

Kantar Media (KM) Directory Enquiry Services Report

An Independent Review

Executive Summary

1. This paper provides an independent review of the Directory Enquiry survey undertaken by KM on behalf of Ofcom. In doing so I have reviewed three documents downloaded from the Ofcom website: the Directory Enquiry Services Research Report; the technical report; and the data tabulations; and also information provided by Ofcom to TNUK about project quality controls (following a Freedom of Information (FOI) request). The survey involved interviewing on six waves of the KM face-to-face Omnibus and ten follow-up qualitative telephone interviews.
2. The Omnibus is a well-recognised and respected method for interviewing nationally representative samples, and as in this case achieving a representative sample of a low incidence consumer group (those who called 118 DQ services in the last 12 months who personally paid for the call). However, the research project did not have the quality controls that I would expect to see for a regulatory survey of this importance: cognitive testing and/or piloting of the questionnaire before main fieldwork starts, and Executive monitoring of some interviews to ensure that fieldwork is of the required standard.
3. In my view the KM quantitative survey has only limited evidential value as it includes many respondents whose views are potentially less informative i.e. those whose last DQ call was made more than three months ago (and who therefore might struggle to recall the detail necessary to answer many of the survey questions), and those who do not know about the payment of their telephone services (and therefore have no knowledge of their bills and/or the cost of DQ calls). I use the term “limited” to mean that the quantitative work has evidential value in some parts, but not others. So, for example, in my view the quantitative work is robust in determining the extent to which consumers use 118 DQ services for personal reasons (the first information objective described in the KM quantitative report). However, it is not robust in evaluating expectations of 118 DQ prices and consumer “bill shock”, nor does it provide reliable evidence about the availability of alternative information sources available to consumers.

4. The qualitative work also has little evidential value in my view, due to its limited scope and unrepresentative coverage. I use the term “little” to mean that the survey findings do not carry much evidential weight, as in this case the qualitative work does not provide much understanding of DQ users experiences when using the service for the reasons described above.
5. Furthermore, I have specific concerns about some of the survey findings, in particular:
 - a) There should be a more balanced summary of the profile of service users, to say that DQ usage is reasonably consistent across all age, working status and SEG groups, excepting lower DQ usage among the 16-34 years age group.
 - b) The survey does not provide reliable evidence about the availability of alternative information sources, as the wording for this question is vague and may have been interpreted in different ways. Indeed, the reasons why respondents thought they had no alternative tended to reflect the circumstances that the respondent found themselves in at the time (e.g. no internet at the time, location-based circumstances, being in a rush) or the limitation of other sources (could not find the number using an alternative method), rather than necessarily a lack of other options.
 - c) There is insufficient emphasis in the report on consumer satisfaction with the service, and the benefits to the consumer of using the service. The quantitative survey shows that the majority of DQ users were satisfied with their last experience of calling a DQ service (just 14% indicated that they were dissatisfied). Consumer benefits of using the service include both speed of access and the ease of use/convenience of using the service, and these should be given much greater weight in the research findings.
 - d) The “bill shock” evidence based on perceptions of last call cost against prior expectation is potentially unreliable, as many may struggle to recall or have knowledge of the specific detail required to answer the relevant questions. There is also robust evidence within the survey (Q1a) which indicates that “bill shock” is much lower.
6. A much more robust approach for the analysis of the quantitative survey would be to re-run the tables, based on those whose most recent call was within the last three months and exclude those who said they “do not know about the payment of my telephone services”. This would provide a dataset from a group of respondents who would be more likely to recall specific 118 DQ calls they had made and had the knowledge of the call costs. However, the problem then would be sample size, as it is likely that this group from whom we could obtain more reliable data would be too small to withstand rigorous statistical analysis, which is a fundamental design flaw with the KM survey.

Background

7. Ofcom launched a Call Cost Review in May 2017 to examine the costs for consumers of calling 118 Directory Enquiry (DQ) services, and to consider whether these costs are causing consumer harm. As part of this review, Ofcom commissioned Kantar Media (KM) to conduct quantitative and qualitative research to determine: a) the extent to which consumers use 118 DQ services for personal reasons and; b) consumer reliance on 118 DQ services and the reasons why these services are used and; c) consumer awareness and expectations of price.
8. The purpose of this paper is to provide an independent review of the surveys undertaken by KM. In doing so I have reviewed three documents downloaded from the Ofcom website:
 - a) The Directory Enquiry Services Research Report
 - b) The technical report
 - c) The data tabulations

I have also reviewed documentation about the quality controls that KM had in place for the survey, supplied by Ofcom to TNUK following a Freedom of Information (FOI) request.

Methodology Review

Quantitative Research

9. Data for the quantitative research has been collected through the Kantar face-to-face Omnibus. This Omnibus is a well-recognised and respected method for interviewing nationally representative samples, and as in this case achieving a representative sample of a low incidence consumer group (those who have called 118 DQ services in the last 12 months and who personally paid for the call), and therefore I do not have any specific observations to make on the Omnibus vehicle as a data collection method.
10. However, the survey did not have the quality controls that I would expect from a regulatory survey of this importance. The Competition and Markets Authority (CMA) recently published a document entitled: “Good practice in the design and presentation of consumer evidence in merger cases” (23 May 2018). Whilst this document has been written primarily for an audience involved in designing surveys for merger cases, the CMA makes clear that the good practices described within are equally applicable to other surveys such as Market Investigations, and hence is relevant for this review of the KM survey conducted for Ofcom.
11. Firstly, for a regulatory survey of this importance, I would expect to see some testing of the questionnaire before the main fieldwork to ensure that the questionnaire works properly; that respondents understand the questions being asked; and interpret and can answer the questions in a consistent fashion. Good practice would typically involve: a) undertaking a small number of cognitive interviews to identify any problems with the questionnaire and; b) piloting the entire survey process with members of the Executive team closely involved with the piloting process i.e. conducting some of the pilot interviews themselves or listening in to the pilot interviews and taking notes about how well each question works. My review of the survey findings highlights issues with specific questions and raises doubts about how to interpret and use the results. These are the kinds of issues that should have been examined prior to fieldwork starting via cognitive interviews and/or piloting.

12. KM relied on an “internal” review of the questionnaire before it went into the field, in that it was “tested” by colleagues within the agency not working on the project and was then sent to a person within their Marketing Sciences team to provide feedback “including looking at reducing any potential bias based on question design”. There is no information from KM on what changes if any were made to the questionnaire following this internal review.
13. In my view this “internal” testing of the questionnaire is inadequate. Indeed, KM in their proposal for the work stated that: “It might be advisable to conduct some cognitive testing of the questionnaire if it is felt to be new or particularly complex”. Some of the material covered in the questionnaire is complex (e.g. “bill shock”, the interplay between customer satisfaction and the cost of the service, the impact on household finances) and in my view should have been examined properly via a cognitive test and a pilot of the survey questionnaire.
14. The reliance on colleague testing is flawed for two reasons: first, it is not a proper test of consumer understanding and interpretation as the questionnaire is being assessed by market researchers not consumers; secondly, it is very unlikely that any of those testing the questionnaire would have been DQ users eligible to take part in the survey and therefore in a position to answer the questions. This is acknowledged by KM in their proposal for the work, when they say: “we would happily conduct a ‘friend and family’ test for no additional charge, albeit it is unlikely that we will find many users of the service within our immediate circles given the low incidence”.
15. Secondly, I would expect to see rigorous Fieldwork quality controls used to ensure strict adherence to the questionnaire script. Good practice would typically involve: a) A full briefing of all interviewers scheduled to work on the survey before fieldwork starts, with interviewers personally briefed – either by telephone or face-to-face - by an Executive(s) involved in the design of the survey and; b) The Executive(s) who conducted the survey to observe (face-to-face) a selection of the interviews initially conducted post briefing, and to continue to monitor a proportion of the interviews.

16. It appears that KM relied solely on written interviewer briefing instructions. (There is reference to such instructions in Ofcom's reply to the FOI request but not having seen these instructions I am unable to comment on their content). Whilst such written instructions are useful to help document the briefing, this is not adequate in my view to put across the key messages to interviewers about script adherence, where and how to prompt and probe specific questions etc. My view is that the personal briefing of interviewers and supervisors (by telephone in group sessions) is a much more effective way of ensuring that interviewers know what is required of them and follow the questionnaire script and instructions precisely.
17. Looking at the controls used by KM to monitor fieldwork quality, it seems that KM again relied on standard survey procedures which in my view are inadequate for a regulatory survey of this importance. KM appear to have relied on standard checks to monitor the quality of interviewing undertaken, validating 10% of the survey interviews undertaken. The validation checks involved re-contacting 10% of respondents to confirm that: the interview took place, the interviewer ID card was shown, the MRS leaflet was handed out, the interviewer was previously unknown to the respondent, and checking answers to specific questions pertinent to the survey (although it is not specified which questions were checked in this way and the feedback from these checks). Such validation checks are useful but not adequate in my view as they do not provide a comprehensive assessment of interviewer performance, in particular they do not verify whether or not there was strict adherence to the questionnaire script, appropriate use of prompts and probes, proper use of pre-code lists (with lists read out or not read out as instructed). There is no mention of bespoke quality checks to ensure that interviewing is of the required standard for a survey of this importance such as: Executive requests for feedback or analysis of responses to specific questions by interviewer, Executive accompaniment of fieldworkers to listen to interviews, any additional supervisor accompaniment of interviewers (over and above standard KM fieldwork requirements).

Qualitative Research

18. The qualitative research involved the re-contacting by telephone of respondents from the quantitative phase who had used 118 DQ services and said they had paid more than they expected to when using the service. In my view this research has little evidential value. Only ten qualitative interviews were conducted in total, with a heavy emphasis on social grade DE respondents with only four interviews (all C1s) from other SEG groups. Furthermore, it is likely that the more financially challenged (less time poor) respondents will have agreed to take part in a follow-up interview which gave them a £20 incentive, having already taken part in a reasonably long Omnibus interview. It would be interesting to see: a) the response rate to the re-contact question (Q25), as a low response rate would indicate an even greater concern about the representativeness of the qualitative sample and; b) the “financial situation” profile (Q23) of the qualitative sample and how this compares with DQ users. I would be particularly concerned if the qualitative sample was skewed towards those saying they were finding it difficult on present income.

Review of Survey Findings

19. Overall, the quantitative questionnaire reads reasonably-well with clear routing, prompt instructions, and inverting of scales where appropriate. However, as a general observation, the questionnaire requires respondents to read a lot of possible response codes, which can be tiring during a long interview, particularly as this bank of questions will be accompanied by some other question sections on completely different topics (in an Omnibus environment). I have some concerns therefore about the quality of answers to questions with long or complex pre-code response lists.
20. Also, some of the questions are asking respondents to recall details of an event that may have taken place several months ago (e.g. the last DQ call). I am doubtful about the validity of findings taken from questions which require respondents to recall specific detail of an event that took place many months in the past (in some cases), and where the event in question (a DQ call) may be of only limited importance to the respondent in the broader context of their lives.
21. A further key concern is that the Omnibus survey provides a representative sample of UK adults (aged 16+ years) and as such it will include many respondents who are not bill payers. There does not appear to be any question in the survey about bill payer status, which means we cannot do any specific analysis in this area. However, those who are not bill payers may not have much idea of how much they paid for a DQ call. For this sub-group, it is difficult therefore to see how concepts such as “bill shock” can be measured in any meaningful way.
22. The other main concern is the way that specific quotes from the qualitative research are used by way of illustration on some of the quantitative charts, as this can give a misleading view of the interpretation taken away from the chart. With such a limited number of interviews amongst a very specific sub-population of the DQ user base, it is potentially misleading to use the qualitative research in this way.
23. Turning now to the detail in the presentation charts, I have concerns about some of the questions and commentary, as follows.

24. Chart 17-18 – Q1: In the last 12 months, have you made a call to a directory enquiry 118 service on either a fixed line or mobile phone? (This includes calls even if you did not manage to speak to an operator or get the information you needed from them). Base: All
- a) Assuming the significant testing has been applied to percentages calculated to decimal points (and not integers), the differences indicated on the charts will be statistically significant due to the large base sizes (the base is all adults aged 16+ from one wave of the Omnibus). However, whilst it is true that DQ services usage is significantly higher among adults aged 65+ years compared with the total, my observation is that usage is reasonably consistent across all age, working status and SEG groups, excepting lower usage among 16-34 year olds. If we removed responses from 16-34 year olds from the survey, the total DQ usage would be almost exactly 3%, and using this as the benchmark we see very little difference in DQ usage by demographic groups. The key finding is the lower DQ usage among 16-34 year olds, which is not unsurprising as this is the age group that is probably most likely to use the internet as an alternative information source for 118 DQ services.
25. Chart 26 - Q9: Why did you decide to use that specific directory enquiry service on that occasion? Prompted. Multi-code. Base: All.
- a) This question asks about reasons for use (on that occasion) but the question is prompted, which raises concerns that it potentially leads respondents, particularly if they are having difficulty recalling the event if it took place many months previously. Such concerns would have been assuaged by asking the question unprompted (with spontaneous response categorised to pre-codes). Also, the question does not make it clear whether they are asking a brand choice question, or whether they are asking a question about using DQ services rather than another information source. The prompted list of responses makes this confusion worse, as it is a mix of brand related and category related reasons. Finally, there appears to be no probe of "why else?" in the question wording which there should be for a multi-response question, and the lack of multi-mentions indicates that this question may not have been probed properly.

26. Chart 30-31 – Q12a: And overall, how satisfied were you with your last experience of calling a directory enquiry service? Prompted. Single-code. Base: All; Qualitative findings on customer satisfaction.

- a) The quantitative survey shows that the majority of DQ users were satisfied with their last experience of calling a DQ service. Indeed, just 14% indicated that they were dissatisfied. This is statistically robust evidence and demonstrates that only a small minority were dissatisfied with the experience.
- b) The commentary on chart 31 attempts to steer the interpretation of this evidence by saying that “the qualitative evidence suggests that those who answered satisfied and very satisfied on the survey may have disassociated their view of the service received from the cost and seemed to treat these as two different experiences”. My view is that the qualitative work has little evidential value due to the low number of interviews conducted, and the skew towards a particular respondent type. The qualitative work does not represent the views of the great majority of those who said they were satisfied and very satisfied with their experience in the quantitative survey, and therefore it is misleading to use the qualitative findings in this way.
- c) Furthermore, the third bullet point on chart 31 talks about qualitative participants describing “bill shock” in varying degrees, and yet none of the verbatim comments show customers using the term “bill shock”, which raises some doubt about whether “bill shock” is the consumer or researcher language.
- d) It is difficult to see why consumers in the quantitative survey would disassociate their view of the service received from its cost. The satisfaction question is phrased in terms of satisfaction with their “last experience”, and typically when rating a service experience a consumer will consider all the factors that are important to them, including the cost of the service. Indeed, we often find in other markets that price or cost is a key driver of satisfaction ratings with a service provision.
- e) The evidence from examination of the data tables indicates that respondents do take account of call cost in coming to a view about their satisfaction with the experience. Satisfaction is significantly lower among those who said the cost of their last call was more than expected (53%) compared with the total sample (71%).

27. Chart 34 - Q14a: On that occasion, were there any other ways available to you to get the telephone number you needed rather than calling a directory enquiry service? Unprompted into code frame. Multi-code. Base: All

- a) The question wording - "any other ways available to you" - is vague and may have been interpreted in different ways. It is possible therefore that an alternative option had been available but was not considered by the respondent when replying to this question, particularly if they were struggling to recall the detail of an event many months past. Also, there are very few multiple responses to this question, which seems odd given the different information sources that are available. This would further indicate that the question may not have been interpreted in a consistent fashion and may not have been probed fully by interviewers to ensure that all possible options were recorded.
- b) Also, the qualitative comment on the chart ("you would desperately call 118") suggests this is a typical situation for many DQ users, and there is no evidence of this from the quantitative research.

28. Chart 35 - Q15: Why did you call a directory enquiry service to get the telephone number you were looking for instead of using an alternative method? Unprompted into code frame. Multi-code. Base: All

- a) The commentary in the title heading misses another key reason for using a DQ service - the convenience/ease of getting number benefit (19% saying it was more convenient to get number by calling DQ and 6% saying an alternative was too much hassle). Also, the fact that 23% said they could not obtain the number in any other way shows how some respondents misinterpreted the previous question (Q14a - see above), evidence that Q14a did not work well as a question.

29. Chart 37 - Q14b: Can you tell me why there weren't any other options available to you? Coded verbatim response. Base: Those saying they did not have any other option.

- a) Looking at the breakdown of responses to this question, we see that only 9% of those answering this question (just 12 DQ users) say they had no alternative. The other reasons mentioned reflect the circumstances that the respondent found themselves in at the time (e.g. no internet *at the time* (my italics), location-based circumstances, being in a rush) or the limitation of other sources (could not find the number using an alternative method), rather than a lack of other options. It is reasonable to suggest that some of those who said they did not have an alternative because they had no internet at the time, or were based somewhere else out of the house, could have used the internet (or the phone book) at another time to find the telephone number if they had so wanted. It could be argued that

these people did have an alternative option, but they preferred not to use this alternative because it was quicker and easier to use DQ services at that time.

- b) It is noteworthy that 39% of those who said they had internet access claimed not to have any other options available, supporting the idea that some of these respondents could have used the internet at another time to find the telephone number.

30. Charts 39-43 - Q3/3a: Can you tell me approximately how much you think it costs to call a 118 Directory enquiry service from: a) a personal mobile phone and; b) a home fixed/landline phone? This is for a call that would last just under 1 minute. Unprompted. Base All. If Don't know response, follow-up question: Using this list approximately how much do you think it costs to call a 118 Directory enquiry service from a personal mobile phone? Base: Those saying DK spontaneously.

- a) Use of prompted ranges for those who say they do not know spontaneously can potentially lead respondents to an answer that they might not otherwise have given, as some respondents who are not sure and then prompted with a range list are likely to have been guessing and may have picked a range option in the middle of the list to try to appear sensible. My view is that these results should be reported solely on those who answered spontaneously.

31. Charts 44-47 - Q4a/b/c: Some directory enquiry services provide an onward call connection. In other words, if you choose to use the onward call connection service you are put through to the number you asked for. Which one of the following do you think applies to using an onward call connection service? Prompted. Single-code. Base: All.

- a) This question is asked of everyone, but only some have used the call connection service. As the price of a connected call is announced just before the call connection, those who have not used the call connection service will not have been given the same price information as those who have used the service. Given this, my view is that the results would be of more value if reported from those who had used the call connection service in the last 12 months. The data tables show that understanding of connection call charges was higher among those who have used the service, for example just 8% of those who say they had used the call connection service at Q4a answered "don't know" compared with 24% of those who had not used the service.

- b) There may also be some confusion here with the fixed fee for using the DQ service, with some interpreting this as being part of the onward call connection service. So, the "Net Correct" figure in chart 46 may be understating the real price knowledge of service users (the 12% saying "fixed charge + pay per minute at the same rate as calling DQ services" may include some who are thinking about the entire call cost including the DQ service).
32. Chart 45 - Q4b: Which one of the following do you think applies to the per minute call charges once you have been connected to the number you requested? Prompted. Single-code. Base: All.
- a) The qualitative comment on this chart is potentially misleading, as it is illustrating a scenario where pricing information is not given for connected calls and suggesting that this is a reason why some are unaware of the connected call cost. There is no evidence of this from the quantitative research; the key finding from this chart is that the majority of those aware of the pay-per-minute charge for connected calls knew that it is charged at the same rate as the call to DQ services.
33. Chart 47 - Q4c: Which of the following best describes what you know about the cost of using an onward call connection service through directory enquiries? Prompted. Single-code. Base: All personal director enquiry users who don't know about the cost of onward call connection (at Q4a/Q4b).
- a) This question uses a hybrid scale, mixing value for money with cheap-expensive. It is perfectly possible for something to be considered expensive but good value for money (e.g. premium brands). This makes it difficult to interpret responses to this question, as we do not know whether those who answered expensive still thought it was good value. This could have been avoided by asking two separate questions: a) Whether considered Cheap, Neither cheap nor expensive, Expensive and: b) Whether considered Good value for money, Reasonable value for money, Poor value for money.
34. Chart 48 - Q16: On that occasion, which of the below best describes what you knew about the cost of calling the directory enquiry service before you made the call? Prompted. Single code. Base: All.
- a) This is an example of a question with long and complicated pre-codes (covering two different dimensions) which need to be read in their entirety and considered accordingly before answering. I have concerns about the quality/accuracy of the response with these types of pre-codes, it would be better to ask two separate but simpler questions: a) whether they knew the cost of the call before making it (Yes/No) and; b) whether considered Good value for money, Reasonable value for money, Poor value for money. Note also the same

concern as at Q4c (Chart 47 above) about mixing two different concepts - value for money and expensive - in the same scale.

b) Note also that the proportion in the title heading is wrong – it should read one in five (not two in five).

35. Chart 50 - Q18/18a: Thinking about your last call to directory enquiry services, how much more did it cost compared to what you had expected? Base: All. If Don't know response, follow-up question: Using this list approximately how much more did it cost compared to what you had expected? Base: Those saying DK spontaneously.

a) This question requires a respondent to know: a) how much they expected the call to cost before making it; b) how much the call cost in fact and then; c) to calculate the difference. There is considerable doubt about whether such a task can be undertaken by all respondents accurately. Some respondents will be trying to recall an event that happened many months ago, are they really going to be able to remember what they had expected the call to cost before they made the call? Many respondents will not be the bill payer and therefore will not have known how much the call cost. Some respondents would have struggled to do the arithmetic in an interview situation, even if they did remember these details.

b) Furthermore, some of those respondents who were not sure and then prompted with a range list are likely to have been guessing and may have picked a range option in the middle of the list to try to appear sensible.

c) Given all these issues, in my opinion these results have little evidential value.

36. Chart 51 - Q2/Q18b: How many times in the last 12 months have you made a call to a directory enquiry service from your home fixed landline or personal mobile phone (i.e. not paid for by a business)? Prompted. Single Code. IF TWO OR MORE: Q18b: You said you had called directory services more than once in the last 12 months. Which of the following applies to any of these calls (i.e. other than your most recent one just answered about?). Prompted. Single-code. Base: Those saying they had made two or more DQ calls in the last 12 months.

- a) The read-out scale is not complete, as there is no option for: "At least one call was a lot less/a little less than expected (and all the others were about what I expected)". Also, it is potentially confusing to have a scale which has one option talking about "at least one call" and another option talking about "all (calls)". Furthermore, in my view it is doubtful that all respondents can remember for ALL their DQ calls: a) what they had expected it to cost (before making the call) and; b) what it cost in fact. This is the same observation as for Q18/18A (chart 50 above).
37. Chart 52 - Q17/18b/Q1a: Q17 And after the call, how much did the cost of this call compare to your expectations? Prompted. Base: All; Q18b You said you had called directory services more than once in the last 12 months. Which of the following applies to any of these calls (i.e. other than your most recent one just answered about?). Prompted. Single-code. Base: Those saying they had made two or more DQ calls in the last 12 months; Q1a. In the last 12 months, have you spent more than you expected to for any telephone services, and if so for what reason? Prompted. Multiple code. Base: All
- a) This chart shows the net responses from Q17 and Q18b, that is those saying either that their last DQ call cost a lot/a little more than expected or that at least one other DQ call cost a lot more/a little more than expected. As such, these results are subject to the same concerns as discussed above in charts 50 and 51, namely that it is doubtful whether all respondents can accurately recall their prior price expectation and the actual cost of the call, for all the DQ calls they had made in the last 12 months.
- b) There is an apparent conflict of evidence in this chart. On the one hand the chart is reporting that 39% of DQ users claim they had a DQ call cost more than expected in the last 12 months, and on the other hand that only 10% mentioned DQ calls when asked whether they had spent more than expected on any telecoms services in the last 12 months. The commentary in the chart attempts to reconcile the apparent conflict by saying that "We consider these (the 10%) to represent those that either had a particularly memorable DQ "bill shock" experience or had not experienced other more significant "bill shocks" through their use of telecoms in a 12-month period". However, there is no evidence to support this interpretation.
- c) In my view the 39% net "bill shock" figure (from Q17 and Q18b) has little evidential value as it is based on questions that require respondents to recall a level of detail that will be beyond many of them, for the reasons described above. The 10% figure (from Q1) has greater evidential value in my view, as the data comes from a question that is

straightforward to answer, is based on current perceptions, and is within the broader context of all telecoms spend.

38. Charts 54-56 – Q22: Now thinking about your overall spend on directory enquiry services in the last 12 months, in what way, if at all, has this impacted you or your household finances? Have you experienced any of the following as a result? Please note that I am only interested in times when you think your experience was a result of using a directory enquiry service. Prompted. Multi-code. Base: All.

- a) In charts 54-55 we see that 35% of DQ users “do not know about the payment of my telephone services”. This emphasises the point made above (see chart 50), that many of the respondents interviewed on the Omnibus survey are not bill payers and therefore will not know about the cost of their telephone services. For these respondents all questions about “bill shock” are of little evidential value, as they do not have the knowledge of the cost of telephone services.
- b) In charts 55 and 56, the title heading should say: 1-in-6 (17%) DQ users whose call cost more than expected SAY THEY have been affected financially, as this question is based upon the respondent’s perception of their situation. Also, as this figure is based upon the price expectation/actual price of the last call, there are the same concerns about the evidential value of this figure (as chart 50).
- c) The survey does not provide any data on total consumer spend on 118 DQ services in the last 12 months, which would have been useful when trying to assess whether the claimed impact of 118 DQ service usage on household finances was realistic or not. However, the average number of 118 DQ calls in the last 12 months was just two calls, and just 12% had made five or more DQ calls in the last 12 months, and therefore it seems odd that so many respondents should have claimed to have been financially affected given the likely level of spend on 118 DQ services across the year.
- d) Also, both the qualitative comments on chart 56 refer to the “shock of the bill”, which colours the interpretation of the quantitative results in a way that cannot be proven to be representative.

Summary of Key Findings in the Presentation Deck

Directory enquiries usage in the last 12 months

39. Whilst the figures reported in this section are correct, I think a more balanced commentary would say that usage is reasonably consistent across all age, working status and SEG groups, excepting lower DQ usage among the 16-34 years age group.

Circumstances of last DQ call to 118 DQ services

40. A key finding from the quantitative survey, that is not highlighted in the summary, is that the majority of DQ users were satisfied with their last experience of calling 118 DQ services, with only 14% saying they were dissatisfied. Whilst the % rating satisfied with the experience is reported by KM, it comes with a major caveat about DQ users disassociating their experience with the service and the cost of calling the service. The evidence cited for this caveat is taken from the qualitative survey, which does not represent the views from the great majority of the quantitative respondents, and therefore should not be used in this way, and is contradicted by evidence from the data tables.

Alternatives to using 118 DQ services

41. In my view the evidence cited in the KM report - that four in ten DQ users claimed they had no alternative - is potentially flawed, as the question wording appears to be vague (Q14a). It has generated results that are inconsistent with responses at subsequent questions and relies on respondent's being able to recall specific detail of an event that may have happened many months past.
42. Second, whilst the report cites speed of access as the key consumer benefit, there is little mention of another consumer benefit – the convenience/ease of getting the number - which should not be overlooked in evaluating how consumers benefit from the service.

Knowledge and expectations of call costs

43. The key concern is the “bill shock” evidence reported in this section, namely that “over one third (35%) of users said the cost of their last call to 118 DQ services was more than expected, and two in five (39%) DQ users said they had made at least one call in the last 12 months to 118 DQ services that cost more than they expected”.
44. In my view the 39% net “bill shock” figure (from Q17 and Q18b) has little evidential value as it is based on questions that require respondents to recall a level of detail that will be beyond many of them. The 10% figure (from Q1) has greater evidential value in my view and should be highlighted within this section.

Review of the KM “Data Checks and Rationale”

High Level of “Don’t know” responses

45. There is some discussion in the technical report about data checks and additional analysis that was carried out “to ensure reported results were accurate”. The first observation made in the technical report is that “the time lapse between making the call and answering the questionnaire did not elicit any concern over the reliability of the data” (page 11). The evidence used in the technical report to support this observation is that:
- a) Just two out of seven (unspecified) questions with relatively high levels of “don’t know or related” responses showed higher levels of “don’t knows” among those with less recent calls.
 - b) The high level of “don’t knows” were not based on a large consistent group throughout.
 - c) Results for those who made the call within the last six months did not yield any significant differences from the total results.
46. I have examined the data tables to see whether the proportion of “don’t know” responses varied by time lapse and have found evidence of some differences. The results in the tables are shown in three different time bands, calls made: within the last three months, 4-6 months ago, 7-12 months ago. Looking at the % of “don’t know” responses among those whose last call was more than three months ago compared with those saying it was within the last three months we see markedly higher proportions of “don’t know” responses among the former on:
- a) Q13: What type of call were you looking for when you made your last directory enquiry service call? Unprompted. Base: All
 - b) Q14a: On that occasion, were there any other ways available to you to get the telephone number you needed rather than calling a directory enquiry service? Unprompted. Base: All
 - c) Q18: Thinking about your last call to directory enquiry services, how much more did it cost compared to what you expected? Unprompted. Base: Those who said the call cost more than expected**

** Whilst this difference is statistically not significant (due to low base sizes) it repeats the pattern of higher “don’t know” responses among calls made more than three months ago.

47. The data from these unprompted questions shows that respondents whose last DQ was more than *three months ago* were less able to answer. These are all unprompted questions which, as was acknowledged in the technical report, are the type of questions which may have decreased a respondent's ability to respond after a longer time-period. The critical point, however, is that the lack of difference in "don't know" responses to *prompted* questions does not prove that respondents with a less recent DQ call could answer these questions; it could just mean that respondents were more likely to guess an answer from a prompted list (if unable to remember the detail of the call).
48. The technical report says that the high level of don't knows were not based on a large consistent group throughout the survey. It is not possible to check this statement from the data tables, but as discussed above it seems a reasonable hypothesis that the 35% who said they "do not know about the payment of my telephone services" (from charts 54 and 55) would form a large part of the group who answered "don't know" to many of the questions about the knowledge and expectation of call costs.
49. The absence of any information in the survey about the bill payer status in the survey is a flaw in my view, as it seems reasonable to suggest that a high proportion of "don't know" responses would come from those who are not responsible for paying the bill and/or do not see the bill. However, without information on bill payer status, it is still possible to take those who say they "do not know about the payment of my telephone services (from chart 54/55)" as a proxy for non-bill payers.
50. In my view, the KM "data quality" approach - re-running of the tables filtered on those whose last 118 DQ was made six or less ago - is an insufficient data check. A much more robust approach would be to re-run the tables, based on those whose most recent call was within the last three months and exclude those who said they "do not know about the payment of my telephone services" (from chart 54/55). This would provide a dataset from a group of respondents who would be more likely to recall specific 118 DQ calls they had made and had the knowledge of the call costs.
51. However, the problem then would be sample size, as it is likely that this group from whom we could obtain more reliable data would be too small to withstand rigorous statistical analysis, which is a fundamental design flaw with the KM survey.

Rationale for focussing on Q17/Q18b to calculate the incidence of bill shock due to DQ

52. The technical report notes that “DQ calls are relatively infrequent and as such it is reasonable to expect them not to be particularly memorable”. It then goes on to say that “Mindful of this, the main survey was designed to aid respondent’s recall of their last DQ call experience” and that “We consider the more direct questions, following on from more detailed questioning aimed to prompt recall of a specific DQ call, provides a more robust measure of the proportion of DQ users who paid more than they had expected to, for a DQ call”. This is clearly a matter that requires judgement, but if we accept that DQ calls are not particularly memorable it seems unrealistic to expect that respondents will be able to remember specific detail about prior price expectations and the actual cost of their last DQ call if it was many months ago; and the danger is that by prompting respondents with detailed questions it is leading them to give an answer when the reality is that they will not know or remember.

Overall Summary

53. Overall, the KM quantitative survey has only limited evidential value in my view as it includes many respondents whose views are potentially not informative i.e. those whose last DQ call was made more than three months ago (and who therefore might struggle to recall the detail necessary to answer many of the survey questions), and those who do not know about the payment of their telephone services (and therefore have no knowledge of their bills and the cost of DQ calls). The qualitative work has little evidential value in my view, due to its limited scope and unrepresentative coverage, and therefore should not be used to interpret or provide a deeper understanding of the quantitative results.
54. Specifically, with regards to the detailed findings in the KM survey, my view is that:
- a) There should be a more balanced summary of the profile of service users.
 - b) The survey does not provide reliable evidence about the availability of alternative information sources.
 - c) There is insufficient weight placed on consumer satisfaction with the service, and consumer benefits of using the service.
 - d) The “bill shock” evidence based on perceptions of last call cost against prior expectation is unreliable and conflicts with other more robust evidence from the same KM survey which indicates that “bill shock” is much lower.
55. A much more robust approach for the analysis of the quantitative survey would be to re-run the tables, based on those whose most recent call was within the last three months and exclude those who said they “do not know about the payment of my telephone services”. However, there is probably insufficient interviews to be able to do this, which is a fundamental design flaw with the KM survey.

David Rodgers

16 August 2018