Smartphone Cities

Technical annex

Research Document

Publication date: 31 March 2016

Section 1

Our metrics

1.1 Introduction

This section is included as an explanation of how the results shown in the main body of this report have been captured. It covers the following topics:

- What parameters were measured
- How those parameters were measured;
- What equipment was used for the measurements;
- How the measurement results relate to how users of mobile broadband and voice calls experience the services available in the UK;
- Important points to note when reviewing the results obtained from our testing; and
- The reasoning behind the exclusion of some data from our results

1.2 The parameters we measured

We measured and reported fifteen distinct metrics:

- HTTP Download speed;
- HTTP Download success rate;
- "Ping" Latency;
- HTTP Upload speed;
- HTTP Upload success rate;
- Web browsing speed;
- Web browsing success rate;
- YouTube;
    - Time to first picture
    - Freezing time
    - Jerkiness
    - Playback Resolution
    - Video Quality
- Voice Call Setup Time;

- Voice Call Success Rate; and
- Voice Quality

It is considered that these metrics capture important aspects of mobile broadband network performance. For this reason the results are intended to be used as part of the information required by the consumer when making choices concerning mobile broadband provision. The following sub-sections explain each of the metrics in more detail:

HTTP Download speed

HTTP (hypertext transfer protocol) is the method commonly used to transfer information over the internet; for example, in delivering web pages or a video stream. Download speed indicates the rate at which a connection is able to transfer data from the internet to the consumer.

A connection with a higher download speed would take less time to transfer the same data than a connection with a lower download speed. For example, at a constant speed of 20Mbit/s (20 million bits per second), the theoretical time taken to download a 10MB (83,886,080 bits of data) file would be just over four seconds, while on a constant speed of 10Mbit/s, it would be just over eight seconds.

The rate of HTTP download can vary according to the location and time of day, even during a short session of use, for a multitude of reasons. Some of these effects may be related to the mobile network that is providing the connection to the internet such as contention on the cell the device is registered to, or the operator's use of traffic management systems, the handset being used to access the service or the SIM card tariff used by the consumer.

It can also be limited by factors outside the control of the mobile network operator. For example, if multiple users are attempting to access content from the same server at the same time, and that server lacks the capability to serve them all at the same time, the download rate could be limited.

HTTP is used to deliver many types of content, including web pages, audio, video, and images, as well as for downloading applications to a consumer's device. While HTTP is used for delivering various types of content, this content may be treated in different ways by the content providers and by the networks that transmit the information.

HTTP Download Success Rate

This provides an indication of the number of times a request to download a file was made through the test equipment versus the number of instances that the request to download was successful. This is expressed in the results as both a percentage and a count: e.g. if twenty requests for a download were made and nineteen were successful, the success rate is 19/20 or 95%.

Response Time

Response time (referred to technically as round-trip time or latency) indicates the delay between a request for information and the response. A connection with low latency will "feel" more responsive and certain applications perform far better with lower latency.

Latency was measured by sending a series of ICMP (internet control message protocol) "ping" tests. Latency refers to the responsiveness of a network and is measured as the time between sending a signal and receiving a response from the targeted system. An example to show the effect of latency is demonstrated through live satellite television news broadcasts, where a delay is sometimes seen between a presenter asking questions in a studio in the UK and the response from the reporter in a distant location.

Low latency is important for applications that require information to be delivered with as little delay as possible. In particular, low latency is most important when using services such as video calling, VoIP (voice over internet protocol) and online gaming.

HTTP Upload speed

As with download speed, upload speed indicates the rate at which a connection is able to transfer data from one device to another, although with upload speed this represents the rate at which data can be transferred from the handset to the remote Data Server.

HTTP Upload speed success rate

This is defined in the same manner as the HTTP Download success rate, however in this case referencing the success of the upload requests.

Web page loading speed

Web browsing speed indicates the amount of time it takes to completely load a given page hosted on a website. We chose to use two different pages for testing:

- The BBC homepage as an example of a commonly used webpage with dynamic content i.e. the actual content of the page changes on a regular basis.

- A standard HTML reference web page. The page used for testing is based on an ETSI (European Telecommunications Standards Institute) “mKepler” standard reference page, designed for smartphones to represent a typical static (i.e. unchanging) HTML web page.

For testing, a URL redirection took place as part of the page loading sequence for both websites. Each webpage was loaded completely, with no cache held on the devices to test network performance. Consumer devices typically cache parts of webpages to contribute towards a faster download experience.

Each approach has its own advantages and disadvantages. Using a standardised reference web page hosted on a dedicated server means that the conditions for downloading this page remain stable for each test, however as this page type is not typically accessed by consumers it could be viewed as unrepresentative.

Loading a web page with dynamic content (e.g. the BBC Homepage) does provide a more typical example of how a consumer would normally use a mobile device, however this also introduces another variable into the equation as the page content size is not guaranteed. There is also a possibility that unusually heavy use of the website (such as during a major breaking news story) may also affect performance. Regardless of the web page chosen, there was the possibility that the mobile operators would have optimised the data transferred to the handset as part of their normal operations.

Web page loading success rate

In the same manner as the HTTP Download and HTTP Upload success rate, this is calculated by comparing the number of successful requests against the total number of requests made.

YouTube (Video streaming)

The video clip used for testing was one of the most popular videos played in the UK at the beginning of testing in November 2015 - the film trailer for “Spectre”. It is important to note that the size and quality of a video file streamed is determined by the content provider depending on a number of factors including the device capabilities and the network capacity at the time that the content is being delivered. The network may choose to re-encode this content before providing it to the end-user. This is in order to create a more responsive experience by minimising the volume of data transferred and use of its network capacity. The decision by the operator to perform this will be reflected in the metrics recorded.

<u>Time to first picture</u>

This is defined as the time between making a “play” request to the content provider and the time the first image of the requested video clip is displayed on the device.

<u>Freezing time</u>

This is defined as the cumulative time that the requested video clip freezes during playback. As this is a cumulative statistic, a result of 1.5 seconds may be derived from one pause of that length, ten pauses of 0.15 seconds or three pauses of 0.5 seconds etc. Minor freezes that would not be noticed by a consumer are ignored by our measurement system.

<u>Jerkiness</u>

Jerkiness is a perceptual value that measures the loss of information from one frame to the next due to a freezing period or a low frame rate. The value reported is derived from the following measures:

- Freezing;
- Loss of information, estimated by the inter-frame difference;
- Dominating Frame Rate; and
- The amount of time an image remains visible until the image information changes in the next update. In the case of a constant frame rate, the dominating frame rate is equal to the constant frame rate

<u>Playback Resolution</u>

During a video session, there is a negotiation between the mobile device and the content provider to determine the optimum video resolution delivered to the mobile device. The resolution of the first picture to be displayed (for example 360p or 720p) is then recorded, giving an indication of the video resolution during the session.

<u>Video Quality</u>

This metric reports a number between 1.0 (bad) and 5.0 (excellent), derived from the ITU approved J.343.1 ‘RS-T-VModel’ algorithm which analyses the incoming video data for the video client and observes the resultant video at the same time. A video quality score is then calculated based on information taken from the video data, such as the size of a compressed frame, and the visual impression of the decoded and displayed video frame.

Voice Call Setup Time

This parameter is defined as the time taken between the mobile device initiating a call and the connection to the dialled number being completed, with the call being answered immediately on receipt.

Voice Call Success Rate

All our voice call tests were for a duration of ninety seconds. In the same manner as the other success rate calculations, this metric makes reference to the number of completed voice calls made versus the total number of attempts.

We have defined a "dropped" call as any call that terminates before the ninety second duration, excluding calls that stop due to intervention by our test engineers (i.e. stopping a measurement during a call).

Voice Quality

Similar to the Video Quality metric, Voice Quality is reported as a MOS (Mean Opinion Score) between 1.0 (bad) and 5.0 (excellent), using an implementation of the ITU T P.863 POLQA algorithm approved in January 2011. POLQA (Perceptual Objective Listening Quality Assessment) is known as a full-reference model: the quality estimation is based on comparing the transmitted signal with the high quality original reference signal.

We chose to run these tests in "half duplex" mode- measuring the voice quality in both directions in sequence between the measurement equipment and identical handsets controlled by two dedicated Voice Quality Servers situated at Ofcom's Engineering Hub in Baldock, Hertfordshire.

1.3 How these metrics were gathered

We collected our data using a proprietary software testing application, running directly on the consumer handsets we used for testing. A tablet device was also used to control the handsets and to observe tests as they were running.

1.4 Comparing our metrics to real services and applications

Our tests measure network performance. These types of metrics are often described as Quality of Service (QoS) parameters. Consumers use different applications on their mobile devices, and their perception of the experience of using their applications and services is what matters most to them. How a consumer rates their experience of a mobile service is referred to as Quality of Experience (QoE).

To make some of the metrics easier to understand we would ideally like to translate our QoS metrics into QoE metrics. A simple demonstration of this might be:

QoS: "this service has a latency of 32ms on average",

QoE: "this service is good for VoIP traffic on average".

In reality, it is not always possible to produce this type of simple translation. This is because many factors can affect the performance of a particular application: e.g.:

- Device capability- certain types of "demanding" applications such as gaming or playback of HD video may be limited by the device.

- Network Optimisation- some Mobile Operators may have taken the decision to optimise certain types of data traffic through their network to manage capacity demands and deliver a better service to more customers. The type and level of optimisation applied by the Operators may change for different circumstances and for different users.

- Application Servers- The servers that provide data supporting applications may lack the capacity to deal with high demand, resulting in reduced performance.

There has not been an attempt to translate our results into an explicit indication of how our metrics might represent consumer QoE for these reasons. The ways in which online services and content are provided over networks differ by provider, and the way the traffic is handled differs by network. To create a consistent approach to modelling our results, which is fair to all providers, is not feasible.

An example of these differences is in streaming video. This is a service area that attracts significant investment by MNOs, to deliver the best possible service to as many consumers as possible. Video streaming demands high bandwidth (i.e. fast data rates) and MNOs undertake a variety of activities to improve the service they offer to their customers. This includes re-encoding video to reduce the bit rate, caching popular videos at the edge of the network to deliver faster response times, and optimising content for device screen size and resolution.

This all means that it is not possible to assess video performance by simply looking at the maximum download rate achievable on a network, hence our decision to test the YouTube application.

In general terms, a higher download rate should deliver a better video streaming experience. For most mobile video applications a data rate beyond 2 to 3Mbit/s should be enough to ensure a high quality viewing experience, and increases in speed beyond this will generally have limited impact on viewing experience. As videos are viewed for a period of time it is not just the speed that is important, but the ability of that rate to be sustained over the duration of the video, and potentially while the user is on the move.

1.5 General Rules for using our results

When reviewing our results it is important to note the following points:

- Indication of better performance: the highest number is not necessarily always the best for all mobile device applications, as the other factors discussed in previous sections can also affect performance.

- Relative difference: even a statistically significant difference may have no user-perceivable impact on the consumer’s QoE.

- Performance can vary between devices.

- The number of 4G subscribers is increasing, which may lead to higher network load and a reduction in performance. Conversely, as 4G networks are further optimised, performance may improve.

1.6 Exclusion of data from our calculations

All the tests were carried out using the test equipment specified earlier to gather data as well as process the results. The outcomes of the tests were broadly classified into

a) Successful – where the test completed successfully.

b) Failed – where the test failed due to various reasons pertaining to the network

c) Aborted – where the tests were unsuccessful in completion due to user intervention or test equipment behaviour.

d) System release – where the individual session was terminated by the control software as part of setting up a particular test.

Both the "Aborted" and "System Release" outcomes were excluded from our results as their inclusion would adversely affect our results.

Section 2

Testing Methodology

2.1 The principles of our methodology

Our methodology has been designed to measure metrics relevant to the consumer experience of using mobile broadband and voice services. It has also been designed to produce a statistically robust dataset that treats each MNO equally. This is to allow us to compare the performance of each MNO's network on a fair and equivalent basis.

In previous reports on measuring mobile broadband performance in the UK (the first report was published in November 2014 and the second report in April 2015.) we specifically measured the 4G and 3G networks separately at a fixed number of indoor and outdoor locations across the test city. Whilst this was appropriate during the early stages of 4G rollout we now consider that there is widespread 4G coverage, along with a significant number of multiband handsets and 4G tariffs in use. This enables us to move to a 'Best Bearer' approach where the handset/network will choose the best service automatically between 4G, 3G or 2G depending on what is available in that area. However this does mean that comparisons with results gathered for this report are not comparable with those presented in our first two reports.

We also moved away from purely static measurements to also include measurements made whilst driving and walking. These changes were made to provide more representative tests akin to actual consumer use and experience. Nominally, each city comprised of 50% driving, 40% walking (both indoor and outdoor) and 10% static measurements.

Driving measurements were conducted with the measurement equipment placed in a roofbox attached to a vehicle. Walking measurements were a mixture of indoor (e.g. Shopping Centres) and outdoor measurements along public streets. Static measurements were also a mixture of indoor and outdoor measurements.

2.2 Device-based testing

MNOs undertake testing of handsets on their networks before including them in their product range, and so it was important to use a device that would be compatible with all of the networks.

The testing for this report was carried out using Samsung Galaxy S5 (Model G900F) handsets. This was one of the the most popular consumer devices at the time we carried out our tests and was thus reflective of the typical consumer experience of mobile broadband and voice performance.

Handset choice is of course one factor that can affect the network performance experienced by consumers. Early adopters with the latest handsets are likely to experience improved performance as they can benefit from the latest network developments, such as the deployment of carrier aggregation technology from EE and Vodafone and Three's introduction of VoLTE services. Similarly, consumers using devices with older technology may experience poorer performance.

The devices that we used were not MNO-branded devices; i.e. they were not purchased directly from the MNOs. MNO-branded devices generally have firmware pre-installed which is tailored to their network. There is also the possibility of minor customisation to hardware specifications. To allow us to achieve comparable measurements, and to test every network under the same conditions, we did not want to use MNO-branded devices as this would mean that each handset could have been modified in a different way, and would therefore perform in a slightly different way. Using MNO-branded devices would also have precluded us from rotating SIMs across the handsets.

This SIM rotation is important for removing differences between handsets due to manufacturing tolerances. We recognise that there may be small differences in network performance between a branded and unbranded handset, but we consider that the benefits of handset rotation and treating each network equally outweigh any benefits of using branded handsets.

We did not set out to measure the relative performance of different makes or types of device. We wanted to test network performance, therefore as many elements as possible, including the devices used, were set as constants across the testing to maintain comparability across networks.

2.3 Backend data services

To ensure that upload and download transfers from the test handsets to the media server was not the constricting factor in our testing, a high performance server with a 10 Gbit/s IP transit connection was commissioned and tested.

2.4 Data collection tool

The “QualiPoc” application developed by Swissqual was loaded onto each of the handsets and the control tablet to carry out our testing. Swissqual is a provider which specialises in services and systems for measuring, analysing and reporting performance of mobile devices and network services.

The system used for the collection of our results consisted of a backpack containing four handsets and a scanning receiver. The control tablet is used to configure the test setup of the slaves and scanner and to monitor test progress.

At the end of each measurement day, the result data was transferred from the handsets to the control tablet and then copied to cloud storage for retrieval and analysis.

2.5 Measurement locations

We chose to carry our tests in 5 major cities across the UK: Cardiff, Edinburgh, Liverpool, London and Norwich

In each city, our testing area was defined by a 4km radius, using the major rail station as the centre point, with Charing Cross station used for London. A 4km radius was chosen to provide commonality with previous research.

Figure 1 - Test city centre points and dates

	Centre point	Dates of testing
Cardiff	Cardiff Central	16 November – 27 November
Edinburgh	Edinburgh Waverley	2 November – 12 November, 9 December - 12 December
Liverpool	Liverpool Lime Street	23 November – 4 December
London	Charing Cross	30 November – 11 December
Norwich	Norwich Station	9 November – 20 November

The 4km radius for each city was divided into 10 segments to enable an even distribution of measurement results where possible across the city, with each segment covered by a measurement team per day.

Figure 2: Cardiff test area

Figure 3: Edinburgh test area

Figure 4: Liverpool test area

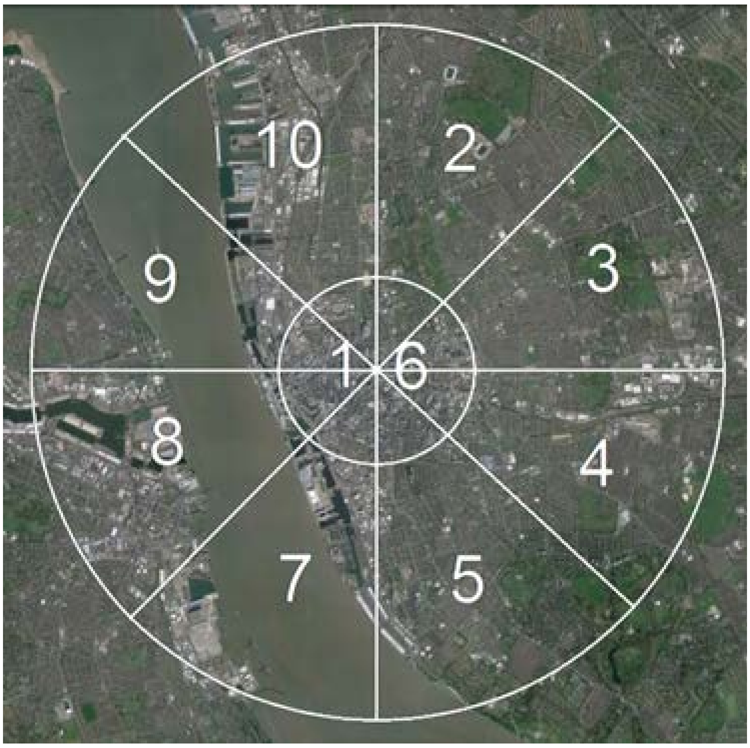


Figure 5: London test area

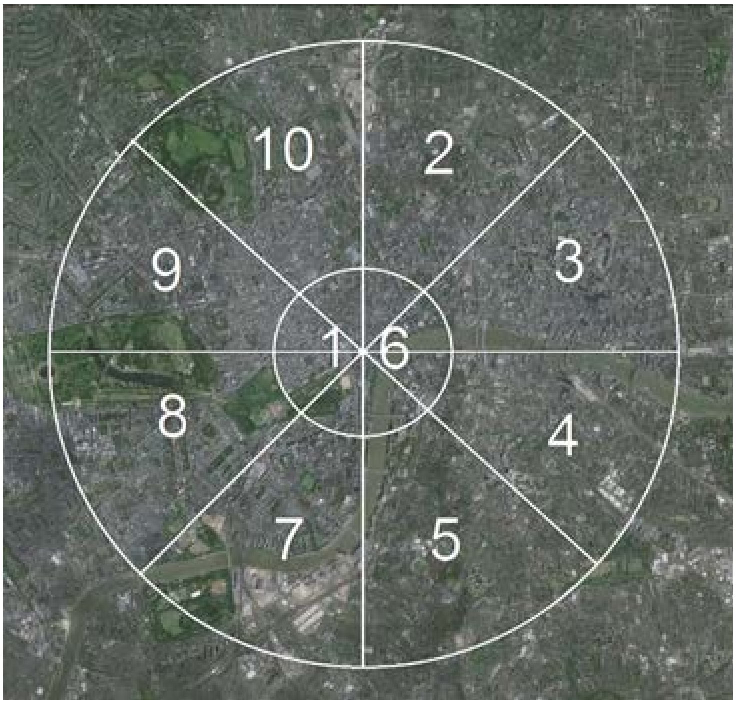
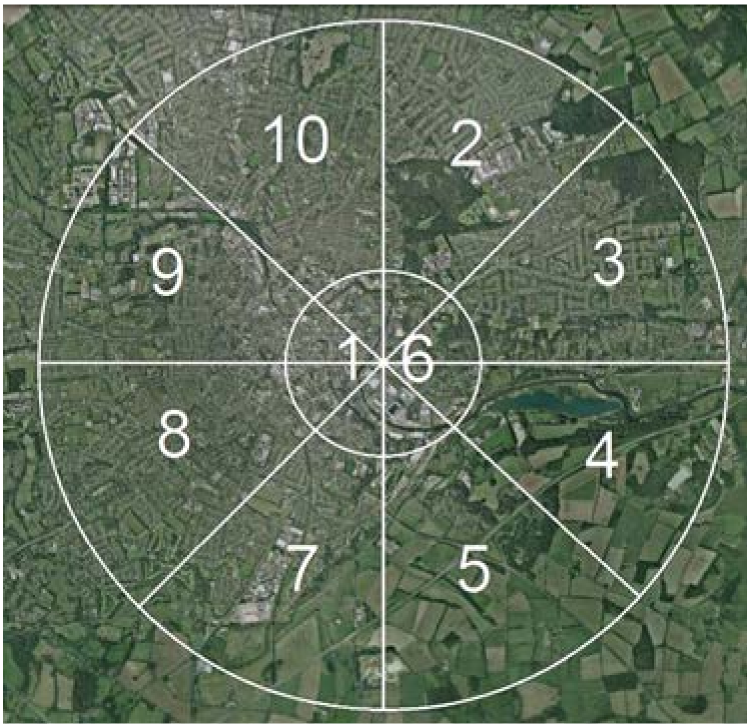


Figure 6: Norwich test area



2.6 Testing in a fair and consistent manner

Our test methodology was designed to ensure consistency and fairness for all operators.

- Each network was tested concurrently to ensure that environmental conditions were the same for each operator.

- Identical handsets were used for each network: the Samsung Galaxy S5 handsets. Additional tests were performed using Samsung Galaxy S6 Edge+ handsets.
- The handset position was rotated in the measurement equipment daily to evenly distribute the physical location of the handsets in the measurement backpacks.
- The SIM Cards were rotated between devices every four days to eliminate any bias that might occur from variations in individual handset performance.
- The measurement period was between 7am and 8pm each day, including weekends.

2.7 Test schedule

The measurement schedule was set up in the order shown in the table below. The test shown in bold fonts were used for internal testing purposes and were not used included in our metrics.

HTTP Download (30s)	Max test duration 30s
	Max setup time 30s
Ping 32 bytes (5 pings)	100ms Interval
	Timeout 1s
	Max test duration 5s
Ping 1300 bytes (5 pings)	**100ms Interval**
	Timeout 1s
	Max test duration 5s
PAUSE 15s	
HTTP Upload (15s)	Max test duration 15s
	Max setup time 30s
PAUSE 15s	
HTTP Browser mKepler	Max test duration 15s
PAUSE 15s	
HTTP Browser 'small'	**Max test duration 15s**
PAUSE 15s	
HTTP Browser BBC	Max test duration 15s
PAUSE 15s	
YouTube	Display Duration 20s

	Max Duration 30s
	Connection timeout 25s
	Stream loss timeout 20s
PAUSE 15s	
Voice Call	Call duration 90s
	Max call setup time 20s
Voice Quality Test (POLQA)	Sample frequency 12.5s
	Call duration 90s
	Max call setup time 20s
PAUSE 15s	

The 1300 byte test and the HTTP browser ‘small’ test were included for additional analysis but not published in our final report.

2.8 SIM cards

All of the operators we tested supplied SIM cards to Ofcom for use for the duration of our measurements. Due to the amount of data to be used during these tests (in the order of 2.5-3 GB per phone per day), the SIMs provided were standard consumer examples with the volumetric data caps removed.

The performance of the MNO-supplied SIM Cards were compared to the performance of consumer SIM Cards to ensure that that they were performing in the same way and had not been “optimised” for our testing. The consumer SIM cards were purchased on Ofcom’s behalf via a 3rd party and supplied on a twelve month contract basis, to ensure that they were representative of available consumer tariffs.

2.9 Quality control: during measurement and post measurement

The following checks were conducted during testing to ensure the integrity of our results prior to processing:

- Prior to the start of testing each day, the arrangement of SIM Cards, Handsets and their physical location in the backpack were checked according to the predetermined schedule.
- The correct measurement schedule was loaded onto the devices.
- All testing was conducted by Ofcom engineers operating in pairs to ensure that the tests could be frequently observed via the control tablet regardless of whether the measurements were static, walking or driving.
- Markers were added to the generated data files to confirm what type of test was running and to denote whether the tests were being conducted indoors or outdoors.

- A paper log was maintained by the engineers in case clarification was required during our analysis.
- Replacement SIM Cards were held by the teams in case of problems with the removal of the volumetric cap on both Voice and Data usage and as a general backup.

2.10 Data quality control

The measurement data from each city was retrieved from cloud storage to Ofcom's internal data server daily. This was then processed using NQDI and Excel to produce a report for each test conducted. The reports were then analysed by an external statistician and the final report produced.

After each day's measurement, the data were uploaded to local servers and then imported to the database using NQDI. A report was produced after the tests every day showing the number of samples collected for each test and for each technology.

Additional checks were made on the collected data prior to publication:

- The daily dashboard was used to check the number of samples collected for each test and for each operator was reasonable.
- The number of samples per city was monitored after each day of testing to see if enough samples were collected for 3G only / Best Bearer and the test campaigns were set accordingly to make up the numbers.
- Checks were performed to see if enough samples were collected per city. Should more samples be required, additional testing was scheduled.
- Any measurements that were aborted by the engineer or by the measurement equipment were identified and rejected
- Any data collected out of the predetermined core measurement hours (7am to 8pm) and out of the 4 km radius were filtered out
- Markers added to the data were double checked with paper logs for accurate position (indoor or outdoor) and state (either Walk/Drive/Static)
- The data was manually checked for missing data by looking for gaps in either date or time alongside the paper logs to ensure all the data is taken into account
- Each team was visited by a Quality Control assessor during their tests to ensure compliance with our predetermined processes

Section 3

Data Processing

3.1 NQDI

Once the results were copied to Ofcom's data servers, it was then imported into our post-processing system, NQDI (Network Quality Data Investigator). NQDI is a Swissqual-supplied data analysis and report generating tool which enabled us to analyse and report on the data collected to compile the various metrics and statistics used in this report.

3.2 Data quality control

The measurement data from each city was retrieved from cloud storage to Ofcom's internal data server daily. This was then processed using NQDI and Excel to produce a report for each test conducted. The reports were then analysed by an external statistician and the final report produced.

After each day's measurement, the data were uploaded to local servers and then imported to the database using NQDI. A report was produced after the tests every day showing the number of samples collected for each test and for each technology.

Additional checks were made on the collected data prior to publication:

- The daily dashboard was used to check the number of samples collected for each test and for each operator was reasonable.
- The number of samples per city was monitored after each day of testing to see if enough samples were collected for 3G only / Best Bearer and the test campaigns were set accordingly to make up the numbers.
- Checks were performed to see if enough samples were collected per city. Should more samples be required, additional testing was scheduled.
- Any measurements that were aborted by the engineer or by the measurement equipment were identified and rejected
- Any data collected out of the predetermined core measurement hours (7am to 8pm) and out of the 4 km radius were filtered out
- Markers added to the data were double checked with paper logs for accurate position (indoor or outdoor) and state (either Walk/Drive/Static)
- The data was manually checked for missing data by looking for gaps in either date or time alongside the paper logs to ensure all the data is taken into account
- Each team was visited by a Quality Control assessor during their tests to ensure compliance with our predetermined processes

3.3 Weighting

In order to make comparisons between networks fair, it is desirable that the same number of readings for each network are taken across any variable, otherwise a network might benefit in the comparison by having more readings at quieter times of day, outdoor, or in easier locations. The relevant variables in our test were across cities, indoor and outdoor locations, and the time of day.

Obviously the practicalities of fieldwork make this difficult to achieve, and so we weight the results to provide this equality. Put simply, if fewer readings are made in Cardiff outdoor during the weekend for O2 than for Vodafone, each Vodafone reading is weighted up by a factor which equalises this.

This is done after excluding unsuccessful and invalid tests, but not those that failed due to the network. It is done for all BB (Best Bearer) readings, regardless of the proportion that are carried on 4G as this is a measure of network capability rather than equalising the trial.

With the sample available for BB, this can be achieved with each city and network weighted to be 5% of the total, and within this to have identical profiles by time of day and indoor/ outdoor. For 3G, the sample is sparser and we cannot achieve the same balance by both network and city, so each network is weighted to the identical (average) profile by city, indoor/ outdoor and time of day.

All significance tests take account of this weighting, which slightly reduces the accuracy compared to an unweighted sample, but is an acceptable approach given the “fairness” it introduces into comparisons.

3.4 Means or medians

Ideally, we would use means to show an “average” score for each network, and conduct significance tests using the mean and its standard error. The problem is that such a test is only accurate if the underlying distribution is (close to) normal. In fact, if the sample is very skewed the mean can be considered a poor reflection of the “typical” customer experience, since we interpret it as the service the typical customer experiences. But, for example if 9 customers experienced a delay of 0.5 seconds and 1 a delay of 95.5 seconds, the mean score of 10 seconds would very poorly reflect “average” service.

With skewed distributions the median is considered a fairer comparison, as it is the value which 50% of customers do better than, and 50% do worse than. Using medians allows us to conduct significance tests without worrying about the underlying distribution, as follows:

- Take all the readings for the two subgroups we’re comparing.
- Calculate the overall median
- Calculate the % for each subgroup that fall below the overall median – if the two networks performed the same, this would be 50% for each
- Test whether the observed percentages are consistent with this, using the standard test for comparing two percentages

To decide whether to use means or medians, we examined the distribution – if this were normal or close to normal, we would find one eleventh (9.1%) of the readings in each of the following ranges:

- More than 1.335 standard deviations less than the mean (R1)
- Between 1.335 and 0.908 standard deviations less than the mean (R2)

- Between 0.908 and 0.605 standard deviations less than the mean (R3)
- Between 0.605 and 0.349 standard deviations less than the mean (R4)
- Between 0.349 and 0.114 standard deviations less than the mean (R5)
- Between 0.114 less than and 0.114 standard deviations greater than the mean (R6)
- Between 0.114 and 0.349 standard deviations greater than the mean (R7)
- Between 0.349 and 0.605 standard deviations greater than the mean (R8)
- Between 0.605 and 0.908 standard deviations greater than the mean (R9)
- Between 0.908 and 1.335 standard deviations greater than the mean (R10)
- More than 1.335 standard deviations greater than the mean (R11)