

New collection entitled: [Residential Fixed Broadband Services Testing and Measurement](#)

Part B: Collections of Information Employing Statistical Methods:

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

The target population for this study is American household with broadband. The survey will seek reliable national estimates of broadband speed performance for the main US ISP (list below).

The research will provide a reliable national estimate of actual broadband speed performance per ISP for 34 services tiers, and a reliable estimation of performance by ISP packages within geographical region.

Based on recent experience in the UK, SamKnows will develop a panel of 10,000 broadband consumers' panelists covering 15 ISP, 34 packages within 4 geographical regions. Broadband speed will be collected through a hardware devices placed in consumers' panelists.

Internet Service Provider	Type	Estimated Subscribers [†]
Comcast	Cable	15,930,000
AT&T	DSL	15,789,000
Time Warner	Cable	9,289,000
Verizon	DSL	9,220,000
Cox	Cable	4,200,000
Charter	Cable	3,062,300
Qwest	DSL	2,974,000
Cablevision	Cable	2,568,000

Internet Service Provider	Type	-	Estimated Subscribers [†]
CenturyLink	DSL		2,236,000
Windstream	DSL		1,132,100
Mediacom	Cable		778,000
Frontier	DSL		635,947
Insight	Cable		501,500
Cable ONE	Cable		392,832
RCN	Cable		312,000
FairPoint	DSL		295,000
Cincinnati Bell	DSL		244,000

[†] Estimated at Q4 2009.

The various media channels used for recruitment have been selected so as not to correlate with the broadband performance (as long as the media channels reach the population without pre-selection from the media itself). Media used will be kept on record to ensure at the entirety of the respondents from a given media does not differ from the performances from respondents from other medias

- Note that broadband speed performance may be considered to correlate with
- Distance between premises and exchange,
- Contention in the ISP's own network particularly at peak time,
- Location density, such as rural or urban; typically in urban areas there are greater availability of higher speed services in urban area, further more in rural areas the average line length from local exchange to premises is longer than in urban areas.
- The technology used. There are different technologies available to delivered broadband services.

- Therefore although this sample is based on pool of volunteers, there is no good reason to believe that this particular sample would be different to a random sample as SamKnows control the region, U/S/R split and services tiers. The panelist behavior, as an individual, has no impact on the technical performance of broadband received nor on the main influencers of performances.

Source Quality and Size of Frame

The size of 10,000 has been designed to provide broadband speed performance at

- geographic region level;
- ISP's level; and
- service tiers level by region.

- To achieve requirement at regional service tiers level, we require a minimum of 125 panelists for each service tier present at regional level. There are 34 services tiers, which equates for the 4 regions to 80 services tiers, since not all 34 services are present in the 4 regions. Therefore the total sample of 10,000 is required (80 x 125).

Strata Definition and Proposed Allocation

- The sample is split using census 4 regions: West, Midwest, Northeast and South. For each services tiers, SamKnows will target 125 panelists per region when the service tiers is present (see table below, a bucket=125 panelists). Quotas will also be set on Urban-Suburban-Rural (U/S/R) continuum in order for each region to be representative of the Urban-Suburban-Rural universe per region. The classification of the data will be based on an analysis of the location of the exchange using the US Census bureau definition Urban-Suburban-Rural. Using the exchange location, each panelist will be categorized in terms of Urban-Suburban-Rural.

Within that primary split, we subgroup the service tiers further by region, with other subgroups relating to Density and Geographical region.

Region breakdowns are:

Northeast Region (including the New England division): Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont; and the Middle Atlantic division: New Jersey, New York, and Pennsylvania.

Midwest Region (including the East North Central division): Illinois, Indiana, Michigan, Ohio, and Wisconsin; and the West North Central division: Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota.

South Region (including the South Atlantic division): Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia; the East South Central division: Alabama, Kentucky, Mississippi, and Tennessee; and the West South Central division: Arkansas, Louisiana, Oklahoma, and Texas.

West Region (including the Mountain division): Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming; and the Pacific division: Alaska, California, Hawaii, Oregon, and Washington.

Overall Precision Requirement

Our overall precision requirements are an error margin of ±4.4% or ±8.8% (depending on whether or not the sample size is 125 or 500) with 95% confidence. (Typical minimum market research standards are a 90% confidence with an error margin of ±7.5%).

Sample error at service tiers by region, assuming the sample is comparable to a random sample, is 8.76% for 125 panelists, and 4.38% for 500 panelists at 95 % confidence.

This is calculated using the formula:

$$e = z\sqrt{(p\%(100-p\%))/s}$$

Where:

e = sampling error (the proportion of error we are prepared to accept)

s = the sample size

z = the number relating to the degree of confidence required

p = an estimate of the proportion of people falling into the group in which we are interested in the population

2. Describe the procedures for the collection of information including:

- *Statistical methodology for stratification and sample selection,*
- *Estimation procedure,*
- *Degree of accuracy needed for the purpose described in the justification,*
- *Unusual problems requiring specialized sampling procedures, and*
- *Any use of periodic (less frequent than annual) data collection cycles to reduce burden.*

Building on its experience of completing a similar project in the UK, SamKnows will recruit a US-based panel and deploy its technology throughout all fifty states to cover all regions.

Sample quotas will be set for 4 regions, West, Midwest, Northeast and South. The quotas, defined by geography, technology and service level, have been named buckets.

Within each buckets, soft quota will be set on density: large metro areas (A), suburban and mid market areas (B), exurban (C) and rural counties (D). The density classification of the data will be based on an analysis of the exchange location and Neilson Breakdown definition unless FCC uses a different definition.

SamKnows will use best practice in panel recruitment, which will be open to audit by 3rd party at anytime during the project. The panel recruitment will consist of two steps, using a multi mode recruitment effort to build a large pool of volunteers. This will provide flexibility in building the panel to ensure geographical representation, excluding outliers and as much as possible, maximizing the multi mode recruitment selection to avoid pitfalls such as selection bias, coverage error, and panel attrition.

SamKnows will report 'estimates' when accurate enough to be useful, accuracy is reflected by sample size and variance. In order to reflect the limitation of the sample SamKnows will show the speed results as an interval using the 95% interval around the mean. With a sample of up to 500 national wide panelists, the associated sample error is 4.4% and from experience the confidence interval for broadband speed interval is expected to be less than 0.5 Mbit/s.

It is to note that estimates for sub groups – including buckets – will be subject to higher sampling error margins based strictly on the smaller number of panellists in a specific subgroup e.g. for each bucket - 125 panelists the sample error is 8.8% and the expected confidence interval is around 0.8 Mbit/s.

Weighting

To insure representativeness of the US broadband population, and to obtain national average speed, the results will be weighted by density: large metro areas, suburban, exurban and rural counties and by region to reflect the penetration in region. Please note that we need to find the penetration data by region from third parties.

To compare ISPs' performances for DSL, we would normalise the data by distance to exchange.

Normalisation from distance to exchange is critical for DSL, because with this technology, speed degrades as the length of the line from the exchange to the premises increases. Therefore operator that have a higher proportion of customer in rural area where distance to exchange is typically higher, may be expected to deliver lower speed than those who focus on towns and cities because they have different customer profile. In order to normalise the distance, we will use the distance between the panelist location and their exchange.

A weight adjustment will be calculated to the contribution to the average speed made by each respondent based on their distance to exchange by matching the percentage observed in distance bands to the percentage in the distribution of the total sample distribution.

To reduce burden on the panelist, information technology is used extensively, all data collection will be automated after the initial installation of a hardware devices. The speed and performance will be monitored through the hardware devices in consumers' homes, to accurately measure the performance of fixed line broadband connections based on real-world usage. These hardware devices are controlled by a cluster of servers, which host the test scheduler and reporting database. The data is collated on the reporting platform and accessed via a reporting interface and secure FTP.

Further Information on Weighting Methodology

To compare ISP, SamKnows will normalize the data using the distribution of the straight line distance (as the crow flies) and then match this distribution to each ISP. This process will take in consideration outliers that may be excluded above a certain threshold. Normalization from distance to exchange is critical for DSL, because with this technology, speed degrades as the length of the line from the exchange to the premises increases. Therefore operators that have a higher proportion of customers in rural area where distance to exchange is typically higher, may be expected to deliver lower speed than those who focus on towns and cities.

Detailed Description of the Weighting Procedures and Formulas

In order to normalize the distance, we will use the distance between the panelist location and their exchange. A weight adjustment will be calculated to the contribution to the average speed made by each respondent based on their distance to exchange by matching the percentage observed in distance bands to the percentage in the distribution of the total sample distribution.

3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

Panel Recruitment

The recruitment strategy will solicit volunteers primarily through a media campaign using social and traditional media, such as consumer and technology press, alongside Twitter and independent bloggers and opinion formers. The FCC and SamKnows will both conduct this outreach. We are confident in doing this based on significant public interest in this study and past success that SamKnows had with this approach when conducting a similar project in the UK. These efforts would outline the project and direct interested volunteers to a URL where they could sign-up.

Once the volunteer panel is recruited and the SamKnows 'Whiteboxes' deployed, the tests will run on a pre-configured schedule, subject only to changes in the schedule and the volunteers own use of their broadband connection. Therefore, 'response rates' will always be at a 'maximum'. Details on the proprietary framework are provided below:

Test Scheduling

Tests will be run every hour on a randomized timing in the hour - 24 hours a day, 7 days a week. Within a month period, the software within the white box unit will perform over 700 separate speed tests. The unit is installed immediately in front of their home internet connection. This ensures that tests can be run at any time, even if all home computers are switched off.

Testing locking

No two tests may operate simultaneously. The first test to begin will have priority, and the other will block until the first finishes (or a timeout is reached).

Threshold Manager

Both before and during testing, the Threshold Manager analyzes the UDP & TCP packet data going over the WAN interface on the unit to check whether the Internet connection is in active use, The Threshold Manager is configurable by design, but is set by default at 400Kbps. If this threshold is breached before or during the tests, activity is and the threshold test is then repeated for one minute, for a maximum of 5 times until either the amount of traffic returns to a level below the threshold or the tests are suspended and that time period is marked as having a busy line.

Order of tests

The tests are run in the following order:

- ping - for all target hosts serially
- dns - for all target hosts serially
- www - for all target hosts serially
- single-threaded http get
- multiple-threaded http get
- single-threaded http post

Required number of data points

For reporting purpose, the data will be aggregated first per hour for the reporting period and then overall to minimize the effect of missing data. If less than 5 data points are recorded for one time slot within a month, the data will be discarded.

Tests

Latency / Ping and packet loss

This test uses ICMP pings to measure latency and packet loss to popular in country websites. Three target hosts will be tested, each receiving three pings (the first of which will be ignored). The round trip time for each is recorded individually, as well as the number of pings that were not returned.

Whilst ICMP packets may be dropped by routers under heavy load, their simplicity still provides one of the most accurate measures of latency. Extended periods of packet loss over many units on a single ISP may indicate a congested network, which would become another metric that could be tracked.

DNS resolution time and failure rate

DNS resolution of the ISPs' recursive DNS resolvers is tested by querying their DNS servers for popular USA websites. A standard "A Record" query is sent, and resolution time and success/failure results are recorded.

Two of the ISP's recursive DNS resolvers are tested directly from the monitoring unit. Queries are sent for three popular USA websites, with each DNS server being tested independently.

Note that these tests do not rely on the DNS servers set on the user's router.

Web page loading

This test fetches the main HTML body of a website. Note that additional resources, such as images, embedded media, stylesheets and other external files are not fetched as a part of this test.

The time in milliseconds to receive the complete response from the web server is recorded, as well as any failed attempts. A failed attempt is deemed to be one where the web server cannot be reached, or where a HTTP status code of something other than 200 is encountered.

We will use the home page of three popular country-hosted websites and tests will be run every hour against these. Note that tests were designed to ensure that pages were not cached.

HTTP test methodology

All of the tests are run against two or three in country managed servers dedicated purely to this task.

All servers to have at least 1Gbps connectivity and have diverse routes through a multiple of transit providers. All servers reside on networks that peer directly or indirectly within one hop to the core network

Units attempting to perform a speed test will request the target speedtest server from the data collection server, this allows us to remove speedtest servers from the equation should they be temporarily overloaded or down for maintenance. Under normal operation the process will round robin between the servers.

The speed test servers are all configured to return immediate content expiry in the HTTP headers, ensuring that compliant proxy servers should not cache the speedtest content.

The HTTP tests make use of a SamKnows designed test that includes a built in HTTP compatible client following RFC 2616

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.

The SamKnows tests and methodology have been evaluated and approved by Government Regulators, Internet Service Providers, leading Academics and researchers.

The stated service level name/identifier, headline speed, price and download limit will be used to check and assign panelist to the correct broadband packages. Different tests and checks will be held to confirm ISP and flag incorrect packages in the following manner:

At the recruitment stage

1. The ISP allocation will be validated using customer's IP against those held in ISP IP libraries. If inconsistent the record is flagged
2. Actual speed will be validated in relation to package and distance from premises to exchange distribution. Outliers will be excluded
3. Exclusion of household when distance to exchange is untypical to maximize the effective sample size after normalization

Test on continuous basis

4. Panelists have the ability to update their information though SamKnows web interface
5. Automatic flag of change of ISP and geographical location though the reporting system
6. Automatic flag of untypical maximum download speed

Independent Review: As there are a large number of stakeholders in this project, SamKnows has dedicated resource to liaising with interested 3rd parties who wish to participate in validating/enhancing the SamKnows methodology. These discussions are open to anyone who wishes to participate.

5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

[Jean-Michel Chapon](#)

Analyst, SamKnows

Email: jeanmichel@samknows.com

Cell: ~~+44 7710 126 525~~ [Dave Vorhaus](#)

[Expert Advisor, Federal Communications Commission](#)

Email: dave.vorhaus@fcc.gov

Phone: [202-418-3641](tel:202-418-3641)

Nathalie Sot

Statistician, SamKnows

Email: nathalie@samknows.com

Cell: +44 7715 484 803

[Dave Vorhaus](#)