Statistical methodology

Key statistical concepts used in this report

This report presents the findings from research which has involved the collection and interpretation of 1.94 million data points. It has been a complex process, both technically and statistically.

The glossary in the interactive report provides definitions of the technical terms we use throughout the report. However, knowledge of the following is important in order to understand how we have analysed the performance data collected.

- We present data in the report only in cases where there are sufficient data points to deliver a statistically sound result. This means that we report performance only when statistical analysis indicates that our findings are accurate enough to be useful. Accuracy is determined by the number of measurement tests undertaken, the size of the sample (number of panellists) and the variation (spread or range of results) between panellists.
- In order to acknowledge the limited accuracy of the estimates, and to ensure that we highlight only those differences that are statistically significant, for many charts we do not show a value but instead show a range around the mean value which indicates the statistical confidence we have in our results. The range we use is called a 95% confidence interval, which is a statistically derived range calculated from the standard error (which is itself calculated from the sample size and the variation within the sample). A 95% confidence interval means that if we repeated the research with a different sample, assembled in the same way, there would be a 95% probability that the mean value would be in the range shown. Where we have large samples and/or little variation within the sample, the confidence interval is much narrower than where we have smaller samples and/or large variation within the sample. Differences are reported as significant if they are significantly different as judged by a two-tailed 5% test of statistical significance. In the tables where we present differences which are statistically significant, we present differences which are significant to a 95% level of confidence, but also highlight those which are significantly different to a 99% level of confidence by using an asterisk.
- In order to ensure that the national data we present are representative of UK residential broadband users as a whole, we have weighted the data by ISP package, nation and rural/urban split, market classification, distance from the exchange for ADSL packages and max attainable speed for FTTC packages.
- We have similarly weighted the data where we are comparing the performance of individual ISPs' packages, in order to ensure that the analysis provides a fair comparison of actual performance rather than reflecting random differences in the ISP package customer profiles in the sample.
- To ensure comparisons between the performance of different technologies in urban and rural areas are representative of the urban and rural market, we applied a separate weight which gave a representative sample by technology, speed tier and urbanity when making these comparisons.

- A difficulty in comparing ADSL and FTTC broadband providers is that with this technology, speed varies by the length and quality of the specific consumer's telephone line. Therefore, providers which have a higher proportion of customers in rural areas, where line lengths are typically longer, may be expected to deliver lower speeds on average than those which focus on towns and cities, simply because they have a different customer profile. For FTTC customers, the critical part of the line is that between the customer's house and the cabinet – this section of the line is copper and subject to line degradation.
- To address this issue, we have taken the following steps:

  - For all ISP comparisons, we have included only consumers who live in an area where the exchange has been 'unbundled' by at least one LLU operator. This means that ISPs using wholesale services (such as BT Wholesale's IP stream or Wholesale Broadband Connect products) can be compared on a like-for-like basis with LLU operators.
  - We have excluded all ADSL customers where the straight-line distance from their home to the local telephone exchange is more than 5km, in order to limit the impact of outliers when weighting, and normalised data to straight-line distance distributions.
  - Straight-line distance weighting was applied only to ADSL operators in this report and not to cable or FTTC services, where performance is less influenced by distance from the exchange.
  - For FTTC customers, we do not have adequate information on distance between cabinet and customer premises. We therefore approximate this distance by normalising data using the maximum attainable speed. The maximum attainable speed is the best speed which a line can carry and is therefore a suitable proxy for measuring quality of the line.
  - No weights are applied to Virgin cable packages or FTTP packages, as these circuits are not affected by distance from the exchange or supplier cabinets.

Sample methodology

A panel of UK residential broadband users was drawn from a pool of volunteers following a recruitment campaign by SamKnows. The objective was to obtain a representative panel, in order to monitor the performance of residential fixed-line broadband in the UK over a two-year period of research. In addition to obtaining a panel sufficient for monitoring changes in overall performance, the panel was recruited to enable specific analysis of the performance of the most common ISP packages in the UK, in particular higher-speed packages.

The main purposes of this recruitment have been to:

- To replace panellists who leave the panel due to natural attrition, such as moving house or losing interest in participating in the research.
- Ensuring adequate samples for all ISPs and replacing panellists who although remaining on the panel have decided to switch their operator and/or package. As the Ofcom panel tends to be comprised of people with an interest in telecoms, there is a strong focus on ensuring coverage of lower speeds packages.
- To enable the ISP-level reporting of new packages such as high-speed fibre and cable as soon as sufficient numbers can be recruited.

Due to shortfalls in some areas, SamKnows provided additional data from its independent global platform, SamKnows One. At the moment, the total active panel is 2,945 – active meaning contributing results to either the ISP, or national analysis. 2,160 of these respondents belong to the Samknows panel and 362 to SamKnows' enabled embedded panel. Their data will be used for this report, but Ofcom seeks to recruit panellists unique to the Ofcom panel to fill these gaps.

The current active panel also excludes customers with packages with headline speeds of 2Mbit/s and less, because of the current low share of these connections (less than 0.1% of the total in November 2013). In our first round of research, conducted between October 2008 and April 2009, we found that the speeds delivered by 2Mbit/s and less packages were consistent over time and between providers. In this report we have excluded data from 2Mbit/s and less packages, due to their low market share.

Prior to dispatch of the monitoring units, volunteers were screened and preliminary speed measurements and checks on IP addresses were undertaken, in order to reduce the impact of respondent misconceptions regarding which package they were using on the sampling.

Definition of valid panellists and test volumes

All measurement data were collated and stored for analysis purposes as a monthly trimmed average of the measurements obtained for each respondent for the relevant time interval (e.g. 24 hours, 8 to 10pm weekday, 9am to 5pm Monday to Friday). Only panellists who provided a minimum of five valid measurements across all the download speeds, upload, latency, DNS and webpage tests for the narrow 8 to 10pm peak measurement period, and a good spread of tests throughout the day, were included in the monthly analysis. A trimmed mean was used because, for a small proportion of respondents, the occasional test result was far in excess of what was achievable on the line. The top and bottom 1% of results per respondents did not count towards the average.

The average number of measurements per respondent for the 24-hour multi-thread download speed tests between 1 November 2020 and 30 November 2020 was 292, from a theoretical maximum of 360 per respondent on the SamKnows panel, and a theoretical maximum of 720 per respondent on the embedded panel (i.e. if all panellists had their monitoring unit connected on 1 November and all scheduled tests were run - tests were not run when the monitoring unit detected concurrent use of the bandwidth).

Average download speeds are generally very accurately measured, so the main factors limiting the accuracy of the analysis reported here are the number of panellists and the average number of measurements.

Quotas and weightings

Quotas were set before the exact package market shares for operators were available, but results were weighted to be representative at national level. To recruit ISP packages to match the specific quota criteria above, and to achieve 100-150 panellists per package, only those ISP packages with more than 250,000 subscribers in total were targeted, although we do include ISP packages with less

than 250,000 subscribers where we can recruit sufficient panellists, and where we believe a package is important enough to the future development of the market to warrant inclusion in the report.

The results and analysis of the 2,945 panellists' measurement results were divided into three separate datasets, each weighted to targets.

- National panel (over 2Mbit/s packages): 2,147 panellists. All with at least five valid test measurements across all download tests, with a validated IP address, single measurement speed check, and distance and geographic market classification data.
- ISP package panel: 2,522 panellists. Respondents for this panel consist of panellists from geographic markets 2 and 3 only. Panellists from LLU operators Sky and TalkTalk, and cable provider Virgin Media were on-net only. There was a target of 100 valid panellists for each ISP package, but the criterion for inclusion in the reporting was an effective sample minimum of c.50 valid panellists (those with a base of fewer than 75 should be treated with caution). Additional validation for the ISP package panel included a review of measured speed against straight-line distance from the exchange to the panellist's premises, and a review of outliers. Any package reassignment identified was made to both the ISP package panel and the national panel datasets.
- Rurality panel: 2,180 panellists. All with at least five valid test measurements across all download tests, with a validated IP address, single measurement speed check, and distance and geographic market classification data. The sample composition of this panel is the same as the national panel.

Sample weighting

National panel:
- Weighting by ISP market and package shares by LLU/ non-LLU connections supplied by ISPs as at November 2020, urban/rural, geographic market classification, ADSL distance to exchange (fitted to UK representative exchange line distribution provided by BT Openreach) and max attainable normalisation for FTTC lines.

ISP package panel:
- Weighting to distance from exchange (those panellists with an unrecorded or straight-line distance to the exchange of more than 5km were excluded);
- ADSL2+ packages were normalised by distance from exchange, to the aggregated distribution of straight-line distance between premises and exchanges of all panellists on those headline packages
- FTTC packages were normalised to the appropriate max attainable speed curve that matched the headline package speed (36Mbit/s, 50Mbit/s or 63-67Mbit/s).
- Sky's 59Mbit/s package was weighted to the distribution of max attainable speed lines serving this package
- Cable or FTTP packages are not weighted, as speed of services is not directly related to distance from the exchange

Rurality panel:
- Weighting by package speed tier take-up within technology and within rurality. The weighting structure was formulated using June 2020 data provided to Ofcom by internet service providers.

- As mentioned previously, our measurement approach does not take account of respondent-specific issues, such as wiring, which may influence the speed of connection. Such variations have most impact on high-speed services where a respondent has a short line length. We assessed several methods of accommodating this issue and asked Saville Rossiter-Base for guidance.

- The conclusion was that allowing for variance across the sample based on line length would not necessarily lead to the widening of confidence intervals to build in this element of respondent variability. This is because the calculation of confidence intervals requires a constant mean and standard error across the sample or sub-sample under review. If we allow variance to differ by band, we would also need to allow the mean to differ by distance band. Leaving aside the increased complexity of the calculation, allowing the mean to differ by distance band to reflect respondent difference would reduce the variance in each band and reduce the confidence intervals for pooled estimate of the mean across the whole sample. The following calculation, based on all non-cable 20Mbit/s packages in May 2012, shows this to be the case.

Figure 1: Variation of mean and variance, by distance band

Distance Band	Sample	Mean	Variance	Standard Deviation
1	62	12.91482	13.95910	3.73619
2	68	11.60854	9.42604	3.07019
3	74	8.73505	10.31055	3.21101
4	78	5.87748	9.55572	3.09123
5	67	2.90284	5.73256	2.39428

Source: Ofcom

The average variance across the five cells is 9.8, giving a standard deviation of 3.1, giving a confidence interval of 8.48 +/- 0.3Mbit/s. But the overall standard deviation, if mean is held constant, is 4.7, which would give a confidence interval of 8.48 +/- 0.5Mbit/s. The current methodology therefore overestimates the variance in the sample and hence the confidence intervals.

Assigning panellists to ISP and broadband package

The following process was applied, to select panellists and assign them to the correct ISP package:

- Volunteer panellists were required to provide their ISP, package name, headline speed and download limit from drop-down menus and/or text boxes provided in an online form. This was used as initial categorisation of potential candidates against the target quotas.

  – The stated package name and headline speed (where they allowed identification of the correct ISP package) were used to assign panellists to an ISP package.

- Volunteers who matched the sample criteria were screened by ISP package, and an average speed-reading estimate was obtained to screen actual versus stated package. Those who were successfully screened were sent monitoring units.

  - The stated ISP allocation was validated against IP address. When an IP address and stated ISP were inconsistent or missing, the volunteer was rejected. When an average speed measurement was outside the feasible range, the volunteer was flagged, and a monitoring unit box dispatched if sample required for the assessed package.

- Once the volunteer correctly connected the monitoring unit and test measurements were received, straight-line distance from home to exchange and geographic market classification were added to the measurement data.

- A further stage of ensuring that respondents were assigned to the correct ISP package took place before the analysis stage. Four steps were undertaken:

  - The initial assumption was that the package assignment, recorded in the panel data file, was correct. However, the ISPs were asked to verify that respondents were on the correct package.

  - However, those participants whose stated and measured package assignments or ISP were not consistent, and could not be definitively reconciled, were excluded from the comparison data. Only those panellists with an ADSL connection, who were connected to an ADSL2+ enabled exchange, were considered for an ADSL2+ package allocation. The above modification (upload speed assignment) was necessary to identify those customers using ADSLMax on an ADSL2+ exchange.

Weighting to distance from exchange

As performance of ADSL broadband is significantly affected by the length of the line between a consumer's premises and the local exchange, any comparison between ISPs or technology could be affected by the distribution of distance among the sample.

It was therefore necessary to weight the data by distance from exchange in order to provide like-for-like comparison between the previously published data, to ensure that any differences identified were due to differing performance and not due to a differing distribution of line lengths. BT Openreach provides 3 curves which indicate the national distance profile of ADSL1, ADSL2+ and all ADSL lines for all lines in the UK. Each relevant ADSL2+ ISP package in the ISP panel is adjusted to match this national profile. ADSL packages in the national panel are adjusted to match the profile of all ADSL lines as provided by BT Openreach.

Distance from premises to local exchange was captured as the straight-line ('as the crow flies') distance, measured from the full postcodes of premises to the local exchange.

Weighting fibre packages

Although fibre technologies show little speed degradation between the local exchange and the final point where fibre is present, most respondents with fibre have FTTC only. This means that the length of the co-axial cable between the cabinet and the consumer premises can have a significant impact

on speed. As the FTTC network is being rolled out into more rural areas, the distribution of distance from the cabinet becomes important, as rural lines tend to be longer than urban.

In a similar manner as weighting to distance from exchange for ADSL, Ofcom has decided to normalise for distance from cabinet for FTTC products, to ensure a like-for-like comparison. An identical model to ADSL, based on straight-line distances from the cabinet, is not possible, as the relevant cabinet for many premises will be in the same postcode. Therefore, a proxy for distance from cabinet was used – this is maximum attainable speed. This is a network metric which assesses the line and determines the maximum speed it can carry. BT Openreach provided the maximum attainable speed for each panellist, and also the profile of fibre lines in the UK. Each ISP's respondent profile is adjusted to match the national profile and weighted accordingly to ensure like-for-like comparisons.

Ofcom uses a single curve for each speed, which does not discriminate between respondents with self and engineer installed lines.

Weighting efficiency

Overall, against the entire weighting framework, the national panel achieved a weighting efficiency of 41%. Since there are five factors making up the national weights, the weighting efficiency is often lower than the broadband provider and rurality panel weighting. It is also affected by the composition of our panel and can therefore vary year on year. The under-0.5s are primarily driven by the over-representation (against current market shares) both of some FTTC packages, and panellists in rural areas, and Scotland and Wales. The over-2s are driven by market shortfall for some packages, in particular those on ADSL technology, and for under-sampling in market A.

Figure 2: National panel range of weights

Weights	Count	Percentage
Less than 0.5	921	42.9%
0.5 to 1	481	22.4%
1 to 1.5	289	13.5%
1.5 to 2	107	5.0%
More than 2	349	16.3%

Source: Ofcom

Overall, against the entire weight frame, the ISP package panel achieved a weighting efficiency of 88%. This is because Virgin Media cable and FTTP packages are not weighted, as distance from exchange does not impair download speeds.

Figure 3: ISP package range of weights

Weights	Count	Percentage
Less than 0.5	148	5.6%
0.5 to 1	2009	75.5%
1 to 1.5	378	14.2%
1.5 to 2	90	1.4%
More than 2	36	3.4%

Source: Ofcom

Figure 4: Weighting efficiency, by ISP package

ISP Package	Weighting Efficiency	ISP Package	Weighting Efficiency
BT 36 Mbit/s	62.0%	TalkTalk 76 Mbit/s	95.6%
EE 36 Mbit/s	82.4%	BT 76 Mbit/s FTTP	100%
Plusnet 36 Mbit/s	71.1%	BT 160 Mbit/s FTTP	100%
TalkTalk 26 Mbit/s	91.7%	BT 330Mbit/s FTTP	100%
BT 52 Mbit/s	54.5%	Virgin Media 100 Mbit/s	100%
Sky 59 Mbit/s	80.3%	Virgin Media 200 Mbit/s	100%
BT 76 Mbit/s	96.9%	Virgin Media 350 Mbit/s	100%
EE 76 Mbit/s	93.0%	Virgin Media 500 Mbit/s	100%
Plusnet 76 Mbit/s	98.6%		

Source: Ofcom

Overall, the rurality panel achieved a weighting efficiency of 68%. The sub-samples for both urban and rural ADSL1, and for rural FTTC 50 and FTTC 63-67 packages were boosted to ensure these panellists did not need to be over-weighted too much.

Figure 5: Rurality panel range of weights

Weights	Count	Percentage
Less than 0.5	278	12.8%
0.5 to 1	1,107	50.8%
1 to 1.5	446	20.5%
1.5 to 2	259	11.9%
More than 2	90	4.1%

Source: Ofcom

Comparison of urban and rural speeds

Using Bluewave Geographic's Locale dataset, it is possible to segment all UK postcodes into one of seven urban-rural groupings. This dataset, widely used in market research design and sampling, allocates postcodes to a category based on their population density and how close the settlement they live within is to a larger one. The seven groupings range from A (large cities such as London and Birmingham), to G (isolated rural areas such as the Western Isles and Dartmoor).

To simplify the analysis, the groupings have been banded together into two broad groups: urban and rural (population less than 2.5k and in open countryside). This grouping enables us to compare rural and urban areas over time.