

A2. Statistical methodology

The 2021 Mobile Matters report analyses findings collected from panellists from the umlaut crowdsourced data between 1 January and 31 March 2021.

Overall panellists' records were included in the analysis if they met the following criteria:

- the wholesale mobile network was either EE, O2, Three or Vodafone;
- the location was in the UK and the UK nation was either England, Northern Ireland, Wales or Scotland;
- the location was also classified as a rural or urban area¹; and
- the data technology was known – either 2G, 3G, 4G, 5G² or wi-fi, depending on the metric.

Records that did not meet the above criteria were excluded from the analysis.

To estimate the number of panellists, and to check that there was a sufficient number of people behind each sub-group analysis, the distinct number of unique ids were counted for each dataset, each metric and sub-group analysed. We wanted to make sure there were at least 100 unique ids for the smallest sub-group. The actual number of unique ids for the smallest sub-group was higher than this for every metric.

For an overall count of the number of people in the data that had records that met the criteria above, the count of unique ids in the connection test dataset were looked at. The distribution of those counts are as follows:

Panellist counts

- Panellists with mobile records – 162,295
- Panellists with either mobile and/or wi-fi records – 286,523
- By nation (those with mobile records)
 - 137,783 panellists with records in England
 - 4,916 panellists with records in Northern Ireland
 - 13,825 panellists with records in Scotland
 - 8,627 panellists with records in Wales
- By rurality (those with mobile records)
 - 58,591 panellists with records in rural areas
 - 154,512 panellists with records in urban areas

¹ These locale classifications are derived from [Bluewave Geographics](#), which harmonises the rural and urban classifications between ONS (for England and Wales), Northern Ireland Statistics and Research Agency (NISRA, for Northern Ireland) and the National Records of Scotland (for Scotland). Records classed as 'SubUrban' in the 2021 dataset were grouped into the urban category to allow for year-on-year comparisons.

² For a small number of panellists who had a 5G enabled android smartphone and tariff.

- By nation and rurality (those with mobile records)
 - 47,123 panellists with records in rural England
 - 132,171 panellists with records in urban England
 - 2,663 panellists with records in rural Northern Ireland
 - 4,303 panellists with records in urban Northern Ireland
 - 6,332 panellists with records in rural Scotland
 - 12,583 panellists with records in urban Scotland
 - 4,228 panellists with records in rural Wales
 - 7,671 panellists with records in urban Wales
- By mobile network (those with mobile records)
 - 52,728 panellists on the EE network
 - 58,155 panellists on the O2 network
 - 28,660 panellists on the Three network
 - 26,926 panellists on the Vodafone network

Metrics and analysis methodology

For the 2021 Mobile Matters report the following metrics were analysed:

- Network share
- 5G connections
- Reliability of data connections
- Video experience
- Response times for data requests
- Data use
- Mobile voice calls

Network share

The data in the connection test dataset enabled us to look at what percentage of the time panellists were connected to each technology – 2G, 3G, 4G, 5G and wi-fi. Of particular interest is the proportion of data connections to cellular technologies versus wi-fi.

This was calculated simply by totalling the count of records for each technology and looking at this by location and by mobile network. Comparisons were carried out by using a two-tailed test to look for statistically significant differences which are shown in the report.

5G connections

Out of the 286,523 panellists, 1918 had a 5G-enabled device and made at least one connection to the 5G network during the data collection period when not also connected to wi-fi. Of the approximately 44,000 5G records generated, 99% were in urban areas, with just under half (45%) of all records being in London.

Average cellular network share was calculated for 5G users using the same methodology described in the network share section above. Comparisons were made between the total number of 3G, 4G and 5G records. The reliability of data connections and the response times for data requests were also looked at specifically for 5G. The methodology used was the same as stated in the relevant sections in this annex.

Reliability of data connections

A test runs every 15 minutes where the device attempts to download a small file and logs whether this can be completed successfully. To assess the reliability of data connections, we applied the filters stated above, plus an additional filter which selects records where the device screen is either on or off, and grouped the test results according to whether they were successful or unsuccessful. By using records where the mobile screen was on during this test, we can assess the percentage of cases when the user was able to connect to data services when the device was in active use. By using records where the mobile screen was *off* during this test, we can assess the percentage of cases when the user was able to connect to data services when they were not actively using their phone but may have still required a data connection e.g. to receive emails or stream music.

When there is such a large number of readings, even very small differences between averages or proportions can be statistically significant. However, this may not equate to a noticeable or practical difference for the consumer in their experience of using their phone. Differences were only highlighted in the report if they were statistically significant and the difference between sub-groups was at least 2 per cent.

The percentage of successful tests was calculated, and comparisons were made, using two-tailed statistical tests. Analysis was carried out comparing 3G, 4G and 5G data access technologies, mobile network operator, rurality, and nation to look for any statistically significant differences in the performance of data services.

Video experience metric

Estimated video experience is calculated by utilising a multitude of passively collected metrics per user, and categorised on a five-point scale from 'bad' to 'excellent'³. There are 618,223 records in the data from 83,347 users, this was filtered to include only records where the user had a 4G capable device and where a categorisation is given for the video experience, resulting in 493,639 records from 71,559 users. All four of the MNOs had sufficient records and sufficient users to make robust comparisons within nation, and within selected cities.

The percentage of records for each classification was calculated for each MNO within each classification. Comparisons of each classification were analysed by running two-tailed tests to see if any differences were statistically significant.

³ See [Annex 1: Technical Methodology](#) for further detail.

Response time for data requests

Response time, which is the delay in milliseconds between a consumer making a request to their mobile network for information and the network providing this information to the device, were compared by data access technology, rurality and nation. A two-tailed test of median response times was carried out on all sub-group comparisons.

Comparisons by mobile network were carried out using the same methodology to filter the data as for the reliability of data connections. A two-tailed test of the median response times was carried out on this subset of data.

Data use

We first calculated the average data downloaded for mobile technologies per person per day. As panellists were present in the data for a variable amount of days, we then multiplied the average daily amount by 30 to estimate a monthly data usage figure. Panellists were classified as Light, Medium or Heavy users dependent on their estimated monthly data usage. The following thresholds were used:

- Light usage: < 500MB of data downloaded per month
- Medium usage: \geq 500MB to 2.5GB of data downloaded per month
- Heavy usage: > 2.5GB per month

The amount of data downloaded was present in the network throughput dataset.

Comparisons by network operator were analysed by carrying out two-tailed tests to see if there were any statistically significant differences.

Mobile phone calls

We analysed data on the average duration of outgoing voice calls and the proportion of outgoing calls by call length. This data is captured in the voice call dataset. Data for each mobile call made by panellists during the fieldwork period was analysed.

The average call length for the fieldwork period was calculated and these averages were also then grouped to show the distribution of values by the percentage of calls. Differences in average call length were compared between nations, cities, urban and rural areas, and urban and rural areas within nations. These comparisons were analysed by running two-tailed tests to see if any differences were statistically significant.

Mapping Connection Test Success Rates

We mapped 3G, 4G, and 5G connection test success rates over the entire UK with data aggregated to a 10km x 10km grid, and 1x1km mapping for all UK cities whose population exceeds a population of 250 000, producing both static and interactive web maps. These maps cover the period 1 January to 31 March. We only used records where the mobile screen was active during the connection test, to assess the percentage of cases when the user was able to connect to data services when the

device was in use. The .csv data represents geographic data in the form of Ordnance Survey National Grid coordinate strings, whereas the .shp file also holds the geometry for use in GIS software.

We noticed that the data distribution is very skewed towards higher success percentage values, and for this reason we used a colouring scheme with non-uniform intervals. We decided not to colour areas where the number of tests was lower than 10 to avoid those areas to be characterised by a small number of tests and hence to skew our representation.

Mapping Response Time

We mapped 3G, 4G, and 5G median response times over the entire UK with data aggregated to a 10km x 10km grid, and to 1x1km mapping for all UK cities whose population exceeds a population of 250 000, producing both static and interactive web maps. These maps cover the period 1 January to 31 March. The .csv data represents geographic data in the form of Ordnance Survey National Grid coordinate strings, whereas .shp file also holds the geometry for use in GIS software.

As with the connection tests success percentage maps, we decided not to colour areas where the number of tests was lower than 10.