# No WAN's Land: Mapping U.S. Broadband Coverage with Millions of Address Queries to ISPs

David Major
Princeton University

Ross Teixeira
Princeton University

Jonathan Mayer
Princeton University

## ABSTRACT

Accurate broadband coverage data is essential for public policy planning and government support programs. In the United States, the Federal Communications Commission is responsible for maintaining national broadband coverage data. Observers have panned the FCC's broadband maps for overstating availability, due to coarse-grained data collection and a low coverage threshold.

We demonstrate a new approach to building broadband coverage maps: automated large-scale queries to the public availability checking tools offered by major internet service providers. We reverse engineer the coverage tools for nine major ISPs in the U.S., test over 19 million residential street addresses across nine states for service, and compare the results to the FCC's maps.

Our results demonstrate that the FCC's coverage data significantly overstates the availability of each ISP's service, access to any broadband, connection speeds available to consumers, and competition in broadband markets. We also find that the FCC's data disproportionately overstates coverage in rural and minority communities. Our results highlight a promising direction for developing more accurate broadband maps and validating coverage reports.

## CCS CONCEPTS

• **Networks → Public Internet**; • **Social and professional topics → Broadband access**; **Governmental regulations**.

**ACM Reference Format:**
David Major, Ross Teixeira, and Jonathan Mayer. 2020. No WAN's Land: Mapping U.S. Broadband Coverage with Millions of Address Queries to ISPs. In *ACM Internet Measurement Conference (IMC '20), October 27–29, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 27 pages. https://doi.org/10.1145/3419394.3423652

## 1 INTRODUCTION

Broadband internet access is an essential instrument for advancing public policy goals, including economic opportunity, educational attainment, and public health [1, 2]. The United States and the European Union—along with many other nations and international organizations—are prioritizing initiatives to increase broadband availability [3, 4]. Governments worldwide are investing in network infrastructure and closing the "digital divide" between broadband haves and have-nots. These policies are especially critical today: the COVID-19 pandemic has compelled a global shift to working, attending school, and socializing through broadband connections.

Accurate coverage data is the foundation of efforts to close the digital divide. Broadband maps guide both overall funding levels and resource targeting, by identifying areas where infrastructure investments are most needed. Coverage data also informs related telecommunications policies, such as net neutrality, broadband privacy, municipal broadband, and unbundling, by identifying possible market failures that substantiate regulatory intervention.

In the United States, the Federal Communications Commission (FCC) is responsible for maintaining nationwide broadband coverage data [5]. Stakeholders have criticized the FCC's maps for methodological shortcomings that would overstate coverage. There is wide agreement—including by the FCC's leadership—that the current maps are not adequately accurate. But there is little clarity on what the errors are, and the FCC is just now beginning to revise its broadband coverage data collection methods.

We investigate a new direction for developing broadband coverage maps: automated large-scale measurement of the representations that ISPs make to prospective customers. Every major ISP in the U.S. offers a broadband availability tool (BAT) to check whether an address is eligible for service. We systematically submit residential address queries to BATs and derive a dataset of U.S. fixed broadband coverage. We make two main contributions: a rigorous methodology for generating a coverage dataset from BATs, and a comparative analysis that estimates errors in the FCC's maps.

Our methodology begins with reverse engineering the BATs for nine major ISPs. We examine each BAT with test queries, developing a preliminary taxonomy of distinct response types and corresponding outcomes (e.g., the address is covered, not covered, or unrecognized). Next, we conduct automated data collection, querying the BATs with residential addresses from U.S. Department of Transportation and U.S. Postal Service data. In total, we query over 19 million addresses across nine states. We iteratively add new BAT response types to our taxonomy, and we evaluate the accuracy of our final taxonomy by placing test calls to ISPs. We then use our taxonomy to convert the BAT responses into a coverage dataset.

We analyze our dataset in comparison to the FCC's coverage data, contributing the most rigorous large-scale evaluation of the FCC's maps to date. Our results demonstrate that the FCC's data significantly overstates coverage by *each* ISP, access to *any* broadband service, access to *higher-speed* broadband, and access to *competing* providers. We also find the FCC's data disproportionately overstates coverage in rural and minority communities.

The rest of the paper proceeds as follows. We provide background on broadband maps in the U.S., then summarize and compare related work (Section 2). We then describe our BAT methodology


*IMC '20, October 27–29, 2020, Virtual Event, USA*

ACM ISBN 978-1-4503-8138-3/20/10…$15.00
https://doi.org/10.1145/3419394.3423652

for generating broadband coverage data (Section 3). Next, we analyze our dataset in comparison to the FCC's and discuss the results (Section 4). We conclude with directions for future work (Section 5).

## 2 BACKGROUND AND RELATED WORK

In this section, we provide background on the FCC's broadband maps and their shortcomings. We then describe and compare related work that evaluates the FCC's maps and other broadband studies.

### 2.1 Background on U.S. Broadband Maps

For over two decades, the FCC has been responsible for encouraging broadband deployment and competition in the U.S. [6]. As part of that mission, the FCC requires ISPs to file coverage reports, via a process termed Form 477 [7].[1] The FCC then compiles the reports into a dataset, which it uses to allocate infrastructure subsidies and evaluate potential regulation. The FCC also makes its coverage data and an interactive national broadband map available online [9].

Our work focuses on the Form 477 data for fixed broadband, such as fiber, cable, and DSL. These ISPs are required to submit biannual reports about service that exceeds 200 kbps in either direction [10].

There are several notable shortcomings to the FCC's current fixed coverage methods. First, the FCC collects data at the census block level. While census blocks are the most granular geographic units tracked by the U.S. Census Bureau, these areas can still encompass nearly a thousand housing units (especially in urban areas) or hundreds to thousands of square miles (especially in rural areas) [11]. Second, if an ISP reaches *one* address in a census block, it reports coverage for the *entire* census block [12]. Third, even if an ISP does not *currently* reach a census block, it still reports coverage if it *could soon* provide service to an address in the block.[2]

The FCC's current data collection methods necessarily lead to inaccurate coverage analysis, because the resulting broadband maps do not account for incomplete coverage within census blocks and do not distinguish between actual and potential coverage. When the FCC conducts analysis with its current maps—for annual reports on broadband deployment [13], subsidy programs [14], policy planning [15], and myriad other purposes—it treats coverage for the geography, residences, and individuals within each census block as all-or-nothing. Independent analysis relying on the FCC's coverage data similarly assumes—because it must—that reported coverage in a census block means complete coverage for the block [16–18].

Erroneous coverage reports are another source of inaccuracy in the FCC's data. Unfortunately, the FCC's ability to detect these errors is very limited. In 2019, for example, FCC staff conducted on-the-ground tests of the Verizon, T-Mobile, and U.S. Cellular mobile networks in 12 states [19, 20]. The staff report concluded that Form 477 data from all three providers so "generally overstated" coverage and performance that the data was not "generally reliable." Earlier this year, AT&T notified the FCC that it had mistakenly reported fixed (i.e., non-mobile) coverage since 2017 in over 3,500 census blocks across 20 states [21]. Also this year, the FCC initiated penalty proceedings against BarrierFree—a local fixed ISP in New York—for submitting years of inaccurate coverage data [22]. In one Form 477 filing, BarrierFree claimed to provide service from Connecticut through Virginia. In another filing, the ISP claimed to serve an impossible number of customers in an area—over ten times the number of housing units. The FCC relied on BarrierFree's reports until a civil society group spotted the errors [23].

For clarity throughout our work, we delineate two types of inaccuracy that result from the FCC's broadband coverage data. An "overstatement" is an instance where the FCC data indicates a geography, residence, or individual is covered, but broadband service is not actually available. An overstatement may be attributable to the FCC's methods or to an ISP's error. "Overreporting" is the specific type of overstatement caused by a mistaken ISP filing.

The FCC's maps have been widely criticized for inaccurately reflecting U.S. broadband availability, performance, and competition, including by members of the FCC, legislators, executive officials, civil society groups, and ISPs [5, 24–27]. In 2019, the FCC began taking steps to overhaul its broadband coverage data collection methods and restructure the Form 477 process as a new Digital Opportunity Data Collection [28]. Congress required additional updates in the Broadband DATA Act earlier this year [29], and the FCC issued further revisions in response [30]. In the updated FCC data collection process, ISPs will be required to submit fixed coverage data by providing either geospatial polygons or address lists. The methods for generating these polygons or lists—and their accuracy—will be left up to ISPs, subject only to lax technology-specific maximum buffer zones (e.g., for fiber, a provider may have latitude to report service within 35 miles of its optical terminals).

Meanwhile, ISPs already maintain broadband coverage data at much finer granularity than the census blocks that the FCC currently uses and the maximum buffer zones that the FCC plans to use. For example, ISPs keep exact address-level records for current and past subscribers. Moreover, every major ISP offers a public broadband availability tool (BAT) on its website that responds to address-level queries. ISPs have meaningful economic incentives to maintain accurate BATs, because a false negative is a missed potential customer and a false positive is a waste of resources attempting to provide service. While we cannot definitively determine the granularity of coverage data that BATs rely on when responding to queries, we have consistently observed that BATs provide coverage data at finer granularity than the geographic scale of both census blocks and the FCC's new maximum buffer zones.

### 2.2 Related Work

There have been three notable efforts to examine the accuracy of the FCC's fixed broadband coverage data. In 2019, Microsoft studied the related problem of U.S. broadband *usage*. Microsoft measured the network speed for users of its services, geolocated users to ZIP codes based on IP addresses, and then compared the resulting usage data to the FCC's maps [31]. The Microsoft study estimated that, as of November 2019, about 157 million Americans were not using broadband service at the FCC's benchmark minimum of 25 Mbps download and 3 Mbps upload. Microsoft inferred from its results that the FCC's maps pervasively overstate coverage [32]. In comparison to Microsoft's work, we collect data on broadband

[1] The original Form 477 involved mailing paperwork and a floppy disk [8]. While the FCC's data collection is now online, the FCC has retained the name.

[2] The FCC ambiguously directs ISPs to report coverage if they could provide service "within a service interval that is typical for that type of connection—that is, without an extraordinary commitment of resources" [12].

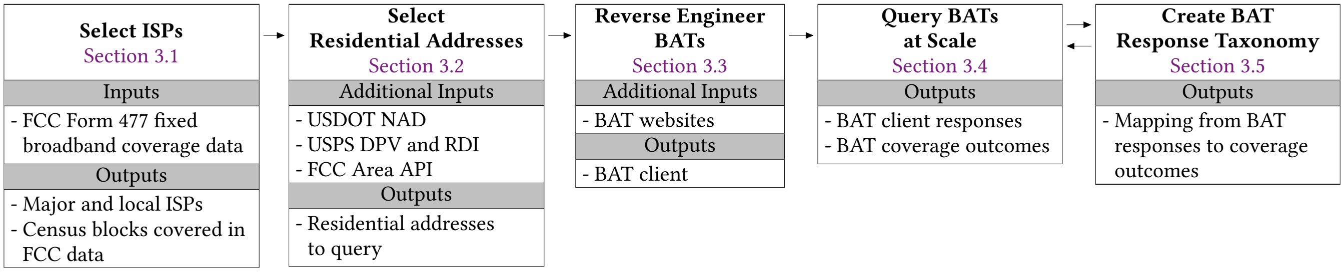


**Figure 1: An overview of our methodology for generating a coverage dataset from ISP broadband availability tools (Section 3).**

availability rather than usage, we do not rely on IP geolocation, and our data is much more granular than ZIP codes.

BroadbandNow, a portal for consumers to compare broadband service offerings, conducted a concurrent study of the FCC's coverage data [33]. The study methodology was similar to our own: BroadbandNow manually queried BATs for nine ISPs using 11,663 addresses, totaling 20,065 address-ISP combinations. BroadbandNow found that 19.6% of address-ISP combinations resulted in a BAT response other than service available, and 13.0% of addresses did not have any relevant BAT indicating service. The report inferred from these results that the number of Americans without broadband access is double FCC estimates. BroadbandNow also hypothesized that overstatements in Form 477 data disproportionately impact rural states, because a larger population proportion in those states already lacks coverage according to the FCC's data. While the BroadbandNow report is a valuable contribution, it differs from our work in several important methodological respects: we rigorously source and format comprehensive residential address lists, we develop and evaluate a full taxonomy mapping BAT responses to coverage outcomes, and we collect orders of magnitude more data by using automation. Our analysis also includes evaluation of overstatements for speed and competition, as well as disproportionate overstatements for rural and minority communities.

CostQuest Associates, a telecommunications consulting firm, conducted a 2019 pilot study of the FCC's maps in Missouri and Virginia [34].[3] CostQuest developed a "fabric" of serviceable building structures using proprietary methods that synthesized property tax records, parcel records, address data, road data, satellite imagery, and crowdworker labeling. The report then used coverage data from an unspecified number of collaborating ISPs to evaluate the accuracy of the FCC's broadband maps. CostQuest concluded that the FCC's data overstated participating ISP coverage for 16% of structures in Missouri and for 19% of structures in Virginia. We cannot more directly compare our work to CostQuest's, because we do not have access to the proprietary methods and data.

In addition to the FCC, other government agencies are compiling broadband coverage data. The National Telecommunications and Information Administration is developing a National Broadband Availability Map, which integrates FCC, state, and commercial data [36]. The California Public Utilities Commission also has an initiative to collect state-level broadband coverage data [37].

[3]The study was commissioned by USTelecom, the U.S. ISP trade group [35].

Our work is situated in the networking literature on broadband connectivity. Prior work has measured the broadband performance that subscribers experience [38–45], investigated demographic disparities ("digital redlining") in broadband coverage through manual speed tests and mapping [46–48], and examined how ISPs can efficiently connect underserved communities [49, 50].

## 3 METHODOLOGY

In this section, we describe our methods for creating a dataset of U.S. fixed broadband coverage using automated large-scale address queries to ISP BATs. We begin by selecting nine major ISPs to study, based on the FCC's Form 477 data (Section 3.1). We then generate a set of over 19 million residential addresses across nine states to query for coverage, using U.S. Department of Transportation (USDOT) and U.S. Postal Service (USPS) address data (Section 3.2). Next, we reverse engineer the BAT for each ISP we study, building both a client and a preliminary taxonomy for interpreting responses (Section 3.3). We use our BAT client to accomplish large-scale data collection, querying ISP-address combinations where, according to FCC data, an address is covered by an ISP (Section 3.4). We iteratively add new BAT response types to our taxonomy during our data collection (Section 3.5), and we evaluate the final taxonomy by manually checking that addresses exist and placing test telephone calls to ISPs (Section 3.6). Fig. 1 presents an overview of the steps in our methodology. We are intentionally conservative at each step, because BATs are black-box systems from our perspective; we present results with less conservative methods in Appendix I.

### 3.1 Selecting ISPs

We study nine broadband providers, which we term "major" ISPs: AT&T, CenturyLink, Charter, Comcast, Consolidated, Cox, Frontier, Verizon, and Windstream. We choose these providers for two reasons. First, the major ISPs represent most of the broadband market: according to the FCC's Form 477 coverage data and population estimates, these ISPs have large geographic footprints and together reach over 90% of the population that has access to broadband in the states that we study. Second, each of the major ISPs we study has a BAT that we can reliably query with residential addresses [52–60].

The broadband market includes a number of small providers, which we term "local" ISPs. These providers typically do not have a public BAT, so we cannot measure coverage with our methods. In the states that we study, we consider every provider that is not a

| State | Census ACS Housing Units | NAD Addresses | Excluding NAD Incomplete and Non-Residential Addresses | Excluding USPS Undeliverable and Non-Residential Addresses | Excluding FCC Census Blocks with No ISP Coverage | Excluding FCC Census Blocks with No Major ISP Coverage |
|---|---|---|---|---|---|---|
| Arkansas | 1,389,129 | 1,420,282* | 953,094 | 803,869 | 779,653 | 707,119 |
| Maine | 750,939 | 628,320 | 601,016 | 454,261 | 452,892 | 443,522 |
| Massachusetts | 2,928,732 | 3,505,320 | 2,991,125 | 2,791,829 | 2,783,539 | 2,782,284 |
| New York | 8,404,381 | 6,250,440 | 6,250,400 | 4,744,097 | 4,714,062 | 4,622,305 |
| North Carolina | 4,747,943 | 4,769,561 | 4,184,488 | 3,166,112 | 3,139,441 | 3,033,847 |
| Ohio | 5,232,869 | 4,669,233* | 4,316,165 | 3,790,397 | 3,766,547 | 3,684,408 |
| Vermont | 339,439 | 313,950 | 254,291 | 195,188 | 192,505 | 182,855 |
| Virginia | 3,562,143 | 3,622,803 | 3,620,877 | 3,037,980 | 2,977,503 | 2,831,837 |
| Wisconsin | 2,725,296 | 1,426,684* | 1,426,655 | 1,195,863 | 1,191,339 | 1,114,896 |
| Total | 30,080,871 | 26,606,593 | 24,598,111 | 20,179,596 | 19,997,481 | 19,403,073 |

**Table 1: We compile a dataset of residential addresses by starting with the USDOT NAD, excluding addresses that lack essential fields or are categorized as non-residential, validating against USPS data, and filtering for broadband coverage according to FCC data (Section 3.2). We include housing unit data from the 2019 U.S. Census Bureau American Community Survey as context [51]. NAD address counts denoted with a ∗ indicate missing county data.**

major ISP to be a local ISP.[4] We also treat major ISPs as local ISPs in states where they have limited market presence, owing to how we prioritized data collection (see Appendix A for detail).

We conservatively assume that local ISPs have 100% availability within census blocks that they report to the FCC as covered. Local ISPs do (collectively) have broad coverage (Appendix C), so our analysis of access to broadband is very sensitive to this assumption. We present results from relaxing the assumption in Appendix I.

We use the FCC's Form 477 data from June 2019 [12] and the FCC's census block population estimates from 2018 [61]. Both datasets were the most recent available when we began this study.

### 3.2 Selecting Residential Addresses

After identifying ISPs for study, we compile a dataset of residential addresses for our queries and analysis. We rely on USDOT, USPS, and U.S. Census Bureau data. Table 1 presents counts for each step in our address validation process. In total, our dataset includes nearly 20 million addresses that are served by at least one ISP and 19.4 million addresses that are served by at least one major ISP.

The USDOT National Address Database (NAD) is an ongoing initiative to create a comprehensive federal public database of addresses and associated locations [62, 63]. USDOT receives data from state, county, and municipal agencies, then consolidates and standardizes records [64]. The NAD includes basic address fields (e.g., address number and street name), latitude and longitude coordinates, and an optional address type (e.g., residential or commercial).

We limit our study to nine states where the NAD includes address data and where the major ISPs are the predominant providers.[5]

We process the NAD addresses into a dataset of residential query addresses with four steps. First, we filter for and standardize essential fields. We begin with the set of NAD addresses in each of the nine states, and we exclude addresses that are missing an address number, street name, municipality/community, or ZIP code, since these fields are typically required by BATs. We also normalize street suffixes according to USPS address standards [65], because we find that certain BATs require properly formatted addresses.[6]

Second, where the NAD includes an address type, we filter out non-residential categories. We retain addresses categorized as multi-use, unknown, or other because we find many of these addresses are residential and USPS data provides an effective further filter.

Third, we verify the remaining addresses against two USPS datasets, which we access through a commercial provider [66]. We confirm that each address is able to receive ordinary postal mail with USPS Delivery Point Validation [67]. We further confirm that each address is residential by checking the USPS Residential Delivery Indicator [68], which labels whether an address is subject to residential rates for mail delivery.

Finally, we identify addresses that are served by at least one major or local ISP according to the FCC's data, as well as addresses that are served by at least one major ISP. We associate each remaining address with a census block using the address's NAD location and U.S. Census Bureau shape data (via the FCC Area API [69]).

This methodology gives us high confidence in the resulting addresses, though we note that the dataset may be biased toward urban areas. Rural addresses may not receive residential mail (e.g., rural routes), for example, or may be less likely to appear in USDOT or USPS data. Our work is, to our knowledge, the first to develop a residential address list based on the NAD; future work could explore additional methodological refinement.

### 3.3 Reverse Engineering BATs

We next reverse engineer the BATs for the ISPs that we study. We send test address queries to the BATs to understand how they function and develop a preliminary taxonomy mapping BAT responses to coverage outcomes. We then build a client for querying the BATs.

[4] We treat Altice as a local ISP in New York, for reasons explained in Appendix B.

[5] The NAD is missing data from counties in three of the states, which we note in Table 1. We confirmed that the gaps do not cross census blocks, so we do not weight our analysis for incomplete census blocks.

[6] In the NAD, for example, "ALLEY" might appear as "ALLY" or "ALY." We address this issue by substituting in the correct suffix based on keyword matching.

**Manual Reverse Engineering.** We identify the public BAT for each major ISP by navigating its website. Next, we input test queries for a set of coverage outcomes: addresses that are residential and covered, residential and not covered, and nonexistent. We log HTTP(S) traffic for each query. We then manually identify the sequence of requests and responses that begins with the query address and ends with the BAT response.

BAT implementation details vary by ISP. Some BATs are RESTful APIs, while others are ordinary webpages. We find that some BATs require a multi-step querying process, where the browser issues a request with an address, receives a response with an ID for that address, then issues further requests with the ID. Some BATs require a session cookie from a previous webpage. We describe particularly unusual BAT behaviors in Appendix D.

**Building a BAT Client.** We build a Python client for each BAT that submits queries and parses responses, based on our reverse engineering and BAT response taxonomy. The client stores the query address and either a response type (if parsing succeeds) or an error (if parsing fails) in a MySQL database.

The client opportunistically parses additional information from BAT responses. Four ISP BATs (AT&T, CenturyLink, Consolidated, and Windstream) provide speed tier data. The client parses and stores that data, which we use to evaluate coverage overstatements by connection speed (Section 4.2).

The BATs for four ISPs (AT&T, CenturyLink, Charter, and Verizon) also respond with an address. The client parses the response address, and if it does not match the query address (e.g., the BAT substituted in a similar but distinct address), the client categorizes the response as an unknown type.[7]

**Handling Apartment Units.** BATs differ in how they handle apartments. The same unit might, for example, appear as "APT 15G," "#15G," or "15 G" across ISPs. We test apartment addresses during our reverse engineering, and we incorporate logic for handling apartments into the BAT client.

When a BAT prompts for a unit number, it includes a list of suggestions.[8] The client parses this list and randomly selects a unit, making the assumption that broadband availability is uniform within the building.

## 3.4 Querying BATs at Scale

We use our BAT client to collect coverage data for the 19.4 million residential addresses in our dataset that are covered by at least one major ISP (Table 1). The client issues queries for combinations of a major ISP and an address that are covered according to the FCC's data, totaling nearly 35 million queries (Appendix F).[9]

When the client encounters a response that it cannot parse, we iteratively add the new response type to our taxonomy (as described in the following section), and the client then re-queries coverage for the address.

The data collection period for our study is January through August 2020.[10] We rate limit BAT queries to ensure that our data collection does not interfere with public availability.

## 3.5 Creating a BAT Response Taxonomy

BAT responses are diverse, and many either do not clearly indicate whether there is coverage or reflect an error. We create an initial taxonomy that maps response types to coverage outcomes when reverse engineering each BAT, as described in Section 3.3. When the BAT client encounters a new type of response, we manually inspect the response and begin from a presumption that the information visually presented to the user reflects the coverage outcome. We then submit test queries and reverse engineer how the BAT triggers and handles the response, which can surface additional information that indicates a different coverage outcome is appropriate. Finally, we identify unique attributes for the response and integrate parsing for those attributes into the BAT client.

The implementation details for parsing BAT responses vary by ISP. Some BATs are RESTful APIs, for example, that return straightforward JSON values. Other BATs are webpages, where we identify unique strings or DOM elements for the client to parse.

We map each BAT response type to one of five coverage outcomes: the address is covered, the address is not covered, the address is not recognized, the address is a business, or the response is unknown (i.e., we cannot interpret it). Appendix D provides detail on ISP-specific response interpretation challenges that we encounter, Appendix E presents our final taxonomy, and Appendix F gives BAT response counts by coverage outcome. Our final taxonomy includes 74 response types across the nine ISPs that we study.

**Non-Covered Addresses.** We are able to reliably categorize non-covered addresses for seven of the nine major ISPs, because there are clearly distinct response types for when the query address is not covered. For the remaining two ISPs (CenturyLink and Cox), however, we encounter challenges distinguishing non-covered addresses from unrecognized addresses. We are able to infer the distinction in one case based on further examination of the response type (CenturyLink) and in the other case by querying an affiliated availability tool (Cox).

As an illustration of this issue, and more generally the complexity of interpreting BAT responses, Fig. 2 presents a pair of response types from CenturyLink. At first glance, both responses appear to indicate that the address is not covered. The first response occurs for known nonexistent addresses, however, and the response consistently appears when the BAT cannot autocomplete an address and has an internal address ID set to `null`. Also, the JavaScript that triggers the response includes the status string "We were unable to find the address you provided." Based on these factors, we treat the response as an unrecognized address rather than a non-covered address. Our evaluation of BAT responses in the following section confirms that many of these addresses are nonexistent rather than non-covered. Appendix G provides screenshots of all the CenturyLink BAT response types, further illustrating the challenge in interpreting responses.

[7] In this step, the BAT client checks the query address against both the response address and the response address with a normalized street suffix (as described in Section 3.2).
[8] See Appendix D for discussion of a Cox special case.
[9] See Appendix A for limited circumstances where we treat major ISPs as local ISPs.

[10] Our BAT queries began over six months after the reporting date for the Form 477 data we analyze. This difference in time period likely introduces a slight bias against our analysis identifying understatements. ISP service areas usually expand over time, so an address might have been an overstatement in mid-2019 but covered by 2020.

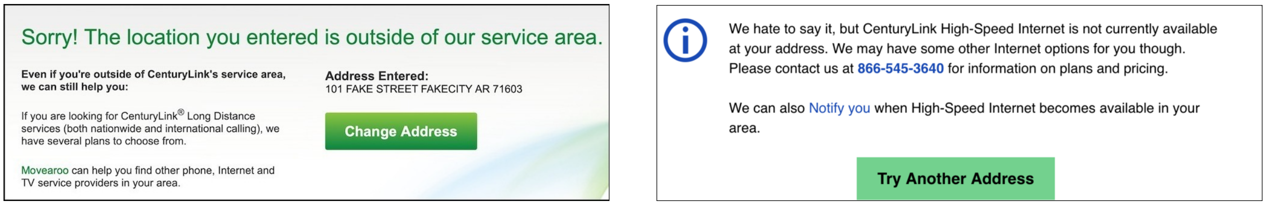


**Figure 2: Two example response types from CenturyLink: $Ce_0$ (left) and $Ce_3$ (right). Both response types appear to indicate that the ISP does not cover the address. $Ce_0$ is, however, a response for a nonexistent address ("101 FAKE STREET") while $Ce_3$ is a response for a manually verified residential address. We conservatively interpret $Ce_0$ as meaning the BAT does not recognize the address and $Ce_3$ as meaning the address is not covered.**

**Unknown Responses.** We categorize certain response types in our taxonomy as unknown, because we cannot map the response to a coverage status. These responses are website errors, for example, or instructions to call a telephone number for further information.

For two ISPs (Charter and Frontier), we are not able to distinguish between unrecognized addresses and unknown responses: we find that querying with nonexistent addresses results in either a generic request to call customer service (Charter) or a generic error (Frontier). In both cases, we follow our presumption of labeling based on the information provided to the user, and we treat the response types as unknown.

A limitation in our BAT client also requires us to categorize certain Charter responses as unknown, even though the website might have shown a different coverage outcome to the user. We built the BAT client to parse key coverage fields in an API, and we subsequently found that when the key fields were absent, the BAT could still visually present coverage or non-coverage to the user. This finding indicates that our BAT client did not fully parse the information available in Charter BAT responses. Because our BAT client did not store information beyond the main coverage fields, we treat all responses missing the fields as unknown.

As noted in Section 3.3, we treat a mismatch between the query and the response address (when available) as an unknown response.

## 3.6 Evaluating the BAT Response Taxonomy

We further evaluate two dimensions of our BAT response taxonomy: addresses that are unrecognized, and addresses that have a coverage status (i.e., either covered or not covered).

**Unrecognized Addresses.** BAT responses indicating an unrecognized address are common in our dataset—nearly a million address-ISP combinations. An unrecognized address could be a real residence that the ISP serves, but the ISP's BAT formats the address differently from our client. Alternatively, an unrecognized address could reflect a residence that is entirely missing from the BAT.[11] An unrecognized address also might not be a residence at all.

We conduct a small-scale manual evaluation to understand the relative frequency of these scenarios. We begin by randomly sampling 40 unrecognized addresses for each major ISP, with the exception of Charter and Frontier (because those ISPs have no BAT responses in our taxonomy that map to an unrecognized address).

Next, we determine if the address is formatted differently in the BAT's database by manually querying the BAT with the address. If the BAT suggests the address in a format that our BAT client did not recognize, but that we can verify is the same address (e.g., a slightly different spelling of the street name or suffix), we request coverage for the reformatted address.

If a clear coverage status is not available for a reformatted address, we continue our evaluation by attempting to identify what occupies the address. We search real estate websites, property records, Google Street View, and Google Maps satellite imagery.

We assign one of the following labels to each address in our evaluation: incorrect format (when the BAT provides a coverage status for a reformatted address), residence exists (when we confirm a house or apartment building occupies the address), residence does not exist (when we confirm there is a non-residential occupant for the address, such as a business), residence could exist (when we confirm there is a vacant lot or mobile home at the address, and we are uncertain if it is currently being used as a residence), and could not determine (when we fail to find additional information about the address). Table 2 presents counts by ISP for each label type.

The results of our evaluation are mixed. We find that most unrecognized addresses reflect real residences, but we cannot obtain a clear coverage status from the BAT. We also find that many unrecognized addresses are not, or might not be, an actual residence. Because of these ambiguities, we conservatively omit unrecognized addresses from our main analysis in Section 4. We present results from relaxing this assumption in Appendix I.

**Covered and Non-Covered Addresses.** In our study, we lack conventional ground truth: we measure what major ISPs *represent* about service availability for a large set of addresses. We do not measure whether service *actually* is available, because conducting a rigorous evaluation (i.e., arranging and following through on service appointments for a sample of addresses across the U.S.) would be prohibitively complex.

There are many reasons why an ISP's BAT might not accurately reflect service availability. The latest local coverage data might not have propagated to the national BAT, for example, or coverage might only be available after further evaluation by a local service center. It is also possible that the ISP's BAT simply contains erroneous data for an address.

[11] We hypothesize these addresses are likely not covered. We do not, however, conduct further evaluation on the subset of unrecognized addresses we verify are residences.

| ISP | Incorrect Format | Residence Exists | Residence Does Not Exist | Residence Could Exist | Cannot Determine |
|---|---|---|---|---|---|
| AT&T | 0 (0%) | 9 (22.5%) | 30 (75%) | 0 (0%) | 1 (2.5%) |
| CenturyLink | 7 (17.5%) | 17 (42.5%) | 7 (17.5%) | 5 (12.5%) | 4 (10%) |
| Comcast | 0 (0%) | 32 (80%) | 3 (7.5%) | 4 (10%) | 1 (2.5%) |
| Consolidated | 3 (7.5%) | 26 (65%) | 4 (10%) | 5 (12.5%) | 2 (5%) |
| Cox | 3 (7.5%) | 28 (70%) | 3 (7.5%) | 2 (5%) | 4 (10%) |
| Verizon | 9 (22.5%) | 20 (50%) | 9 (22.5%) | 0 (0%) | 2 (5%) |
| Windstream | 0 (0%) | 31 (77.5%) | 2 (5%) | 4 (10%) | 3 (7.5%) |
| Total | 22 (7.9%) | 163 (58.2%) | 58 (20.7%) | 20 (7.1%) | 17 (6.1%) |

**Table 2: Results from a small-scale manual evaluation of $N = 40$ unrecognized addresses per ISP (Section 3.6). Charter and Frontier are absent because our taxonomy does not include unrecognized address responses for those ISPs (Section 3.5).**

Our taxonomy of BAT responses, described in the prior section, is an additional possible source of error. The BAT may itself be accurate, but how we interpret the BAT's output could be mistaken.

We conducted another small-scale manual evaluation to address these possible issues.[12] Every major ISP in our study is reachable by telephone, and attempting to arrange service by phone provides another source of coverage data.

For each ISP, we randomly sampled a minimum of 8 addresses from our BAT response dataset: 4 addresses that were covered and 4 that were not covered.[13] We then attempted to identify the telephone number for a local service center or store for each address, and when we could not (or when local representatives were unavailable), we fell back to a national sales number. When we called for each address, we requested service and noted the response.

For 5 of the 9 major ISPs in our study, we did not find a single instance where the coverage status offered by phone differed from the coverage status returned by the BAT. Multiple ISP representatives noted that we should just check the ISP's website for coverage.

In our calls to Comcast, 2 of 6 covered addresses required further evaluation by a local service center. 2 of 9 non-covered addresses were actually served according to a representative, but there was an unpaid balance at each address. Neither representative could explain why an unpaid balance caused the BAT to report no coverage.

For Cox, a representative responded that a local service center would have to follow up on 3 of the 4 non-covered addresses.

In our calls to Charter, a representative reported that a local service center would have to evaluate 1 of 4 non-covered addresses.

For Consolidated, a representative indicated that service was available at 1 of the 4 non-covered addresses.

In total, we checked 83 addresses by telephone. The response we received by phone matched the coverage outcome in our dataset for 74 addresses (89%), and the response by phone disagreed with our dataset (as opposed to requiring follow-up) for only 3 addresses (4%). The results from our evaluation give us general confidence that our taxonomy correctly interprets BAT responses and that our dataset is consistent with the representations ISPs make by phone.

Our evaluation has several limitations. The number of addresses that we test is relatively small, since placing telephone calls is time-consuming. It is also likely that some ISPs share an address database between their website and their telephone representatives, so placing calls is not an independent measurement. Furthermore, telephone responses still are not conventional ground truth—they are another type of representation to prospective customers.

## 3.7 Limitations

Before turning to analysis of our dataset, we reemphasize two important limitations of our methodology. First, each step in our methods—especially selecting addresses and creating a BAT response taxonomy—is a possible source of measurement error. We discuss and evaluate these issues in the prior section.

Second, BATs are black-box systems from our perspective. We can submit address-level queries and examine the responses. But we do not have certainty about the granularity or on-the-ground accuracy of the coverage databases underlying the BATs.

Nevertheless, we believe that ISP representations about coverage are an important *type* of ground truth for public policy purposes—especially when coverage is reportedly unavailable. If an ISP informs a prospective customer both online and by telephone that service is unavailable, we hypothesize that the customer will likely take the information at face value and not obtain service.

# 4 RESULTS AND DISCUSSION

Based on the coverage dataset we assemble from ISP BAT responses, we examine the extent to which the FCC's Form 477 data overstates broadband availability across nine states.[14]

We begin our analysis with coverage overstatements for each ISP (Section 4.1). Next, we assess speed overstatements for four ISPs where our client collected speed data (Section 4.2). We then examine overstatements at the state level, aggregating across ISPs to understand overstatements of access to *any* broadband (Section 4.3) and access to *competing* providers (Section 4.4). Finally, we conduct a regression analysis to understand the relationship between overstatements and rural areas, poverty, and race (Section 4.5).

[12] We sought and obtained approval from the Princeton University Institutional Review Board before conducting our evaluation of covered and non-covered addresses.

[13] We sampled 15 addresses for Comcast (6 covered and 9 not covered) because of the responses we received during our evaluation, 10 for AT&T and Verizon (5 covered and 5 not covered), and 8 (4 covered and 4 not covered) for the remaining ISPs.

[14] Note that our methodology precludes discovery of potential coverage underreporting (i.e., census blocks that ISPs *should* have reported as covered but did not), since we only query an ISP's BAT for addresses that are covered according to Form 477 data. We present a small-scale evaluation of possible underreporting in Appendix L.

| | | Residential Addresses Covered by | | | | | | Population Covered by | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Provider ≥ 0 Mbps | | | Provider ≥ 25 Mbps | | | Provider ≥ 0 Mbps | | | Provider ≥ 25 Mbps | | |
| ISP | Area | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ |
| AT&T | All | 4,516,190 | 3,894,238 | 86.23% | 3,368,933 | 3,110,667 | 92.33% | 13,195,479 | 11,350,764 | 86.02% | 9,772,773 | 8,999,811 | 92.09% |
| | Urban | 4,003,429 | 3,587,448 | 89.61% | 3,124,254 | 2,924,184 | 93.60% | 11,686,871 | 10,477,473 | 89.65% | 9,069,172 | 8,484,813 | 93.56% |
| | Rural | 512,761 | 306,790 | 59.83% | 244,679 | 186,483 | 76.22% | 1,508,608 | 873,291 | 57.89% | 703,601 | 514,998 | 73.19% |
| CenturyLink | All | 1,643,526 | 1,521,772 | 92.59% | 1,099,526 | 1,058,604 | 96.28% | 4,970,303 | 4,582,302 | 92.19% | 3,349,144 | 3,213,848 | 95.96% |
| | Urban | 867,774 | 846,600 | 97.56% | 637,543 | 628,378 | 98.56% | 2,546,197 | 2,480,876 | 97.43% | 1,902,233 | 1,869,742 | 98.29% |
| | Rural | 775,752 | 675,172 | 87.03% | 461,983 | 430,226 | 93.13% | 2,424,106 | 2,101,426 | 86.69% | 1,446,911 | 1,344,106 | 92.89% |
| Charter | All | 8,680,140 | 8,495,464 | 97.87% | 8,680,137 | 8,495,461 | 97.87% | 26,822,638 | 25,885,022 | 96.50% | 26,822,611 | 25,884,995 | 96.50% |
| | Urban | 7,027,339 | 6,941,933 | 98.78% | 7,027,339 | 6,941,933 | 98.78% | 21,328,689 | 20,886,613 | 97.93% | 21,328,689 | 20,886,613 | 97.93% |
| | Rural | 1,652,801 | 1,553,531 | 93.99% | 1,652,798 | 1,553,528 | 93.99% | 5,493,949 | 4,998,409 | 90.98% | 5,493,922 | 4,998,382 | 90.98% |
| Comcast | All | 3,645,212 | 3,567,459 | 97.87% | 3,645,212 | 3,567,459 | 97.87% | 10,239,740 | 9,926,535 | 96.94% | 10,239,740 | 9,926,535 | 96.94% |
| | Urban | 3,217,999 | 3,169,744 | 98.50% | 3,217,999 | 3,169,744 | 98.50% | 8,929,385 | 8,737,817 | 97.85% | 8,929,385 | 8,737,817 | 97.85% |
| | Rural | 427,213 | 397,715 | 93.10% | 427,213 | 397,715 | 93.10% | 1,310,355 | 1,188,718 | 90.72% | 1,310,355 | 1,188,718 | 90.72% |
| Consolidated | All | 433,078 | 392,824 | 90.71% | 339,916 | 312,057 | 91.80% | 1,485,551 | 1,297,204 | 87.32% | 1,184,151 | 1,040,767 | 87.89% |
| | Urban | 207,209 | 200,376 | 96.70% | 167,862 | 162,191 | 96.62% | 672,754 | 655,078 | 97.37% | 553,205 | 539,402 | 97.50% |
| | Rural | 225,869 | 192,448 | 85.20% | 172,054 | 149,866 | 87.10% | 812,797 | 642,126 | 79.00% | 630,946 | 501,365 | 79.46% |
| Cox | All | 1,132,153 | 1,095,762 | 96.79% | 1,132,153 | 1,095,762 | 96.79% | 3,154,414 | 3,041,010 | 96.40% | 3,154,414 | 3,041,010 | 96.40% |
| | Urban | 1,061,582 | 1,033,844 | 97.39% | 1,061,582 | 1,033,844 | 97.39% | 2,968,081 | 2,884,740 | 97.19% | 2,968,081 | 2,884,740 | 97.19% |
| | Rural | 70,571 | 61,918 | 87.74% | 70,571 | 61,918 | 87.74% | 186,333 | 156,270 | 83.87% | 186,333 | 156,270 | 83.87% |
| Frontier | All | 1,125,636 | 1,045,938 | 92.92% | – | – | – | 3,532,716 | 3,305,064 | 93.56% | – | – | – |
| | Urban | 715,621 | 692,988 | 96.84% | – | – | – | 2,098,515 | 2,034,223 | 96.94% | – | – | – |
| | Rural | 410,015 | 352,950 | 86.08% | – | – | – | 1,434,201 | 1,270,841 | 88.61% | – | – | – |
| Verizon | All | 8,015,081 | 6,915,507 | 86.28% | 4,272,541 | 4,175,882 | 97.74% | 27,321,134 | 23,989,510 | 87.81% | 16,550,269 | 16,140,157 | 97.52% |
| | Urban | 7,146,747 | 6,520,060 | 91.23% | 4,134,925 | 4,050,362 | 97.95% | 24,840,964 | 22,888,159 | 92.14% | 16,179,899 | 15,808,543 | 97.70% |
| | Rural | 868,334 | 395,447 | 45.54% | 137,616 | 125,520 | 91.21% | 2,480,170 | 1,101,351 | 44.41% | 370,370 | 331,614 | 89.54% |
| Windstream | All | 475,527 | 451,512 | 94.95% | 385,836 | 374,796 | 97.14% | 1,424,271 | 1,349,481 | 94.75% | 1,146,346 | 1,114,466 | 97.22% |
| | Urban | 311,063 | 301,199 | 96.83% | 260,833 | 254,652 | 97.63% | 893,813 | 867,281 | 97.03% | 747,301 | 731,490 | 97.88% |
| | Rural | 164,464 | 150,313 | 91.40% | 125,003 | 120,144 | 96.11% | 530,458 | 482,200 | 90.90% | 399,045 | 382,976 | 95.97% |
| Total | All | – | – | 92.29% | – | – | 96.80% | – | – | 91.95% | – | – | 96.04% |
| | Urban | – | – | 94.85% | – | – | 97.62% | – | – | 94.66% | – | – | 97.19% |
| | Rural | – | – | 80.00% | – | – | 91.90% | – | – | 79.20% | – | – | 89.35% |

**Table 3: Analysis of address and population coverage overstatements for each ISP (Section 4.1).**

Throughout our analysis, we rely on several additional datasets: census block population estimates (from a 2018 FCC release [61]), census block urban and rural classifications (from the 2010 U.S. Census [70]), and census tract demographic and poverty data (from the 2018 U.S. Census Bureau American Community Survey with five-year estimates).[15]

## 4.1 Overstatements of Coverage by ISP

In this section, we examine coverage overstatements for each major ISP. We assess overstatements as a whole, overstatements in rural areas, overstatements by available speed (according to FCC data), and the distribution of overstatements among census blocks. We illustrate how pronounced overstatements can be with a case study on Wisconsin, and we also demonstrate the viability of our methods for detecting overreporting with a case study on AT&T.

**Coverage Overstatements.** Table 3 presents a comparison of coverage by each ISP according to the FCC's data and our dataset of BAT coverage outcomes. We arrive at these results as follows.

For each ISP, we begin with the set of census blocks in our dataset where the ISP provides coverage according to Form 477. We then filter out census blocks where the ISP's BAT returns address unrecognized or response unknown for every address. This step excludes census blocks where the BAT responses are entirely ambiguous.

We next generate ISP-specific labels and counts for covered addresses. For each address in the remaining census blocks, if the ISP's BAT returns coverage, we label the address as covered by the FCC and BAT data. If the ISP's BAT returns no coverage, we label the address as covered only by the FCC data. If the ISP's BAT returns address unrecognized or an unknown response, we do not label the address as covered.[16] We then compute a count of FCC covered addresses, a count of BAT covered addresses, and an address overstatement ratio of the two counts.

Intuitively, this method provides a conservative address count by making no assumption (covered or not covered) about an address when an ISP's BAT does not provide a clear coverage status. The left-hand side of Table 3 presents the results of this method.

Next, we estimate the population covered by ISP $i$ according to the FCC's data $P_{FCC}(i)$ and the population covered by $i$ according to our BAT dataset $P_{BAT}(i)$ by computing:

$$P_{FCC}(i) = \sum_{c \in C_i} p_c \quad \text{and} \quad P_{BAT}(i) = \sum_{c \in C_i} p_c * \frac{B_i(c)}{F_i(c)}$$

where $C_i$ is the set of census blocks used to generate covered address counts for ISP $i$, $p_c$ is the population of each census block $c$, $B_i(c)$ is the BAT covered address count for ISP $i$ in census block $c$, and $F_i(c)$ is the FCC covered address count for ISP $i$ in census block $c$. Finally, we calculate a population overstatement ratio: the BAT covered population divided by the FCC covered population.

Intuitively, our method adjusts the covered population for each ISP at the census block level by applying the address overstatement ratio for the block derived from BAT responses. The right-hand side of Table 3 presents the results of this method.

We find that overstatements are a significant problem across ISPs, and especially for AT&T and Verizon coverage. Our analysis

[15] Census tracts are larger U.S. Census Bureau units of analysis than census blocks.

[16] Throughout our analysis, we treat business address responses as unknown responses.

suggests that in just the nine states that we study, the FCC's data overstates coverage by specific ISPs for millions of Americans.

**Overstatements in Rural Areas.** We additionally present per-ISP overstatements by urban and rural area classification in Table 3. The proportional overstatement of each ISP's coverage is consistently larger in rural areas than in urban areas. Our data suggests that Verizon is a particular outlier for overstated coverage in rural areas—we estimate that Verizon service is not available to *over half* of the rural population that is covered according to FCC data.

**Overstatements at Lower Speeds.** Table 3 also presents overstatements by connection speed. For this analysis, we add a speed threshold to our criteria for excluding census blocks, where the speed for an ISP in a block is the maximum speed the ISP reports to the FCC in that block. We examine two speed thresholds: any service (≥ 0 Mbps) and service at benchmark speed (≥ 25 Mbps).

We find that the FCC's data is consistently less accurate for addresses in lower-speed census blocks. This effect is proportionally much larger in rural areas in comparison to urban areas, and the effect is particularly significant for rural addresses that are covered by AT&T and Verizon in the FCC's data. Verizon, for example, is an outlier in rural coverage overstatement when considering all speed tiers—but its overstatement ratio is roughly consistent with other ISPs whe considering only blocks with benchmark broadband.

We hypothesize that the disproportionate coverage inaccuracy for lower-speed service in rural areas is attributable, at least in part, to network technology. In areas where Form 477 data indicates that the maximum service speed is below 25 Mbps, AT&T and Verizon report offering ADSL almost exclusively; in areas with higher maximum speeds, the providers offer newer VDSL and fiber connectivity. These ISPs may have more accurate coverage data for the newer network technologies because building the network infrastructure provided an opportunity and incentive to map possible service addresses. For the legacy network technologies, by contrast, the ISPs may be investing less in mapping and might rely on simple models for coverage when reporting to the FCC (e.g., distance from central office DSLAMs for DSL connectivity).

**Distribution of Overstatements.** Coverage overstatements are unevenly distributed among census blocks. Fig. 3 shows, for each major ISP, the cumulative distribution function for the address overstatement ratio in each block. Our analysis demonstrates that the median coverage within census blocks is 100% for every ISP, and the 25th percentile coverage is also 100% for every ISP except AT&T and Verizon. At lower percentiles, however, the address overstatement ratio drops significantly. Thus, the majority of coverage overstatements are confined to a minority of census blocks.

**Possible Overreporting.** Our results surface census blocks that are covered by an ISP in Form 477 data, but in which we do not observe any coverage based on BAT responses. We cannot definitively determine whether these blocks reflect ISP overreporting, because we lack conventional ground truth and the FCC allows reporting coverage where an ISP could soon provide service [12]. Nevertheless, these results raise concerns about whether ISPs have accurately reported coverage and the clarity of FCC reporting guidelines.

For purposes of examining possible overreporting, we are especially conservative in our analysis. We do not consider a census block as possible overreporting if our dataset has few addresses in the block (less than 20). We also do not consider a census block as possible overreporting for an ISP if there is even *one* BAT response for the ISP that is anything other than a not covered address.

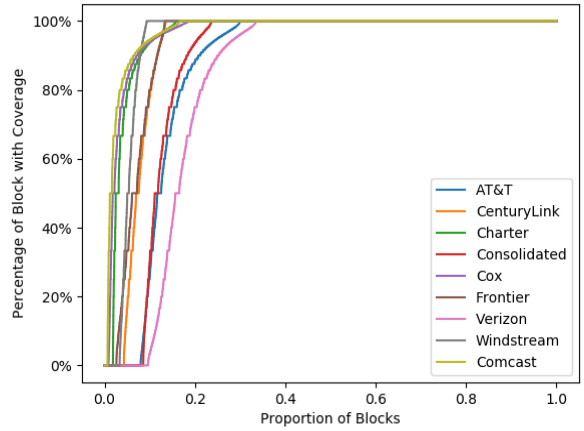


**Figure 3: The cumulative distribution function, for each major ISP, of the address overstatement ratio across census blocks (Section 4.1).**

| ISP | # Blocks (≥ 0 Mbps) 0% Coverage | Total | # Blocks (≥ 25 Mbps) 0% Coverage | Total |
|---|---|---|---|---|
| AT&T | 2,196 | 440,079 | 211 | 360,933 |
| CenturyLink | 112 | 549,542 | 10 | 446,135 |
| Charter | 81 | 712,831 | 81 | 627,159 |
| Consolidated | 21 | 41,885 | 15 | 38,588 |
| Cox | 61 | 156,708 | 61 | 109,976 |
| Comcast | 69 | 260,387 | 69 | 209,897 |
| Frontier | 97 | 594,209 | — | — |
| Verizon | 1,403 | 401,046 | 20 | 349,660 |
| Windstream | 54 | 369,511 | 8 | 299,621 |

**Table 4: Counts of census blocks where ISPs claim coverage to the FCC, but we found no coverage (Section 4.1).**

Table 4 presents the count of census blocks, for each ISP and by speed tier, where we observe no covered address and the block passes the filtering just described. The table also includes, as context, the total number of census blocks that the ISP covers according to the FCC's data. We find evidence of possible overreporting by every provider, and especially by AT&T and Verizon at lower service speeds. We emphasize again that we cannot definitively determine whether these are instances of overreporting, and while our results call into question reported coverage for thousands of census blocks, that number of census blocks is relatively small in comparison to overall provider coverage.

**Case Study: Census Blocks in Wisconsin.** In order to clearly visualize the problem of coverage overstatements, Fig. 4 maps census blocks with acute overstatement. We show eight blocks in Wisconsin: four covered by AT&T and four by CenturyLink according to Form 477 data. In each block, nearly every address lacks coverage by the relevant ISP according to its BAT. Because of the Form 477 data collection methods, however, these blocks appear to be fully covered in FCC data.

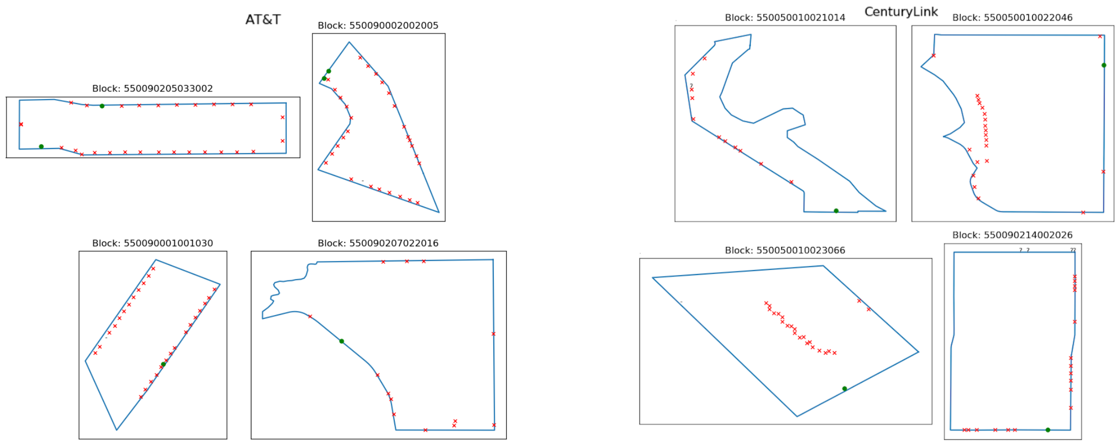


**Figure 4: Census blocks in Wisconsin that, according to FCC data, are covered by AT&T and CenturyLink. • indicates an address where coverage is available in our BAT dataset, X where coverage is not available, and ? where the address is unrecognized.**

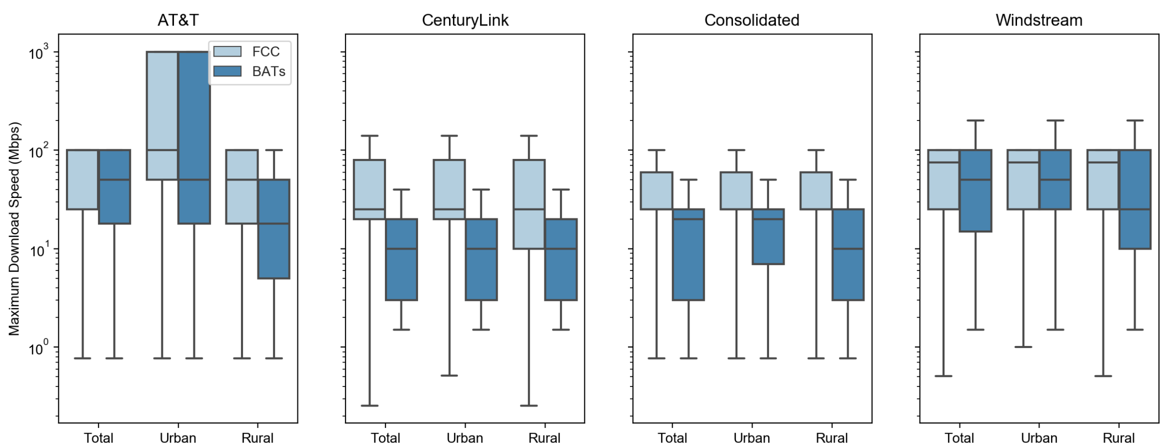


**Figure 5: Distributions of maximum service speed across addresses, according to FCC data and BAT responses (Section 4.2).**

**Case Study: AT&T Overreporting.** As discussed in Section 2, AT&T notified the FCC this year that it had mistakenly reported fixed coverage with speeds ≥ 25 Mbps in over 3,500 census blocks [21]. We randomly select 20 blocks from AT&T's notice and manually examine the blocks in our BAT coverage dataset to understand whether our methods could have identified the overreporting. For 12 of the blocks, our analysis dataset does not include any addresses. For 5 of the blocks, AT&T's BAT responded that all the addresses were either not covered or were covered at speeds below 25 Mbps. In the remaining 3 blocks, we find at least one address where the AT&T BAT indicated service of at least 25 Mbps (though in one of the blocks, 95% of addresses either lacked service or had coverage at speeds below 25 Mbps). Based on this limited analysis, we believe our methods have promise as a means of validating ISP coverage filings: in this known instance of overreporting, our dataset indicated problems with 17 of the 20 census blocks we evaluated. We note, though, that this case study also suggests that our method for identifying possible overreporting may be too conservative—our method would have filtered out nearly all the blocks in our sample.

## 4.2 Overstatements of Speed by ISP

In this section, we examine speed overstatements for the four ISPs from which our client collects speed data. We continue with the method described in Section 4.1 for labeling addresses with coverage. For each address in our dataset that we label as covered by the ISP according to the FCC, we set the FCC maximum download speed to the maximum speed for that ISP in that block according to Form 477 data. For each address that we label as covered according to our BAT data, we set the BAT maximum download speed to the maximum speed our client observed when querying the ISP's BAT for the address. Fig. 5 shows the resulting distributions of address maximum speeds, further subdivided by urban and rural areas.

For every ISP, the Form 477 data overstates the broadband speeds provided to residential addresses. Across ISPs, the median speed

| | | Residential Addresses Covered by | | | | | | Population Covered by | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Any Provider $\geq$ 0 Mbps | | | Any Provider $\geq$ 25 Mbps | | | Any Provider $\geq$ 0 Mbps | | | Any Provider $\geq$ 25 Mbps | | |
| State | Area | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ |
| Arkansas | All | 738,930 | 729,909 | 98.78% | 644,909 | 639,270 | 99.13% | 1,931,568 | 1,876,454 | 97.15% | 1,659,973 | 1,636,439 | 98.58% |
| | Urban | 461,854 | 460,834 | 99.78% | 443,263 | 441,672 | 99.64% | 1,143,777 | 1,139,924 | 99.66% | 1,102,379 | 1,097,947 | 99.60% |
| | Rural | 277,076 | 269,075 | 97.11% | 201,646 | 197,598 | 97.99% | 787,791 | 736,530 | 93.49% | 557,594 | 538,492 | 96.57% |
| Maine | All | 430,377 | 429,611 | 99.82% | 409,362 | 407,412 | 99.52% | 1,236,455 | 1,221,562 | 98.80% | 1,203,289 | 1,181,068 | 98.15% |
| | Urban | 184,630 | 184,626 | 100.00% | 181,098 | 181,010 | 99.95% | 483,478 | 483,467 | 100.00% | 482,639 | 482,369 | 99.94% |
| | Rural | 245,747 | 244,985 | 99.69% | 228,264 | 226,402 | 99.18% | 752,977 | 738,094 | 98.02% | 720,650 | 698,699 | 96.95% |
| Massachusetts | All | 2,717,975 | 2,714,493 | 99.87% | 2,650,770 | 2,645,527 | 99.80% | 6,576,766 | 6,556,509 | 99.69% | 6,537,295 | 6,522,212 | 99.77% |
| | Urban | 2,535,311 | 2,534,470 | 99.97% | 2,476,893 | 2,472,768 | 99.83% | 6,055,863 | 6,053,196 | 99.96% | 6,043,541 | 6,032,396 | 99.82% |
| | Rural | 182,664 | 180,023 | 98.55% | 173,877 | 172,759 | 99.36% | 520,903 | 503,313 | 96.62% | 493,754 | 489,815 | 99.20% |
| New York | All | 4,572,327 | 4,564,326 | 99.83% | 4,410,349 | 4,399,536 | 99.75% | 16,720,148 | 16,667,225 | 99.68% | 16,489,377 | 16,439,000 | 99.69% |
| | Urban | 3,790,630 | 3,790,305 | 99.99% | 3,725,591 | 3,723,134 | 99.93% | 14,596,827 | 14,595,598 | 99.99% | 14,581,660 | 14,573,820 | 99.95% |
| | Rural | 781,697 | 774,021 | 99.02% | 684,758 | 676,402 | 98.78% | 2,123,321 | 2,071,626 | 97.57% | 1,907,717 | 1,865,179 | 97.77% |
| North Carolina | All | 2,968,700 | 2,960,050 | 99.71% | 2,822,176 | 2,812,650 | 99.66% | 8,429,747 | 8,373,766 | 99.34% | 8,076,725 | 8,023,358 | 99.34% |
| | Urban | 2,016,259 | 2,016,056 | 99.99% | 1,992,290 | 1,990,910 | 99.93% | 5,565,328 | 5,564,196 | 99.98% | 5,543,163 | 5,537,487 | 99.90% |
| | Rural | 952,441 | 943,994 | 99.11% | 829,886 | 821,740 | 99.02% | 2,864,419 | 2,809,570 | 98.09% | 2,533,562 | 2,485,870 | 98.12% |
| Ohio | All | 3,546,890 | 3,536,254 | 99.70% | 3,294,723 | 3,275,567 | 99.42% | 9,406,714 | 9,359,282 | 99.50% | 8,923,090 | 8,808,589 | 98.72% |
| | Urban | 2,844,832 | 2,843,454 | 99.95% | 2,772,130 | 2,767,529 | 99.83% | 7,426,004 | 7,419,032 | 99.91% | 7,344,869 | 7,313,573 | 99.57% |
| | Rural | 702,058 | 692,800 | 98.68% | 522,593 | 508,038 | 97.21% | 1,980,710 | 1,940,250 | 97.96% | 1,578,221 | 1,495,016 | 94.73% |
| Vermont | All | 188,316 | 187,396 | 99.51% | 180,148 | 177,577 | 98.57% | 583,549 | 575,926 | 98.69% | 566,383 | 551,687 | 97.41% |
| | Urban | 66,540 | 66,537 | 100.00% | 66,191 | 66,159 | 99.95% | 225,165 | 225,146 | 99.99% | 224,624 | 224,474 | 99.93% |
| | Rural | 121,776 | 120,859 | 99.25% | 113,957 | 111,418 | 97.77% | 358,384 | 350,779 | 97.88% | 341,759 | 327,213 | 95.74% |
| Virginia | All | 2,891,937 | 2,868,401 | 99.19% | 2,726,445 | 2,696,231 | 98.89% | 7,985,763 | 7,830,978 | 98.06% | 7,611,696 | 7,477,134 | 98.23% |
| | Urban | 2,160,200 | 2,156,536 | 99.83% | 2,127,082 | 2,111,827 | 99.28% | 6,142,903 | 6,122,642 | 99.67% | 6,091,893 | 6,020,920 | 98.83% |
| | Rural | 731,737 | 711,865 | 97.28% | 599,363 | 584,404 | 97.50% | 1,842,860 | 1,708,335 | 92.70% | 1,519,803 | 1,456,214 | 95.82% |
| Wisconsin | All | 1,162,574 | 1,160,654 | 99.83% | 1,076,221 | 1,072,800 | 99.68% | 3,586,981 | 3,575,944 | 99.69% | 3,417,606 | 3,395,923 | 99.37% |
| | Urban | 870,343 | 870,046 | 99.97% | 855,344 | 854,070 | 99.85% | 2,839,515 | 2,836,426 | 99.89% | 2,821,854 | 2,813,679 | 99.71% |
| | Rural | 292,231 | 290,608 | 99.44% | 220,877 | 218,730 | 99.03% | 747,466 | 739,517 | 98.94% | 595,752 | 582,244 | 97.73% |
| Total | All | 19,218,026 | 19,151,094 | 99.65% | 18,215,103 | 18,126,570 | 99.51% | 56,457,691 | 56,037,646 | 99.26% | 54,485,434 | 54,035,410 | 99.17% |
| | Urban | 14,930,599 | 14,922,864 | 99.95% | 14,639,882 | 14,609,079 | 99.79% | 44,478,860 | 44,439,627 | 99.91% | 44,236,622 | 44,096,665 | 99.68% |
| | Rural | 4,287,427 | 4,228,230 | 98.62% | 3,575,221 | 3,517,491 | 98.39% | 11,978,831 | 11,598,014 | 96.82% | 10,248,812 | 9,938,742 | 96.97% |

**Table 5: Analysis of address and population coverage overstatements for *any* broadband coverage (Section 4.3).**

available is 25 Mbps in our dataset, in comparison to 75 Mbps in the Form 477 data. The discrepancy between BAT speed data and FCC speed data is especially pronounced for CenturyLink and Consolidated. We also observe that for all the ISPs except CenturyLink, speed overstatements are larger in rural areas than in urban areas.

We present coverage overstatements at different speed lower bounds in Appendix H.

### 4.3 Overstatements of Any Coverage

We now turn to estimating overstatements of *any* broadband coverage. Table 5 presents a comparison of the addresses and population in each state that have have access to broadband service, according to the FCC's data and our BAT dataset. We arrive at these figures with a method analogous to the analysis in Section 4.1.

For each state, we begin with the census blocks in our dataset where at least one major or local ISP provides service (see Section 3.2). We then filter out census blocks where there is at least one major ISP and the response to every BAT query, across every address-ISP combination, is address unrecognized or response unknown. We take this conservative filtering step because we have less confidence in our address data for these census blocks. This step excludes 2.2% of the census blocks in our dataset, encompassing 0.8% of the population represented in the dataset.

We next generate coverage labels for addresses. For each address in the remaining census blocks, we label the address as covered by broadband service according to our BAT data if it meets one of the following criteria:

- There is at least one major ISP that covers the address according to its BAT.
- There is a local ISP that provides service to the address's census block according to Form 477 data.

We label an address as covered according to the FCC's data if it satisfies either criterion above, or if it satisfies the following criterion:

- For every major ISP that provides service to the address's census block according to Form 477 data, the ISP's BAT returns not covered for the address.

We then compute a count of FCC covered addresses for each state, a count of BAT covered addresses, and an address overstatement ratio between the counts.

Intuitively, this method generates a conservative estimate of coverage overstatements by not making assumptions about an address's coverage if there is no local ISP and BATs return a mix of unrecognized address, unknown response, or no coverage. The left-hand side of Table 5 presents the results from applying this method. We provide a sensitivity analysis in Appendix I that relaxes the method's strict criteria.

We further estimate the population with and without broadband access. We use the same population adjusting formula as in Section 4.1, except we sum over the census blocks within a state, and we use the address overstatement ratio above.[17] The right-hand side of Table 5 presents the results of this method.

**Coverage Overstatements.** Table 5 presents the results of our analysis by state, urban and rural area designation, and speed tier.

[17] If a census block does not have a single address labeled as covered according to the FCC's data, we omit the block from this analysis.

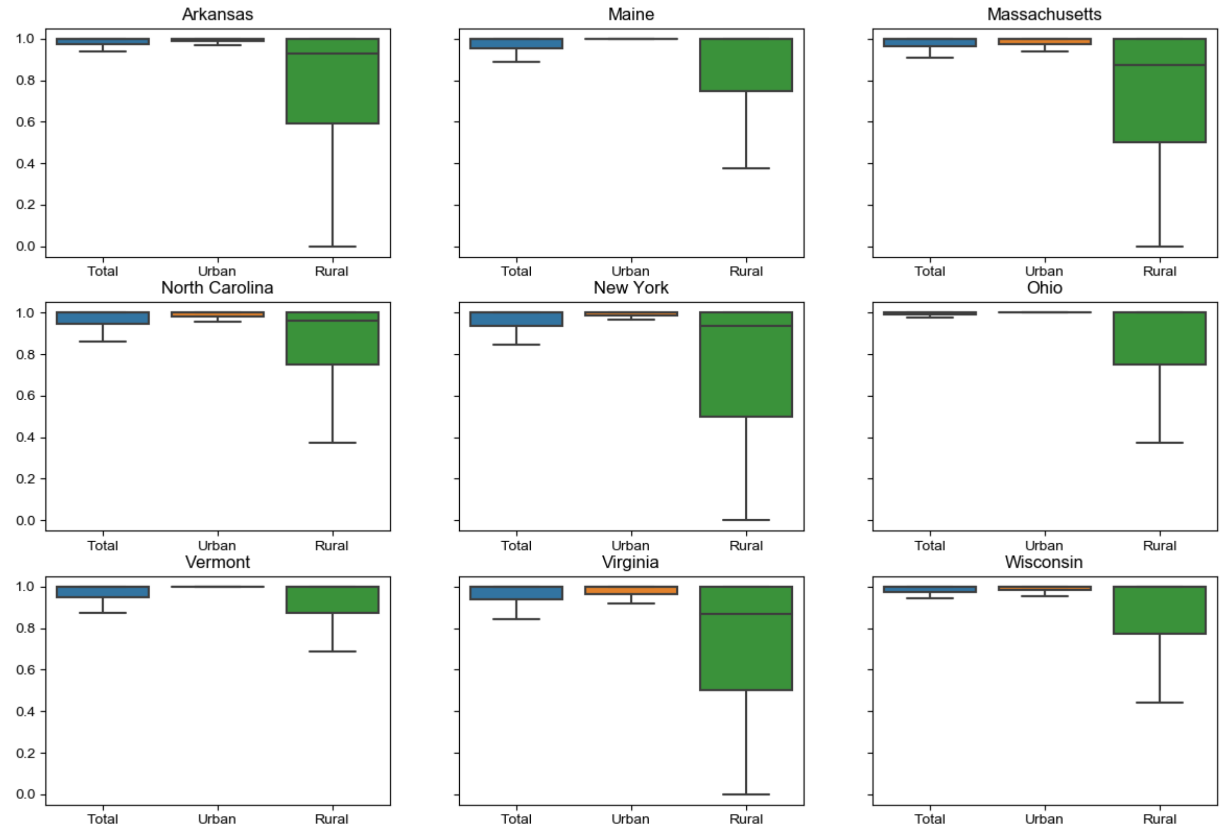


**Figure 6: The distribution of the competition overstatement ratio among census blocks, by state and area type (Section 4.4).**

Across the states we study, 0.83% of Americans that FCC data indicates have access to benchmark broadband service are not covered according to our analysis. There is overstated coverage for every state, speed tier, and area type. If we (naively) assume that our results are nationally representative, our results indicate that 2.5 million Americans in excess of FCC estimates lack access to broadband at benchmark speed.[18]

Our results diverge from the recent BroadbandNow report [33] discussed in Section 2.2, which estimated overstatements for 6.5% of addresses in a sample and 21 million Americans nationally. We hypothesize that the disparity stems from differences in methodology and our deliberate caution in evaluating possible overstatements, at the risk of underestimating inaccuracy in the FCC's data.[19] For example, if we treat BAT unrecognized addresses and unknown responses as meaning an address has no coverage, we estimate benchmark broadband overstatement for 3.80% of addresses in our dataset and 14 million Americans (see Appendix I).

**Overstatements in Rural Areas.** FCC coverage overstatements are both proportionally and absolutely larger in rural areas than in urban areas. Across all nine states, 99.68% of Americans in urban areas that the FCC claims have benchmark broadband are covered according to BAT data, in comparison to 96.97% of Americans in rural areas. Even in absolute numbers, despite the fact that approximately 80% of Americans live in urban areas, the overstatement in rural areas (310,070 people) is larger than in urban areas (139,957 people) at benchmark speeds [71]. The FCC's disproportionate overstatement of coverage in rural areas holds for each state.

## 4.4 Overstatements of Competition

We now estimate the extent to which FCC coverage data overstates broadband competition in each state (i.e., the number of providers available at an address).[20] A broadband market with multiple participants has important consequences for consumers, who benefit from lower prices and greater reliability as a result of competition.

We begin with census blocks in our dataset that are covered by at least one major ISP. For each block, we count the major ISPs in

[18] We calculate our estimate of U.S. broadband coverage by multiplying the FCC's 2019 broadband coverage population estimate [5] by our population overstatement ratio.

[19] Address sampling and weighting methods may be important factors. The BroadbandNow report does not specify the source of its address dataset. We observed that the BroadbandNow website sometimes solicits address information to check broadband coverage. If the report relied on user-provided addresses, that could introduce sampling bias; users who search for broadband coverage on a third-party website might be disproportionately likely to have encountered challenges obtaining broadband service. Furthermore, we weight address overstatements at the census block level to generate population estimates. BroadbandNow directly infers population overstatements from address overstatements, which could interact with any sample bias.

[20] Our competition analysis omits local ISPs, because we lack address-ISP coverage data. Including local ISPs would be effectively adding a constant to the competition overstatement ratio numerator and denominator for addresses with local ISP coverage.

| Variable | Coeff | SE | P-Value |
|---|---|---|---|
| Proportion Minority Population | -0.0065 | 0.002 | 0.00 |
| Proportion Rural | -0.0413 | 0.001 | 0.00 |
| AT&T | -0.0161 | 0.002 | 0.000 |
| CenturyLink | -0.0106 | 0.002 | 0.000 |
| Consolidated | -0.0383 | 0.008 | 0.000 |
| Frontier | -0.0081 | 0.003 | 0.003 |
| Verizon | -0.0160 | 0.003 | 0.000 |
| Windstream | -0.0112 | 0.003 | 0.001 |

**Table 6: Results from regression analysis of coverage overstatements (Section 4.5). We present select independent variables that have statistically significant ($p \leq .05$) correlation with overstatements. Appendix K provides complete results. The dependent variable is the coverage overstatement ratio, so a negative coefficient implies greater overstatement.**

Form 477 data. We filter out addresses in the block where any BAT returned address unrecognized or response unknown, and we set aside the block if it has no remaining addresses. We then count the remaining addresses and the address-ISP combinations where a BAT returned coverage. Next, we calculate an average number of available providers for addresses in the block: the count of address-ISP combinations with coverage divided by the count of addresses. We finally compute a competition overstatement ratio for the block: the average number of available providers according to BAT data, divided by the number of major ISPs according to Form 477 data.

Fig. 6 plots the distribution of the competition overstatement ratio for each state and by urban and rural areas. We provide an analysis by speed tier in Appendix J. We find that in urban areas, the FCC's data does not appear to greatly overstate competition. In rural areas, however, there appears to be very significant overstatement of competition—in Virginia, for example, FCC data may double the number of actual providers for one in four census blocks.

### 4.5 Relationship Between Overstatements and Rural, Low-Income, and Minority Communities

We conduct a regression analysis to examine the relationship between coverage overstatements and rural, low-income, and minority communities. We fit an ordinary least squares model at the census tract level.[21] The dependent variable is the coverage overstatement ratio for the tract, following the method in Section 4.3. The independent variables are tract population, the proportion of the tract that is rural,[22] the proportion of the population in the tract that lives in a household with an income below the federal poverty line, and the proportion of the tract population that is a minority.[23] We include an independent variable for each major ISP with the value set to the proportion of blocks in the tract covered by the ISP according to Form 477 data. We also add a categorical variable for the state where the tract is located.

[21] We use census tracts in this analysis, rather than census blocks, because the U.S. Census Bureau American Community Survey is not available at the block level.

[22] We calculate the rural proportion of a tract as the proportion of addresses in our dataset (filtered as in Section 4.3) that are in the tract and in a rural census block.

[23] We rely on U.S. Census Bureau data for this variable, and we count the population that is categorized as either non-White for race or Hispanic or Latino for ethnicity.

Table 6 shows variables with a statistically significant ($p \leq .05$) correlation with coverage overstatements. We present the full regression results in Appendix K. The results show that coverage overstatements correlate with a greater proportion of rural addresses and a greater proportion of the population that is a minority. The rural address proportion has the most negative coefficient of the independent variables in our analysis. We note, though, that $R^2 = 0.145$ for our regression model, indicating a high level of variability in the data that is not explained by the factors we examined.

## 5 CONCLUSION

In this work, we develop a methodology for using automated large-scale queries to ISP BATs to compile a broadband coverage dataset. We then examine the FCC's broadband maps in comparison to our dataset, contributing the most rigorous large-scale evaluation of the FCC's maps to date. Our results show that the FCC's coverage data significantly overstates the availability of each ISP's service, access to any broadband, connection speeds available to consumers, and competition in broadband markets. We also find that the FCC's data disproportionately overstates coverage in rural and minority communities.

Our research highlights several promising directions for future work.

**Speed Measurements.** In Section 4.2, we examine possible overstatements in Form 477 maximum download speed data for four ISPs. These limited results suggest that overstatements by speed are a fruitful direction for further research. Future work could study overstatements at more granular speed tiers, as well as which network technologies may disproportionately lead to overstatements.

**National and International Measurements.** Our methods enable broadband availability measurements in other states, either as NAD coverage improves or by using other U.S. address databases. Beyond the U.S., our methods allow measuring coverage in any location with a dataset of addresses and a set of ISPs with public BATs. Future work could examine broadband availability in other parts of the U.S. and in other nations.[24]

**Mobile Internet Coverage.** Our work is focused on fixed broadband coverage. The FCC also maintains mobile broadband availability data, which can similarly be inaccurate (see Section 2). Evaluating the accuracy of the FCC's mobile coverage maps is a natural next step, though it would likely require very different methods.

**Evaluating Future FCC Maps.** As discussed in Section 2, the FCC has begun to replace Form 477 with the Digital Opportunity Data Collection, which requires ISPs to report coverage using geospatial polygons or address lists [72]. Our results show that BATs are a promising direction for evaluating both the methods that ISPs use for future FCC coverage reports and whether ISPs are correctly implementing those methods.

## ACKNOWLEDGMENTS

This work is supported by Princeton University and the Digital Divide Roundtable at the University of Nebraska, as well as in-kind contributions by Luminati and SmartyStreets. We thank Jennifer Rexford and our shepherd Zachary Bischof for invaluable feedback.

[24] We have already implemented BAT support for five additional ISPs that serve states beyond those we studied, in anticipation of future measurements.

## REFERENCES


[1] 2012. The Impact of Broadband on the Economy: Research to Date and Policy Issues. International Telecommunication Union, (April 2012), 136.

[2] B. C. Bauerly et al. 2019. Broadband Access as a Public Health Issue: The Role of Law in Expanding Broadband Access and Connecting Underserved Communities for Better Health Outcomes. *J Law Med Ethics*, 47, 2suppl, (June 2019), 39–42.

[3] 2020. Universal service. https://www.fcc.gov/general/universal-service.

[4] 2020. Policies on Broadband Europe. https://ec.europa.eu/digital-single-market/en/policies/75980/3608.

[5] 2019. 2019 Broadband Deployment Report. GN Docket No. 18-238. (May 8, 2019). https://docs.fcc.gov/public/attachments/FCC-19-44A1.pdf.

[6] U.S. House. 104th Congress. 1996. (February 1996). https://www.congress.gov/104/plaws/publ104/PLAW-104publ104.pdf.

[7] 2000. In the Matter of Local Competition and Broadband Reporting. Report and Order CC Docket No. 99-301. Federal Communications Commission, (March 24, 2000). https://transition.fcc.gov/Bureaus/Common_Carrier/Orders/2000/fcc00114.pdf.

[8] 2000. In the Matter of Local Competition and Broadband Reporting. (March 3, 2000). https://transition.fcc.gov/Bureaus/Common%5C_Carrier/Orders/2000/fcc00114.pdf.

[9] 2020. FCC Fixed Broadband Deployment Map. en. Retrieved 09/23/2020 from https://broadbandmap.fcc.gov/.

[10] 2019. FCC Form 477 Local Telephone Competition and Broadband Reporting Instructions. (December 23, 2019). https://us-fcc.box.com/v/Form477Instructions.

[11] 2011. What are census blocks? https://www.census.gov/newsroom/blogs/random-samplings/2011/07/what-are-census-blocks.html.

[12] 2019. FCC form 477 local telephone competition and broadband reporting instructions for filings as of December 31, 2019. https://us-fcc.app.box.com/v/Form477Instructions.

[13] FCC. 2020. 2020 Broadband Deployment Report. FCC 20-50. Federal Communications Commission, (April 20, 2020).

[14] 2020. FCC Launches $20 Billion Rural Digital Opportunity Fund To Expand Rural Broadband Deployment. FCC-20-5. Federal Communications Commission, (February 7, 2020).

[15] 2017. White Paper on Improving the Nation's Digital Infrastructure. Federal Communications Commission, (January 19, 2017).

[16] Kelsey L. Conley and Brian E. Whitacre. 2020. Home is where the internet is? high-speed internet's impact on rural housing values. *International Regional Science Review*, 43, 5, 501–530. DOI: 10.1177/0160017620918652. eprint: https://doi.org/10.1177/0160017620918652. https://doi.org/10.1177/0160017620918652.

[17] Bento J. Lobo, Md Rafayet Alam, and Brian E. Whitacre. 2020. Broadband speed and unemployment rates: data and measurement issues. *Telecommunications Policy*, 44, 1, 101829. ISSN: 0308-5961. DOI: https://doi.org/10.1016/j.telpol.2019.101829. http://www.sciencedirect.com/science/article/pii/S0308596118303823.

[18] Elizabeth A. Mack. 2014. Businesses and the need for speed: the impact of broadband speed on business presence. *Telematics and Informatics*, 31, 4, 617–627. ISSN: 0736-5853. DOI: https://doi.org/10.1016/j.tele.2013.12.001. http://www.sciencedirect.com/science/article/pii/S0736585313000890.

[19] Federal Communications Commission. 2020. Mobility fund phase ii coverage maps investigation staff report. EN-US. (2020). https://docs.fcc.gov/public/attachments/DOC-361165A1.pdf.

[20] 2019. Mobility Fund Phase II Coverage Maps Investigation Staff Report. GN Docket No. 19-367. Federal Communications Commission, (December 4, 2019). https://docs.fcc.gov/public/attachments/DOC-361165A1.pdf.

[21] 2020. Re: Rural Digital Opportunity Fund, WC Docket No. 19-126. (April 10, 2020). https://ecfsapi.fcc.gov/file/104101431124557/RDOF%5C%20Challenge%5C%20Process%5C%20WC%5C%20Docket%5C%20No.%5C%2019-126.pdf.

[22] 2020. In the Matter of Barrier Communications Corporation d/b/a BarrierFree. (September 2, 2020). https://docs.fcc.gov/public/attachments/FCC-20-123A1.pdf.

[23] 2019. Re: GN Docket No. 18-238. (March 5, 2019). https://ecfsapi.fcc.gov/file/10306056687881/Free%5C%20Press%5C%20706%5C%20Report%5C%20Form%5C%20477%5C%20Erroneous%5C%20Data%5C%20ex%5C%20parte.pdf.

[24] 2020. FCC under fire from senators over rural broadband funding limitations. Retrieved 03/10/2020 from https://www.theverge.com/2020/3/10/21173534/fcc-broadband-rural-funding-limits-ajit-pai-chuck-schumer-gillibrand.

[25] 2020. FCC 'Coloring Outside the Lines' on Broadband Mapping, Say Critics at Next Century Cities Event. (January 24, 2020). http://broadbandbreakfast.com/2020/01/fcc-coloring-outside-the-lines-on-broadband-mapping-say-critics-at-next-century-cities-event/.

[26] 2020. Sen. Manchin Pledges to be FCC's Broadband Speed Pen Pal. (January 24, 2020). https://www.nexttv.com/news/sen-manchin-pledges-to-be-fccs-broadband-speed-pen-pal.

[27] 2018. How skeptics called Big Telecom's bluff on broadband coverage maps. (December 18, 2018). https://statescoop.com/broadband-mobility-fund-phase-challenges-fcc-ajit-pai/.

[28] 2019. Establishing the Digital Opportunity Data Collection; Modernizing the FCC Form 477 Data Program. FCC-19-79. Federal Communications Commission, (August 6, 2019). https://docs.fcc.gov/public/attachments/FCC-19-79A1.pdf.

[29] Senate - Commerce Science and Transportation - 104th Congress. 2019. (June 2019). https://www.congress.gov/bill/116th-congress/senate-bill/1822/text.

[30] 2019. In the Matter of Establishing the Digital Opportunity Data Collection, Modernizing the FCC Form 477 Data Program. (July 16, 2019). https://docs.fcc.gov/public/attachments/FCC-20-94A1.pdf.

[31] 2020. US Broadband Usage Percentages. original-date: 2020-04-16T02:40:42Z. (May 2020). Retrieved 05/31/2020 from https://github.com/microsoft/USBroadbandUsagePercentages.

[32] John Kahan. 2019. It's time for a new approach for mapping broadband data to better serve americans. Microsoft on the Issues. (April 8, 2019). Retrieved 02/10/2020 from https://blogs.microsoft.com/on-the-issues/2019/04/08/its-time-for-a-new-approach-for-mapping-broadband-data-to-better-serve-americans/.

[33] John Busby and Julia Tanberk. 2020. FCC Underestimates Americans Unserved by Broadband Internet by 50%. BroadbandNow Research, (February 2020). Retrieved 02/26/2020 from https://broadbandnow.com/research/fcc-underestimates-unserved-by-50-percent.

[34] 2019. Broadband Mapping Initiative: FAQ. https://www.costquest.com/uploads/pdf/broadbandmappinginitiative-summary.pdf.

[35] 2019. Broadband Mapping Initiative: Proof of Concept. https://www.ustelecom.org/wp-content/uploads/2019/03/Mapping-FAQ.pdf.

[36] National Telecommunications and Information Administration. 2019. National broadband availability map | broadband usa. (June 4, 2019). https://broadbandusa.ntia.doc.gov/map.

[37] 2019. Guidelines for broadband data submission. https://www.cpuc.ca.gov/General.aspx?id=2541.

[38] Igor Canadi, Paul Barford, and Joel Sommers. 2012. Revisiting broadband performance. In *Proceedings of the 2012 Internet Measurement Conference* (IMC '12). Association for Computing Machinery, Boston, Massachusetts, USA, 273–286. isbn: 9781450317054. doi: 10.1145/2398776.2398805. https://doi.org/10.1145/2398776.2398805.

[39] Srikanth Sundaresan et al. 2011. Broadband internet performance: a view from the gateway. In *Proceedings of the ACM SIGCOMM 2011 Conference* (SIGCOMM '11). Association for Computing Machinery, Toronto, Ontario, Canada, 134–145. isbn: 9781450307970. doi: 10.1145/2018436.2018452. https://doi.org/10.1145/2018436.2018452.

[40] Srikanth Sundaresan et al. 2014. Bismark: a testbed for deploying measurements and applications in broadband access networks. In *2014 USENIX Annual Technical Conference (USENIX ATC 14)*. USENIX Association, Philadelphia, PA, (June 2014), 383–394. isbn: 978-1-931971-10-2. https://www.usenix.org/conference/atc14/technical-sessions/presentation/sundaresan.

[41] Marcel Dischinger et al. 2007. Characterizing residential broadband networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement* (IMC '07). Association for Computing Machinery, San Diego, California, USA, 43–56. isbn: 9781595939081. doi: 10.1145/1298306.1298313. https://doi.org/10.1145/1298306.1298313.

[42] Bischof, Zachary S. and Otto, John S. and Sánchez, Mario A. and Rula, John P. and Choffnes, David R. and Bustamante, Fabian E. 2011. Crowdsourcing isp characterization to the network edge. In (Proceedings of the 1st ACM SIGCOMM Workshop on Measurements Up the Stack, W-MUST1́1). (September 2011), 61–66. isbn: 9781450308007. doi: 10.1145/2018602.2018617.

[43] Zachary S. Bischof, Fabian E. Bustamante, and Nick Feamster. 2017. Characterizing and improving the reliability of broadband internet access. *CoRR*, abs/1709.09349. arXiv: 1709.09349. http://arxiv.org/abs/1709.09349.

[44] Aruna Dunna, Zachary Bischof, and Romain Fontugne. 2020. Sanitizing a view of consumer broadband in the united states. In *2020 Network Traffic Measurement and Analysis Conference (TMA)*, 1–9.

[45] Ramakrishna Padmanabhan et al. 2019. Residential links under the weather. In *Proceedings of the ACM Special Interest Group on Data Communication* (SIGCOMM '19). Association for Computing Machinery, Beijing, China, 145–158. isbn: 9781450359566. doi: 10.1145/3341302.3342084. https://doi.org/10.1145/3341302.3342084.

[46] Hattie Joanne Elkins and Rachelle Lee Lanfair. 2017. ECFS filing detail. (September 25, 2017). Retrieved 05/31/2020 from https://www.fcc.gov/ecfs/filing/1092355565047.

[47] The Daily Yonder. 2018. Broadband speed: fcc map vs. experience on the ground. Retrieved 06/01/2020 from https://www.dailyyonder.com/broadband-speed-fcc-map-vs-experience-ground/2018/07/25/.

[48] Institute for Local Self-Reliance. 2018. Broadband competition in the rochester region: reality vs federal statistics. EN-US. https://muninetworks.org/sites/www.muninetworks.org/files/rochester-mn-policy-brief-2018.pdf.

[49] Ramakrishnan Durairajan and Paul Barford. 2016. A techno-economic framework for broadband deployment in underserved areas. In *Proceedings of the 2016 Workshop on Global Access to the Internet for All* (GAIA '16). Association for Computing Machinery, Florianopolis, Brazil, 7–12. isbn: 9781450344234. doi: 10.1145/2940157.2940159. https://doi.org/10.1145/2940157.2940159.

[50] Mariya Zheleva et al. 2013. The increased bandwidth fallacy: performance and usage in rural zambia. In *Proceedings of the 4th Annual Symposium on Computing for Development* (ACM DEV-4 '13) Article 2. Association for Computing Machinery, Cape Town, South Africa, 10 pages. isbn: 9781450325585. doi: 10.1145/2537052.2537060. https://doi.org/10.1145/2537052.2537060.

[51] 2019. American Community Survey (ACS). https://www.census.gov/programs-surveys/acs.

[52] 2020. AT&T Internet Plans - Home Internet - AT&T® Official Site. en. Retrieved 09/23/2020 from https://www.att.com/internet/.

[53] 2020. CenturyLink | Authentication. Retrieved 09/23/2020 from https://shop.centurylink.com/MasterWebPortal/freeRange/login/shop/addressAuthentication?cx=true.

[54] 2020. Charter Spectrum® Official - Internet Cable TV and Phone Service. Retrieved 09/23/2020 from https://buy.spectrum.com/buyflow/buyflow-localization?expiration=session.

[55] 2020. Order Internet Service | Consolidated Communications. Retrieved 09/23/2020 from https://www.consolidated.com/residential/internet/order-internet.

[56] 2020. Order Cox Services | Cox Communications. Retrieved 09/23/2020 from https://www.cox.com/residential-shop/order-cox-services.cox.

[57] 2020. Order Online. Retrieved 09/23/2020 from https://frontier.com/order-online/address.

[58] 2020. In Home Product Qualification. Retrieved 09/23/2020 from https://www.verizon.com/inhome/qualification?message=visitidempty.

[59] 2020. High Speed Internet Service Provider | Kinetic by Windstream. Retrieved 09/23/2020 from https://www.windstream.com/high-speed-internet#/.

[60] 2020. Xfinity Offers In My Area | Xfinity. en. Retrieved 09/23/2020 from https://www.xfinity.com/locations/in-my-area.

[61] 2016. Staff block estimates. Federal Communications Commission. Library Catalog: www.fcc.gov. (June 29, 2016). Retrieved 04/20/2020 from https://www.fcc.gov/reports-research/data/staff-block-estimates.

[62] 2020. National address database | US department of transportation. Retrieved 02/11/2020 from https://www.transportation.gov/gis/national-address-database/national-address-database-0.

[63] 2015. National Address Database Summit Final Report. Department of Transportation, (June 2015), 49. https://www.transportation.gov/sites/dot.gov/files/docs/1%5C%20NAD%5C_Summit%5C_Report%5C_0.pdf.

[64] 2018. National address database schema. (April 2018). https://www.transportation.gov/sites/dot.gov/files/docs/mission/gis/national-address-database/308816/nad-schema-v2.pdf.

[65] 2020. Postal addressing standards | postal explorer. Retrieved 04/17/2020 from https://pe.usps.com/text/pub28/welcome.htm.

[66] SmartyStreets. 2020. US Street Address API documentation - SmartyStreets. Library Catalog: smartystreets.com. Retrieved 07/14/2020 from https://smartystreets.com/docs/cloud/us-street-api.

[67] 2020. DPV | PostalPro. https://postalpro.usps.com/address-quality/dpv.

[68] 2020. USPS - Residential Delivery Indicator. https://www.usps.com/nationalpremieraccounts/rdi.htm.

[69] 2010. FCC area API. Retrieved 02/23/2020 from https://geo.fcc.gov/api/census/#!/area/get%5C_area.

[70] 2010. 2010 TIGER Line Shapefiles. Retrieved 08/20/2020 from https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2010&layergroup=Blocks.

[71] US Census Bureau. 2016. New Census Data Show Differences Between Urban and Rural Populations. EN-US. Library Catalog: www.census.gov Section: Government. (2016). Retrieved 05/24/2020 from https://www.census.gov/newsroom/press-releases/2016/cb16-210.html.

[72] Federal Communications Commission. 2019. Establishing new digital opportunity data collection. EN-US. (2019). https://www.fcc.gov/document/establishing-new-digital-opportunity-data-collection.

[73] 2020. Patsy - Describing statistical models in Python using symbolic formulas. https://github.com/pydata/patsy.

# A MAJOR ISPS AS LOCAL ISPS IN CERTAIN STATES

While major ISPs together cover over 90% of the population of the states that we study (according to FCC data), not every major ISP provides service to every state. In addition, in 8 of the 9 states, at least one major ISP provides coverage to only a small fraction of the state's total population. Querying BATs at scale is resource- and time-intensive, so for each ISP, we prioritize querying addresses in states where the ISP has the greatest coverage. In a state where a major ISP has lesser coverage, we do not collect BAT coverage data for that ISP in that state. We omit that state-ISP combination from our evaluation of per-ISP coverage overstatements (Section 4.1), and we treat the ISP as a local ISP in that state for our estimate of per-state coverage overstatements (Section 4.3). We use the same approach for the remainder of our analysis in Section 4 and Appendix I.

In Table 7, we present the states where each major ISP provides coverage according to Form 477 data, as well as whether we treat the ISP as a major or local ISP for that state. Where we treat an ISP as a local ISP for a state, we estimate the population covered by that ISP in that state by identifying census blocks that are covered in the Form 477 data, filtering out census blocks that are not in the NAD (and consequently not represented in our dataset), and then summing the population of the remaining census blocks. We also show the ISP's coverage in the state as a proportion of the population that has any broadband access, using the per-state FCC covered population estimates in Table 12 (where we do not filter census blocks based on BAT responses).

The evaluation in Table 7 shows that the gap in BAT data collection is small in comparison to state covered populations, and as a result, likely does not have a significant impact on our estimates of per-state broadband availability overstatements. In 13 of the 15 state-ISP pairs for which we do not collect BAT data, the gap in BAT data represents less than 2.6% of the state population with broadband access, and no gap exceeds 8% of the covered population.

The results in Table 7 also suggest that the gap in BAT data collection does not have a significant impact on our estimates of per-ISP coverage overstatement ratios. For each ISP, the population for which we do not collect BAT responses is small in comparison to the population for which we do collect BAT responses (roughly the population covered by the ISP according to FCC data in Table 3, but the value is slightly larger because the table omits certain census blocks). Furthermore, because we entirely omit state-ISP combinations that lack BAT data collection from our per-ISP coverage analysis—rather than making an assumption about coverage, as we do for our per-state analysis—the consequence for estimating per-ISP overstatement ratios is solely a smaller and less diverse sample of addresses where we measure BAT responses.

# B ALTICE AS A LOCAL ISP

We treat Altice as a local ISP in New York, even though it provides service in many areas of the state and has a public BAT on its website. In our initial testing, we found that Altice's BAT is very limited—it appears to return coverage based solely on ZIP code and only returns that an address is not covered for a minuscule proportion (0.2%) of addresses that are covered according to Form 477 data. Altice's BAT also does not specify when an address is unrecognized and it returns coverage for nonexistent addresses (seemingly based on ZIP code). Given these constraints, we have no reliable means of applying our methods to Altice, so we treat it as a local ISP (i.e., we assume in our analysis that Altice is 100% available in the census blocks it reports as covered to the FCC).

# C COVERAGE OF LOCAL ISPS PER STATE

Table 8 shows the proportion of the population in each state that is covered by local ISPs.

# D ISP-SPECIFIC BAT BEHAVIORS

We describe unusual ISP-specific BAT behaviors that we have encountered while conducting our study. We also describe how we addressed each behavior in our data collection.

**AT&T.** The AT&T BAT is an API that involves technology-specific queries—one query type for DSL and fiber, and another query type for fixed wireless. We submit both query types for our study, and we use the union of the two responses (i.e., if either indicates coverage, we treat the address as covered).

**CenturyLink.** The CenturyLink BAT has a response type ($Ce_4$) where the underlying API returns coverage with very slow speeds ($\leq$ 1Mbps). When this occurs, the website's user interface shows that the address is not covered. Since users would only see the address does not have service available, we treat the response type as not covered.

The CenturyLink response type $Ce_0$ appears to indicate an address is not covered, but we infer from carefully analyzing the response that it actually means the BAT does not recognize the address. We base our assessment on the following factors: the response occurs for test nonexistent addresses, the response appears to only occur when the BAT cannot autocomplete the address and the BAT's internal address ID is set to `null`, and the JavaScript that triggers the response includes the status string "We were unable to find the address you provided."

**Charter.** We are not able to reliably discern when the Charter BAT does not recognize an address. We found in testing that, when queried with nonexistent addresses, the Charter BAT responds with a generic request to call customer service. We are not able to locate additional information in the BAT response indicating *why* the prospective customer should call customer service. As a result, we follow our presumption of labeling response types based on what appears to the user, and we treat these responses as unknown.

As discussed in Section 3.5, a limitation in our Charter BAT client requires us to categorize certain responses as unknown, even though the website might have shown a different coverage outcome to the user when querying for the address. The client that we built for the Charter BAT queries an API endpoint and parses key coverage fields in the response. We found in subsequent testing that when the fields that our BAT client parsed were absent, the BAT could still present coverage or non-coverage to the user. Because our client did not retrieve the full Charter BAT webpage and did not store response information beyond parsing the main coverage fields, we are limited to treating all responses missing the fields as unknown.

**Consolidated.** The visual presentation of Consolidated's BAT changed during our data collection. We found instances where the BAT's underlying API would return that an address was not

| | State | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ISP | Arkansas | Maine | Massachusetts | New York | North Carolina | Ohio | Vermont | Virginia | Wisconsin |
| AT&T | • | | | | • | • | | | • |
| CenturyLink | • | | | 1 (0.00%) | • | • | | • | • |
| Charter | | • | • | • | • | • | 46,550 (7.96%) | 198,339 (2.48%) | • |
| Comcast | • | 97,812 (7.80%) | • | 107,885 (0.64%) | 6,226 (0.07%) | 102,440 (1.08%) | • | • | 74,919 (2.08%) |
| Consolidated | | • | 12,086 (0.18%) | 231,940 (1.38%) | | 40,414 (0.42%) | • | 19,957 (0.25%) | |
| Cox | • | | 3,358 (0.05%) | | | 241,703 (2.54%) | | • | |
| Frontier | | | | • | • | • | | | • |
| Verizon | | | • | • | | | | • | |
| Windstream | • | | | 174,239 (1.04%) | • | • | | | |

**Table 7: Data collection for the ISPs and states in our study. An empty cell is a state-ISP combination where the ISP does not provide service according to FCC data. A cell marked • is a state-ISP combination where the ISP provides service and our BAT client collected data. A cell with numbers indicates a state-ISP combination where we treat the ISP as a local ISP, because it provides comparatively less coverage. The count is the estimated state population in our study that the ISP covers according to FCC data, and the percentage is that count divided by the state population in our study that has broadband access according to FCC data. In the case of CenturyLink in New York, we manually confirmed that CenturyLink provides *residential* service to a single census block in New York (according to Form 477 data) that has a population of 1 (according to FCC estimates).**

| | Residential Addresses Covered by | | Population Covered by | |
|---|---|---|---|---|
| State | Local ISP ≥0 Mbps | Local ISP ≥25 Mbps | Local ISP ≥0 Mbps | Local ISP ≥25 Mbps) |
| Arkansas | 67.81% | 55.90% | 66.85% | 56.32% |
| Maine | 51.34% | 24.82% | 51.15% | 24.30% |
| Massachusetts | 30.43% | 30.40% | 28.31% | 28.26% |
| New York | 61.59% | 53.23% | 72.95% | 67.88% |
| North Carolina | 29.96% | 25.72% | 29.36% | 24.35% |
| Ohio | 53.31% | 43.17% | 54.04% | 44.07% |
| Vermont | 44.69% | 36.90% | 45.20% | 37.73% |
| Virginia | 35.15% | 17.94% | 32.40% | 15.91% |
| Wisconsin | 59.72% | 22.09% | 55.58% | 19.86% |
| All States | 46.53% | 36.05% | 50.30% | 40.88% |

**Table 8: The percentage of the population with access to broadband in each state or across all states in our dataset that is covered by a local ISP. We present coverage at any level of service (≥ 0 Mbps) and at benchmark broadband speed (≥ 25 Mbps). All States is a weighted average according to the covered population of each state.**

covered, but the webpage presented to the user would show either coverage for an unspecified "area" or a "temporary" error. Because the underlying API did not change since our reverse engineering, and we did not otherwise see evidence that the responses from the API had changed in meaning, we maintained our interpretation of the API responses.

**Cox.** When querying Cox's BAT for an apartment it will, in some instances, respond with 'too many suggestions" rather than a list of suggested units. In this scenario, our BAT client iteratively issues requests with common apartment prefixes (e.g., "APT," "1," and "A") to prompt the BAT to respond with suggestions. On the rare occasion when that approach was not successful, the BAT client noted the error, and we excluded the address from our analysis.

The Cox BAT does not clearly distinguish between addresses that are unrecognized and addresses that are not covered. When we query the BAT with either a nonexistent address or an address that is clearly outside of Cox's coverage area, the BAT returns that same response indicating that the address is not covered. We distinguish between addresses that are not recognized and those that are not covered by using SmartMove, a tool that the Cox BAT displays when an address may not be covered. SmartMove is the product of a marketing collaboration among broadband providers, and the tool enables prospective subscribers to check coverage across participating providers. Our BAT client queries SmartMove and evaluates whether the address is recognized. If SmartMove recognizes the

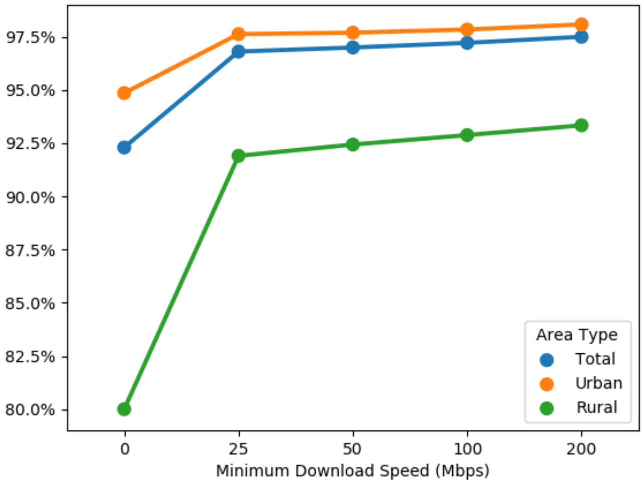


**Figure 7: The average coverage overstatements across ISPs and states for 0, 25, 50, 100, and 200 Mbps lower bounds on coverage speeds reported in Form 477 data.**

address, we treat it as not covered by Cox; if SmartMove does not recognize the address, we treat it as an unrecognized address for Cox. While the SmartMove tool is a step removed from Cox, it is the only available data source we have identified for distinguishing addresses that Cox does not cover from addresses the BAT does not recognize.

**Frontier.** The Frontier BAT, like the Charter BAT, does not provide a way to identify unrecognized addresses. We found in testing that when we supplied nonexistent addresses, the BAT responded with a generic error message and no indication of why the error occurred. As with Charter, we follow our presumption of labeling based on what appears to the user, and we treat these responses as unknown.

**Verizon.** The Verizon BAT, like the AT&T BAT, is an API that involves technology-specific queries—one query type for fiber and another for DSL. As with AT&T, we submit both query types and take the union of the results.

We found that, on rare occasions, Verizon's BAT returned different results for the same query address. We accounted for this issue by querying Verizon's BAT for each address twice, and if the results differed we treated the response as an unknown type.

The Verizon BAT's web user interface does not visually distinguish between addresses that are not covered and addresses that are not recognized—it appears that the address is not covered in either scenario. We are able to distinguish the two coverage outcomes by examining API responses. When an address is not recognized, the API does not suggest an address and corresponding ID (used to retrieve coverage information), and the response field `addressNotFound` is set to `true`.

**Windstream.** During our data collection, Windstream's BAT began returning a specific error message ($w_5$) for addresses that were previously returned as not covered. We could not find a case of an address previously returned as covered that also returns this error message. We called Windstream to check coverage for four addresses returning this error message, and the representative confirmed both that the addresses are not covered and that the BAT error message likely is the result of the addresses not being covered. We consequently treat this error message as an address not covered outcome.

## E FINAL TAXONOMY OF BAT RESPONSES

Table 9 presents our final taxonomy for mapping BAT response types to coverage outcomes. The table includes every BAT response type we identified across the nine ISP BATs we studied, along with the corresponding coverage outcome and an explanation.

## F BAT COVERAGE OUTCOMES BY ISP

Table 10 provides aggregate data about the coverage outcomes that result from our BAT client and response taxonomy during large-scale data collection.

## G RESPONSE TYPES FROM CENTURYLINK

Fig. 8 displays eight example response types from CenturyLink's BAT. We provide these images as an example of the range and variety of response types returned by each ISP's BAT, and to provide context into how we categorized these responses into an overall taxonomy.

## H OVERSTATEMENTS BY SPEED TIER

Fig. 7 plots coverage overstatements at varying speed thresholds according to the four ISP BATs for which we collect speed data (AT&T, CenturyLink, Consolidated, and Windstream). The results are aggregated across ISPs and states.

These results reinforce that overstatements disproportionately occur at lower speeds and lend weight to our hypothesis that network technology is a significant factor in overstatements. At speeds below 25 Mbps, 99% of ISP coverage is provided through DSL according to Form 477 data, in comparison to 21% at speeds greater than or equal to 25 Mbps (with fiber and cable being the main alternatives).

## I SENSITIVITY ANALYSIS OF OVERSTATEMENTS

We replicate the analysis in Section 4.3 while varying assumptions in three different ways, to understand how sensitive our results are to the assumptions we make.

**Unrecognized Addresses.** We repeat our analysis with a slightly less conservative method for counting addresses where at least one ISP's BAT returns that the address is unrecognized. In the main text analysis, we set aside these addresses. Here, we treat these addresses as not covered if BATs return a mix of address not covered and address not recognized responses.

The criteria for labeling an address as covered according to our BAT data are the same as in Section 4.3. The criteria for labeling an address as covered according to the FCC's data, if the address is not already labeled as covered according to our BAT data, expand as follows:

- There is at least one major ISP that does not cover the address according to its BAT.
- For every other major ISP that provides service to the address's census block according to Form 477 data, the ISP's

| ISP | Code | Outcome | Explanation |
|---|---|---|---|
| AT&T | $a_1$ | Covered | AT&T can and does service the address. |
| | $a_2$ | Covered | AT&T can service the address, but currently does not. |
| | $a_0$ | Not Covered | AT&T cannot service the address. |
| | $a_3$ | Unrecognized | AT&T does not recognize the address. |
| | $a_4$ | Unknown | The address in AT&T's response does not match the input address. |
| | $a_5$ | Unknown | AT&T returns the following error message: "Sorry we could not process your request at this time. Please try again later." (We retry this case multiple times). |
| | $a_6$ | Unknown | AT&T returns that it found a close match to the input address, but the returned address does not exactly match the input. |
| | $a_7$ | Unknown | Rare case where the BAT returns no information, which appears to be the result of a bug in the underlying API. |
| | $a_8$ | Unknown | Rare case where the BAT requests a unit selection for the address, but the only option is 'No - Unit', which upon selection, returns the same error. |
| | $a_9$ | Unknown | AT&T returns the following error message: "That wasn't supposed to happen!" |
| CenturyLink | $ce_1$ | Covered | CenturyLink can service the address. |
| | $ce_3$ | Not Covered | CenturyLink cannot service the address. |
| | $ce_4$ | Not Covered | The backend API in this case returns that CenturyLink can service the address, but with very low speeds (≤1Mbps). The browser interface displays that CenturyLink cannot service the address. We interpret this case as CenturyLink cannot service the address. |
| | $ce_0$ | Unrecognized | CenturyLink returns that it cannot service the address, but it is more likely it does not recognize the address (see Section 3.5 and Appendix D). |
| | $ce_2$ | Unrecognized | CenturyLink does not recognize the address. This case often occurs when the BAT suggests several addresses, none of which match the input. |
| | $ce_5$ | Unknown | The address in CenturyLink's response does not match the input address. |
| | $ce_6$ | Unknown | CenturyLink redirects to a page with the headline: "Contact Us." No coverage information is displayed. |
| | $ce_7$ | Unknown | CenturyLink either returns the error message "Our apologies, this page is experiencing technical issues" or returns that the inputted address is not valid. |
| | $ce_8$ | Unknown | Rare case where the page fails to load or redirects to the "Contact Us" page. |
| | $ce_9$ | Unknown | Rare case where the API requests a unit number, but responds to our follow-up request with the headline: "Error 409 Conflict." |
| | $ce_{10}$ | Unknown | Rare case where the API responds that it cannot find the input address, but then suggests the input address with seemingly random letters and numbers attached. |
| Charter | $ch_1$ | Covered | Charter can service the address. |
| | $ch_0$ | Not Covered | Charter cannot service the address. In this case, the BAT returns a simple prompt that the address is not covered. |
| | $ch_6$ | Not Covered | Charter cannot service the address. In this case, the BAT returns a more detailed prompt than in $ch_0$, providing the user with a number to call customer service. |
| | $ch_3$, $ch_4$ | Unknown | Charter prompts the user to call a number to "verify" the address. |
| | $ch_5$ | Unknown | A field in the API response (named "lines of service") is empty, giving an inconsistent output in the user interface. |
| | $ch_7, ch_8, ch_9$ | Unknown | A field in the API response (named "lines of business") is empty, giving an inconsistent output in the user interface. |
| Comcast | $c_1$ | Covered | Comcast can and does service the address. |
| | $c_2$ | Covered | Comcast can service the address, but currently does not. |
| | $c_0$ | Not Covered | Comcast cannot service the address. |
| | $c_3$ | Unrecognized | Comcast does not recognize the address. |
| | $c_4$ | Business | Comcast returns that the address is a business address. |
| | $c_5$ | Unknown | Comcast returns the error message: "Your order deserves a little more attention" and prompts the user to call a phone number. |
| | $c_6$, $c_7$ | Unknown | Redirects the user to the "Xfinity Communities" service. |
| | $c_8$ | Unknown | Comcast returns an error message that the address "needs more attention." |
| | $c_9$ | Unknown | None of the addresses suggested by the BAT match the input address. |
| | $c_{10}$ | Unknown | Rare case where the BAT requests a unit number, but the follow-up request fails. |

| | | | |
|---|---|---|---|
| Consolidated | $co_1$ | Covered | Consolidated can service the address. |
| | $co_0$ | Not Covered | Consolidated cannot service the address. |
| | $co_2$ | Not Covered | Consolidated cannot service the zip code of the inputted address. |
| | $co_3$ | Unrecognized | Consolidated does not recognize the address. |
| | $co_4$ | Unrecognized | None of the addresses that the BAT returns match the input address. |
| | $co_5$ | Unknown | The BAT suggests an address which matches the input, but the response to our follow-up request returns no information. |
| | $co_6$ | Unknown | The BAT repeatedly suggests our input address exactly, but never reports coverage information. Likely a bug. |
| Cox | $cx_1$ | Covered | Cox can service the address. |
| | $cx_0$ | Not Covered | Cox cannot service the address. Cox returns that it does not cover addresses it does not recognize, so we confirm this response type by querying the accompanying SmartMove API (see Section 3.5). |
| | $cx_2$ | Unrecognized | Cox does not recognize the address (confirmed by querying SmartMove API, see Appendix D). |
| | $cx_3$ | Business | Cox returns that the address is a business address. |
| | $cx_4$ | Unknown | An edge case where the BAT requests an apartment number and the client uses one of the suggestions but the BAT continues to request an apartment. |
| Frontier | $f_1$ | Covered | Frontier can and does service the address. |
| | $f_2$ | Covered | Frontier can service the address, but currently does not. |
| | $f_0$ | Not Covered | Frontier cannot service the address. |
| | $f_3$ | Not Covered | Frontier cannot service the address (a similar but distinct message from $f_0$ is returned). |
| | $f_4$ | Unknown | Frontier returns an ambiguous error message: "Don't worry - we'll get this sorted out." |
| | $f_5$ | Unknown | Frontier's API returns that an address is serviceable but does not give speed information. The user interface returns an error message to the user. |
| Verizon | $v_1$ | Covered | Verizon can service the address. |
| | $v_6$ | Covered | Verizon covers the address for Fios. This is a special case, where the BAT returns coverage data on the first request (usually, we need to send a request to another API with an address ID as a parameter to get coverage information). |
| | $v_0$ | Not Covered | Verizon cannot service the address. |
| | $v_3$ | Not Covered | Verizon cannot service the address (indicated after just entering the zipcode and not the full address). |
| | $v_2$ | Unrecognized | Verizon does not recognize the address. |
| | $v_4$ | Unknown | The address in Verizon's response does not match the input address. |
| | $v_5$ | Unknown | The BAT suggests addresses which do not match the input address. |
| | $v_7$ | Unknown | Rare case where Verizon continually prompts the user to "re-enter the address." This is likely a bug in their API. |
| Windstream | $w_0$ | Covered | Windstream can service the address. |
| | $w_4$ | Not Covered | Windstream cannot service the address. |
| | $w_5$ | Not Covered | An error message is returned that likely indicates Windstream cannot service the address (see Appendix D). |
| | $w_1$, $w_2$ | Unrecognized | The BAT returns the following error message: ""We still can't find your address. Contact us to see if you're in our service area." |
| | $w_3$ | Unknown | The BAT returns the following message: "Based on your address, call us to complete your order to receive the $100 online credit." |

**Table 9: The final BAT response taxonomy for our study, including the ISP, response type, coverage outcome, and an explanation.**

| ISP | Area Type | Address Covered | Address Not Covered | % Covered | Address Unrecognized | Business Address | Response Unknown | % Covered (excluding Business) |
|---|---|---|---|---|---|---|---|---|
| AT&T | All | 3,894,238 | 621,952 | 86.2% | 1,270 | 0 | 502,240 | 77.6% |
| | Urban | 3,587,448 | 415,981 | 89.6% | 1,111 | 0 | 434,207 | 80.8% |
| | Rural | 306,790 | 205,971 | 59.8% | 159 | 0 | 68,033 | 52.8% |
| CenturyLink | All | 1,521,772 | 121,754 | 92.6% | 192,453 | 0 | 199,694 | 74.8% |
| | Urban | 846,600 | 21,174 | 97.6% | 59,667 | 0 | 95,288 | 82.8% |
| | Rural | 675,172 | 100,580 | 87.0% | 132,786 | 0 | 104,406 | 66.7% |
| Charter | All | 8,495,464 | 184,676 | 97.9% | 0 | 0 | 1,455,116 | 83.8% |
| | Urban | 6,941,933 | 85,406 | 98.8% | 0 | 0 | 942,422 | 87.1% |
| | Rural | 1,553,531 | 99,270 | 94.0% | 0 | 0 | 512,694 | 71.7% |
| Comcast | All | 3,567,459 | 77,753 | 97.9% | 212,025 | 110,235 | 146,071 | 89.1% |
| | Urban | 3,169,744 | 48,255 | 98.5% | 104,718 | 106,826 | 139,848 | 91.5% |
| | Rural | 397,715 | 29,498 | 93.1% | 107,307 | 3,409 | 6,223 | 73.5% |
| Consolidated | All | 392,824 | 40,254 | 90.7% | 113,947 | 0 | 22,077 | 69.0% |
| | Urban | 200,376 | 6,833 | 96.7% | 41,582 | 0 | 908 | 80.2% |
| | Rural | 192,448 | 33,421 | 85.2% | 72,365 | 0 | 21,169 | 60.3% |
| Cox | All | 1,095,762 | 36,391 | 96.8% | 6,972 | 2,931 | 9,041 | 95.4% |
| | Urban | 1,033,844 | 27,738 | 97.4% | 2,844 | 2,646 | 7,855 | 96.4% |
| | Rural | 61,918 | 8,653 | 87.7% | 4,128 | 285 | 1,186 | 81.6% |
| Frontier | All | 1,045,938 | 79,698 | 92.9% | 0 | 0 | 333,805 | 71.7% |
| | Urban | 692,988 | 22,633 | 96.8% | 0 | 0 | 160,682 | 79.1% |
| | Rural | 352,950 | 57,065 | 86.1% | 0 | 0 | 173,123 | 60.5% |
| Verizon | All | 6,915,507 | 1,099,574 | 86.3% | 419,878 | 0 | 1,626,087 | 68.7% |
| | Urban | 6,520,060 | 626,687 | 91.2% | 312,913 | 0 | 1,545,566 | 72.4% |
| | Rural | 395,447 | 472,887 | 45.5% | 106,965 | 0 | 80,521 | 37.5% |
| Windstream | All | 451,512 | 24,015 | 94.9% | 15,461 | 0 | 74,296 | 79.9% |
| | Urban | 301,199 | 9,864 | 96.8% | 13,343 | 0 | 40,163 | 82.6% |
| | Rural | 150,313 | 14,151 | 91.4% | 2,118 | 0 | 34,133 | 74.9% |
| Total | All | 27,380,476 | 2,286,067 | 92.3% | 962,006 | 113,166 | 4,368,427 | 78.2% |
| | Urban | 23,294,192 | 1,264,571 | 94.9% | 536,178 | 109,472 | 3,366,939 | 81.8% |
| | Rural | 4,086,284 | 1,021,496 | 80.0% | 425,828 | 3,694 | 1,001,488 | 62.5% |

**Table 10: Aggregate coverage outcomes that result from our large-scale data collection and BAT response taxonomy. We present coverage outcomes by ISP and U.S. Census Bureau designation of whether census blocks are urban or rural.**

BAT returns that it either does not cover or does not recognize the address.

We then complete the analysis as in the main text. Table 11 presents the results.

Under the above set of assumptions, we find that the FCC may overstate access to benchmark broadband for 0.87% of the population in census blocks we study. If we naively extrapolate this result to the national population, we estimate that 2.6 million Americans have overstated access to broadband. For comparison, in the main text we provide an estimate of 2.5 million Americans with overstated access.

**Unrecognized Addresses and Unknown Responses.** We repeat our analysis with an aggressive method for treating addresses as not covered. In the main text, we only label an address as not covered if all major ISP BATs return that the address is not covered. Here, we treat an address as not covered if BATs return any mix of not covered, address unrecognized, or an unknown response.

We do not filter out any census blocks for this analysis. We do, however, discard Charter responses that are unknown and have a potential parsing error; we would otherwise count addresses as not covered by Charter owing to a limitation of our BAT client (see Section 3.5 and Appendix D).

The criteria for labeling an address as covered according to our BAT data are the same as in Section 4.3. The criteria for labeling an address as covered according to the FCC's data, if the address is not already labeled as covered according to our BAT data, expand as follows:

- For every major ISP that provides service to the address's census block according to Form 477 data, the ISP's BAT returns that it does not cover the address, that it does not recognize the address, or an unknown response.

We then complete the analysis as in the main text. Table 12 presents the results.

Under the above set of assumptions, we estimate that the FCC may overstate access to benchmark broadband for 4.61% of the population in the blocks we study. Extrapolated nationally, our analysis indicates that 14 million Americans have overstated access to benchmark broadband.

**Excluding Local ISPs.** We repeat our analysis while omitting local ISPs. In the main text, we treat local ISPs as covering all

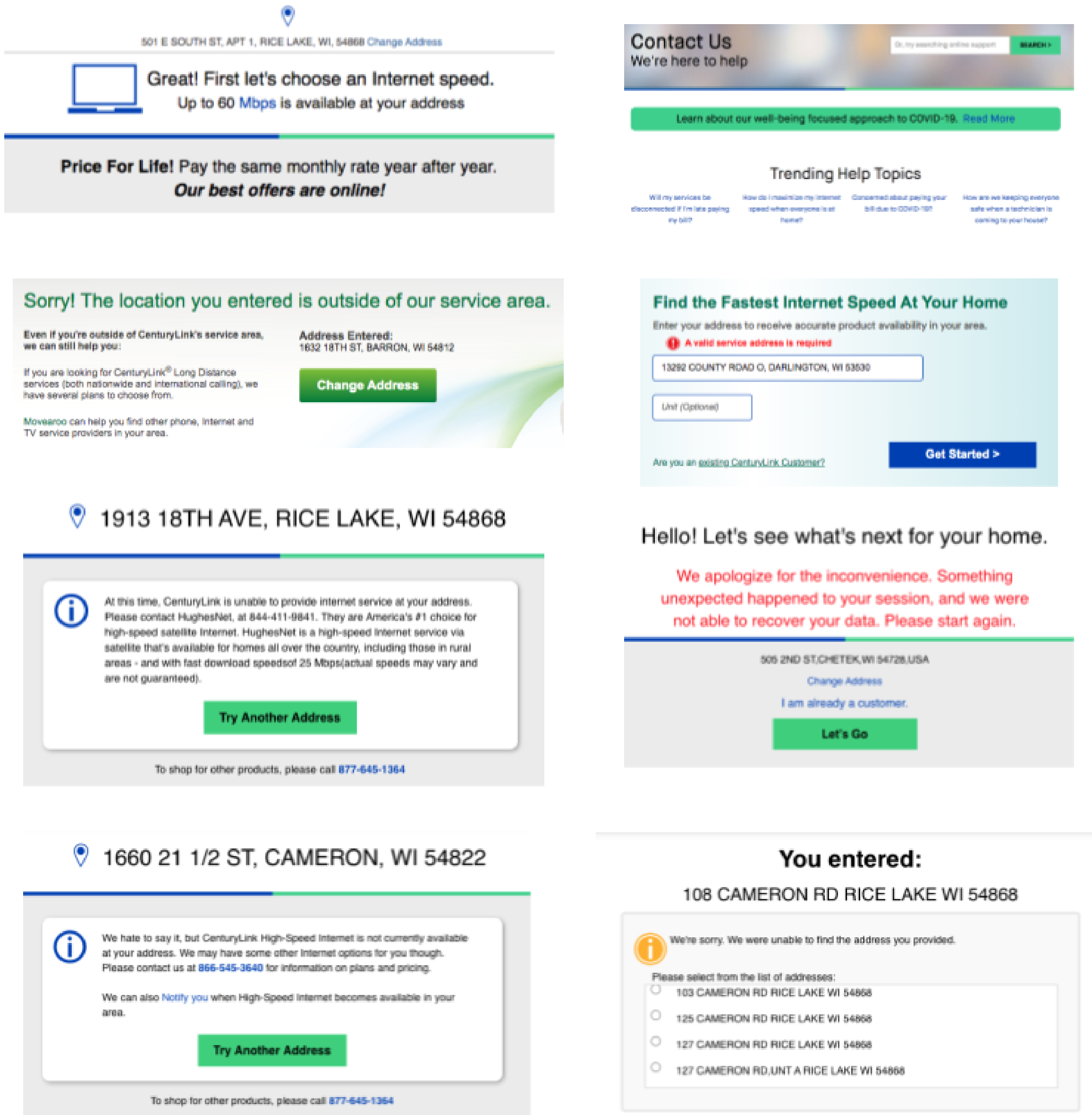


**Figure 8: A range of response types from CenturyLink's BAT. The images, counterclockwise from the top left, correspond to the following response types:** $ce_1$, $ce_0$, $ce_3$, $ce_4$, $ce_6$, $ce_7(a)$, $ce_7(b)$, $ce_2$.

addresses in the census blocks where (according to the FCC's data) they provide service. Here, we ignore local ISP coverage entirely.

The criteria for labeling an address as covered or not covered are the same as in Section 4.3, except that we do not label addresses as covered for having coverage from a local ISP.

We then complete the analysis as in the main text. Table 13 presents the results.

We find that under this set of assumptions, the FCC may overstate access to benchmark broadband for 1.37% of our studied population. This result naively extrapolates to over 4.1 million Americans. Notably, this analysis has a less significant effect than the previous analysis of treating unrecognized and unknown responses as indicating no coverage. These results reflect that the overwhelming majority of local ISP coverage overlaps with coverage by one or more major ISPs.

| | | Residential Addresses Covered by | | | | | | Population Covered by | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Any Provider ≥ 0 Mbps | | | Any Provider ≥ 25 Mbps | | | Any Provider ≥ 0 Mbps | | | Any Provider ≥ 25 Mbps | | |
| State | Area | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ |
| Arkansas | All | 739,805 | 729,909 | 98.66% | 645,726 | 639,270 | 99.00% | 1,932,493 | 1,873,920 | 96.97% | 1,660,074 | 1,634,718 | 98.47% |
| | Urban | 462,007 | 460,834 | 99.75% | 443,756 | 441,672 | 99.53% | 1,143,914 | 1,139,519 | 99.62% | 1,102,402 | 1,097,115 | 99.52% |
| | Rural | 277,798 | 269,075 | 96.86% | 201,970 | 197,598 | 97.84% | 788,579 | 734,400 | 93.13% | 557,672 | 537,602 | 96.40% |
| Maine | All | 430,954 | 429,611 | 99.69% | 410,027 | 407,412 | 99.36% | 1,236,848 | 1,219,474 | 98.60% | 1,203,988 | 1,178,744 | 97.90% |
| | Urban | 184,691 | 184,626 | 99.96% | 181,205 | 181,010 | 99.89% | 483,485 | 483,301 | 99.96% | 482,646 | 482,029 | 99.87% |
| | Rural | 246,263 | 244,985 | 99.48% | 228,822 | 226,402 | 98.94% | 753,363 | 736,173 | 97.72% | 721,342 | 696,715 | 96.59% |
| Massachusetts | All | 2,719,647 | 2,714,493 | 99.81% | 2,651,047 | 2,645,527 | 99.79% | 6,577,547 | 6,552,278 | 99.62% | 6,537,616 | 6,521,497 | 99.75% |
| | Urban | 2,536,347 | 2,534,470 | 99.93% | 2,477,106 | 2,472,768 | 99.82% | 6,056,469 | 6,050,699 | 99.90% | 6,043,862 | 6,031,881 | 99.80% |
| | Rural | 183,300 | 180,023 | 98.21% | 173,941 | 172,759 | 99.32% | 521,078 | 501,578 | 96.26% | 493,754 | 489,615 | 99.16% |
| New York | All | 4,573,104 | 4,564,326 | 99.81% | 4,410,500 | 4,399,536 | 99.75% | 16,721,122 | 16,665,017 | 99.66% | 16,489,475 | 16,438,497 | 99.69% |
| | Urban | 3,790,888 | 3,790,305 | 99.98% | 3,725,723 | 3,723,134 | 99.93% | 14,597,502 | 14,594,763 | 99.98% | 14,581,758 | 14,573,369 | 99.94% |
| | Rural | 782,216 | 774,021 | 98.95% | 684,777 | 676,402 | 98.78% | 2,123,620 | 2,070,254 | 97.49% | 1,907,717 | 1,865,128 | 97.77% |
| North Carolina | All | 2,970,502 | 2,960,050 | 99.65% | 2,823,600 | 2,812,650 | 99.61% | 8,430,254 | 8,368,746 | 99.27% | 8,076,866 | 8,019,544 | 99.29% |
| | Urban | 2,016,882 | 2,016,056 | 99.96% | 1,992,835 | 1,990,910 | 99.90% | 5,565,364 | 5,562,637 | 99.95% | 5,543,196 | 5,536,165 | 99.87% |
| | Rural | 953,620 | 943,994 | 98.99% | 830,765 | 821,740 | 98.91% | 2,864,890 | 2,806,109 | 97.95% | 2,533,670 | 2,483,379 | 98.02% |
| Ohio | All | 3,547,135 | 3,536,254 | 99.69% | 3,294,889 | 3,275,567 | 99.41% | 9,407,041 | 9,358,508 | 99.48% | 8,923,875 | 8,807,722 | 98.70% |
| | Urban | 2,844,848 | 2,843,454 | 99.95% | 2,772,159 | 2,767,529 | 99.83% | 7,426,004 | 7,419,002 | 99.91% | 7,345,456 | 7,313,509 | 99.57% |
| | Rural | 702,287 | 692,800 | 98.65% | 522,730 | 508,038 | 97.19% | 1,981,037 | 1,939,505 | 97.90% | 1,578,419 | 1,494,212 | 94.67% |
| Vermont | All | 188,734 | 187,396 | 99.29% | 180,621 | 177,577 | 98.31% | 583,811 | 574,879 | 98.47% | 566,641 | 550,487 | 97.15% |
| | Urban | 66,567 | 66,537 | 99.95% | 66,211 | 66,159 | 99.92% | 225,209 | 225,058 | 99.93% | 224,668 | 224,404 | 99.88% |
| | Rural | 122,167 | 120,859 | 98.93% | 114,410 | 111,418 | 97.38% | 358,602 | 349,820 | 97.55% | 341,973 | 326,083 | 95.35% |
| Virginia | All | 2,907,721 | 2,868,401 | 98.65% | 2,730,464 | 2,696,231 | 98.75% | 7,988,512 | 7,790,760 | 97.52% | 7,611,908 | 7,466,977 | 98.10% |
| | Urban | 2,164,094 | 2,156,536 | 99.65% | 2,128,788 | 2,111,827 | 99.20% | 6,143,691 | 6,111,510 | 99.48% | 6,092,066 | 6,016,475 | 98.76% |
| | Rural | 743,627 | 711,865 | 95.73% | 601,676 | 584,404 | 97.13% | 1,844,821 | 1,679,250 | 91.03% | 1,519,842 | 1,450,502 | 95.44% |
| Wisconsin | All | 1,162,669 | 1,160,654 | 99.83% | 1,076,300 | 1,072,800 | 99.67% | 3,587,043 | 3,575,743 | 99.68% | 3,417,686 | 3,395,697 | 99.36% |
| | Urban | 870,362 | 870,046 | 99.96% | 855,360 | 854,070 | 99.85% | 2,839,515 | 2,836,399 | 99.89% | 2,821,854 | 2,813,635 | 99.71% |
| | Rural | 292,307 | 290,608 | 99.42% | 220,940 | 218,730 | 99.00% | 747,528 | 739,344 | 98.91% | 595,832 | 582,061 | 97.69% |
| Total | All | 19,240,271 | 19,151,094 | 99.54% | 18,223,174 | 18,126,570 | 99.47% | 56,464,671 | 55,979,325 | 99.14% | 54,488,129 | 54,013,883 | 99.13% |
| | Urban | 14,936,686 | 14,922,864 | 99.91% | 14,643,143 | 14,609,079 | 99.77% | 44,481,153 | 44,422,888 | 99.87% | 44,237,908 | 44,088,582 | 99.66% |
| | Rural | 4,303,585 | 4,228,230 | 98.25% | 3,580,031 | 3,517,491 | 98.25% | 11,983,518 | 11,556,433 | 96.44% | 10,250,221 | 9,925,297 | 96.83% |

**Table 11: Analysis of address and population coverage overstatements for *any* broadband coverage, as in Section 4.3, but treating a mix of non-covered and address unrecognized BAT responses as non-coverage.**

## J OVERSTATEMENTS OF COMPETITION BY SPEED TIER

Fig. 9 presents the same competition analysis as in Section 4.4, but separated by the maximum speed offered in the census block (according to Form 477 data) rather than whether the census is urban or rural. We use speed tiers of any service (≥ 0 Mbps) and benchmark broadband (≥ 25 Mbps). We do not find significant differences in competition overstatements by speed tier.

## K FULL REGRESSION RESULTS

Table 14 shows the full results of the ordinary least squares regression analysis described in Section 4.5). We used the `patsy` statistical model library for analysis [73], which encoded away the categorical value for Arkansas because the regression model contains an intercept.

## L EXPLORATION OF POSSIBLE COVERAGE UNDERREPORTING

We conducted a small-scale evaluation in September 2019 of whether ISPs might *underreport* coverage to the FCC in Form 477 filings. We focused on the four major ISPs in Wisconsin: AT&T, CenturyLink, Charter, and Frontier.

For each ISP, we generated a set of test addresses by starting with the NAD address list, applying our NAD and USPS filtering for residential addresses, filtering by FCC coverage data for where the ISP does *not* provide service (rather than our ordinary filtering for where an ISP *does* provide service), and finally randomly sampling 1,000 addresses. We then queried each BAT with the ISP-specific set of test addresses.

We found that coverage underreporting appears to be rare in comparison to coverage overstatements: BAT responses indicated AT&T broadband service was available for 35 addresses, CenturyLink for 3 addresses, Charter for 0 addresses, and Frontier for 6 addresses. We also note that our evaluation likely magnifies the possibility of underreporting, because the FCC coverage data we rely on predates our BAT data collection. ISPs tend to increase their coverage footprints over time; that service is available now does not mean service was available for the mid-2018 Form 477 reporting deadline.

These results provide confidence that our focus on evaluating overstatements reflects the predominant source of inaccuracy in the FCC's coverage data.

We further note that underreporting and overstatements are likely inversely related due to Form 477 data's coarse census block granularity. An ISP that provides coverage to half the addresses in a census block necessarily overstates coverage if it claims the block, and underreports if it does not claim the block.

| | | Residential Addresses Covered by | | | | | | Population Covered by | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Any Provider $\geq$ 0 Mbps | | | Any Provider $\geq$ 25 Mbps | | | Any Provider $\geq$ 0 Mbps | | | Any Provider $\geq$ 25 Mbps | | |
| State | Area | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ |
| Arkansas | All | 756,992 | 729,909 | 96.42% | 660,519 | 639,270 | 96.78% | 1,964,753 | 1,835,883 | 93.44% | 1,675,744 | 1,601,332 | 95.56% |
| | Urban | 464,487 | 460,834 | 99.21% | 447,109 | 441,672 | 98.78% | 1,146,102 | 1,134,019 | 98.95% | 1,105,237 | 1,090,349 | 98.65% |
| | Rural | 292,505 | 269,075 | 91.99% | 213,410 | 197,598 | 92.59% | 818,651 | 701,863 | 85.73% | 570,507 | 510,982 | 89.57% |
| Maine | All | 441,784 | 429,611 | 97.24% | 428,554 | 407,412 | 95.07% | 1,254,082 | 1,191,933 | 95.04% | 1,233,127 | 1,131,354 | 91.75% |
| | Urban | 186,359 | 184,626 | 99.07% | 185,072 | 181,010 | 97.81% | 483,769 | 480,435 | 99.31% | 483,552 | 473,975 | 98.02% |
| | Rural | 255,425 | 244,985 | 95.91% | 243,482 | 226,402 | 92.99% | 770,313 | 711,498 | 92.36% | 749,575 | 657,378 | 87.70% |
| Massachusetts | All | 2,763,790 | 2,714,493 | 98.22% | 2,725,685 | 2,645,527 | 97.06% | 6,587,924 | 6,456,099 | 98.00% | 6,549,623 | 6,364,977 | 97.18% |
| | Urban | 2,573,640 | 2,534,470 | 98.48% | 2,543,256 | 2,472,768 | 97.23% | 6,062,585 | 5,969,892 | 98.47% | 6,053,248 | 5,894,624 | 97.38% |
| | Rural | 190,150 | 180,023 | 94.67% | 182,429 | 172,759 | 94.70% | 525,339 | 486,207 | 92.55% | 496,375 | 470,353 | 94.76% |
| New York | All | 4,657,443 | 4,564,326 | 98.00% | 4,529,800 | 4,399,536 | 97.12% | 16,772,294 | 16,441,992 | 98.03% | 16,581,939 | 16,132,460 | 97.29% |
| | Urban | 3,832,044 | 3,790,305 | 98.91% | 3,796,216 | 3,723,134 | 98.07% | 14,621,028 | 14,477,590 | 99.02% | 14,618,495 | 14,391,903 | 98.45% |
| | Rural | 825,399 | 774,021 | 93.78% | 733,584 | 676,402 | 92.21% | 2,151,266 | 1,964,402 | 91.31% | 1,963,444 | 1,740,556 | 88.65% |
| North Carolina | All | 3,017,317 | 2,960,050 | 98.10% | 2,880,614 | 2,812,650 | 97.64% | 8,492,648 | 8,234,807 | 96.96% | 8,184,945 | 7,859,526 | 96.02% |
| | Urban | 2,033,195 | 2,016,056 | 99.16% | 2,013,435 | 1,990,910 | 98.88% | 5,590,999 | 5,516,104 | 98.66% | 5,580,605 | 5,478,253 | 98.17% |
| | Rural | 984,122 | 943,994 | 95.92% | 867,179 | 821,740 | 94.76% | 2,901,649 | 2,718,703 | 93.70% | 2,604,340 | 2,381,272 | 91.43% |
| Ohio | All | 3,719,591 | 3,536,254 | 95.07% | 3,523,937 | 3,275,567 | 92.95% | 9,514,226 | 8,925,777 | 93.82% | 9,138,841 | 8,254,602 | 90.32% |
| | Urban | 2,972,571 | 2,843,454 | 95.66% | 2,931,898 | 2,767,529 | 94.39% | 7,491,025 | 7,092,713 | 94.68% | 7,458,473 | 6,913,641 | 92.70% |
| | Rural | 747,020 | 692,800 | 92.74% | 592,039 | 508,038 | 85.81% | 2,023,201 | 1,833,063 | 90.60% | 1,680,368 | 1,340,961 | 79.80% |
| Vermont | All | 190,395 | 187,396 | 98.42% | 183,784 | 177,577 | 96.62% | 585,098 | 569,526 | 97.34% | 569,030 | 541,321 | 95.13% |
| | Urban | 67,202 | 66,537 | 99.01% | 67,044 | 66,159 | 98.68% | 225,437 | 222,591 | 98.74% | 225,262 | 221,286 | 98.24% |
| | Rural | 123,193 | 120,859 | 98.11% | 116,740 | 111,418 | 95.44% | 359,661 | 346,934 | 96.46% | 343,768 | 320,034 | 93.10% |
| Virginia | All | 2,932,644 | 2,868,401 | 97.81% | 2,799,663 | 2,696,231 | 96.31% | 8,007,447 | 7,736,618 | 96.62% | 7,651,398 | 7,308,398 | 95.52% |
| | Urban | 2,175,091 | 2,156,536 | 99.15% | 2,153,083 | 2,111,827 | 98.08% | 6,151,726 | 6,084,169 | 98.90% | 6,116,882 | 5,953,124 | 97.32% |
| | Rural | 757,553 | 711,865 | 93.97% | 646,580 | 584,404 | 90.38% | 1,855,721 | 1,652,449 | 89.05% | 1,534,516 | 1,355,273 | 88.32% |
| Wisconsin | All | 1,182,089 | 1,160,654 | 98.19% | 1,109,522 | 1,072,800 | 96.69% | 3,602,734 | 3,524,764 | 97.84% | 3,449,816 | 3,305,291 | 95.81% |
| | Urban | 882,901 | 870,046 | 98.54% | 875,190 | 854,070 | 97.59% | 2,846,126 | 2,800,322 | 98.39% | 2,841,344 | 2,752,948 | 96.89% |
| | Rural | 299,188 | 290,608 | 97.13% | 234,332 | 218,730 | 93.34% | 756,608 | 724,442 | 95.75% | 608,472 | 552,343 | 90.78% |
| Total | All | 19,662,045 | 19,151,094 | 97.40% | 18,842,078 | 18,126,570 | 96.20% | 56,781,206 | 54,917,399 | 96.72% | 55,034,463 | 52,499,261 | 95.39% |
| | Urban | 15,187,490 | 14,922,864 | 98.26% | 15,012,303 | 14,609,079 | 97.31% | 44,618,797 | 43,777,835 | 98.12% | 44,483,098 | 43,170,103 | 97.05% |
| | Rural | 4,474,555 | 4,228,230 | 94.49% | 3,829,775 | 3,517,491 | 91.85% | 12,162,409 | 11,139,561 | 91.59% | 10,551,365 | 9,329,152 | 88.42% |

**Table 12: Analysis of address and population coverage overstatements for *any* broadband coverage, as in Section 4.3, but treating BAT responses of address unrecognized or response unknown as equivalent to a response of address not covered.**

| | | Residential Addresses Covered by | | | | | | Population Covered by | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Any Provider $\geq$ 0 Mbps | | | Any Provider $\geq$ 25 Mbps | | | Any Provider $\geq$ 0 Mbps | | | Any Provider $\geq$ 25 Mbps | | |
| State | Area | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ | FCC | BATs | $\frac{BATs}{FCC}$ |
| Arkansas | All | 704,098 | 657,375 | 93.36% | 458,612 | 443,473 | 96.70% | 1,737,485 | 1,622,444 | 93.38% | 1,173,245 | 1,133,090 | 96.58% |
| | Urban | 471,014 | 453,875 | 96.36% | 335,087 | 329,283 | 98.27% | 1,124,786 | 1,095,336 | 97.38% | 832,985 | 821,255 | 98.59% |
| | Rural | 233,084 | 203,500 | 87.31% | 123,525 | 114,190 | 92.44% | 612,699 | 527,108 | 86.03% | 340,260 | 311,834 | 91.65% |
| Maine | All | 428,485 | 420,241 | 98.08% | 398,063 | 391,180 | 98.27% | 1,208,548 | 1,179,960 | 97.63% | 1,154,211 | 1,122,931 | 97.29% |
| | Urban | 184,996 | 184,488 | 99.73% | 178,141 | 177,636 | 99.72% | 483,150 | 482,180 | 99.80% | 473,196 | 472,132 | 99.78% |
| | Rural | 243,489 | 235,753 | 96.82% | 219,922 | 213,544 | 97.10% | 725,398 | 697,779 | 96.19% | 681,015 | 650,799 | 95.56% |
| Massachusetts | All | 2,723,575 | 2,713,238 | 99.62% | 2,637,174 | 2,627,466 | 99.63% | 6,574,227 | 6,541,820 | 99.51% | 6,485,398 | 6,462,334 | 99.64% |
| | Urban | 2,538,980 | 2,533,661 | 99.79% | 2,466,345 | 2,458,088 | 99.67% | 6,054,496 | 6,042,907 | 99.81% | 6,000,103 | 5,981,703 | 99.69% |
| | Rural | 184,595 | 179,577 | 97.28% | 170,829 | 169,378 | 99.15% | 519,731 | 498,912 | 95.99% | 485,295 | 480,631 | 99.04% |
| New York | All | 4,601,600 | 4,472,569 | 97.20% | 4,016,657 | 3,977,084 | 99.01% | 16,489,302 | 16,156,683 | 97.98% | 14,923,944 | 14,762,900 | 98.92% |
| | Urban | 3,849,019 | 3,767,741 | 97.89% | 3,448,713 | 3,423,104 | 99.26% | 14,538,152 | 14,313,170 | 98.45% | 13,307,623 | 13,199,835 | 99.19% |
| | Rural | 752,581 | 704,828 | 93.65% | 567,944 | 553,980 | 97.54% | 1,951,150 | 1,843,513 | 94.48% | 1,616,321 | 1,563,064 | 96.71% |
| North Carolina | All | 2,876,195 | 2,854,456 | 99.24% | 2,629,306 | 2,613,488 | 99.40% | 8,170,949 | 8,091,685 | 99.03% | 7,593,573 | 7,527,625 | 99.13% |
| | Urban | 1,997,151 | 1,994,540 | 99.87% | 1,934,655 | 1,931,672 | 99.85% | 5,522,721 | 5,517,597 | 99.91% | 5,395,262 | 5,386,445 | 99.84% |
| | Rural | 879,044 | 859,916 | 97.82% | 694,651 | 681,816 | 98.15% | 2,648,228 | 2,574,088 | 97.20% | 2,198,311 | 2,141,179 | 97.40% |
| Ohio | All | 3,534,240 | 3,454,115 | 97.73% | 3,023,498 | 2,979,665 | 98.55% | 9,185,865 | 9,026,557 | 98.27% | 8,162,895 | 8,002,685 | 98.04% |
| | Urban | 2,869,963 | 2,830,488 | 98.62% | 2,606,938 | 2,587,581 | 99.26% | 7,395,721 | 7,330,815 | 99.12% | 6,908,185 | 6,850,175 | 99.16% |
| | Rural | 664,277 | 623,627 | 93.88% | 416,560 | 392,084 | 94.12% | 1,790,144 | 1,695,742 | 94.73% | 1,254,710 | 1,152,510 | 91.85% |
| Vermont | All | 184,451 | 177,746 | 96.36% | 173,984 | 167,015 | 95.99% | 556,967 | 538,082 | 96.61% | 536,466 | 513,079 | 95.64% |
| | Urban | 65,963 | 65,874 | 99.87% | 65,241 | 65,131 | 99.83% | 222,706 | 222,463 | 99.89% | 221,137 | 220,779 | 99.84% |
| | Rural | 118,488 | 111,872 | 94.42% | 108,743 | 101,884 | 93.69% | 334,261 | 315,619 | 94.42% | 315,329 | 292,299 | 92.70% |
| Virginia | All | 2,861,607 | 2,722,735 | 95.15% | 2,426,428 | 2,386,218 | 98.34% | 7,658,827 | 7,342,664 | 95.87% | 6,897,388 | 6,744,509 | 97.78% |
| | Urban | 2,159,307 | 2,136,456 | 98.94% | 2,029,291 | 2,009,828 | 99.04% | 6,099,763 | 6,049,009 | 99.17% | 5,854,770 | 5,776,725 | 98.67% |
| | Rural | 702,300 | 586,279 | 83.48% | 397,137 | 376,390 | 94.78% | 1,559,064 | 1,293,655 | 82.98% | 1,042,618 | 967,784 | 92.82% |
| Wisconsin | All | 1,102,943 | 1,084,211 | 98.30% | 978,270 | 972,267 | 99.39% | 3,398,953 | 3,357,060 | 98.77% | 3,157,875 | 3,130,063 | 99.12% |
| | Urban | 868,345 | 864,176 | 99.52% | 830,612 | 828,831 | 99.79% | 2,822,779 | 2,810,459 | 99.56% | 2,743,423 | 2,733,378 | 99.63% |
| | Rural | 234,598 | 220,035 | 93.79% | 147,658 | 143,436 | 97.14% | 576,174 | 546,601 | 94.87% | 414,452 | 396,685 | 95.71% |
| Total | All | 19,017,194 | 18,556,686 | 97.58% | 16,741,992 | 16,557,856 | 98.90% | 54,981,123 | 53,856,955 | 97.96% | 50,084,995 | 49,399,216 | 98.63% |
| | Urban | 15,004,738 | 14,831,299 | 98.84% | 13,895,023 | 13,811,154 | 99.40% | 44,264,274 | 43,863,936 | 99.10% | 41,736,684 | 41,442,427 | 99.29% |
| | Rural | 4,012,456 | 3,725,387 | 92.85% | 2,846,969 | 2,746,702 | 96.48% | 10,716,849 | 9,993,017 | 93.25% | 8,348,311 | 7,956,785 | 95.31% |

**Table 13: Analysis of address and population coverage overstatements for *any* broadband coverage, as in Section 4.3, but omitting local ISPs from analysis.**

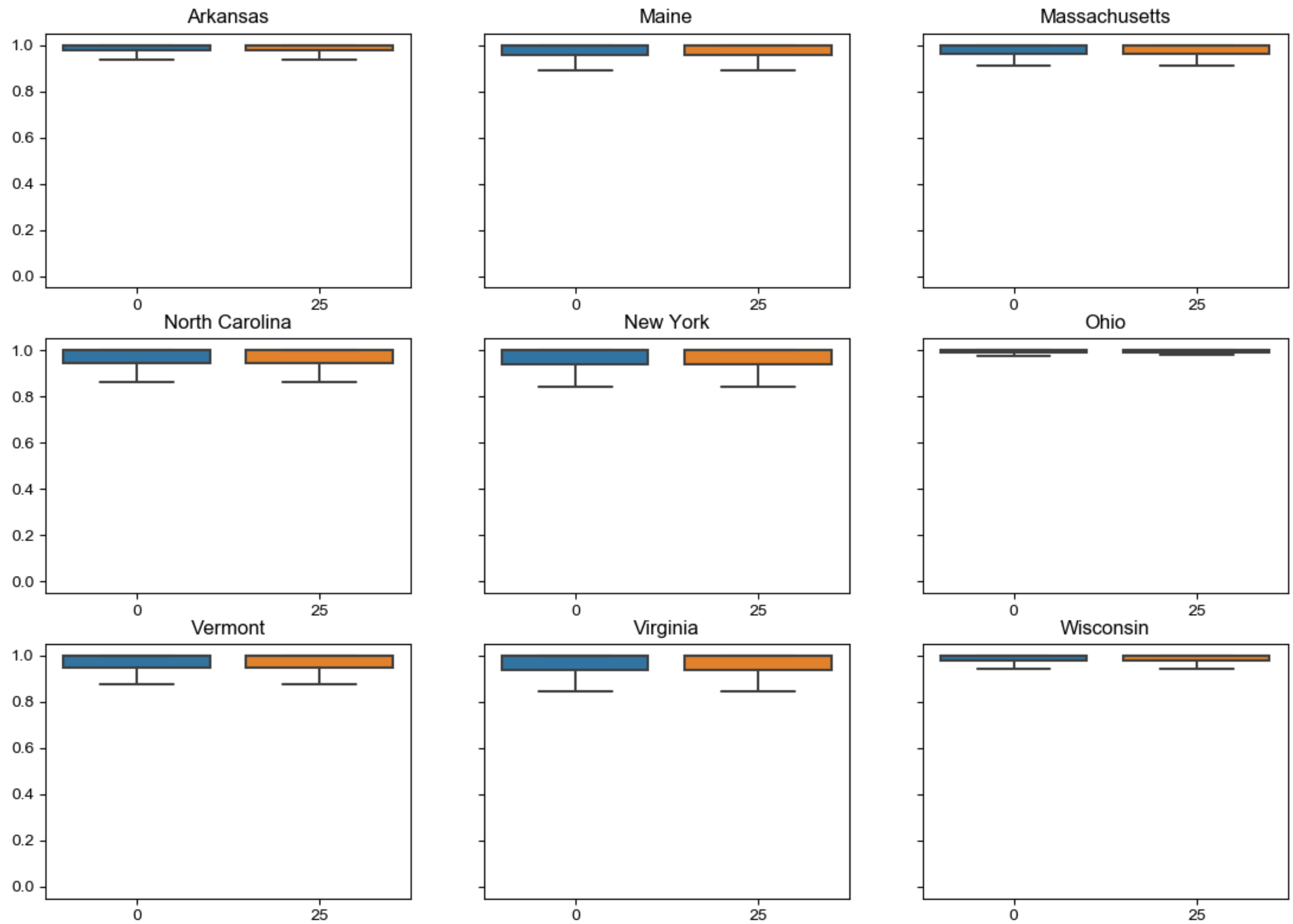


**Figure 9: The distribution of the competition overstatement ratio among census blocks, by state and speed tier.**

| Variable | Coeff | SE | P-Value |
|---|---|---|---|
| Intercept | 0.9891 | 0.004 | 0.000 |
| Maine | 0.0495 | 0.008 | 0.000 |
| Massachusetts | -0.0111 | 0.004 | 0.002 |
| New York | 0.0261 | 0.003 | 0.000 |
| North Carolina | 0.0290 | 0.003 | 0.000 |
| Ohio | 0.0255 | 0.003 | 0.000 |
| Vermont | 0.0306 | 0.008 | 0.000 |
| Virginia | -0.0208 | 0.003 | 0.000 |
| Wisconsin | 0.0299 | 0.003 | 0.000 |
| AT&T | -0.0161 | 0.002 | 0.000 |
| CenturyLink | -0.0106 | 0.002 | 0.000 |
| Charter | 0.0036 | 0.001 | 0.002 |
| Comcast | 0.0425 | 0.002 | 0.000 |
| Consolidated | -0.0383 | 0.008 | 0.000 |
| Cox | 0.0467 | 0.003 | 0.000 |
| Frontier | -0.0081 | 0.003 | 0.003 |
| Verizon | -0.0160 | 0.003 | 0.000 |
| Windstream | -0.0112 | 0.003 | 0.001 |
| Population Count | 3.427e-07 | 2.17e-07 | 0.115 |
| Poverty Rate | -0.0033 | 0.004 | 0.402 |
| Proportion Minority Population | -0.0065 | 0.002 | 0.00 |
| Proportion Rural | -0.0413 | 0.001 | 0.00 |

**Table 14: Full results for the ordinary least squares regression analysis described in Section 4.5 and Appendix K.**