by Internet subscription levels (C)

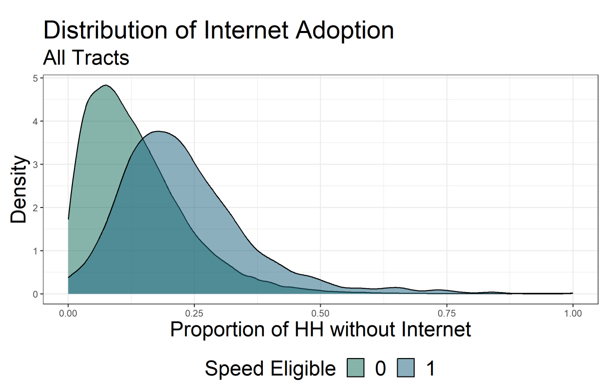
A.



B.



C.



A2. Selecting internet access thresholds for program eligibility

In designing our simulation, we sought to mimic the program guidelines for flagship broadband programs. These programs typically suggest a threshold of Internet speed of 100 / 20 or less to classify an area as underserved and 25 / 3 to classify an area as served. Taking that as our initial cut-off, we then try to find a threshold of Internet adoption for which a similar proportion of tracts will be classified as eligible. We also look at the proportion of tracts classified as eligible for the two different speed thresholds, and by aggregating block to tract speed by taking the minimum advertised speed, the mean advertised speed, or the maximum advertised speed, indicators that are all available in the FCC data. Selecting a threshold of 65% or fewer of households have a broadband subscription, and 100 / 20 mean maximum advertised speed achieves similar proportions of tracts in need (~ 4%), which being broadly consistent with the eligibility criteria that federal broadband programs published in their program documentation.

Table A2.1. Proportion of tracts classified as "in need" under different eligibility guidelines

Internet adoption	
Threshold of need	**% Tracts in need**
65% or less with subscription	4.1%
70% or less with subscription	7.4%
75% or less with subscription	13.0%
80% or less with subscription	22.6%
85% or less with subscription	37.4%
90% or less with subscription	57.1%
Internet speed	
Threshold of need	**% Tracts in need**
25 / 3, Minimum	99.9%
100 / 20, Minimum	99.9%
25 / 3, Mean	1.02%
100 / 20, Mean	4.3%
25 / 3, Maximum	0.28%
100 / 20, Maximum	0.35%

A3. Relationship between internet need measures and need classifications

Our final pre-analysis data exercise was to examine the relationship between the two measures of Internet need. This project is motivated by the observation that different definitions of Internet need and bases of program eligibility would result in prioritization of different types of communities. We adjudicate that hypothesis in the project analysis.

Table A3.1 and Figure A3.1 confirm that these two bases of eligibility generate distinct patterns. Only 0.53% of tracts are eligible under both definitions, while ~ 3.5% of tracts are eligible under one of the two definitions. 91.3% of tracts are not eligible for the program. We also find that while the two variables are positively correlated, the correlation is weak (a correlation coefficient of 0.11 for all tracts, and 0.26 for eligible tracts). There should be many tracts that are eligible for broadband programs under one definition but not the other.

Table A3.1. Proportion of tracts falling under each need classification by eligibility type

Need classification	Number of tracts	Proportion of tracts
Not in Need	77952	91.3%
Adoption Need Only	3076	3.60%
Speed Need Only	2795	3.27%
In Need Both	449	0.53%

Figure A3.1. Relationship between Internet speed and adoption in all tracts (top) and dichotomized to eligibility (bottom)