# Report of the digital preservation expert elicitation workshop held online

## 28 & 29 April 2020

Hannah Merwood

Alex Green

David Underdown

Martine J. Barons

Thais Fonseca

THE NATIONAL ARCHIVES

WARWICK

# Acknowledgements

This workshop was funded by the National Lottery Heritage Fund (NLHF) and the Engineering and Physical Sciences Research Council (EPSRC) Impact Acceleration Fund.

Hannah Merwood is funded by The National Archives

Alex Green is funded by The National Archives

David Underdown is funded by The National Archives

Martine J Barons is funded by University of Warwick

Thais Fonseca is funded by EPSRC Impact Acceleration Fund and University of Warwick

We express our gratitude to the following experts who contributed their expertise:

Lucy Wales, Digital Preservation and Data Manager, British Film Institute

Caylin Smith, Digital Preservation Manager, Cambridge University Library

Jenny Mitcham, Head of Good Practice and Standards, Digital Preservation Coalition

William Kilbride, Executive Director, Digital Preservation Coalition

Sam Johnston, Service Manager for Archives, Dorset History Centre

Claire Collins, Collections Development Manager, Gloucestershire County Council

Viv Cothey PhD, Digital Curation Consultant, Gloucestershire County Council

Heather Forbes, Head of Archives Service, Gloucestershire County Council

Jodie Double, Digital Content and Copyright Manager, Leeds University Library

David Clipsham, Technical Architect in Digital Archiving, The National Archives

Ian Henderson, Digital AV Preservation Specialist, The National Archives

Francesca Mackenzie, Digital Archivist, The National Archives

Alec Mulinder PhD, Head of Service Assurance, The National Archives

Sonia Ranade PhD, Head of Digital Archiving, The National Archives

Anna de Sousa, Senior Digital Archivist, The National Archives

David Underdown, Senior Digital Archivist, The National Archives

Paul Young, Digital Preservation Specialist, The National Archives

Tamara Thornhill, Corporate Archivist, Transport for London

Barbara Taylor, Digital Data and Social Media Coordinator, University of Brighton Design Archives

Sue Breakell, Archivist and Archive Leader, University of Brighton Design Archives

Zoe Fullard, Archives Cataloguer, Independent

Melissa McGreechan, Assistant Archivist, Independent

Simon Wilson, Archives Consultant, Independent

# Contents

# Executive Summary

In 2020, The Applied Statistics and Risk Unit from the University of Warwick worked with Digital Archivists from The National Archives and other UK archival organisations to develop an integrated decision support system based on a Bayesian network to measure digital preservation risk.

Due to there being limited historical or research data available, additional data needed to populate the model was elicited from experts using a structured expert judgement protocol. There was a 2-day facilitated online workshop on 28-29 April 2020 attended by 22 experts, who contributed their knowledge and expertise and provided estimates and uncertainty bounds for the unknown quantities. The results were then mathematically aggregated to produce probabilities which fed into the final model.

This report explains the context, process, methods and results from the expert elicitation exercise that was carried out as part of the research project.

# Introduction

**Project background and the need for expert elicitation**

Within the digital preservation community, there is well-established and widely used guidance available to help identify the risks to digital archives, and mitigating practices are commonly in place. For instance, archival institutions often keep more than one copy of a digital record (there is a well-known mnemonic LOCKSS – 'lots of copies keep stuff safe') and these copies may be kept on different storage media and in different geographical locations to minimise the risk of a single point of failure. However, the threats to digital preservation go beyond storage life and are complex and diverse, and there is a lack of guidance on how to compare and prioritise different risk prevention strategies.

There is a practical need to quantify these risks. Any mitigating actions archivists take will come with their own costs and decision makers need to understand what the expected benefits of these actions are in order to assess them objectively. Archival institutions are often under resourced and smaller archives in particular may struggle to put their case forward. In addition, securing resources for archives may be more challenging than for other sectors as the impact of inadequate preservation practices would not be realised until the far future, so there is unlikely to be any immediate or short-term evidence for the decision maker to take reassurance in or learn from.

At present, there is very limited data available on digital preservation risks. Where it does exist, it is often very specific to storage like the failure rate of commercial NAND solid state drives[1] or focused on end-user services and satisfaction[2], neither of which are useful for our holistic model as we want to look at risks within the archivist's control. There is occasional anecdotal evidence from institutions brave enough to share their findings, although this offers no quantitative evidence on the scale or frequency of these events. Well-established digital archives often have risk mitigation policies so integrated into their practise that errors are not needed to be logged because they are always dealt with, but more importantly, non-errors are not counted either, so again the true frequency is unknown. For smaller archives, may be constantly firefighting the ever-changing risks and do not have time to perform any quantitative analysis to track these in detail. This lack of hard quantitative evidence is why we have needed to use structured expert judgement methods for this project.

---

[1] S. Maneas, K. Mahdaviani, T. Emami, and B. Schroeder (2020) 'A Study of SSD Reliability in Large Scale Enterprise Storage Deployments', *18th USENIX Conference on File and Storage Technologies.* *https://www.usenix.org/system/files/fast20-maneas.pdf*

[2] The Archives and Records Association National Survey group conducts two surveys of UK archive visitors in alternating years. *https://www.archives.org.uk/what-we-do/archive-surveys.html*

In 2017, The National Archives decided that they needed a rigorous method to understand and manage the digital preservation threats to their rapidly growing collection. John Sheridan, the Digital Director, is a mathematician by training and had an idea that Bayesian models could be used for risk analysis. He asked Alec Mulinder and David Underdown to explore whether this technique could be applied to digital preservation. Following the guidance given in Fenton and Neil's Risk Assessment and Decision Analysis with Bayesian Networks, David and Alec used the AgenaRisk software to model storage life to test how the approach could work. This appeared to be a promising technique, so Sonia Ranade and David Underdown met Professor Jim Q Smith at the Turing Institute to discuss how this idea could be taken further. Jim introduced The National Archives to Dr Martine J. Barons, Director of the Applied Statistics and Risk Unit (AS&RU) at the University of Warwick who specialises in integrated decision support systems using Bayesian networks.

**The IDEA protocol**

The use of expert judgement in forecasting and decision-making is well-known. However, when the elicitation of expert judgements is unstructured, the results are un-reproducible and can be unreliable and heavily biased, being subject to cognitive biases and heuristics. Substantial improvements can be made by using structured approaches designed to mitigate the most pervasive and debilitating psychological and contextual frailties of expert judgement.

The IDEA (**I***nvestigate*, **D***iscuss,* **E***stimate,* **A***ggregate*)[3] protocol for structured expert judgement was selected for this expert elicitation since it has been demonstrated to harness the advantages of the longer established Delphi, Cooke's and Sheffield methods and ameliorate their disadvantages.

Aggregation of experts' judgments can be behavioural (seeking consensus) or mathematical or a mixture of the two. The IDEA protocol synthesises specific elements from these approaches to minimise the disadvantages of existing approaches and optimise their respective advantages. The protocol was refined using a US intelligence agency forecasting tournament, The Intelligence Game[4].

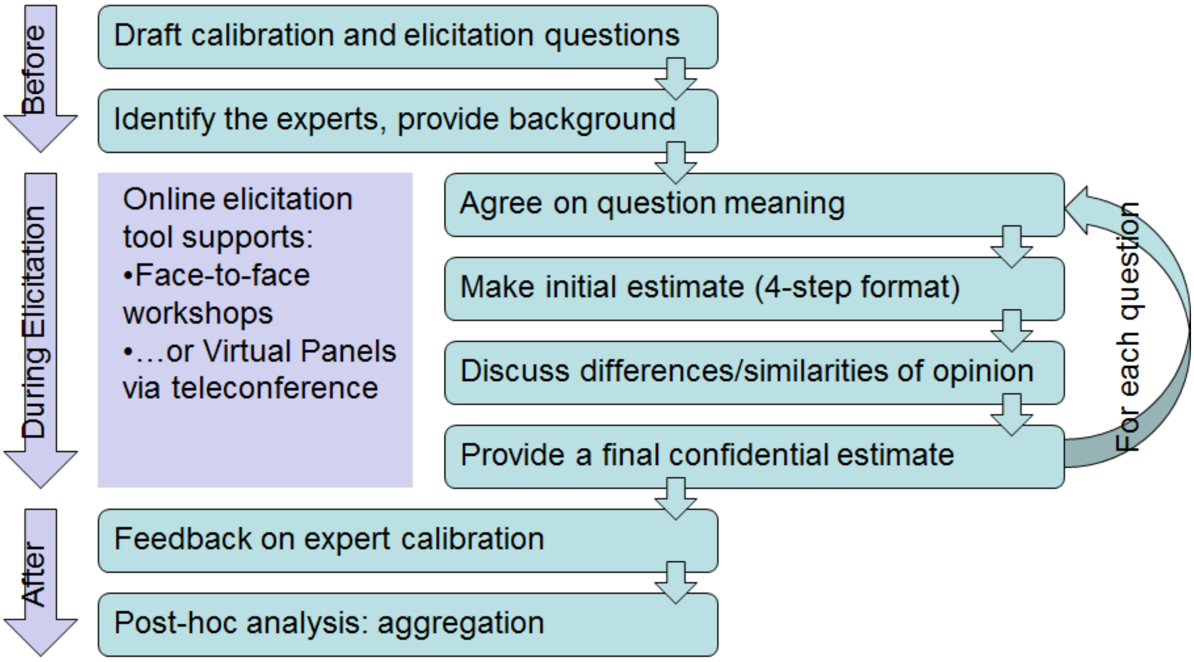
The IDEA protocol consists of 3 stages: the pre-elicitation, the elicitation and the post-elicitation stage.

---

[3] A.M. Hanea, M.F. McBride, M.A. Burgman, B.C. Wintle, F. Fidler, L. Flander, C.R. Twardy, B. Manning, S. Mascaro, 2016, InvestigateDiscussEstimateAggregate for structured expert judgement, International Journal of Forecasting, accepted for publication on 25.02.2016

[4] *http://intelgame.acera.unimelb.edu.au/*

*Figure 1: A flow diagram showing the stages of an elicitation workshop. Thanks to Dr Anca M Hanea for use of this image.*




In the pre-elicitation stage the problem is defined, the experts identified, the validation data found, the questions framed, and the training is delivered to the experts.

The elicitation stage has two rounds. It begins with experts being given suggested resources to investigate the problem and then providing their lowest plausible, highest plausible and best estimate of the quantities of interest on an individual basis. The first estimation round is followed by a facilitated discussion of the anonymised results which irons out any semantic difficulties and allows experts to share their thinking with each other. The elicitation is finalised with a second round of individual estimates, allowing experts to revise (or not) their estimates, based on the discussion.

The post elicitation stage involves post processing the results by mathematically aggregating the experts' judgements. The validation data (that is, the risks for which we have data) is used to construct a set of calibration questions for which experts are asked to give highest plausible, lowest plausible and best estimate followed by a facilitated discussion and second private estimate as before. The estimates on the calibration questions can be compared by the elicitation team to the known values. These comparisons can then be used as measures of discrepancy between experts' estimates and "the truth", therefore they can be used as performance measures which can in turn be used to inform the mathematical aggregation of the individual second estimates.

## The digital preservation risk project

In 2018, a pilot project between The National Archives and the AS&RU took place to investigate whether an integrated decision support system based on Dynamic Bayesian network would be an effective model to quantify the risks to digital archives. The AS&RU held a 2-day workshop with 9 staff from The National Archives where a pilot model was created, and elicitation tested as a technique. From this preliminary work, it was decided that this approach would be well-suited to create a graphical risk assessment model for The National Archives' digital archive.

The AS&RU applied to the Engineering and Physical Sciences Research Council Impact Acceleration Award for a grant in order to fund a post-doctorate researcher for 12 months. This was approved and Dr Thais Fonseca joined the AS&RU team in September 2019. The National Archives recruited for a Research Assistant in Applied Statistics and Hannah Merwood filled this post on secondment from the Department for Digital, Culture, Media and Sport in November 2019.

The project team decided that this research could have a greater impact if the model was built to be useful to all UK digital archives rather than just The National Archives, so they reached out to the wider archival community to see if any other organisations would like to get involved. The archives which responded and became additional project partners are as follows[5]:

- Dorset History Centre
- Gloucestershire Archives
- Transport for London Corporate Archives
- Special Collections, University of Leeds
- Design Archives, University of Brighton

By widening the pool of archives involved, the scope was changed to create a model that would be useful to the whole UK archive sector rather than just The National Archives. To help support the additional costs this would bring, The National Archives was awarded a grant by the National Lottery Heritage Fund (NLHF), in January 2020. NLHF funding also enabled the direct involvement of the Digital Preservation Coalition to conduct the project evaluation, assist with workshops, and to introduce the risk model to the wider digital preservation community.

[5] For the full list of individuals see 'Acknowledgements'

# Preparation for the elicitation

**Model development and data gathering**

On 21st and 22nd November 2019, AS&RU led the first project workshop at The National Archives in Kew, London which was attended by 18 participants from The National Archives, partner archives, and AS&RU. The morning of the first day was spent introducing the archivists to the theory of Bayesian networks. In the afternoon the participants gathered in three groups to brainstorm all the risks involved in Digital Preservation and started to identify the relationships between them. Overnight Jim Smith, Martine Barons and Thais Fonseca combined the archivists' ideas and proposed a first network of digital preservation risks which was then discussed and revised in detail on the second day. What the final output should be was also discussed, with partners saying they needed something simple that they could show and easily explain to senior managers, potentially something similar to a red-amber-green rating.

Afterwards, Hannah Merwood had some more detailed discussions with staff at The National Archives to better understand the network and whether it was accurate. Between 30 November 2019 and 28 January 2020, Thais and Hannah came up with several iterations of the network and expert panel groupings based on the discussions and data, and a new version was presented to partners at the second workshop on 29 January, which was held also in Kew, London. The first part of this day was spent in groups revising the latest network, definitions and key risk outcomes. This information was gathered and processed to inform further iterations of the model, and the structure and definitions were settled by mid-February.

After this first workshop, Hannah Merwood and Thais Fonseca also began exploring potential data sources for the model. It became clear that data on digital preservation risks was very sparse and where it did exist it was often too specific for our vision of a holistic model which considered everything from file format obsolescence to natural disasters. Therefore, expert elicitation would be essential. However, the JISC digital skills survey of UK archivists in 2019 appeared to be a promising data source on the state of the sector's current technical skills, policies and resourcing.

Once the final network was settled, Thais and Hannah compiled a list of all the conditional probabilities for which we needed data. This initially revealed 33 questions where elicitation data would be needed, which was more than could feasibly be elicited in a two-day workshop. Through careful reviews of node states, it was decided that some could be made deterministic in this first edition of the decision support tool as they were less directly influential on the outcomes of interest: that is the renderability and intellectual control of the digital objects. In this way, this list was reduced to 24 questions of interest instead.

For the calibration questions, Hannah reviewed the JISC survey, papers on storage failures and reports on cyber security. 34 potential questions were devised and the project team shortlisted this down to 20. The criteria used in selection were the relevance to digital preservation, the availability of the data i.e. were the answers easily found on internet search engines, whether they require reading through reports or manipulating open data, and whether they were findings that were not published. The full list of calibration questions considered can be found in Appendix E.

In order to ensure the experts gave the same effort and thinking process for the calibration questions and questions of interest, it was decided to present all the questions together, so that the participants would not be encouraged to distinguish between them. The calibration questions and the questions of interest were worded as closely as possible to minimise distinction between them. Whilst the team were confident that the answers to the calibration questions would be difficult to find in most cases, given time they could plausibly have been found. The full list of carefully composed and ordered questions was therefore circulated to participants for the first time on the first day of the workshop to ensure that all the experts were approaching the exercise on the same basis. See Appendix B.

**Selection of experts**

The date for the elicitation workshop was set to 28-29 April based on the availability of the project team and partners. At the start of the planning, we had intended for this to be a physical event which would be held at a conference venue in Kew, London. However, by the middle of March, restrictions were put in place on physical gatherings and travel due to the COVID-19 pandemic. It was decided that we would keep the same dates but move the workshop online instead, which The National Archives would host using the virtual conferencing platform Zoom (https://zoom.us/).

To derive a suitable pool of participants for the expert elicitation session all named project partners were initially invited, as they are all digital archivists. William Kilbride and Jenny Mitcham from the Digital Preservation Coalition were also invited as they have a lot of experience of working with digital archives and are experts on digital preservation risks. At The National Archives, Alex Green, Hannah Merwood and David Underdown discussed which members of staff from the Digital Archiving Department had relevant expertise. They invited all the Digital Archivists and Digital Preservation Researchers, the Digital Archive's Technical Architect and the Head of Service Assurance.

Once it was established that the workshop would happen online, two of the partner organisations asked if additional staff members could join. These two members were both digital archivists and so the project team agreed that their contributions would likely be valuable and additional attendees would not incur any additional costs to the project.

To ensure we had as many experts as possible, The National Archives also invited Caylin Smith and Simon Wilson directly – both UK digital archivists from other organisations who

The National Archives’ team knew from previous projects. Other digital archivists were invited to express interest in participating via an official tweet, which lead to Lucy Wales, a specialist in film archives, also joining.

The project team consulted with The National Archives’ Data Protection Officer to create an appropriate informed consent form for participants to complete in advance, in order to comply with the Data Protection Act 2018. All experts were sent the consent form and an explanation of the project in advance of the workshop, for the full text of these, see Appendix A. Attending experts were sent further details and the slide pack on Monday 27 April and reminded to return their informed consent forms.

# The elicitation workshop

**Description of the day**

Twenty-two experts had agreed in advance to attend the workshop on 28-29 April 2020 and there were no cancellations. Martine Barons and Thais Fonseca were in attendance as facilitators for the elicitation and clarified points where required. Alex Green and David Underdown gave presentations and acted as advisors on terminology and clarifications when needed. Hannah Merwood was the technical assistant and ran the slides, sent emails and managed the overall logistics of the event. A copy of the slides used for the workshop can be found in Appendix B.

The day commenced with presentations of the reasons for the elicitation by Alex Green, Hannah Merwood and David Underdown followed by a presentation of the IDEA protocol by Martine Barons. The nature of biases was explained and the importance of answering the questions in the order they are posed in order to avoid anchoring was also discussed. The definitions of the variables were scrutinised, and the questions discussed in detail to avoid any problems with clarity or meaning. A full list of supplementary definitions shared with participants on the day can be found in Appendix B.

The existence and purpose of calibration questions were clearly explained in the introduction although care was taken not to reveal that the questions set incorporated calibration questions during discussions

During this discussion, three questions were found to be unfit for purpose due to confusion over their interpretation. After the day 1 discussion, David Underdown and Hannah Merwood reviewed the model and original rationale behind these questions and came up with alternatives which were then circulated to the participants.

*Table 1: The questions which were revised during the first day.*

| # | Original Question | Revised Question |
|---|---|---|
| 31 | Out of 1,000 files where a checksum was generated instantly on deposit, at an archive where there is sufficient information management and good system security, how many files can you expect to provide the assurance that the bit-stream is identical to when it was added to the archive? | Out of 1,000 files, how many would you expect to have become corrupted during transfer from a depositor to an archive? |
| 32 | Out of 1,000 where a checksum was generated instantly on deposit, at an archive where there is sufficient information management but poor system security, how many files can you expect to provide the assurance that the bit-stream is identical to when it was added to the archive? | Out of 1,000 files, at an archive where you have poor system security, how many would you expect to have become corrupted and identified this, given you have a checksum to compare to? |
| 33 | Out of 1,000 files which were all deposited with a checksum, at an archive where there is sufficient information management but poor system security, how many files can you expect to provide the assurance that the bit-stream is identical to when it was added to the archive? | Out of 1,000 files, at an archive where you have poor system security, how many would you expect to have become corrupted and not been able to identify this, despite having a checksum to compare to? |

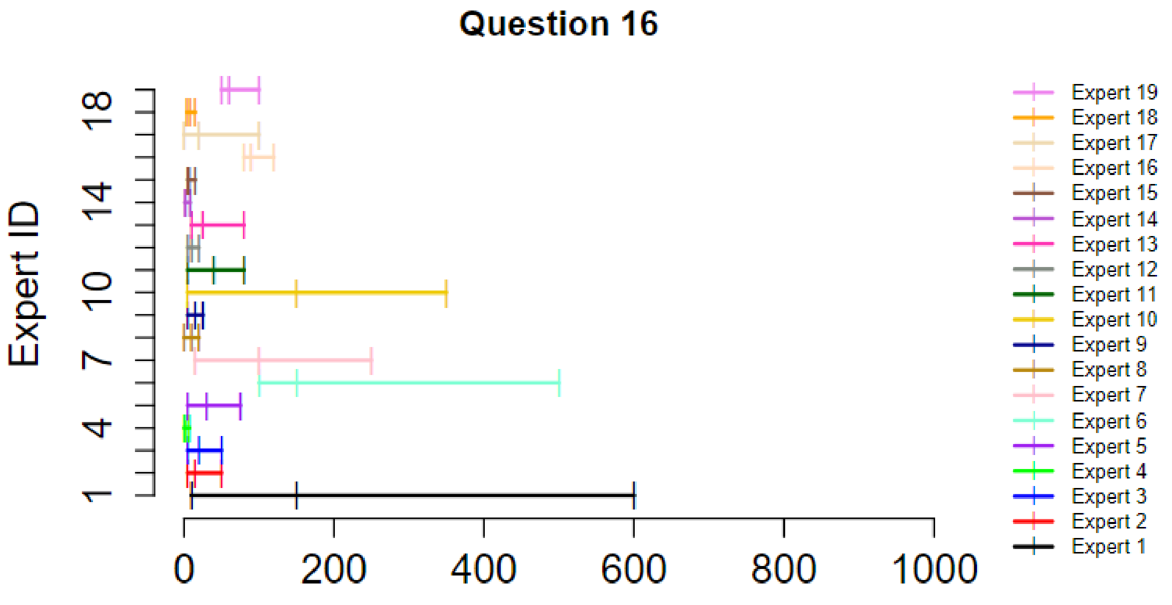
Martine Barons facilitated the process of the IDEA protocol. The experts then gave their individual first-round subjective estimates of the lowest plausible, highest plausible and best estimate of natural frequencies for the questions by 19:00 that evening in a spreadsheet returned to Alex Green. The questions were presented as in table 2. A full list of questions can be found in Appendix B, where the calibration questions are identified by an asterisk.

*Table 2: Extract from the spreadsheet template experts returned their answers on.*

| Number | Question | Your 5%ile | Your 95%ile | Your 50%ile |
|---|---|---|---|---|
| 1 | Out of 1,000 UK archivists, how many would you expect to say that their archive has a list of all the digital file formats collected? | | | |
| 4 | Out of 1,000 files with file formats that are neither ubiquitous nor open, at an archive where staff have good technical skills, how many would you expect to have the tools to render? | | | |

Thais Fonseca produced range graphs for each question overnight, in preparation for a facilitated discussion the next day. Range graphs show the highest plausible, lowest plausible and best estimate for each expert on each question in order to provoke discussion about the likely values, the uncertainty of experts and of the system itself. One graph is produced for each question and the experts are labelled with different numbers for each question to preserve their anonymity until they decide to identify themselves.

*Figure 2: Range graph for the answers to question 16 in the first round as presented to the experts, where there was consensus that the answer was very low. Note that there are 19 responses here, as three experts did not return their answers in time for processing by the end of the first day.*

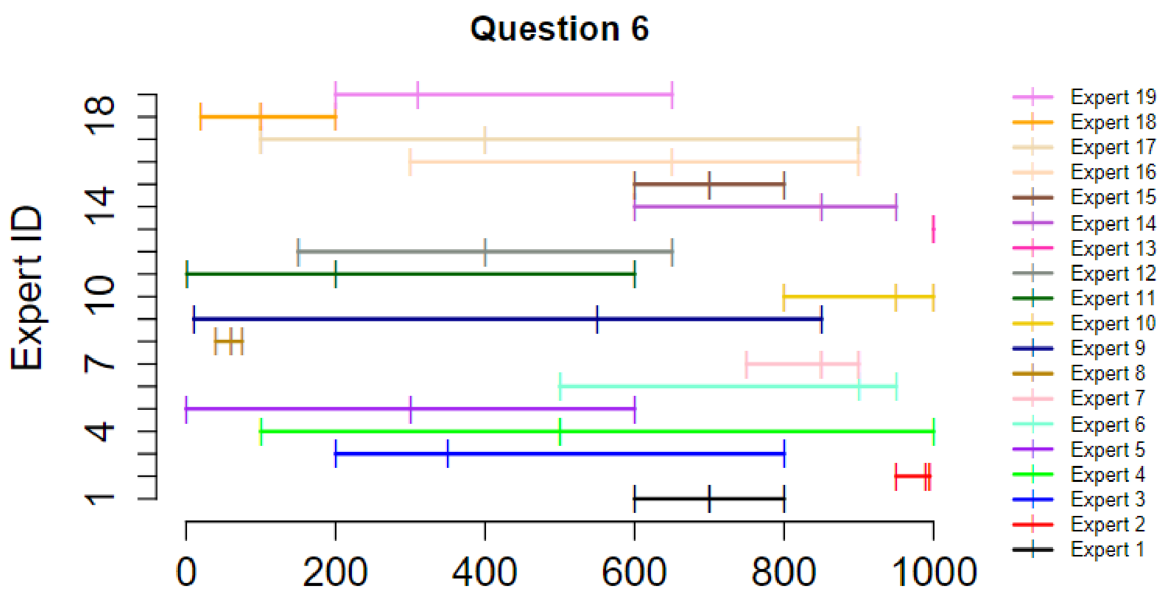


The discussion focused on the reasons for the differences observed – under what circumstances the estimate could be as low as the lower end of the range, under what circumstances the estimate could be as high as the higher end of the range, and why the most likely values fell where they did.

For question 16, there was a significant consensus on where the lower end of the plausible range should lie, but more variability in the upper end. Experts 6, 7 and 10 gave most plausible values which fell well outside most other experts' upper plausible ranges.

For question 6, there was no clear consensus in the first round, with many non-overlapping ranges being given.

*Figure 3: Range graph for the answers to question 6 in the first round as presented to the experts, where there was a lot of variation. Note that there are 19 responses here, as three experts did not return their answers in time for processing by the end of the first day.*




A full set of range graphs can be seen in the Results section.

Martine Barons was the facilitator of the discussions and ensured that each expert had the opportunity to contribute. At the close of the discussions for each question, the experts gave a second set of individual, subjective estimates for those questions, retaining their original estimates for comparison. A transcript of the discussion can be found in Appendix C.

THE NATIONAL ARCHIVES

WARWICK

# Results

### Calibration Questions

Calibration questions are used to access how accurate and informative each expert is. They are questions related to the experts' field, for which the true values are known to the statistician but not to the experts.

The questions in table 3 were the calibration questions used for this project.

*Table 3: The calibration questions and the true answers.*

| # | Question | True answer |
|---|---|---|
| **1** | Out of 1,000 UK archivists, how many would you expect to say that their archive has a list of all the digital file formats collected? | 300 |
| **2** | Out of 1,000 UK archivists, how many would you expect to say that they had at least some knowledge/skill to perform file format analysis of a digital accession? | 510 |
| **3** | Out of 1,000 UK archivists, how many would you expect to say that there was some capability within their organisation to carry out at least one of a) file format migration, b) software emulation or c) data recover from damaged/obsolete media, even in a limited way? | 440 |
| **9** | Out of 1,000 UK archivists, how many would you expect to say that their catalogue management system met the needs of their organisation? | 700 |
| **10** | Out of 1,000 UK archivists, how many would you expect to say that their digital asset management system met the needs of their organisation? | 400 |
| **14** | Out of 1,000 hard drive disks kept in a monitored commercial environment, how many drives would you expect to fail in any 12 month period? | 19 |
| **15** | Out of 1,000 hard drive disks kept in a monitored commercial environment, how many drives would you expect to fail within their first 12 months of use? | 51 |
| **16** | Out of 1,000 NAND solid state drives kept in a monitored commercial environment, how many drives would you expect to fail in any 12 month period? | 2 |

| | | |
|---|---|---|
| **17** | Out of 1,000 NAND solid state drives kept in a monitored commercial environment, how many drives would you expect to experience a persistent read error within any 12 month period? | 50-220 |
| **26** | Out of 1,000 UK archivists, how many would you expect to say that they had at least some knowledge/skill to be able to generate a checksum of a digital file? | 520 |
| **27** | Out of 1,000 UK archivists, how many would you expect to say that their IT provider supports the requirements of their organisation's archival activities to a large extent or a very great extent? | 610 |
| **28** | Out of 1,000 UK businesses and charities which experienced a cyber-security break or attack, how many would you expect to have experienced by viruses, spyware or malware, including ransomware attacks? | 250 |
| **29** | Out of 1,000 global data breaches, how many would you expect to be due to system glitches? | 250 |
| **30** | Out of 1,000 UK archivists, how many would say that their digital collections were fixity checked at regular intervals? | 170 |
| **39** | Out of 1,000 UK archivists, how many would you expect to say that their digital collections were protected with access restrictions/permissions? | 610 |
| **40** | Out of 1,000 UK archivists, how many would you expect to say that they had at least some knowledge to be able to digitally redact part of a document for web publication? | 460 |
| **41** | Out of 1,000 UK archivists, how many would you expect to say that their organisations' born-digital materials were made available to users? | 370 |
| **42** | Out of 1,000 UK archivists, how many would you expect to say that their organisations’ digitised materials were made available to users? | 800 |
| **43** | Out of 1,000 UK non-cyber security incidents, how many would you expect to be due to data being posted, emailed or faxed to an incorrect recipient? | 310 |

| | | |
|---|---|---|
| **44** | Out of 1,000 UK non-cyber security incidents, how many would you expect to be due to loss/theft of paperwork or data left in an insecure location? | 110 |

All calibration questions were presented to be ‘Out of 1,000’ to allow consistency in style. This also meant an additional calculation had to be performed on all the data to convert to this scale, which makes the original data source less likely to be recognised and less easily searchable.

The majority of the calibration questions were based on the JISC Archive Sector Digital Skills Survey from 2019, produced to help inform The National Archives’ digital capacity building strategy[6]. Questions 1, 3, 10, 27, 39, 40 and 42 were derived directly from the raw data which has not been published, although The National Archives staff do have access to the raw data within the corporate electronic document management system. The responses to questions 2, 9, 26, 30 and 41 could be found embedded in the text of the strategy publication, however it was deemed that participants would be unlikely to recognise the question and finding answers in the publication would require very specific searches.

A few of The National Archives’ experts were familiar with the survey as they helped to develop the strategy but given how many questions were in the survey, it was deemed unlikely that they would be able to recall these answers. Scaling the question to be ‘Out of 1,000’ also helped to disguise the source, as the survey was done with approximately 400 participants. This strategy seems to have been successful as one expert commented that the questions seemed very similar to those in the JISC survey but they could not be those because there were not 1,000 responses to that survey.

The other data sources used were as follows:

- Questions 14 and 15 came from statistics published by Backblaze, a cloud storage and computer backup company which uses hard drive disks as their storage medium.[7] These questions were arguably the ones for which the answers could be most easily found, because they are available directly on the Backblaze website and displayed as a percentage.
- Question 16 and 17 were both from papers presented at the USENIX Conference

[6] The National Archives ‘Plugged in, Powered up: a digital capability building strategy for archives’ https://www.nationalarchives.gov.uk/documents/archives/digital-capacity-building-strategy.pdf

[7] Backblaze blog on hard drive statistics *https://www.backblaze.com/blog/hard-drive-stats-for-2019/* and analysis of failure rates by age *https://www.backblaze.com/blog/how-long-do-disk-drives-last/.*

on File and Storage Technologies in 2020 and 2016 respectively.[8]

- Question 28 is based upon data from the Department for Digital, Culture, Media and Sport's Cyber Security Breaches Survey 2019.[9]
- Question 29 was from the IBM Security Cost of Data Breach report 2019.[10]
- Question 43 and 44 were derived from The Information Commissioner's Office reports on data security breaches 2019.[11] To determine the answers to these two questions, participants would have to download several data sets, combine them together and generate summary statistics.

---

[8] S. Maneas, K. Mahdaviani, T. Emami, and B. Schroeder (2020) 'A Study of SSD Reliability in Large Scale Enterprise Storage Deployments', *18th USENIX Conference on File Storage Technologies.* *https://www.usenix.org/system/files/fast20-maneas.pdf*; and

B. Schroeder, R. Lagisetty and A. Merchant (2016) 'Flash Reliability in Production: The Expected and the Unexpected', *14th USENIX Conference on File and Storage Technologies*. *https://www.usenix.org/system/files/conference/fast16/fast16-papers-schroeder.pdf.*

[9] *https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2019*

[10] *https://www.ibm.com/security/data-breach, report from 2019, accessed April 2020.*

[11] All data available for download from *https://ico.org.uk/action-weve-taken/data-security-incident-trends/previous-reports/.*

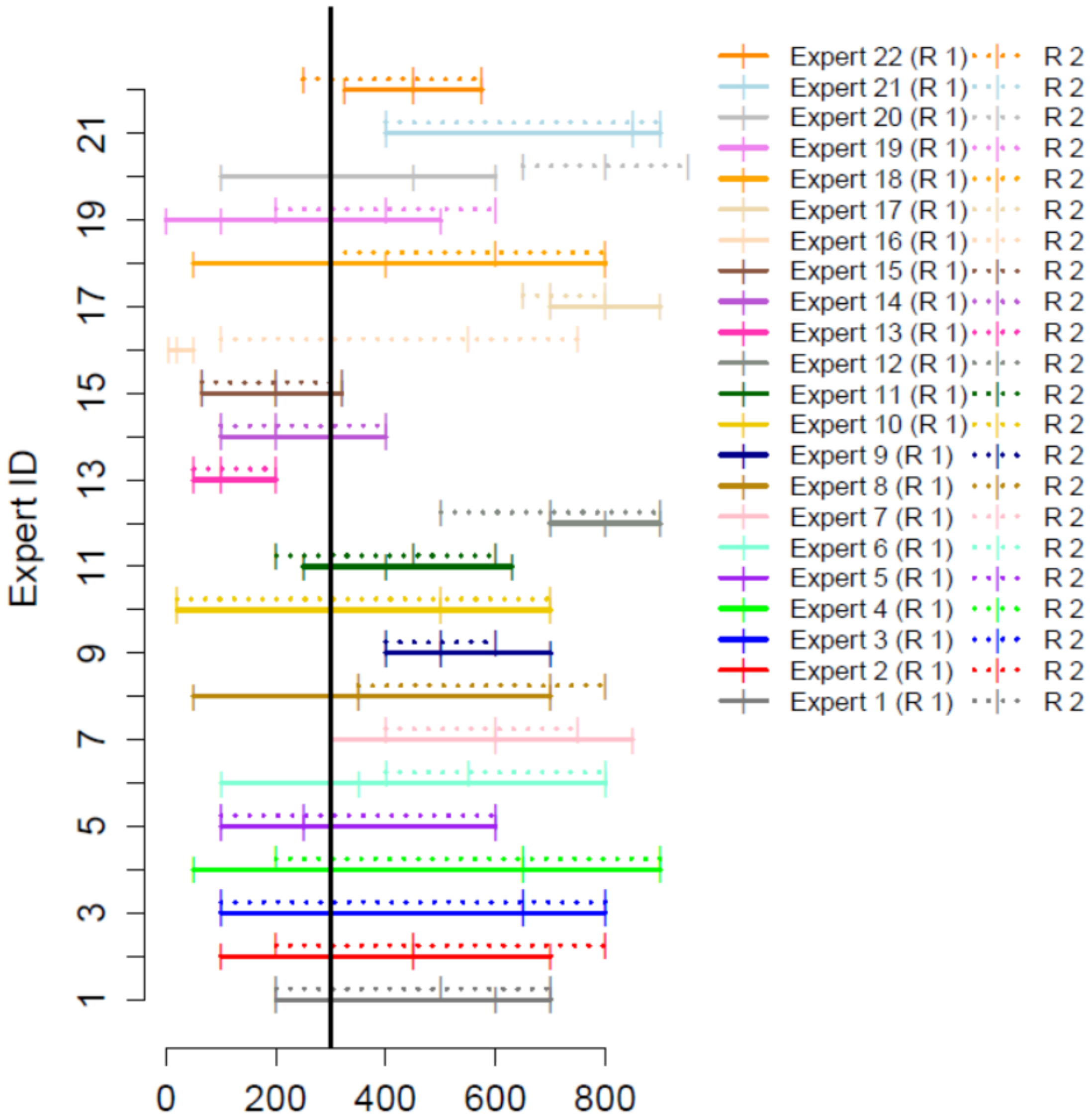
THE NATIONAL ARCHIVES

WARWICK

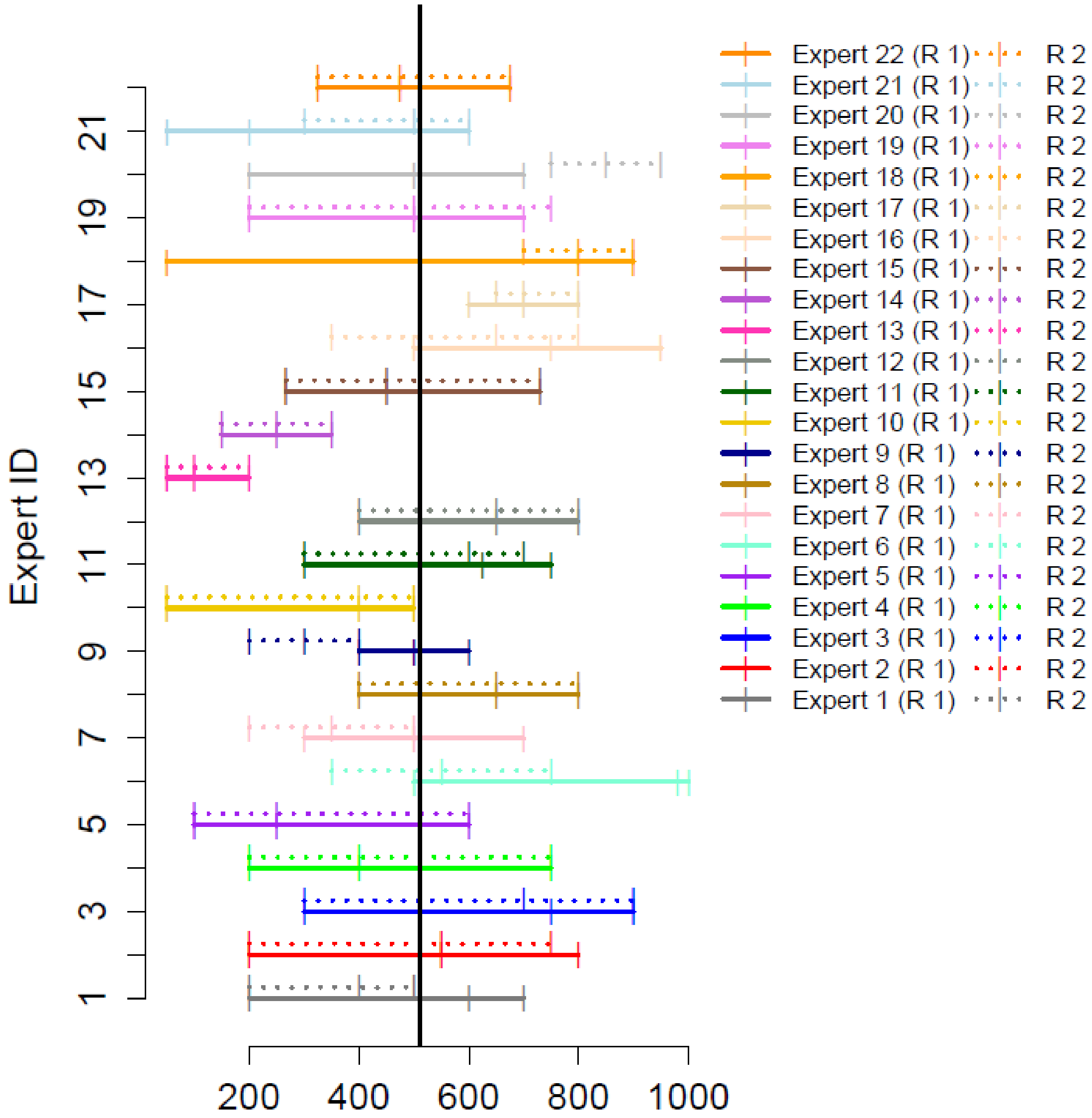
## Plots

For each plot the ranges for each expert are shown for their first round response (solid horizontal line) and second round response (dotted horizontal line) with tick marks indicating 5%-ile, median and 95%-ile values. The true value from the data source is shown as a vertical black line passing through all expert scores.

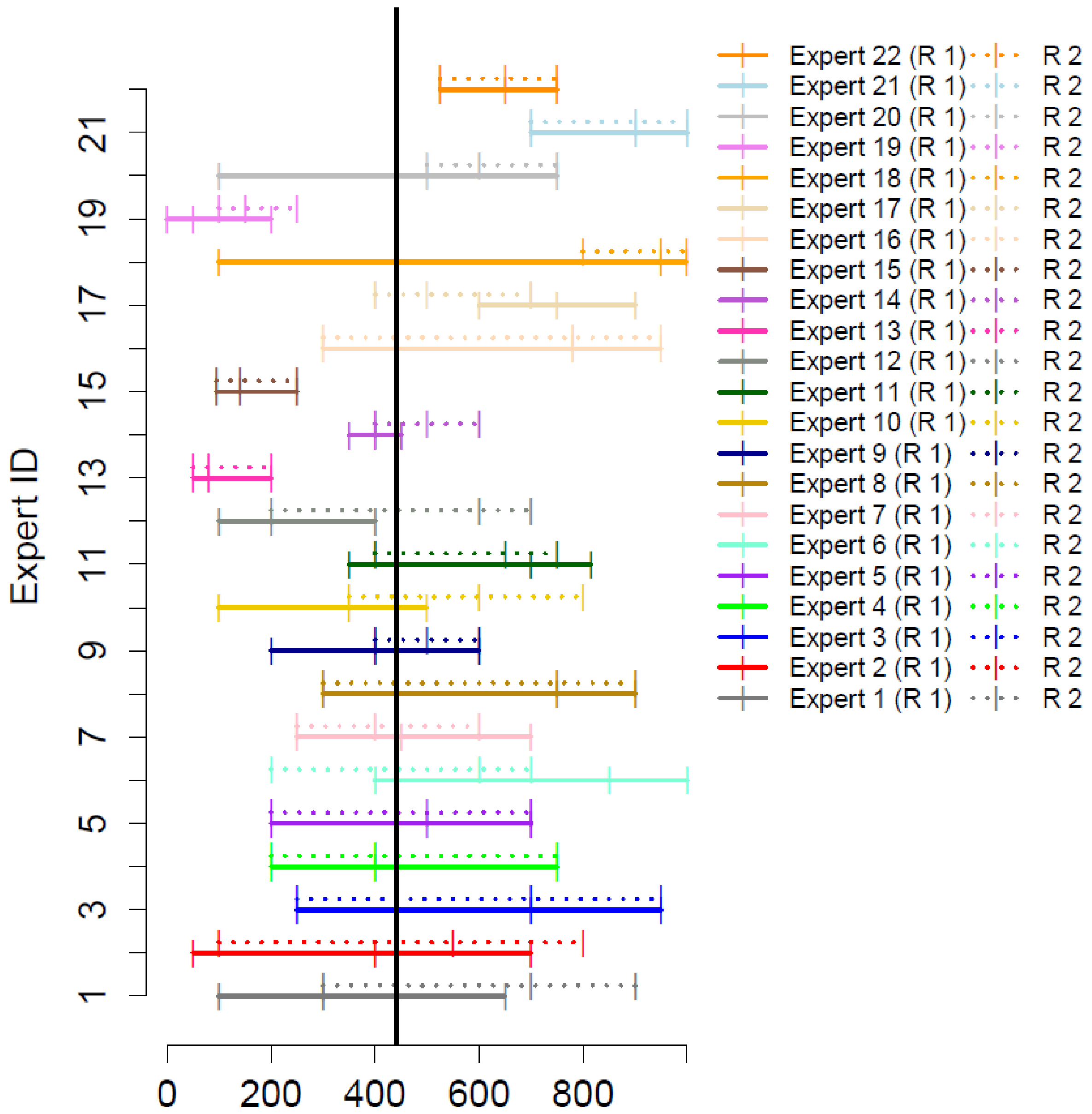
### Plot C1 (Question 1)



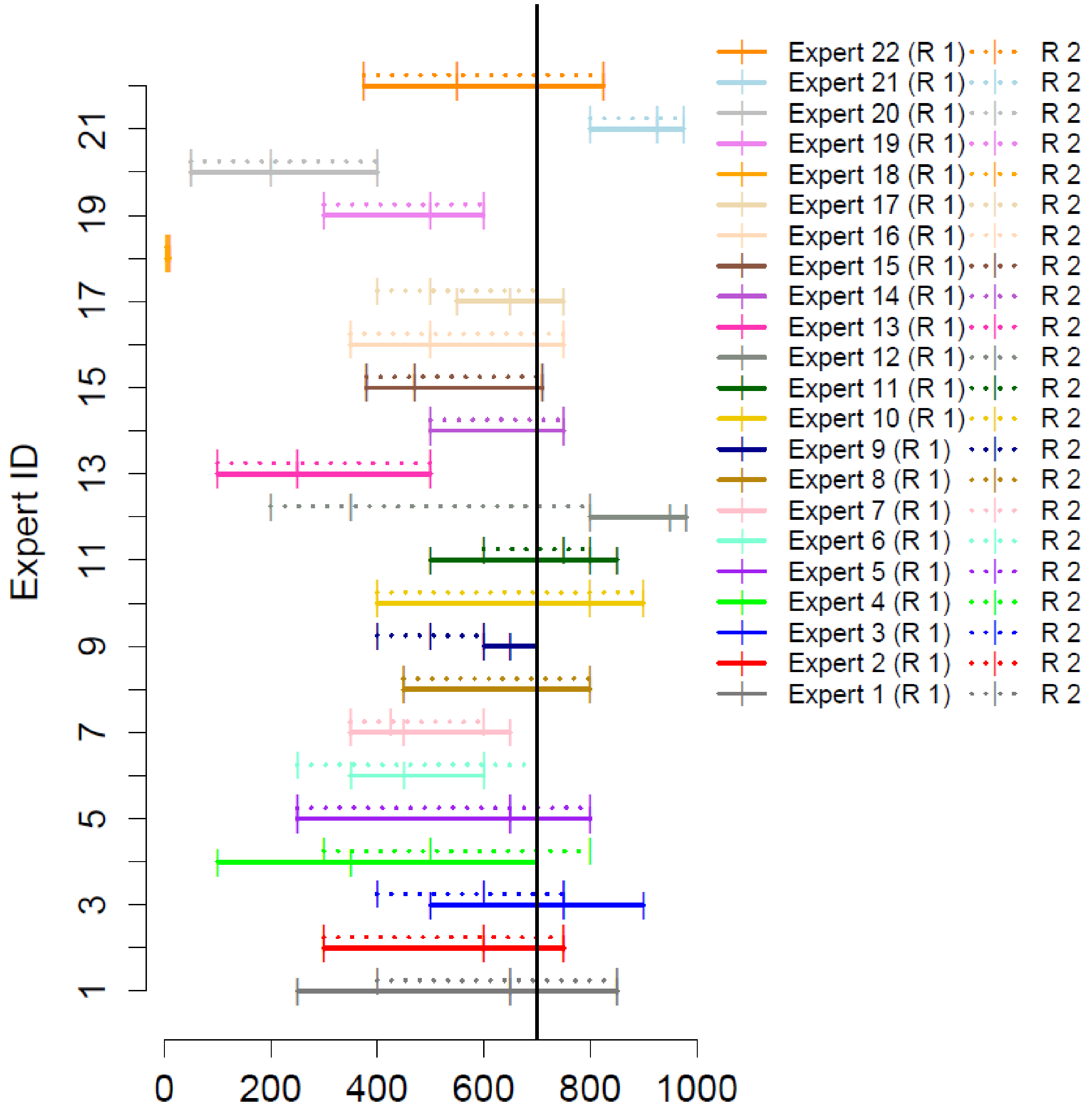
**Plot C2 (Question 2)**



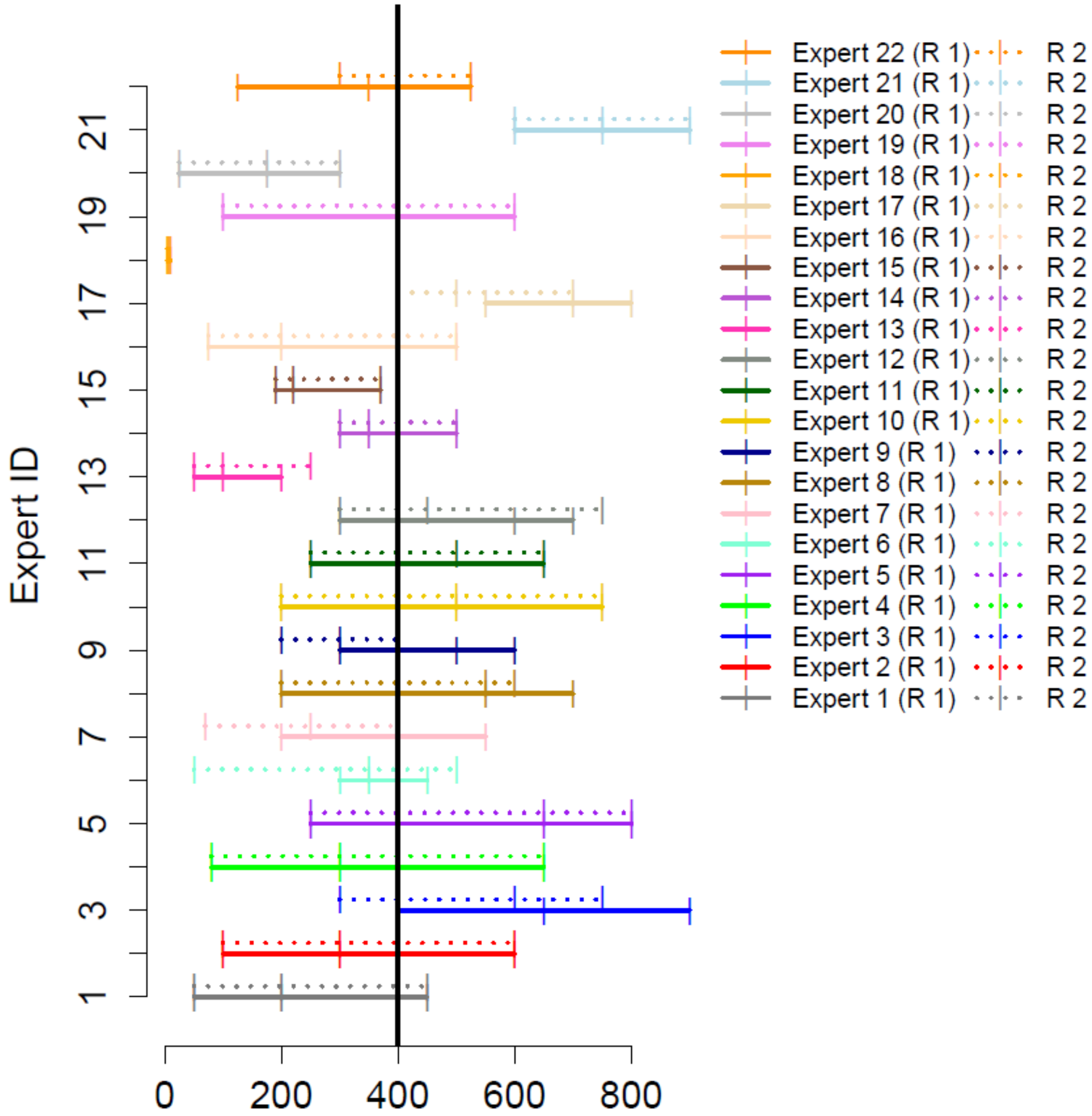
**Plot C3 (Question 3)**



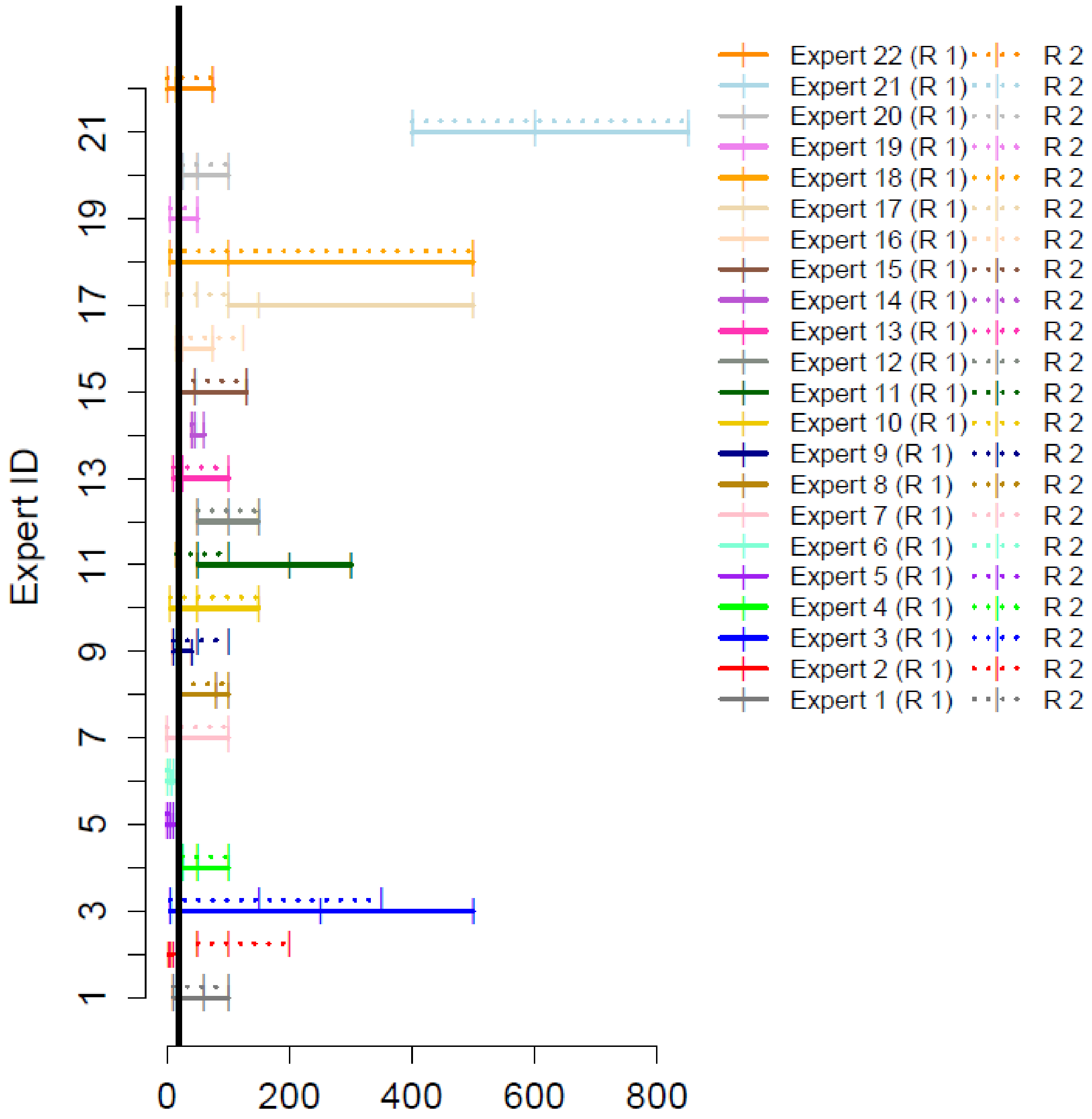
Plot C4 (Question 9)



**Plot C5 (Question 10)**



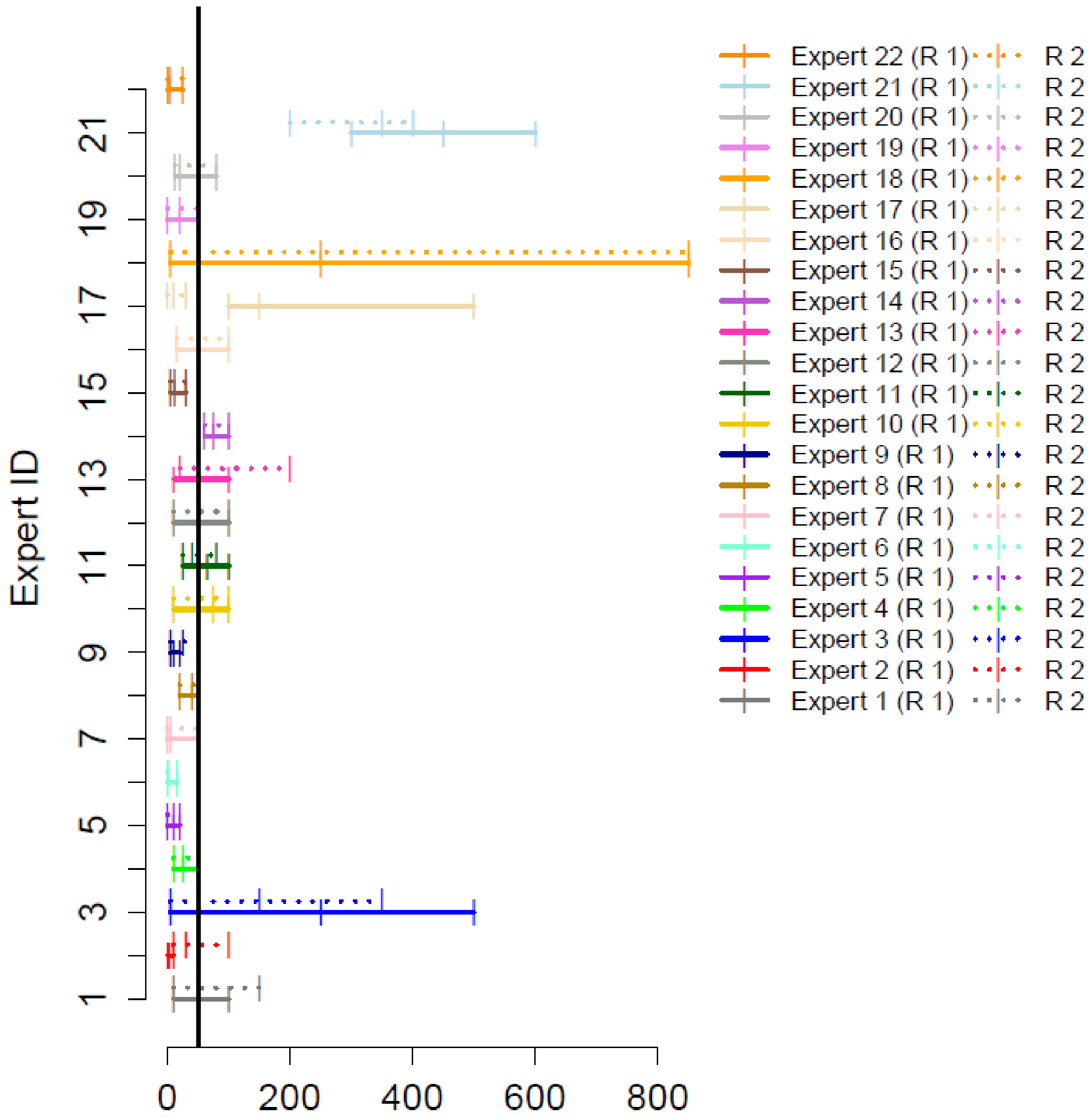
**Plot C6 (Question 14)**



THE NATIONAL ARCHIVES

WARWICK

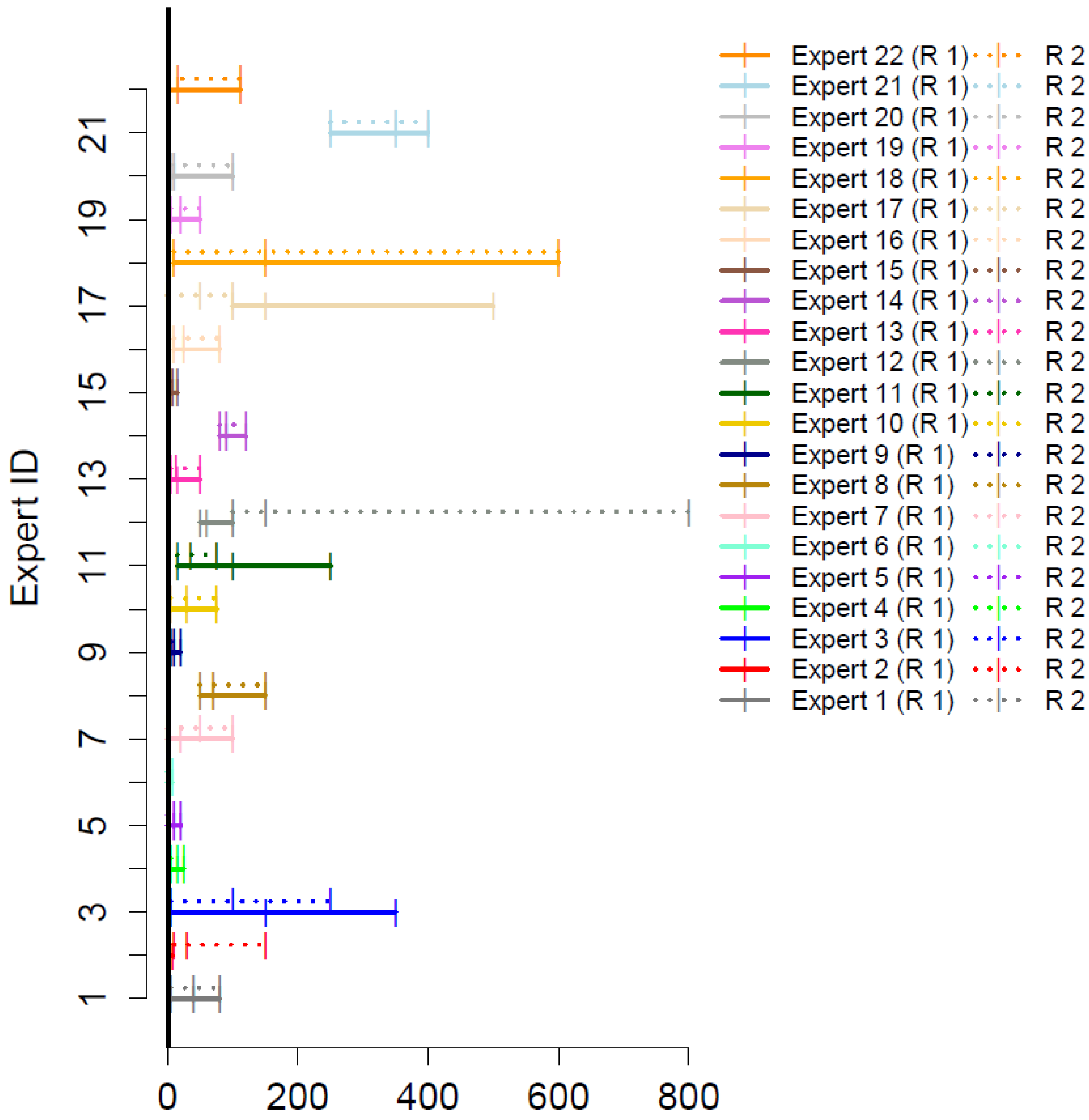
**Plot C7 (Question 15)**



THE NATIONAL ARCHIVES

WARWICK

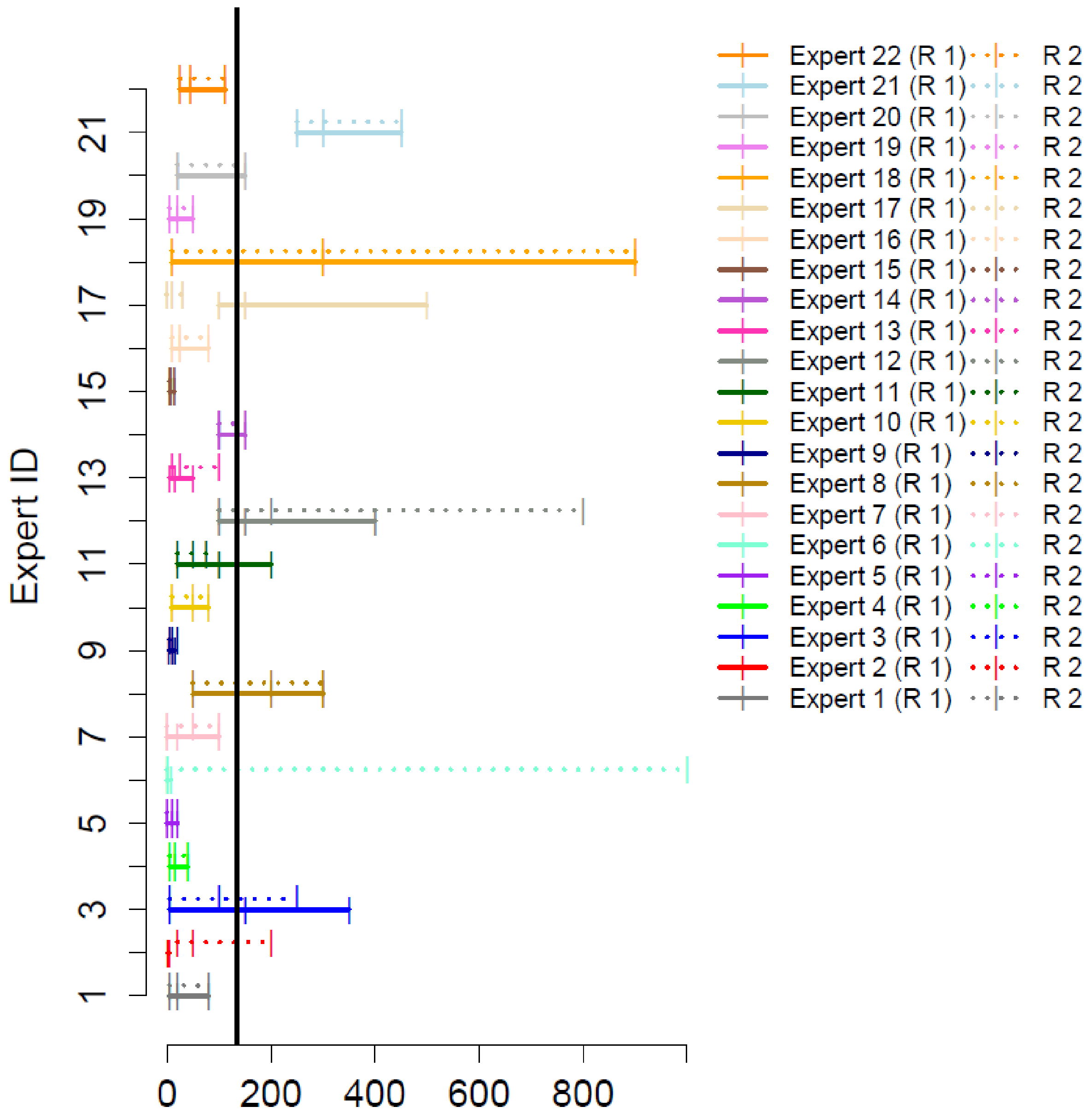
**Plot C8 (Question 16)**



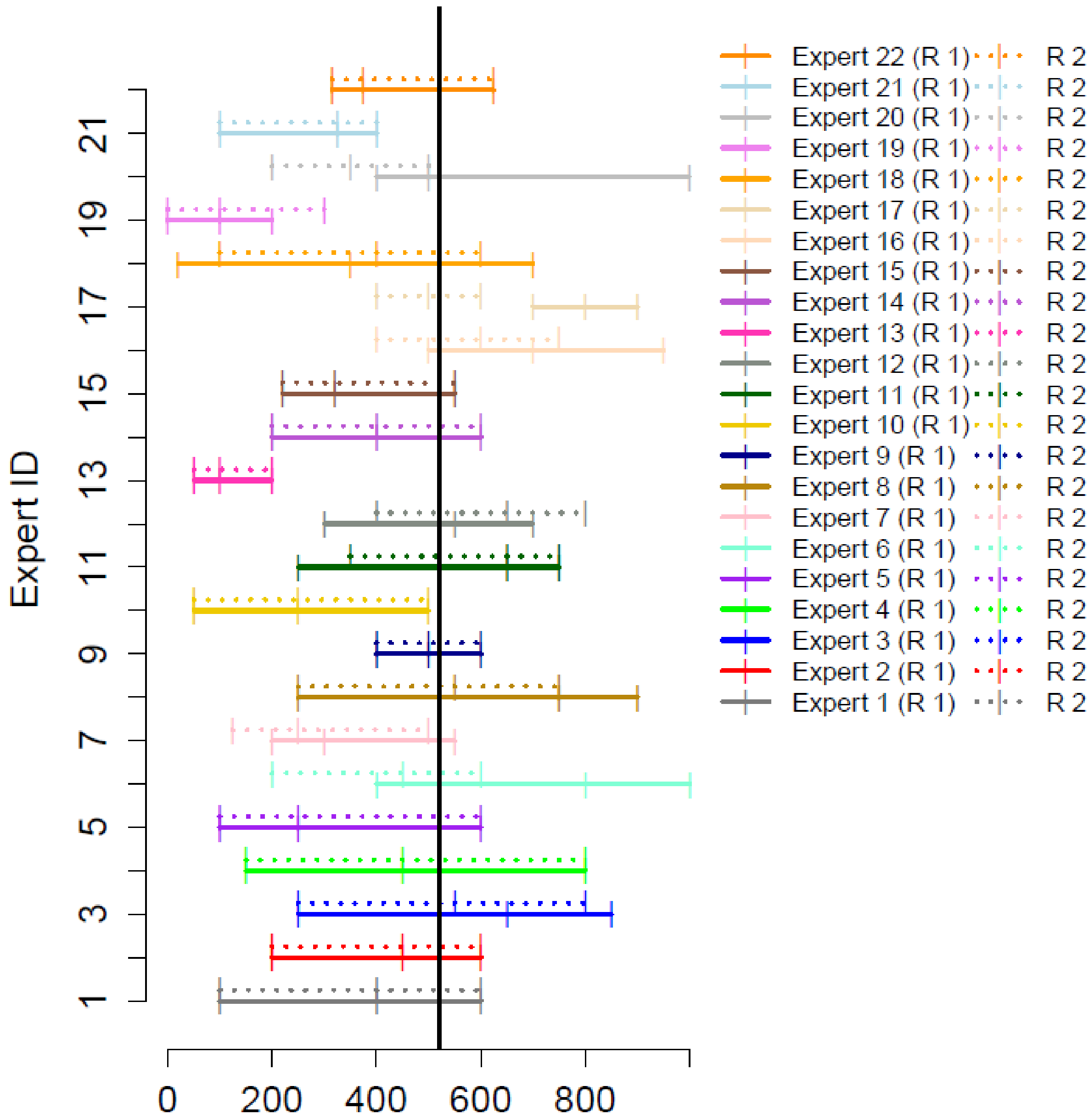
THE NATIONAL ARCHIVES

WARWICK

**Plot C9 (Question 17)**



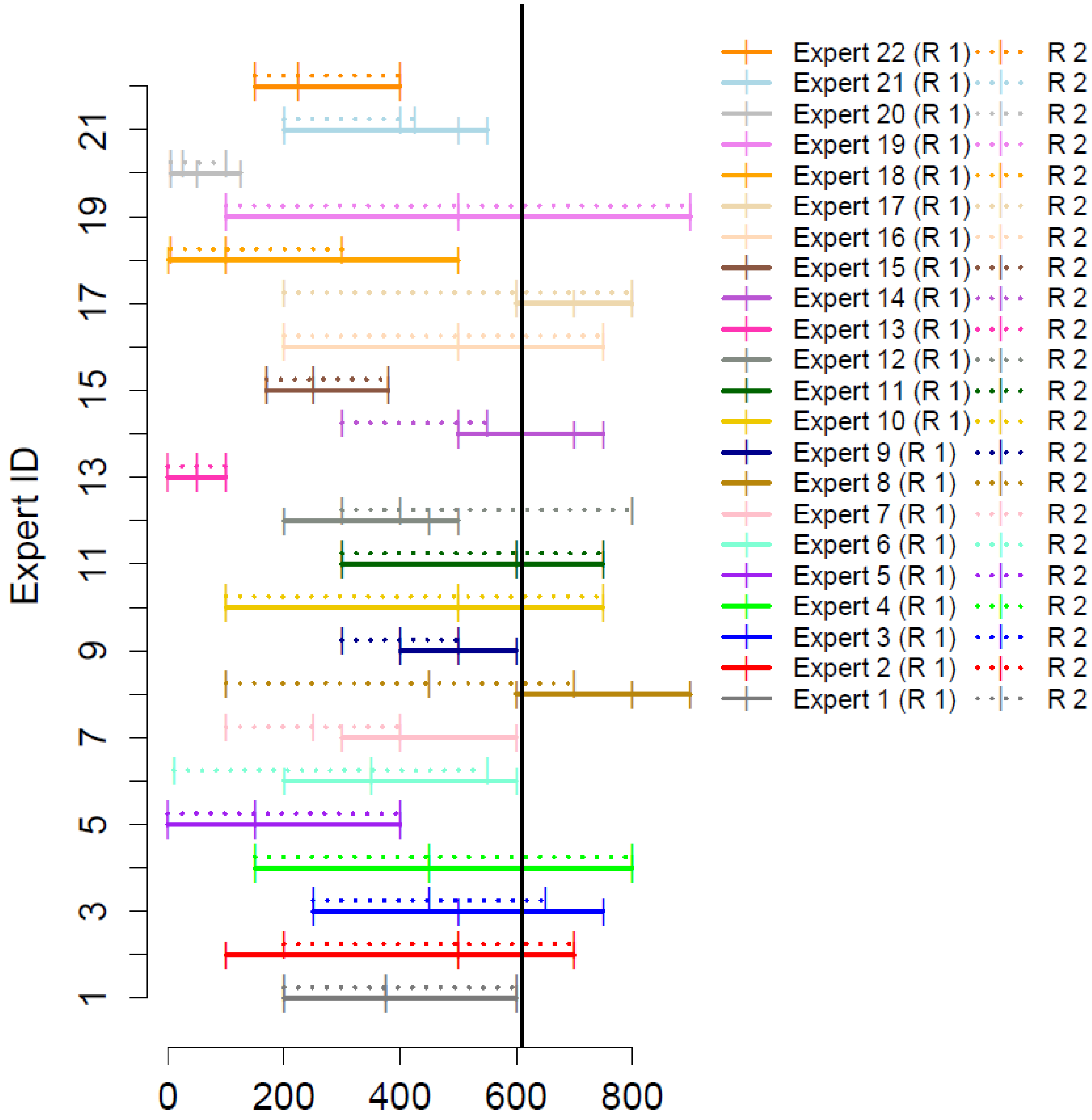
**Plot C10 (Question 26)**



THE NATIONAL ARCHIVES

WARWICK

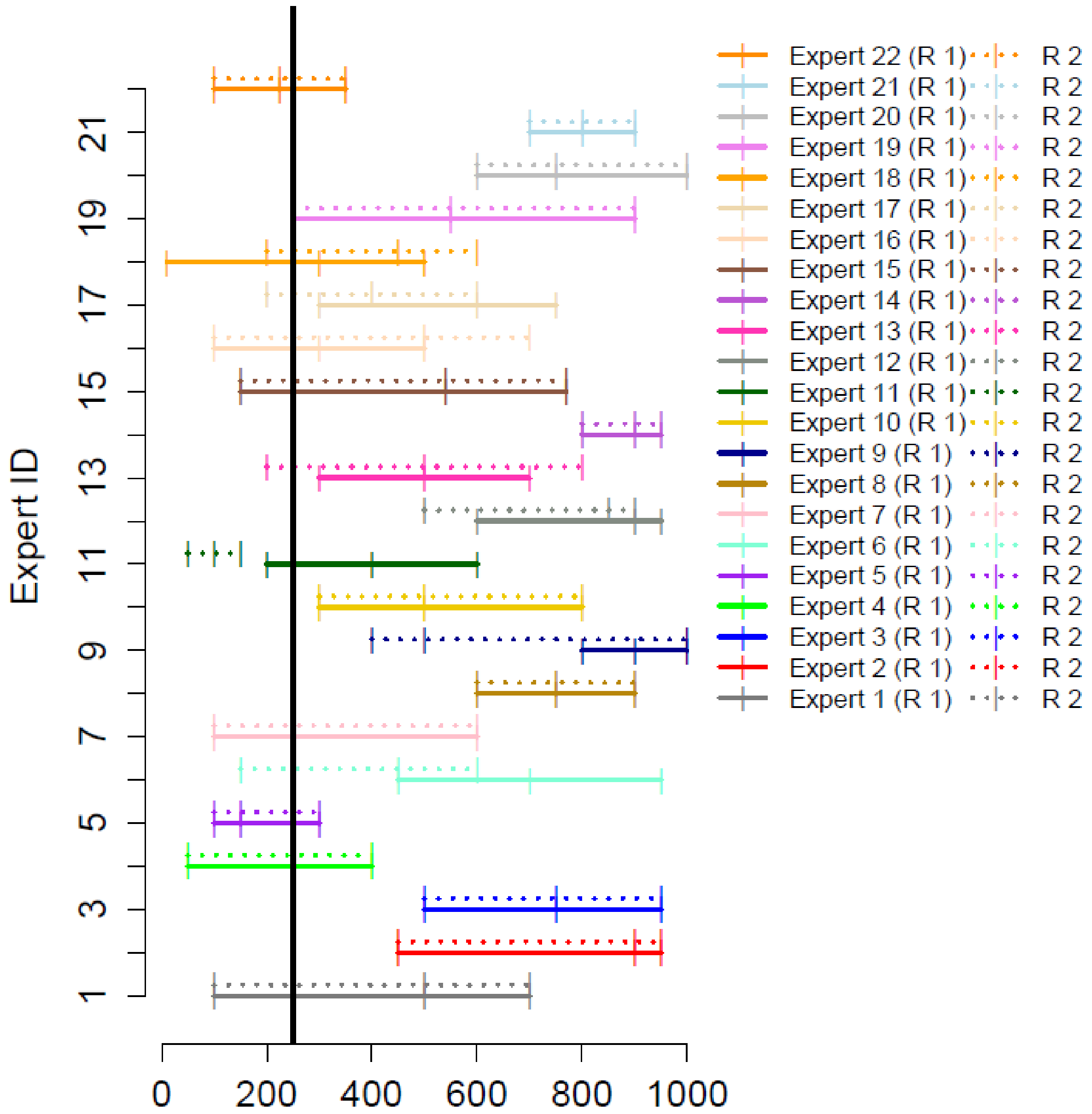
**Plot C11 (Question 27)**



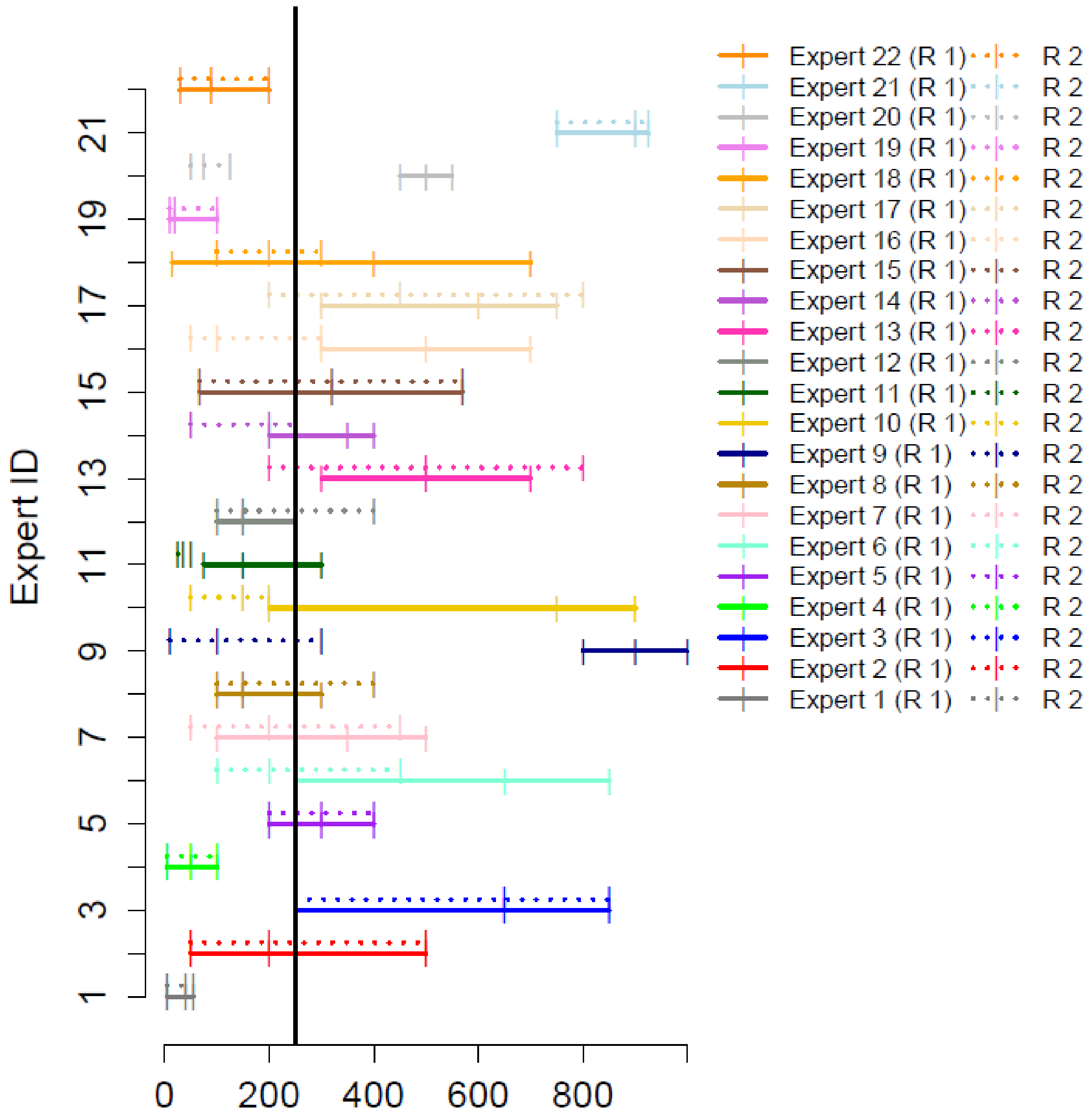
THE NATIONAL ARCHIVES

WARWICK

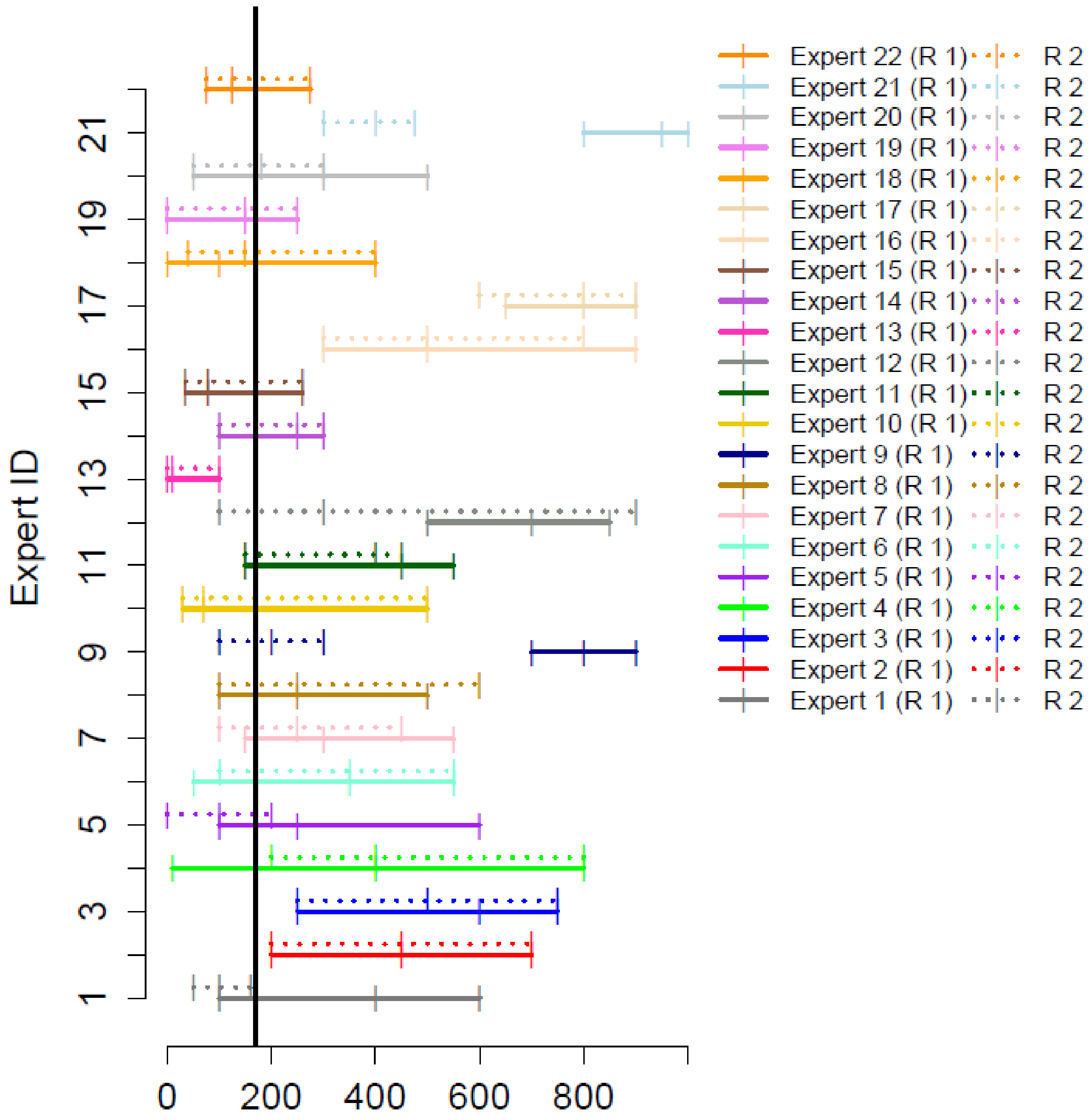
## Plot C12 (Question 28)



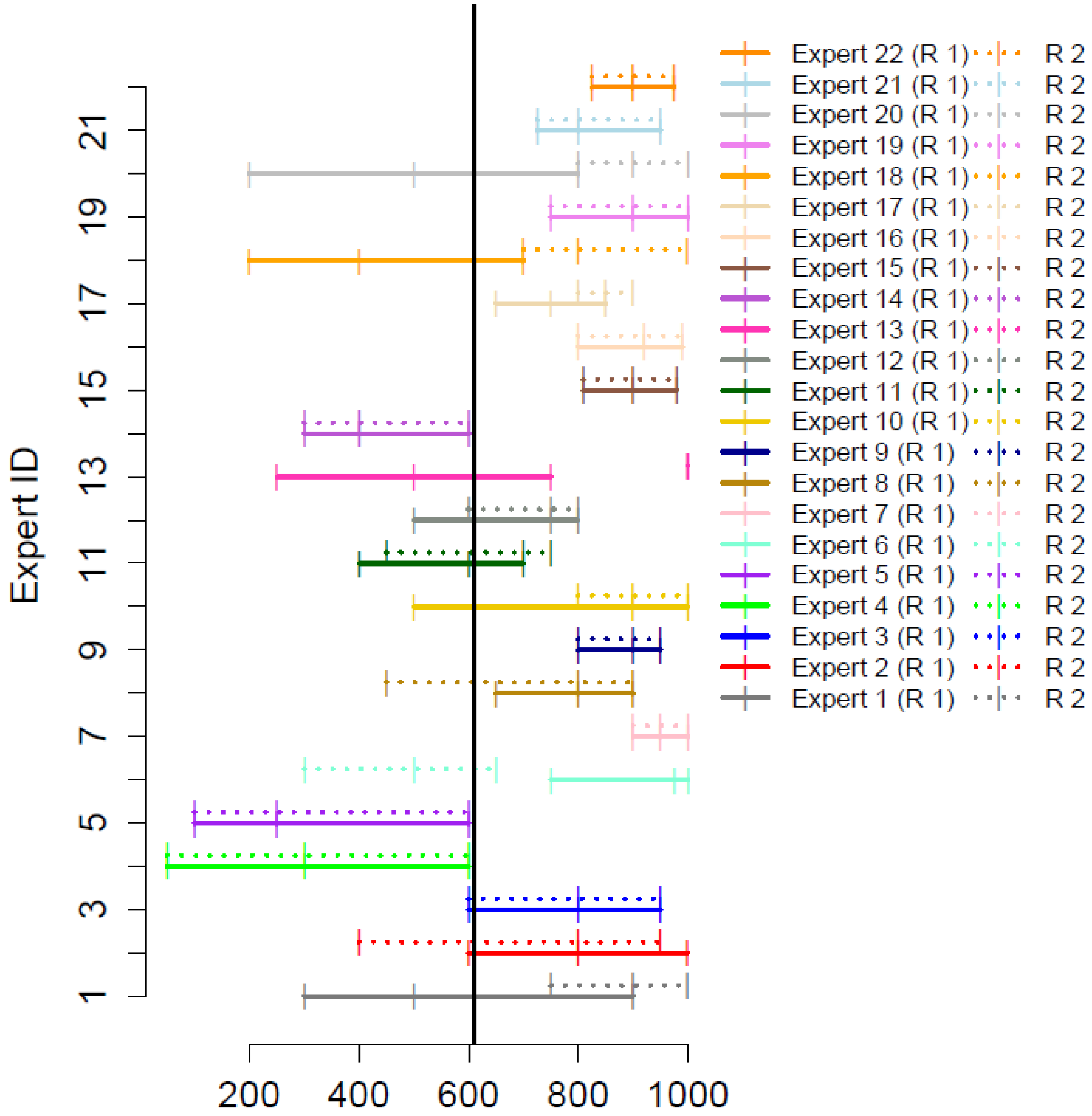
**Plot C13 (Question 29)**



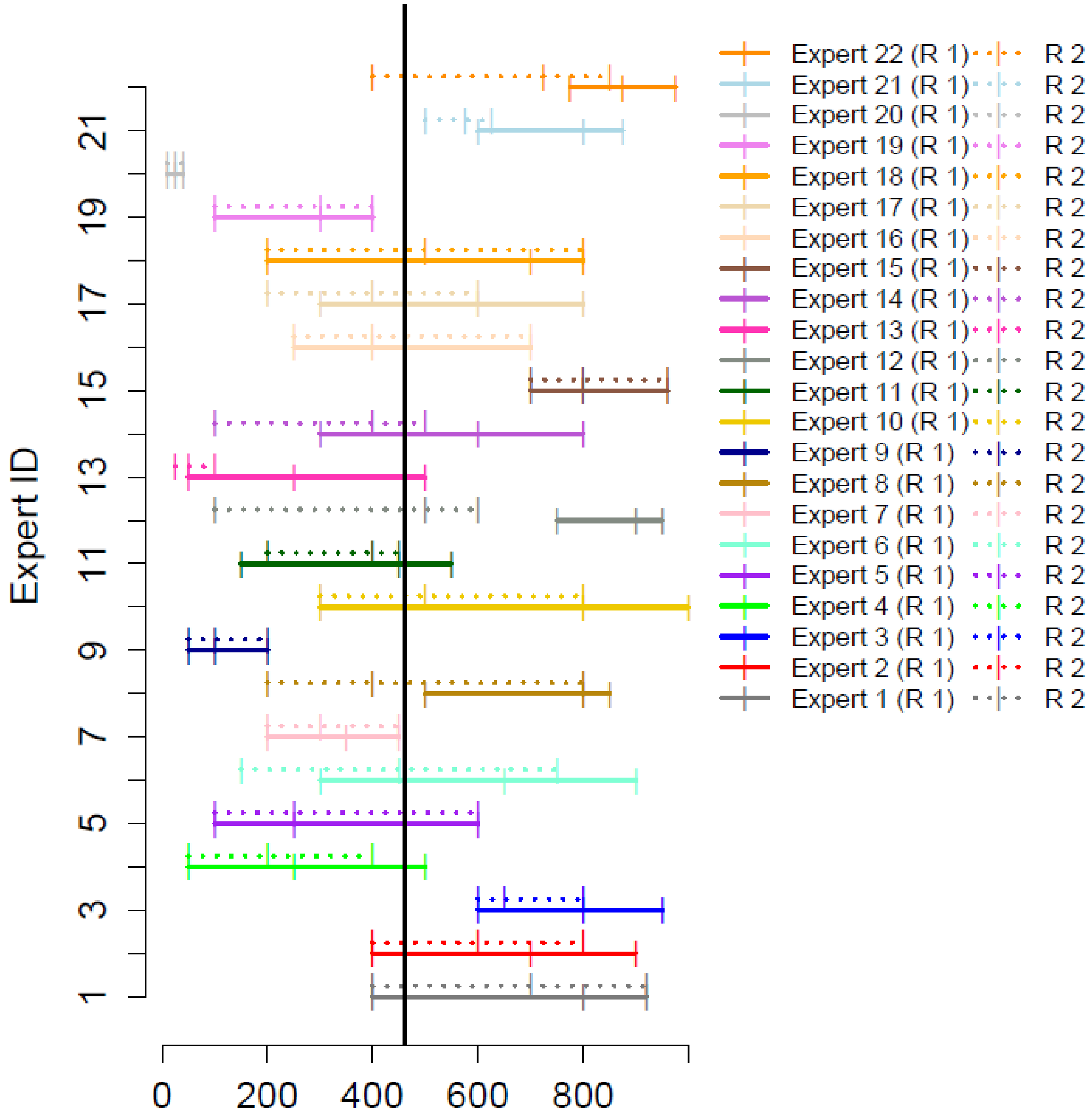
**Plot C14 (Question 30)**



**Plot C15 (Question 39)**



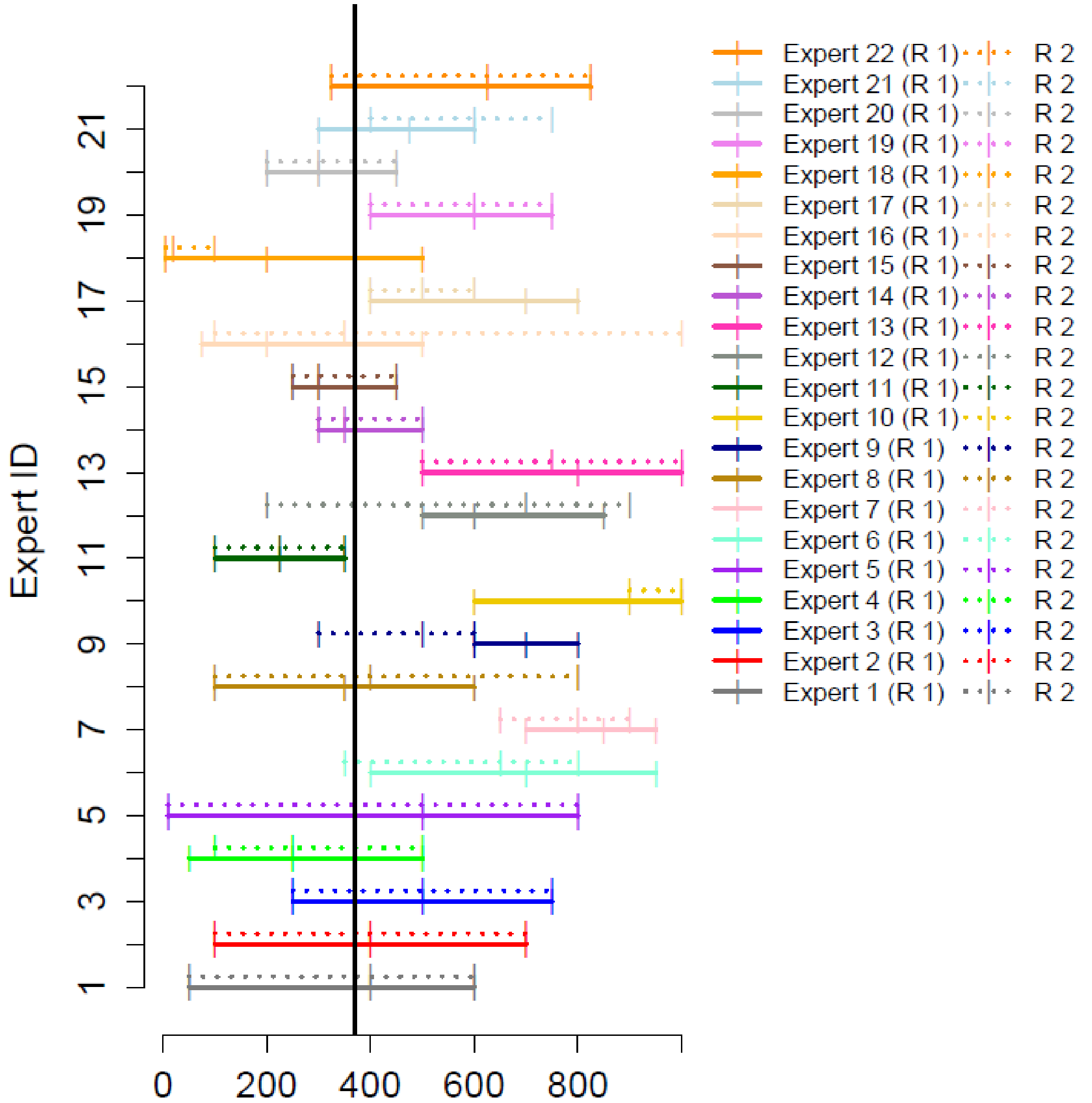
**Plot C16 (Question 40)**



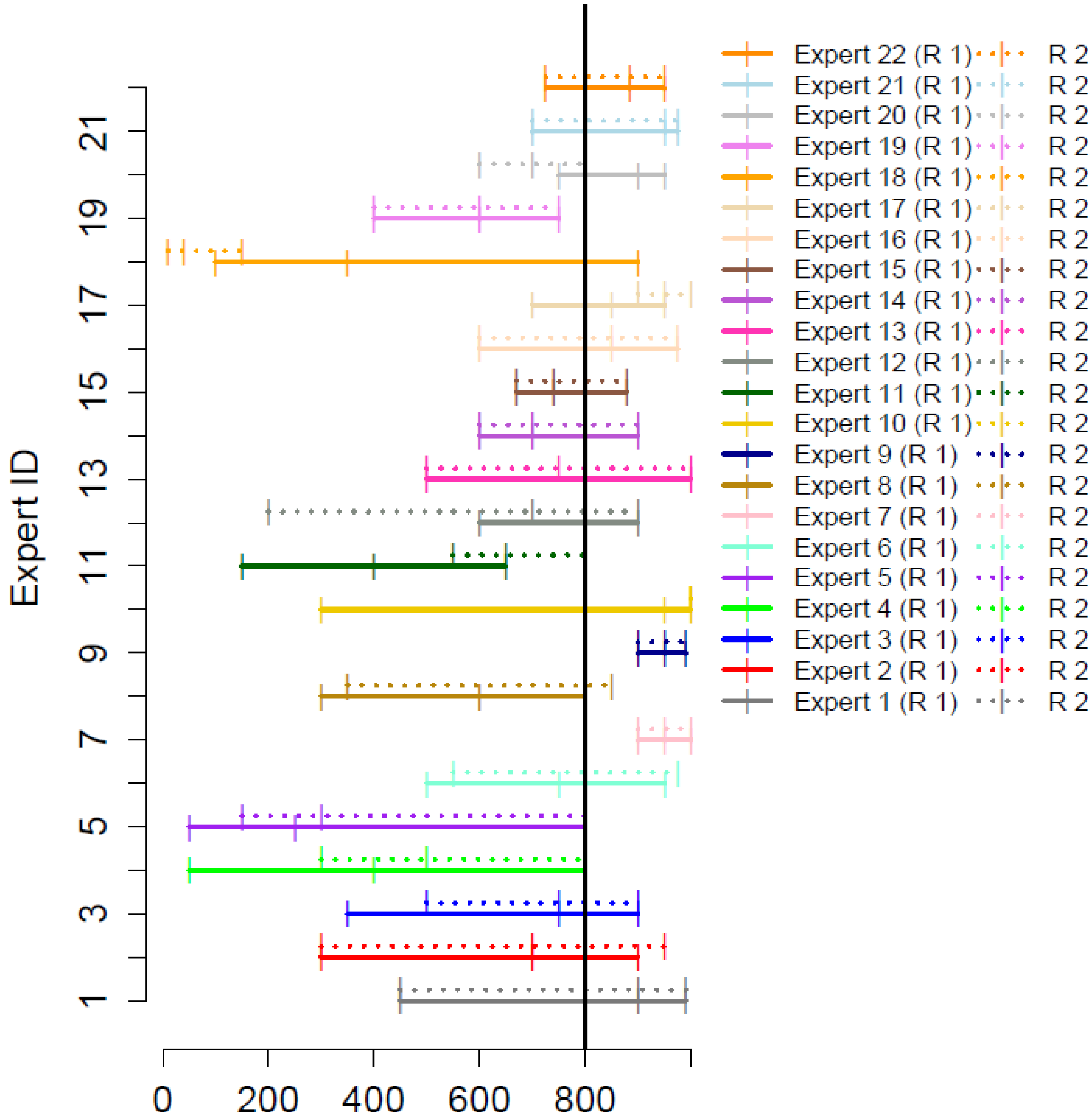
THE NATIONAL ARCHIVES

WARWICK

**Plot C17 (Question 41)**



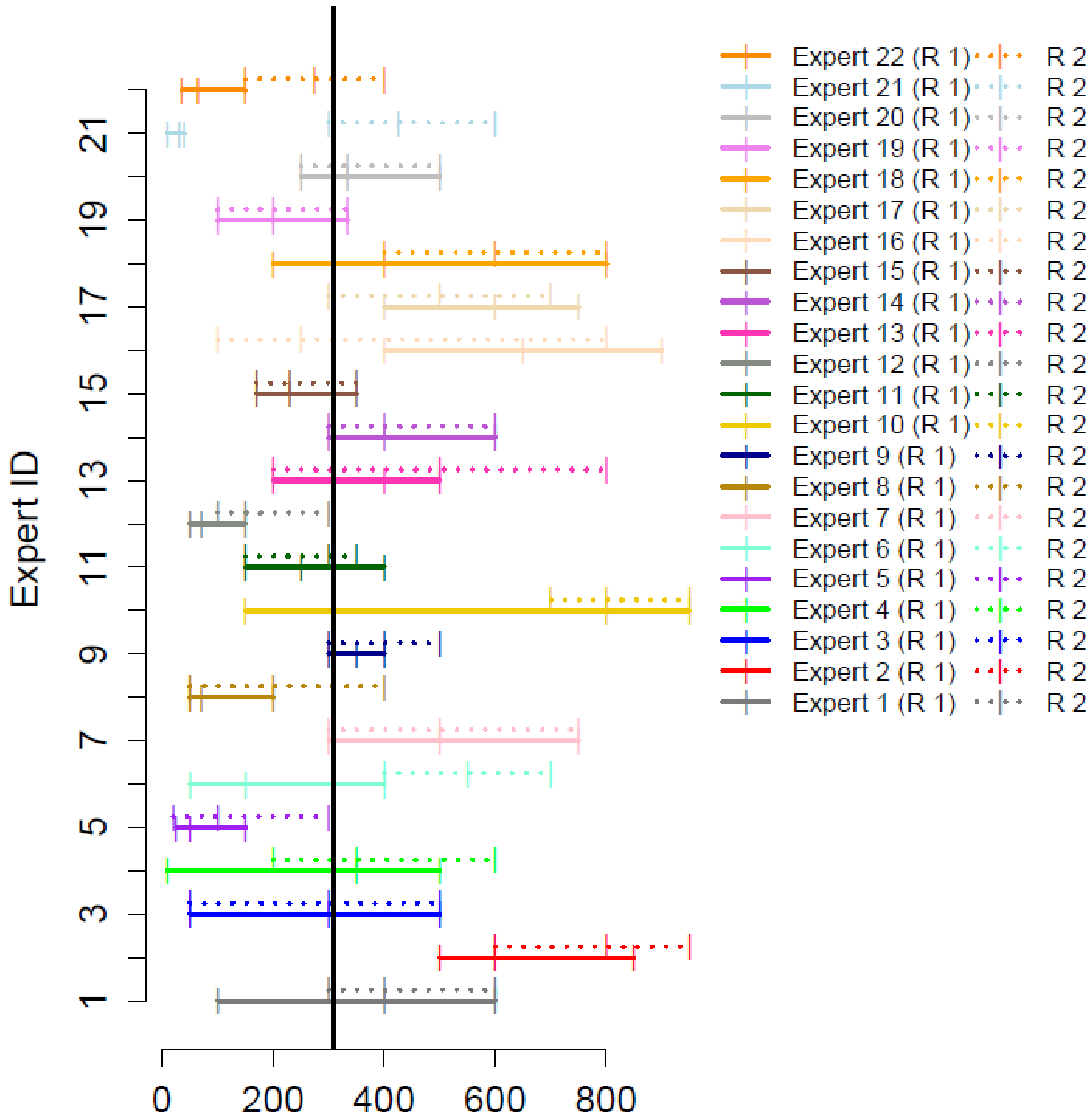
**Plot C18 (Question 42)**



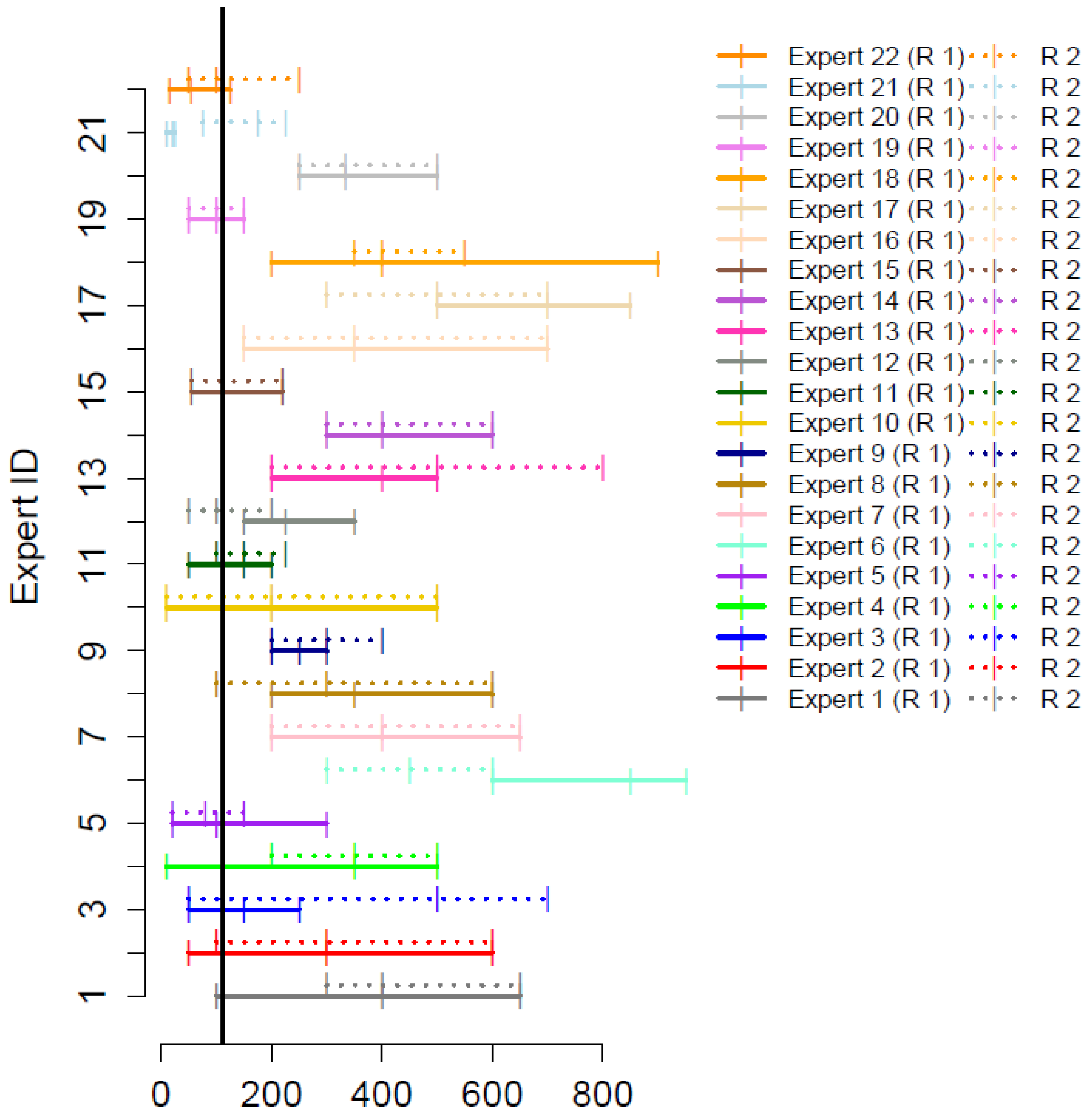
THE NATIONAL ARCHIVES

WARWICK

**Plot C19 (Question 43)**



**Plot C20 (Question 44)**




## Method used to Measure Performance

For the digital preservation elicitation, the classical approach as described by Cooke[12] which is used as standard in the IDEA protocol, was used to pool the judgements of several experts into one common aggregated probability distribution for questions of

[12] R. M. Cooke (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science.* New York: Oxford University Press.

interest. The experts with the best performance are given larger weights in the pooled distribution under this method. The assumption is that ability to perform well on questions of interest is predicted by performance on calibrations questions. This is supported in evidence in the IDEA paper (see Hanea et al under The IDEA protocol). The strategies above to ensure experts' approaches were, as far as possible, the same on the calibration questions and questions of interest is to make every effort to make this assumption valid in this instance.

During the workshop, every expert provided lowest plausible, highest plausible and best estimates for each calibration question which are interpreted as quantiles ie $(0.05, 0.50, 0.95)$. See Appendix D, Figures 2 to 5 for graphs of the experts results compared to the true answers. With the quantiles provided by experts, the proportion of times the true value was observed in each of the following intervals is counted$(-\infty, q_{5\%}), (q_{5\%}, q_{50\%}), (q_{50\%}, q_{95\%}), (q_{95\%}, \infty)$. This is presented in Table 4 as well as Appendix D, Table 1. Note that the expected values for accurate and informative experts are $(5\%, 45\%, 45\%, 5\%)$.

*Table 4: Expert performance in 20 calibration questions for each quantile measured. Expected quantiles for columns 2 to 5 are 0.05, 0.45, 0.45 and 0.05, respectively.*

| Expert ID | Below 5th quantile | 5th to 50th quantile | 50th to 95th quantile | Above 95th quantile |
|---|---|---|---|---|
| 1 | 0.15 | 0.4 | 0.2 | 0.25 |
| 2 | 0.25 | 0.35 | 0.4 | 0 |
| 3 | 0.2 | 0.55 | 0.25 | 0 |
| 4 | 0.2 | 0.1 | 0.35 | 0.25 |
| 5 | 0 | 0.25 | 0.4 | 0.35 |
| 6 | 0.15 | 0.3 | 0.35 | 0.2 |
| 7 | 0.25 | 0.2 | 0.15 | 0.4 |
| 8 | 0.2 | 0.45 | 0.3 | 0.05 |
| 9 | 0.3 | 0.25 | 0.1 | 0.35 |
| 10 | 0.3 | 0.4 | 0.1 | 0.2 |
| 11 | 0.05 | 0.5 | 0.15 | 0.3 |
| 12 | 0.2 | 0.45 | 0.3 | 0.05 |

| | | | | |
|---|---|---|---|---|
| 13 | 0.2 | 0.2 | 0.1 | 0.5 |
| 14 | 0.25 | 0.15 | 0.4 | 0.2 |
| 15 | 0.2 | 0.15 | 0.4 | 0.25 |
| 16 | 0.2 | 0.35 | 0.4 | 0.05 |
| 17 | 0.35 | 0.35 | 0.15 | 0.15 |
| 18 | 0.3 | 0.3 | 0.15 | 0.25 |
| 19 | 0.15 | 0.15 | 0.3 | 0.4 |
| 20 | 0.4 | 0.1 | 0.15 | 0.35 |
| 21 | 0.7 | 0.15 | 0.05 | 0.1 |
| 22 | 0.1 | 0.25 | 0.45 | 0.2 |

To measure how the distribution provided by experts and the theoretical model differ we consider the Kullback-Leibler (KL) divergence, that is,

$$KL(\boldsymbol{o}, \boldsymbol{p}) = \sum_{i=1}^{n} o_i \, log\left(\frac{o_i}{p_i}\right) \qquad (1)$$

with $o_i$ the observed proportion in interval i, $p_i$ the theoretical probability of interval i and $n$ the number of intervals, in our case $n = 4$.

As an example, here is the KL divergence for expert 1,

$$KL(\boldsymbol{o}, \boldsymbol{p}) = 0.15 \, log\left(\frac{0.15}{0.05}\right) + 0.4 \, log\left(\frac{0.4}{0.45}\right) + 0.2 \, log\left(\frac{0.2}{0.45}\right) + 0.25 log\left(\frac{0.25}{0.05}\right) = 0.3578$$

For $q$ calibration questions we have that

$$P(2qKL(\boldsymbol{o}, \boldsymbol{p}) \leq x) \rightarrow \; Q_{n-1}(x) \qquad (2)$$

with $Q_{n-1}(x)$ the cumulative distribution of a $\chi^2_{n-1}$when $q \to \infty$ .[13][14]

We defined the **transformed KL** measure as $2qKL(\boldsymbol{o}, \boldsymbol{p})$**.** This quantity for expert 1 in our example is therefore $2 * 20 * 0.3578 = 14.31$ as displayed in Table 5.

The **calibration score** is defined as: $P(2qKL(\boldsymbol{o}, \boldsymbol{p}) \geq k_{obs})$. If the expert is well calibrated, then $k_{obs} = 2qKL(\boldsymbol{o}, \boldsymbol{p})$ to be close to 0, and $P(2qKL(\boldsymbol{o}, \boldsymbol{p}) \geq k_{obs})$ would be large. While a poorly calibrated expert would have $k_{obs}$ large and the probability $P(2qKL(\boldsymbol{o}, \boldsymbol{p}) \geq k_{obs})$ would be small. In our example, the calibration score for expert 1 is $P(2qKL(\boldsymbol{o}, \boldsymbol{p}) \geq 14.31) \approx Q_3(14.31) = 0.0025$.

To overcome the issue of large calibration score derived simply by large intervals the Cooke method also considers the **information score**. The spread of the experts' intervals is assessed relative to a reference interval $(l_j, u_j)$ for each question j, which is computed based on the observed range for all experts for that question with the addition of an overshooting of 10% of the interval size. Each expert score is measured by the KL divergence relative to a uniform distribution in the reference interval.

For each question j, consider the intervals $\mathbf{I}_j = (l_j, q_{5\%,j}, q_{50\%,j}, q_{95\%,j}, u_j)$. For $U_j \sim Unif(l_j, u_j)$ we have that

$$P(I_{j,i} < U_j < I_{j,i+1}) = \frac{I_{j,i+1} - I_{j,i}}{u_j - l_j} = \tilde{o}_{ji}$$

The information score is computed for each question and expert and is given by

$$I(\boldsymbol{r}, \tilde{\boldsymbol{o}}_j) = \sum_{i=1}^{n} r_i \log\left(\frac{r_i}{\tilde{o}_{ji}}\right) \qquad (3)$$

with $r_{j=}(0.05, 0.45, 0.45, 0.05)$ the theoretical probability in the reference range intervals for question j.

[13] J. Quigley, A. Colson, W. Aspinall, and R. M. Cooke (2018). 'Elicitation in the Classical Model', in L. C. Dias, A. Morton and J. Quigley (eds) *Elicitation The Science and Art of Structuring Judgement*. New York: Springer.

[14] We can approximate with a Chi-squared distribution with 3 degrees of freedom since the Taylor expansion of KL and Chi-squared are similar to the 2nd order, see F. Bavaud, 'Information Theory, Relative entropy and Statistics', p6, https://arxiv.org/pdf/0808.4111.pdf

The final information weight for each expert is

$$I = \frac{1}{q}\sum_{j=1}^{q} I(\boldsymbol{r}, \tilde{\boldsymbol{o}}_j)$$

For each expert, the raw weight is computed as the product of the calibration and the information scores. Finally, the raw weights are normalized across all experts resulting in the final weight for each expert.

The combined distribution is obtained as

$$P(X \leq x) = \sum_{i=1}^{n_e} w_i \, P_i(X \leq x) \qquad (4)$$

where $P_i(X \leq x)$ is the cumulative distribution provided by expert i and $n_e$ is the number of experts.

*Table 5: Experts transformed KL distance, calibration score, information score and final combined weights. Scores close to 0 indicate worse performances and close to 1 indicate better performances.*

| Expert ID | Transformed KL | Calibration score | Information score | Combined weight |
|---|---|---|---|---|
| 1 | 14.31 | 0 | 1.02 | 0.0068 |
| 2 | 10.69 | 0.01 | 0.69 | 0.0246 |
| 3 | 9.63 | 0.02 | 0.57 | 0.033 |
| 4 | 24.69 | 0 | 0.96 | 0 |
| 5 | 19.48 | 0 | 0.64 | 4e-04 |
| 6 | 9.3 | 0.03 | 0.57 | 0.0383 |
| 7 | 36.29 | 0 | 0.71 | 0 |
| 8 | 6.22 | 0.1 | 0.86 | 0.2289 |
| 9 | 36.85 | 0 | 1.54 | 0 |
| 10 | 24.69 | 0 | 0.93 | 0 |
| 11 | 17.02 | 0 | 1.36 | 0.0025 |

| | | | | |
|---|---|---|---|---|
| 12 | 6.22 | 0.1 | 0.78 | 0.2094 |
| 13 | 44.64 | 0 | 1.33 | 0 |
| 14 | 18.71 | 0 | 1.5 | 0.0012 |
| 15 | 18.71 | 0 | 1.49 | 0.0012 |
| 16 | 5.69 | 0.13 | 0.88 | 0.2953 |
| 17 | 23.72 | 0 | 0.86 | 1e-04 |
| 18 | 26.14 | 0 | 1.24 | 0 |
| 19 | 28.41 | 0 | 1.2 | 0 |
| 20 | 47.91 | 0 | 1.59 | 0 |
| 21 | 65.68 | 0 | 1.15 | 0 |
| 22 | 7.99 | 0.05 | 1.29 | 0.1582 |

Note that it is not unusual to have a few experts carry the majority of the weight in an expert elicitation exercise.

In the testing of the IDEA protocol (ref 4), it was shown that the majority of the experts move their second round estimates for calibration questions toward the direction of the truth, following the discussion. This facilitated discussion creates a weak dependence between the experts' scores but also means that all experts contribute to the improvement in scores of the best performing experts. Therefore, whilst an individual expert may have a low or even zero weight in the final aggregation, their contributions are still reflected in the final estimates of the questions of interest.

**Expert performance evaluation**

Table 4 shows the overall performance of 22 experts in terms of the proportion of the realizations observed in the intervals$(-\infty, q_{5\%}), (q_{5\%}, q_{50\%}), (q_{50\%}, q_{95\%}), (q_{95\%}, \infty)$. We observe that a number of experts' responses[15] show narrow intervals which fail to capture the true value on the calibration questions, which we denote as overconfidence. Expert 7's responses show overconfidence, since their intervals are narrow and very often fail to capture the true value (this is clearly shown in plots C9, C11, C15, C17 and C18). The proportion of Expert 7's values outside the 90% interval is 0.65 while the

[15] For the calculation it is only the round 2 answer which is relevant.

expected is 0.10. The same pattern is observed for in Expert 9's responses, with the proportion of values outside the 90% interval being 0.65 while the expected is 0.10 (see plots C7, C9, C12, C15, C16 and C18). For Experts 13, 20 and 21, the proportion of values outside the 90% interval were 0.70, 0.75 and 0.80, respectively (see plots C1, C3, C4, C5, C10, C11 and C16). On the other hand, Experts 2, 8, 12 and 16 had 0.75% of true values inside their 90% credible intervals.

Appendix D Plot 1 displays the KL transformed divergence and score for each expert and indicate that experts 8, 12 and 16 had the best performances based on the 20 calibration questions. Moreover, experts 13, 20, and 21 had the worst performances according to these measures.

Appendix D Figures 2-5 display the assessment of the experts' performance based on the 20 calibration questions for both rounds 1 and 2.

## Questions of interest

These are the questions we needed to elicit from the experts and the performance weighted expert judgement results.

*Table 6: The questions of interest and the performance-weighted expert judgement results. All values stated to 7 significant figures.*

| # | Question | Performance-weighted expert judgement (median) | Lowest plausible (5% quantile) | Highest plausible (95% quantile) |
|---|---|---|---|---|
| 4 | Out of 1,000 files with file formats that are ubiquitous and/or open, how many would you expect to have the tools to render? | **811.0670** | 612.9049 | 909.8701 |
| 5 | Out of 1,000 files with file formats that are neither ubiquitous nor open, at an archive where staff have good technical skills, how many would you expect to have the tools to render? | **434.3472** | 223.5331 | 589.3053 |
| 6 | Out of 1,000 born-digital files, how many would you expect to have the content metadata that meets an archive's requirements? | **453.9015** | 275.7410 | 732.9014 |
| 7 | Out of 1,000 digitised files, how many would you expect to have the content metadata that meets an archive's requirements? | **749.1711** | 528.4230 | 868.0592 |

| | | | | |
|---|---|---|---|---|
| 8 | Out of 1,000 surrogate files, how many would you expect to have the content metadata that meets an archive's requirements? | **726.1470** | 529.5355 | 877.8656 |
| 11 | Out of 1,000 files with insufficient content metadata, at an archive where there is sufficient information management, how many files would you expect to be able to identify (i.e. knowing what they are and where they are from)? | **534.8452** | 286.3836 | 737.3031 |
| 12 | Out of 1,000 files at an archive where staff have good technical skills, how many files would you expect to have sufficient technical metadata? | **772.938** | 606.7368 | 915.9922 |
| 13 | Out of 1,000 files at an archive where staff have poor technical skills, how many files would you expect to have sufficient technical metadata? | **426.9966** | 160.9853 | 652.1798 |
| 18 | Out of 1,000 media of type A (less stable), how many would you expect to reach the end of their life within 12 months? | **241.1851** | 81.22737 | 507.3934 |
| 19 | Out of 1,000 media of type B (more stable), how many would you expect to reach the end of their life within 12 months? | **54.9878** | 20.13336 | 143.8360 |
| 20 | There is a flood at your storage location and the archive has inadequate mitigations. Out of 1,000 media of type A (less stable), how many would you expect to be destroyed? | **678.7532** | 415.8392 | 866.7981 |

| | | | | |
|---|---|---|---|---|
| 21 | There is a flood at your storage location and the archive has inadequate mitigations. Out of 1,000 media of type B (more stable), how many would you expect to be destroyed? | **353.9914** | 185.5885 | 651.2144 |
| 22 | Out of 1,000 files all stored on a storage medium of type A (less stable) at an archive where staff have good technical skills, how many files would you expect the bit-stream to be inaccessible due to obsolescence? | **298.4519** | 106.9200 | 652.0130 |
| 23 | Out of 1,000 files stored on a storage medium of type A (less stable) at an archive where staff have poor technical skills, how many files would you expect the bit-stream to be inaccessible due to obsolescence? | **642.2055** | 407.0915 | 900.3617 |
| 24 | Out of 1,000 files stored on a storage medium of type B (more stable) at an archive where staff have good technical skills, how many files would you expect the bit-stream to be inaccessible due to obsolescence? | **140.5431** | 54.00692 | 284.7728 |
| 25 | Out of 1,000 files stored on a storage medium of type B (more stable) at an archive where staff have poor technical skills, how many files would you expect the bit-stream to be inaccessible due to obsolescence? | **419.0489** | 203.3126 | 613.2554 |
| 31 | Out of 1,000 files, how many would you expect to have become corrupted during transfer from a depositor to an archive? | **90.1241** | 40.40048 | 250.9154 |

| | | | | |
|---|---|---|---|---|
| 32 | Out of 1,000 files, at an archive where you have poor system security, how many would you expect to have become corrupted and identified this, given you have a checksum to compare to? | **329.3772** | 155.5647 | 516.7276 |
| 33 | Out of 1,000 files, at an archive where you have poor system security, how many would you expect to have become corrupted and not been able to identify this, despite having a checksum to compare to? | **138.0666** | 49.16224 | 447.5487 |
| 34 | Out of 1,000 files with their bit-stream stored and accessible but where you cannot guarantee their integrity, how many files would you expect to be bit-preserved? | **715.8353** | 369.2139 | 887.5157 |
| 35 | Out of 1,000 files which are bit-preserved at an archive which has full access to rendering tools but where the files have insufficient technical metadata, how many files would you expect to be able to render? | **599.3193** | 209.5556 | 852.2405 |
| 36 | Out of 1,000 born-digital files, for how many would you expect an archive to know their conditions of use? | **610.9577** | 339.7419 | 861.1080 |
| 37 | Out of 1,000 digitised files, for how many would you expect an archive to know their conditions of use? | **745.0289** | 504.5658 | 919.2316 |

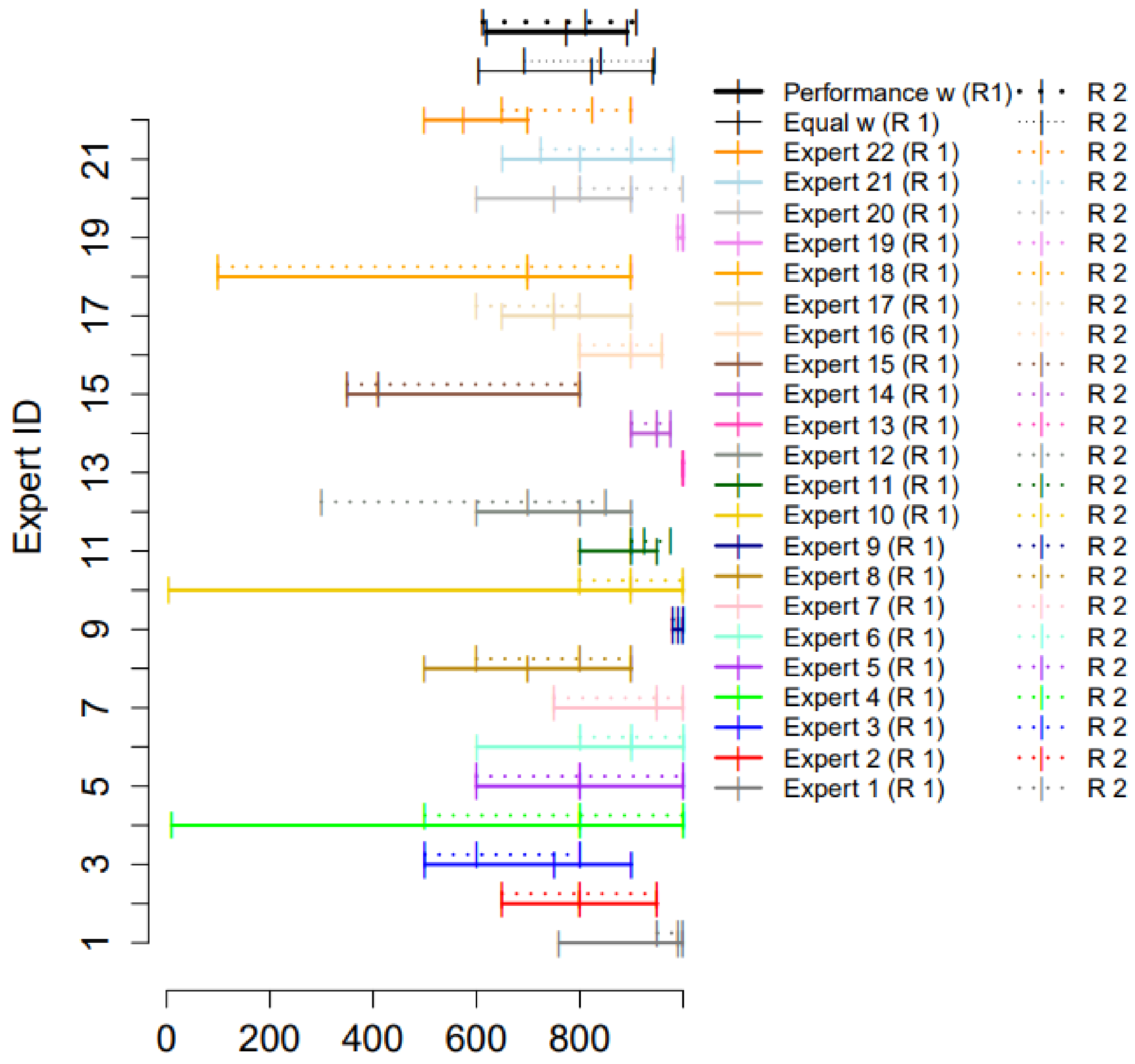
| | | | | |
|---|---|---|---|---|
| 38 | Out of 1,000 surrogate files, for how many would you expect an archive to know their conditions of use? | **789.6147** | 547.0652 | 918.7014 |

The median values from these results have been translated into probabilities (and rounded to 6 decimal places) and used to populate the conditional probability tables in the decision support tool DiAGRAM (https://nationalarchives.shinyapps.io/DiAGRAM/). The tables can be viewed under the 'Advanced customisation' tab.

## Plots

For each plot we have the first (solid line) and second (dotted line) scores for each expert with tick marks indicating 5%-ile, median and 95%-ile values. Above these are indicated the calculated values given by the expert performance weighting method described above and for comparison the calculated values giving all experts equal weight.

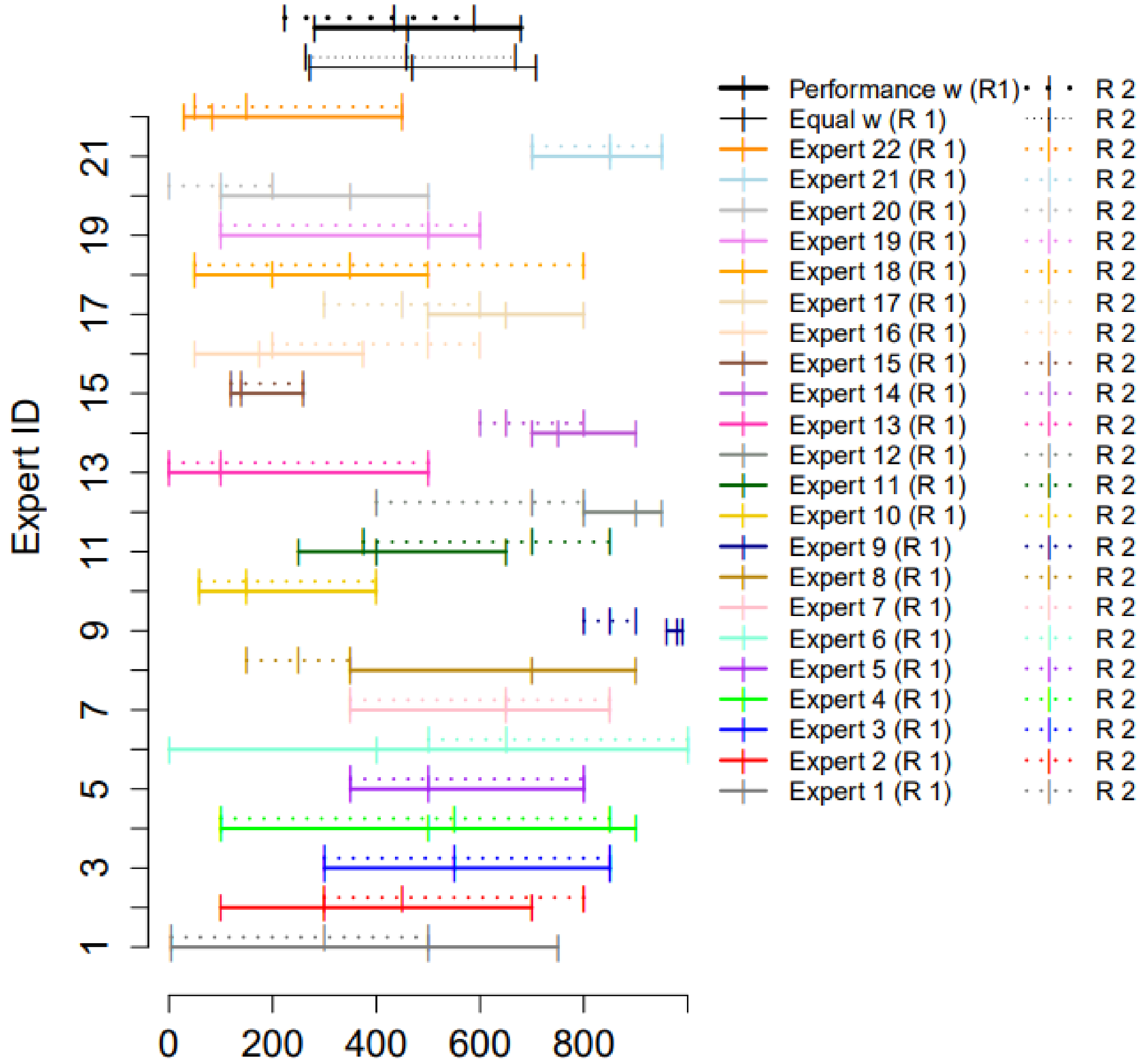
### Plot 1 (Question 4)



THE NATIONAL ARCHIVES

WARWICK

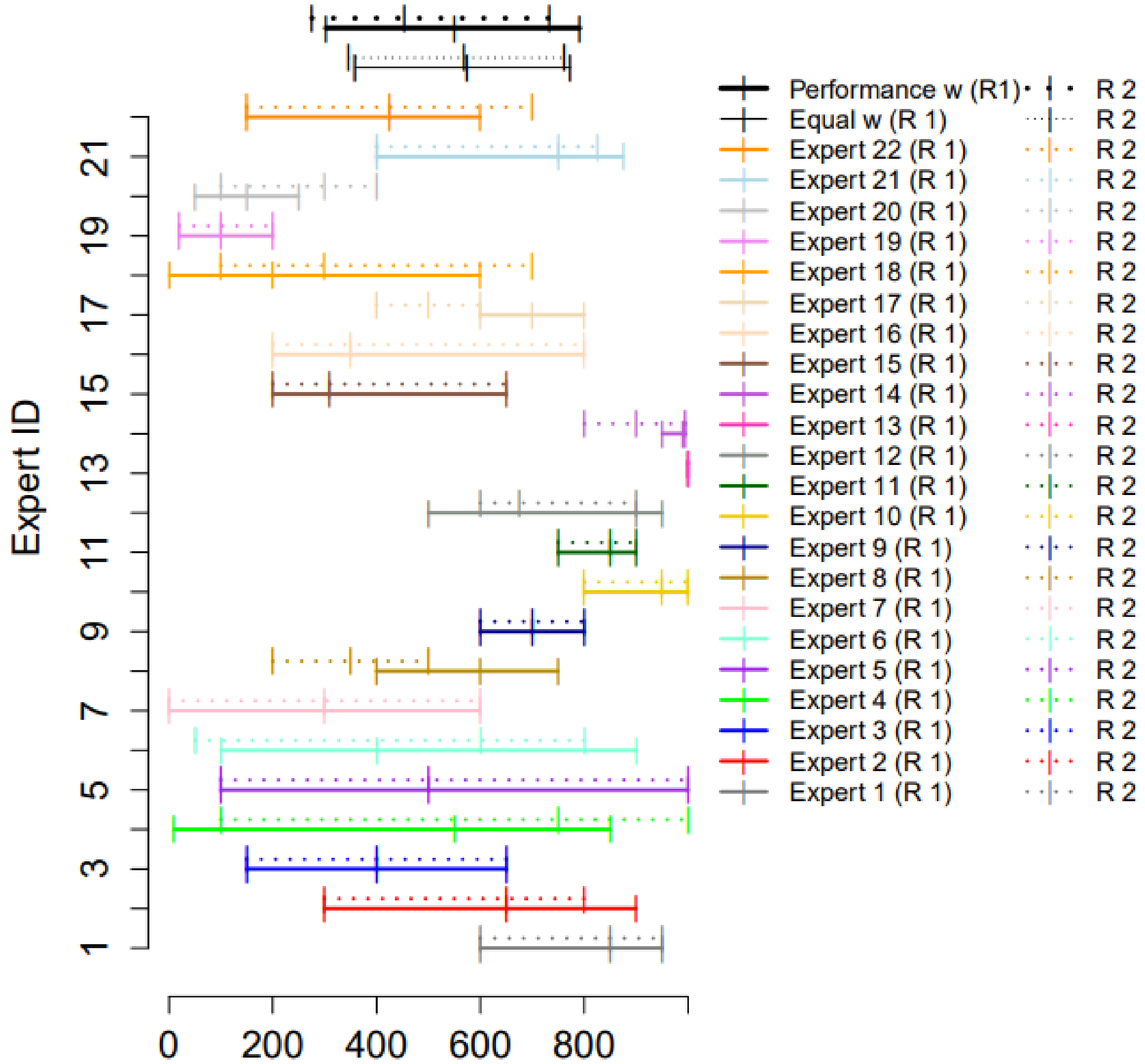
**Plot 2 (Question 5)**



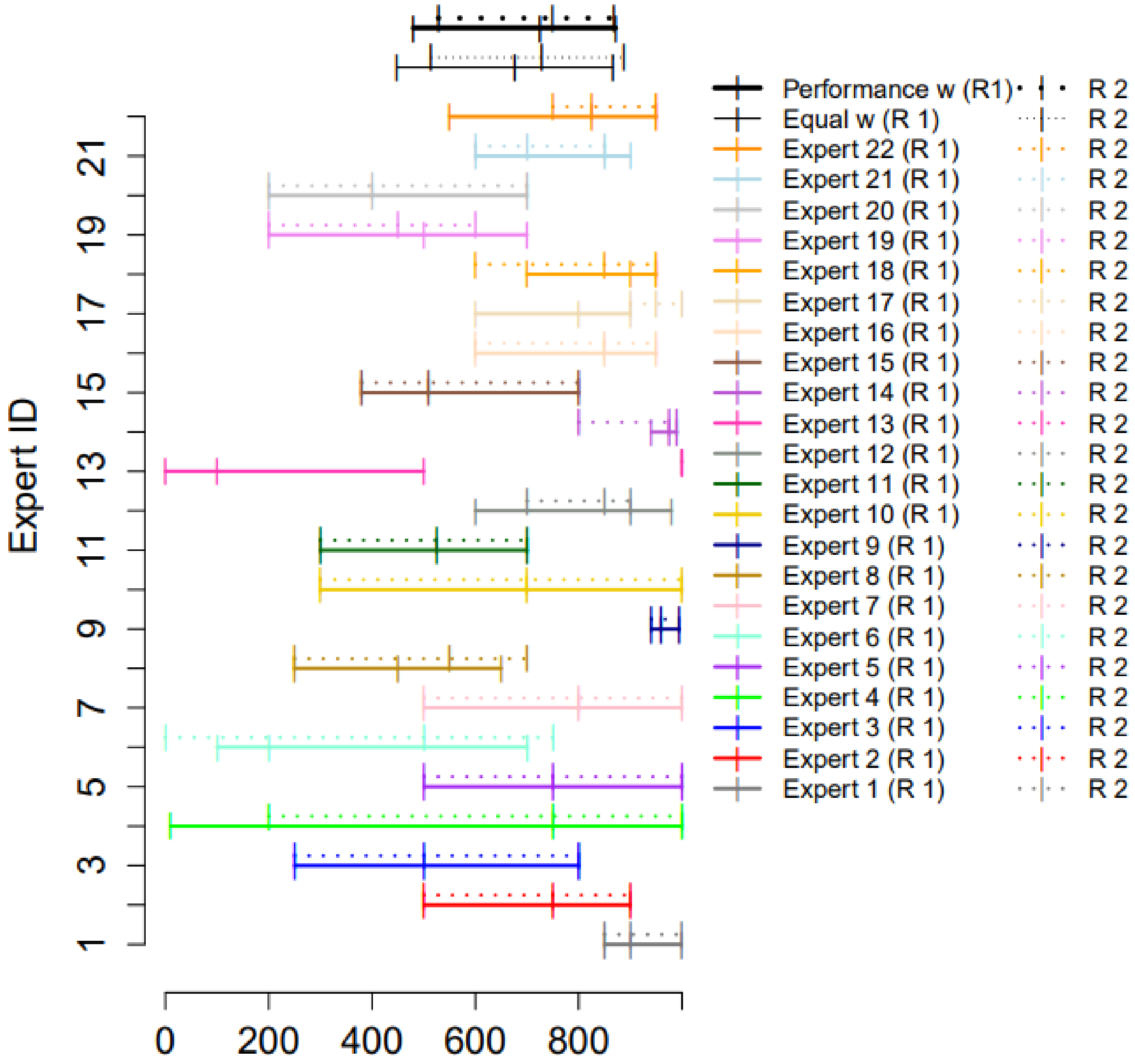
THE NATIONAL ARCHIVES

WARWICK

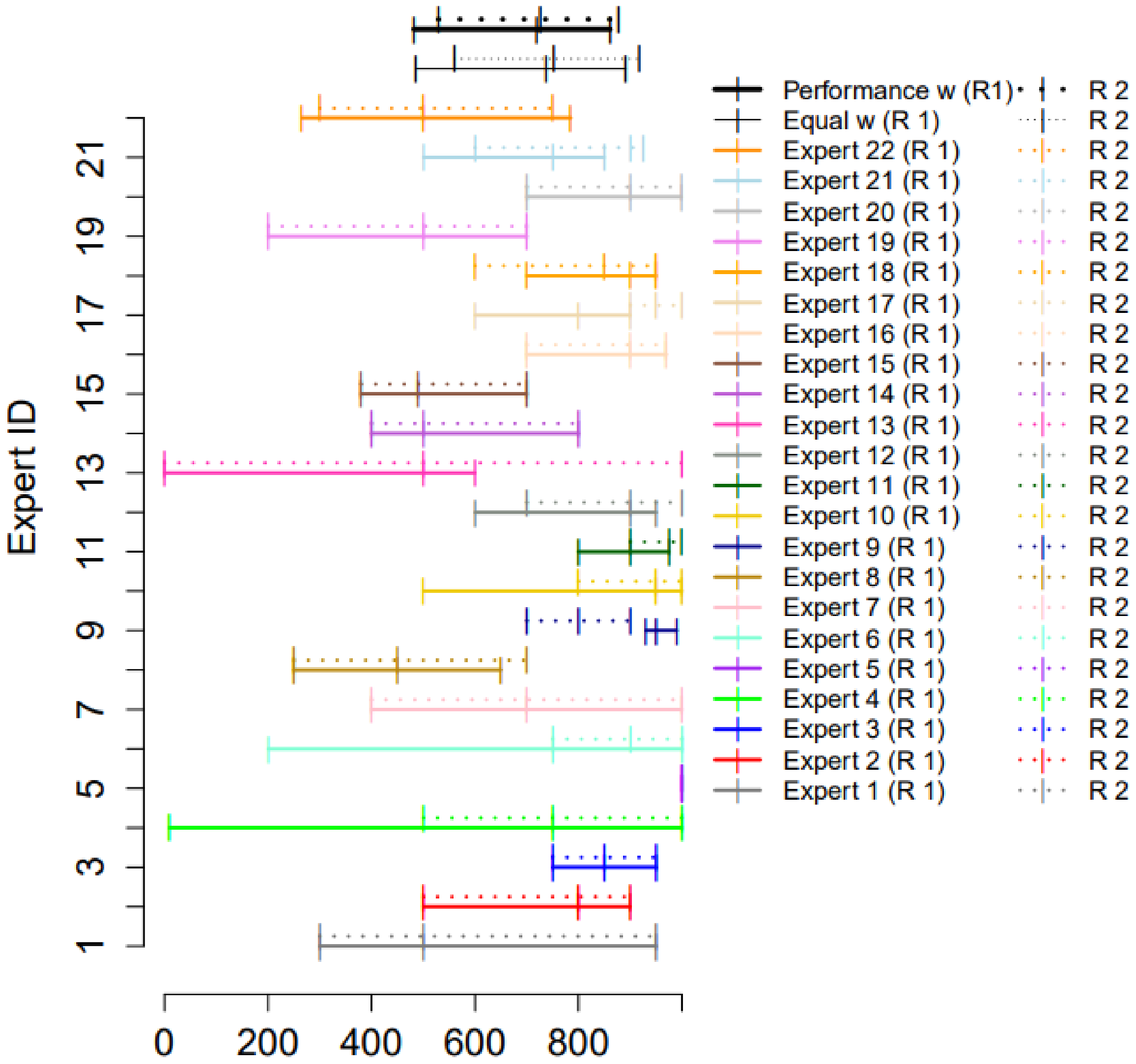
**Plot 3 (Question 6)**



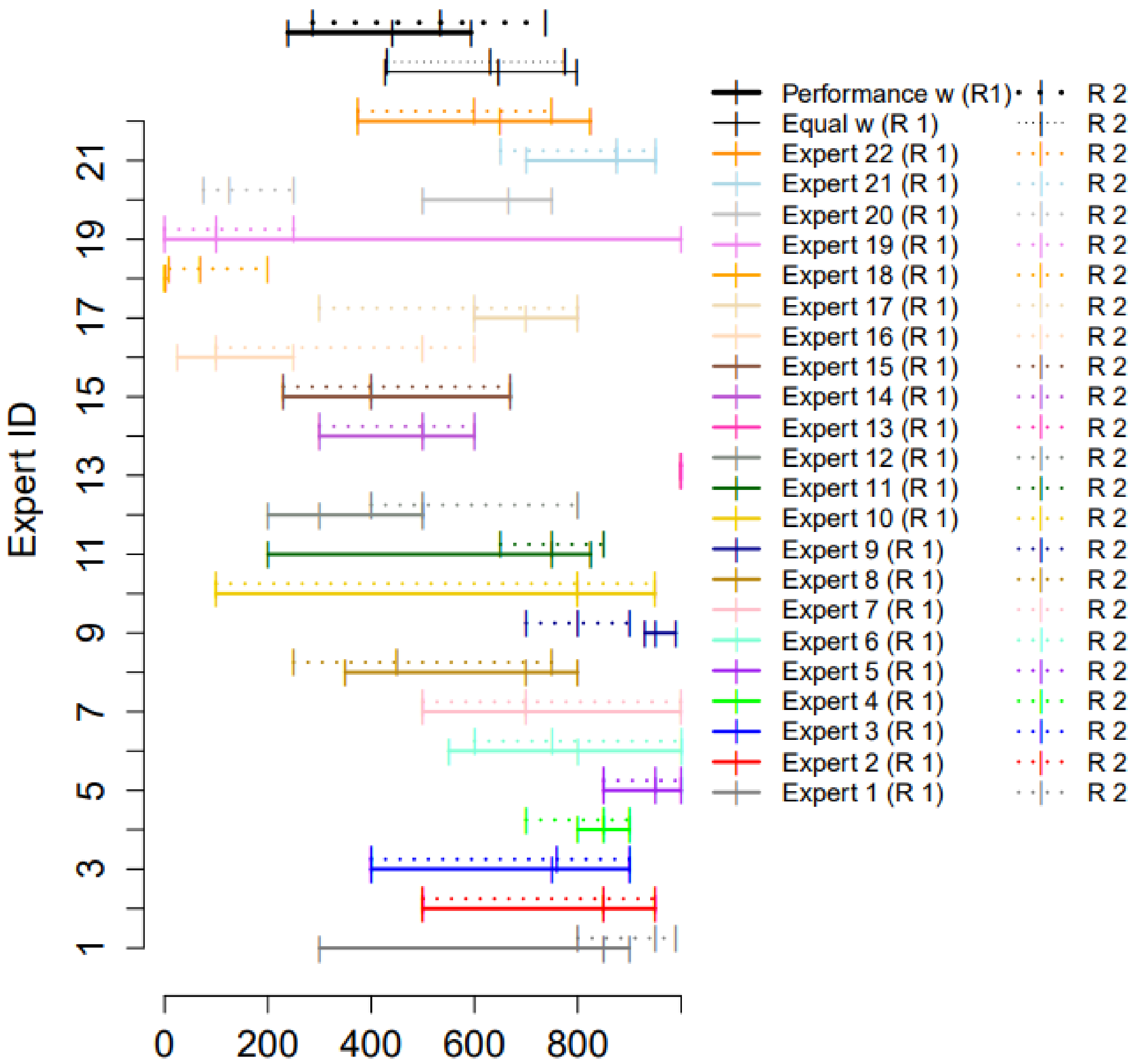
**Plot 4 (Question 7)**



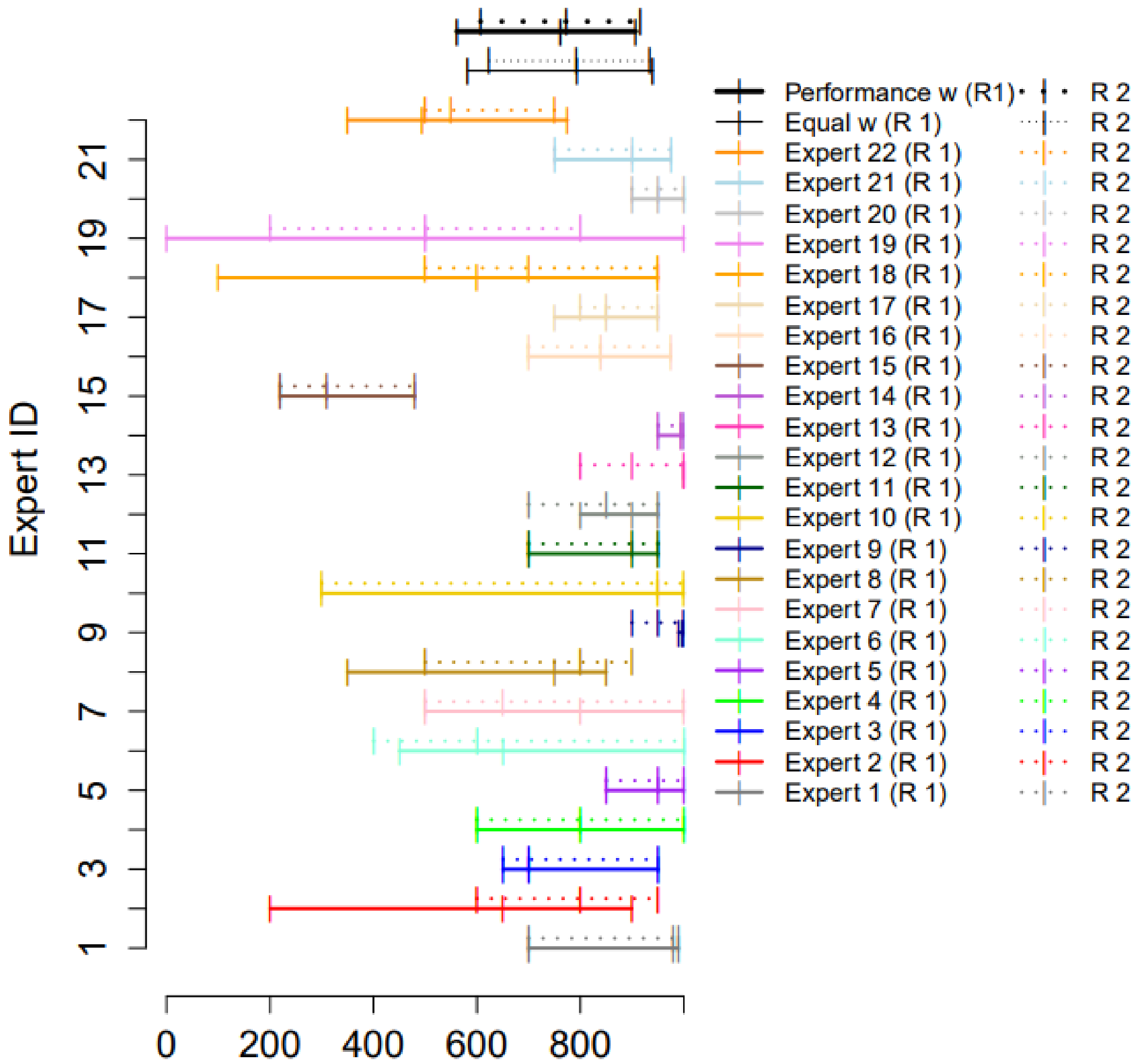
**Plot 5 (Question 8)**



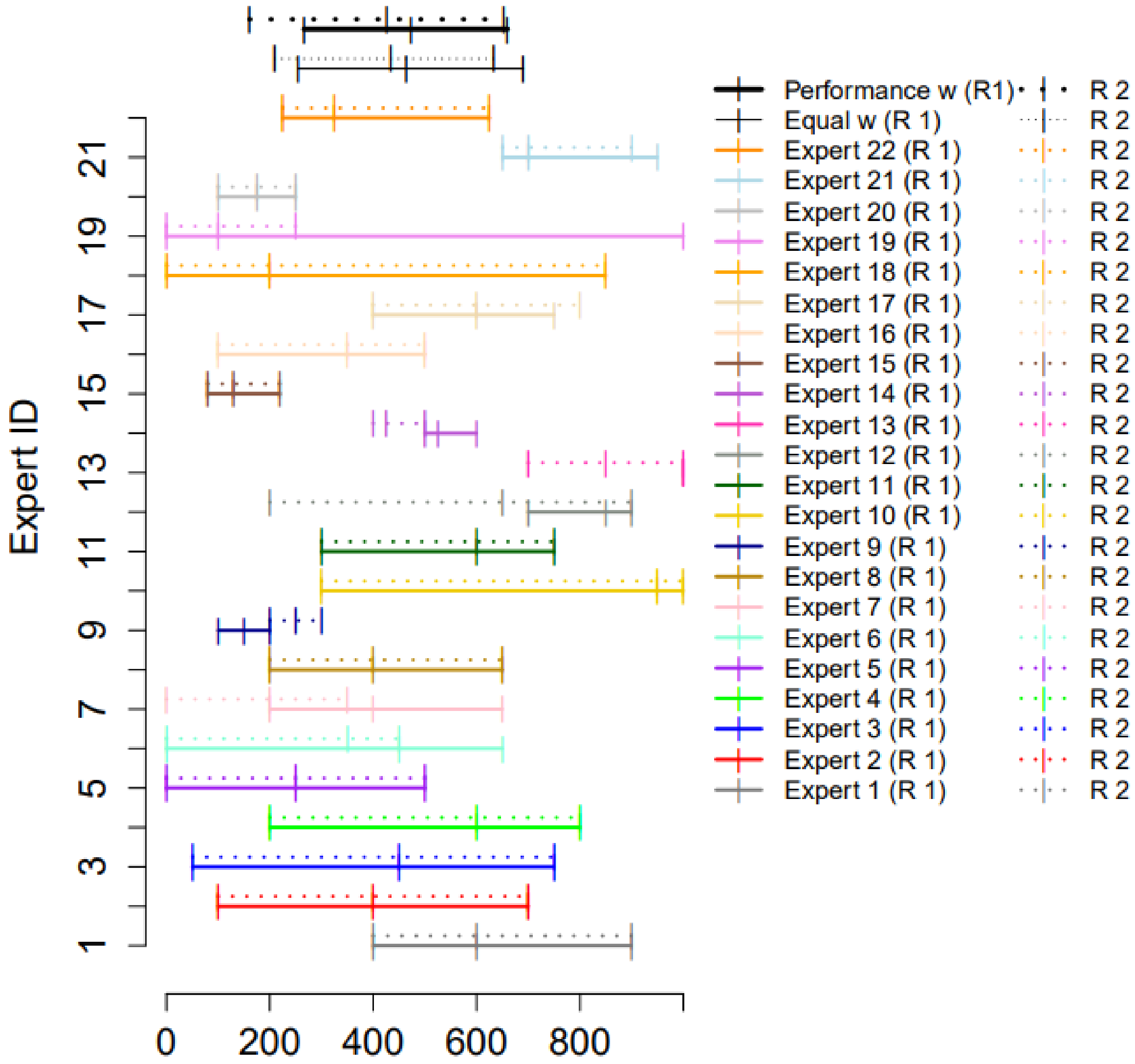
**Plot 6 (Question 11)**



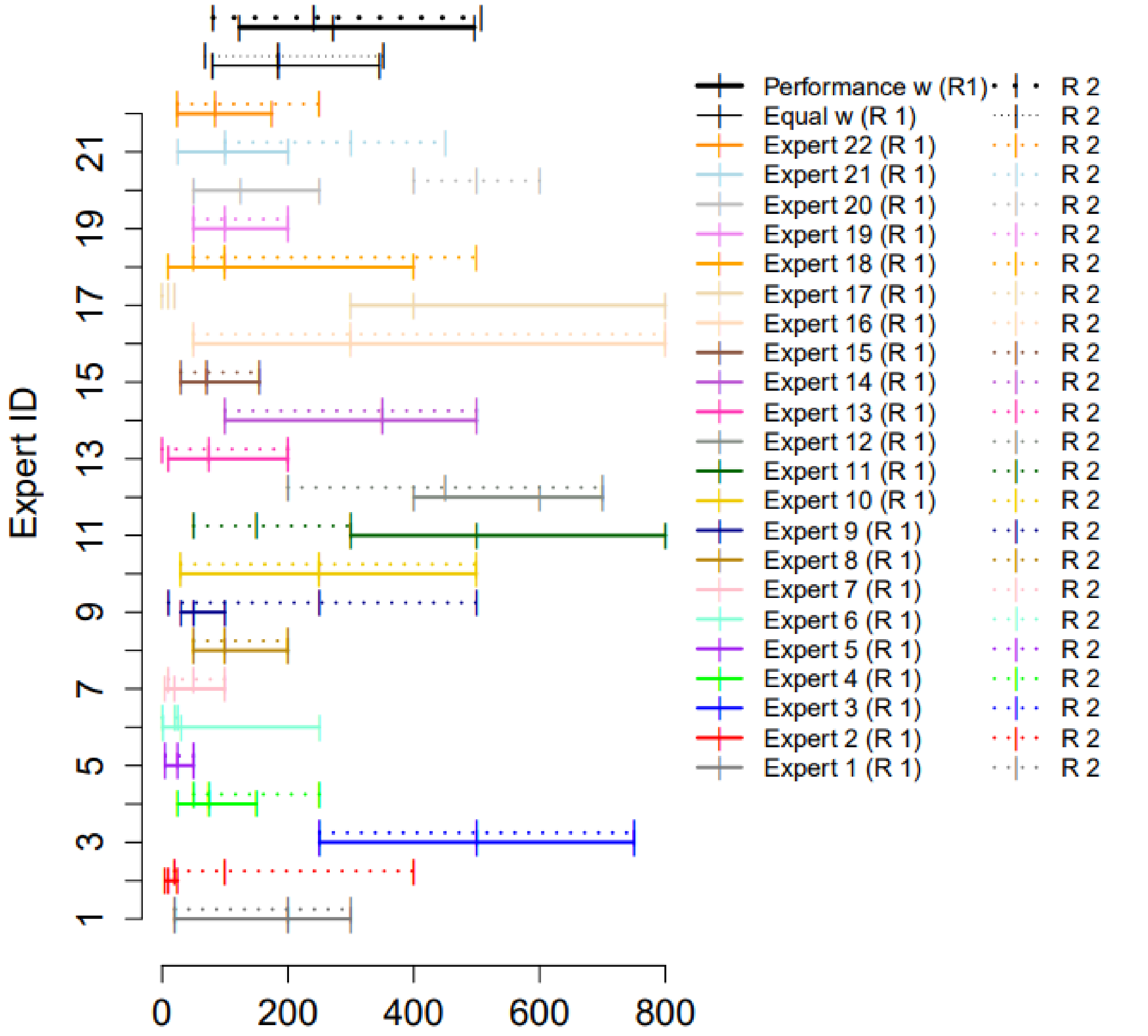
**Plot 7 (Question 12)**



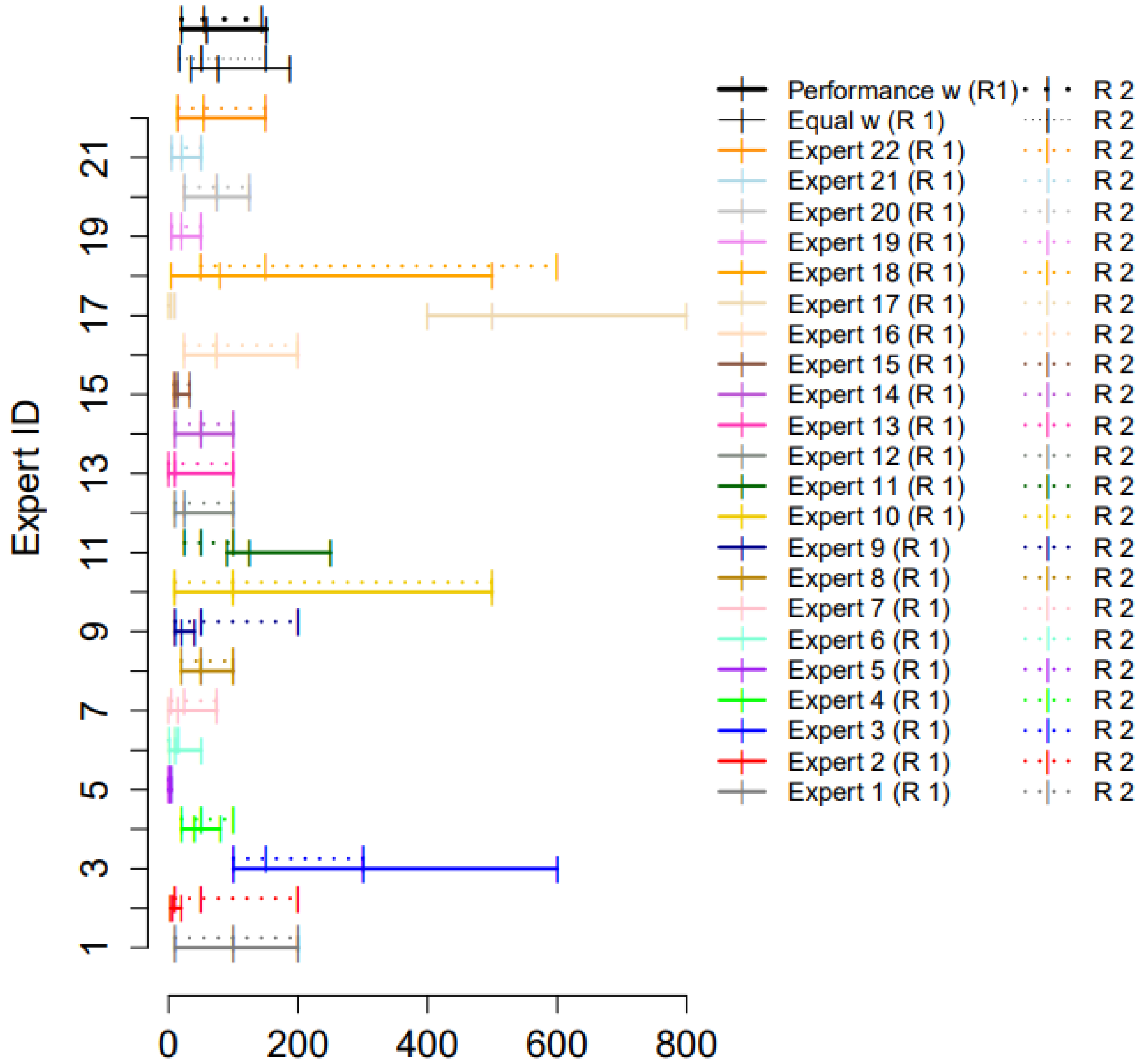
**Plot 8 (Question 13)**



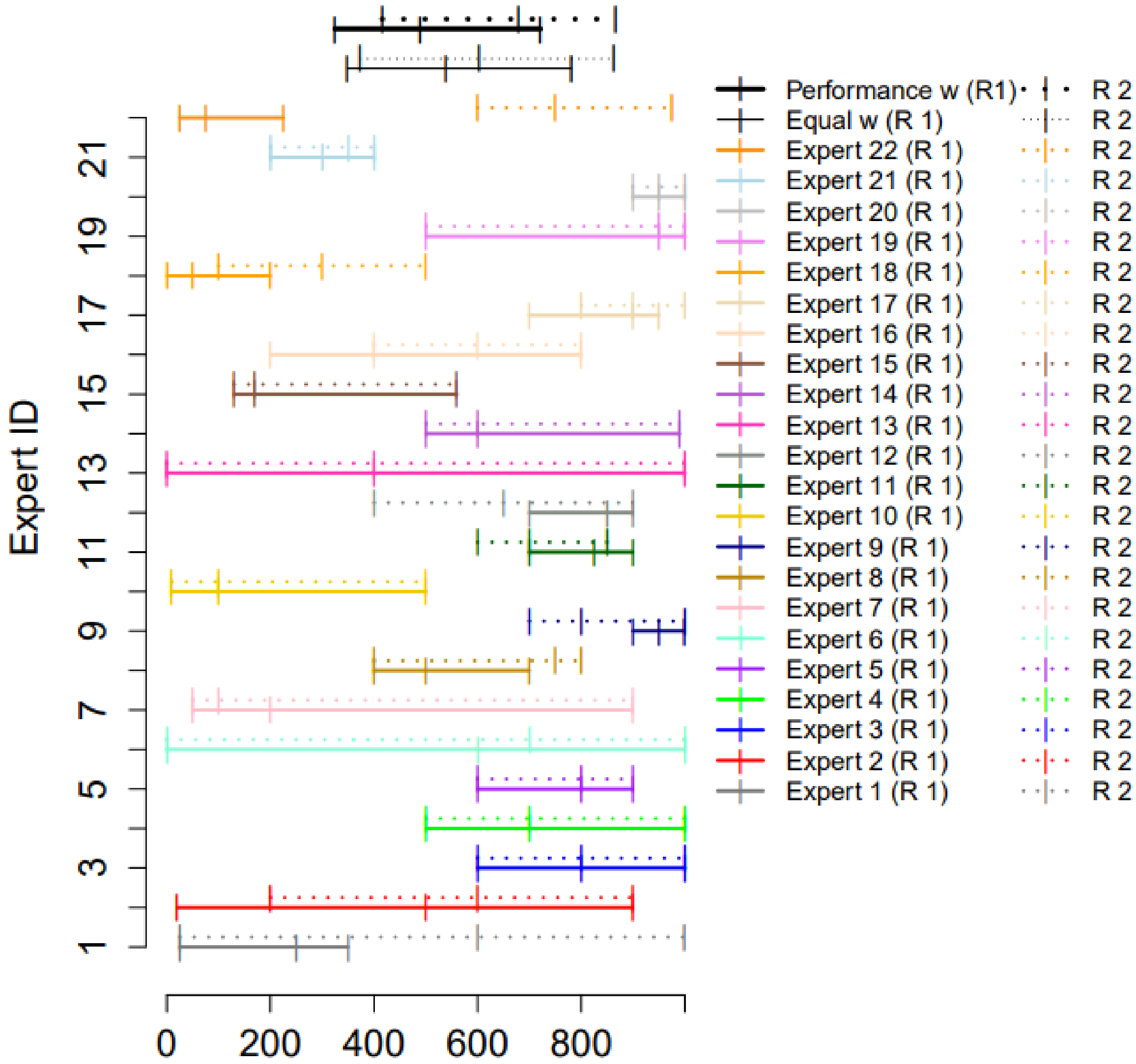
Plot 9 (Question 18)



**Plot 10 (Question 19)**



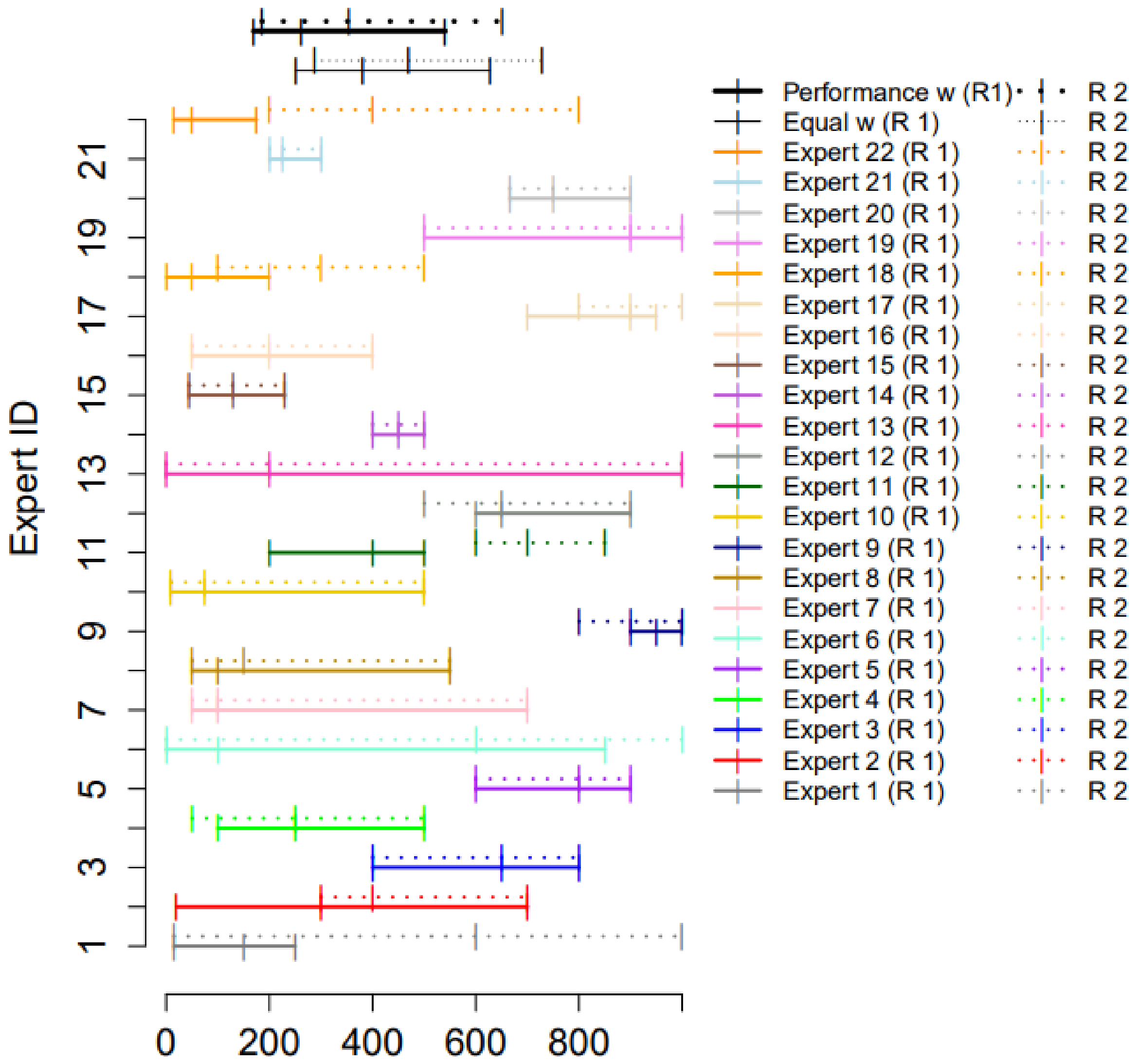
**Plot 11 (Question 20)**



THE NATIONAL ARCHIVES

WARWICK

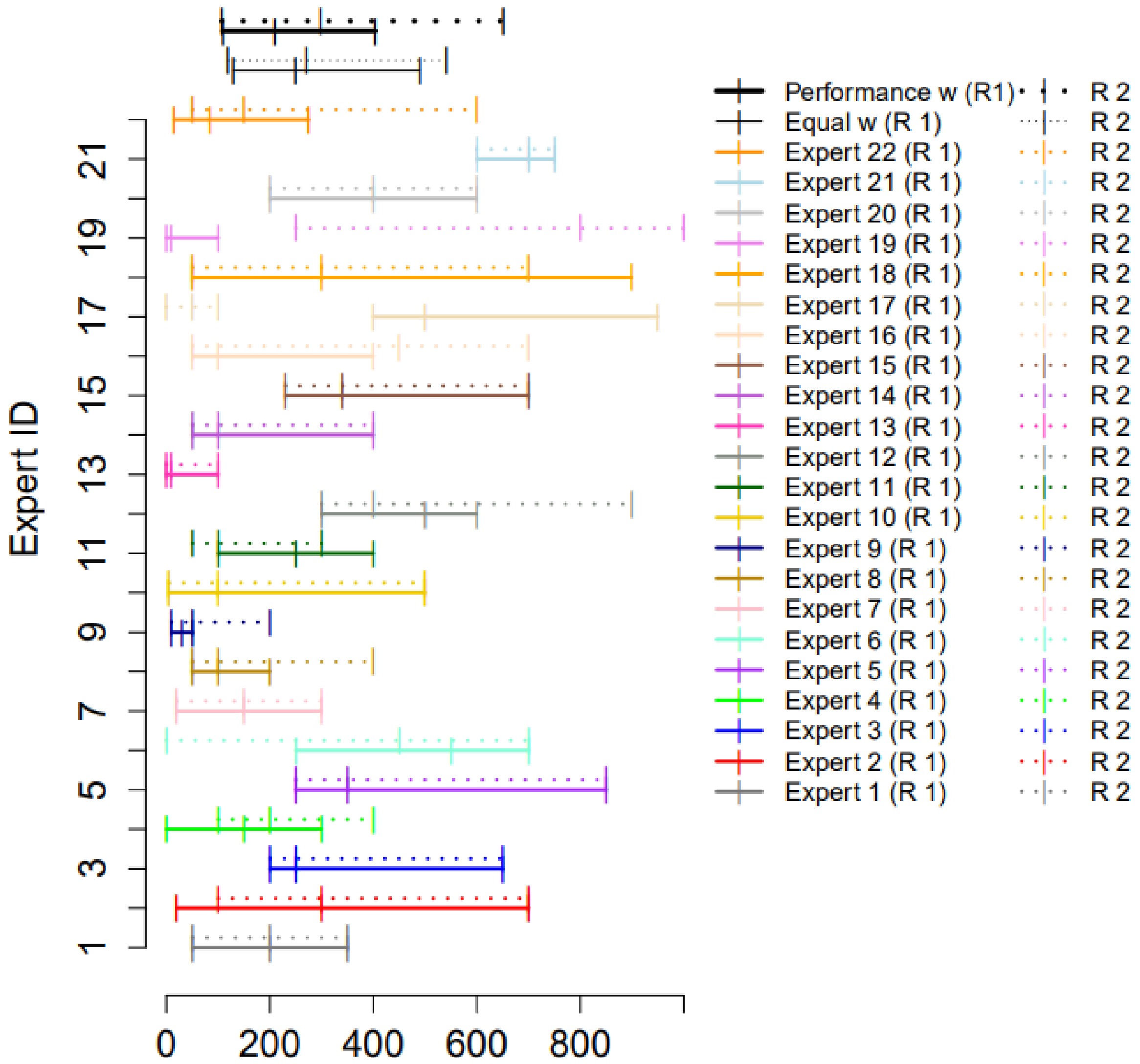
**Plot 12 (Question 21)**



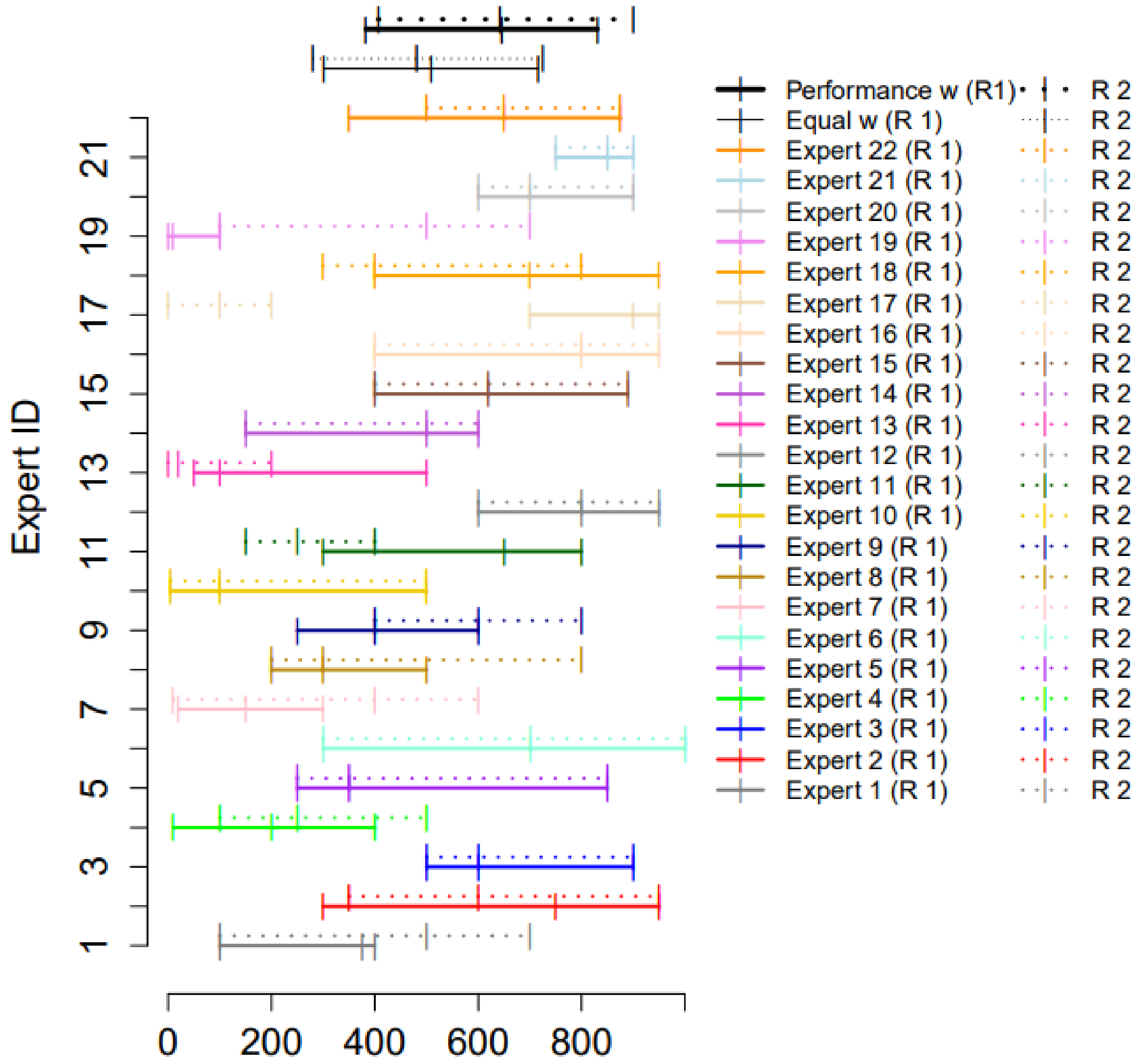
THE NATIONAL ARCHIVES

WARWICK

**Plot 13 (Question 22)**



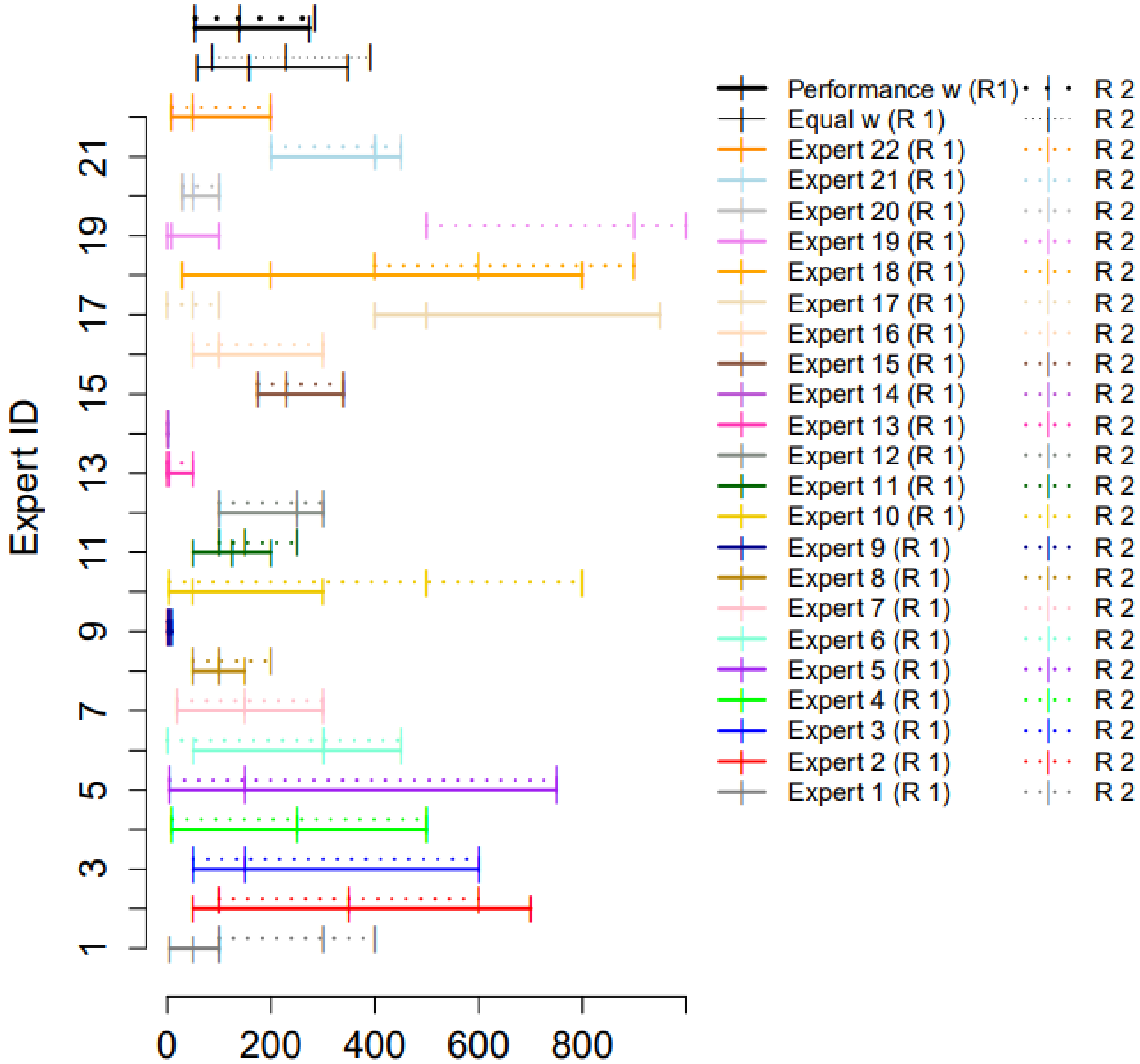
**Plot 14 (Question 23)**



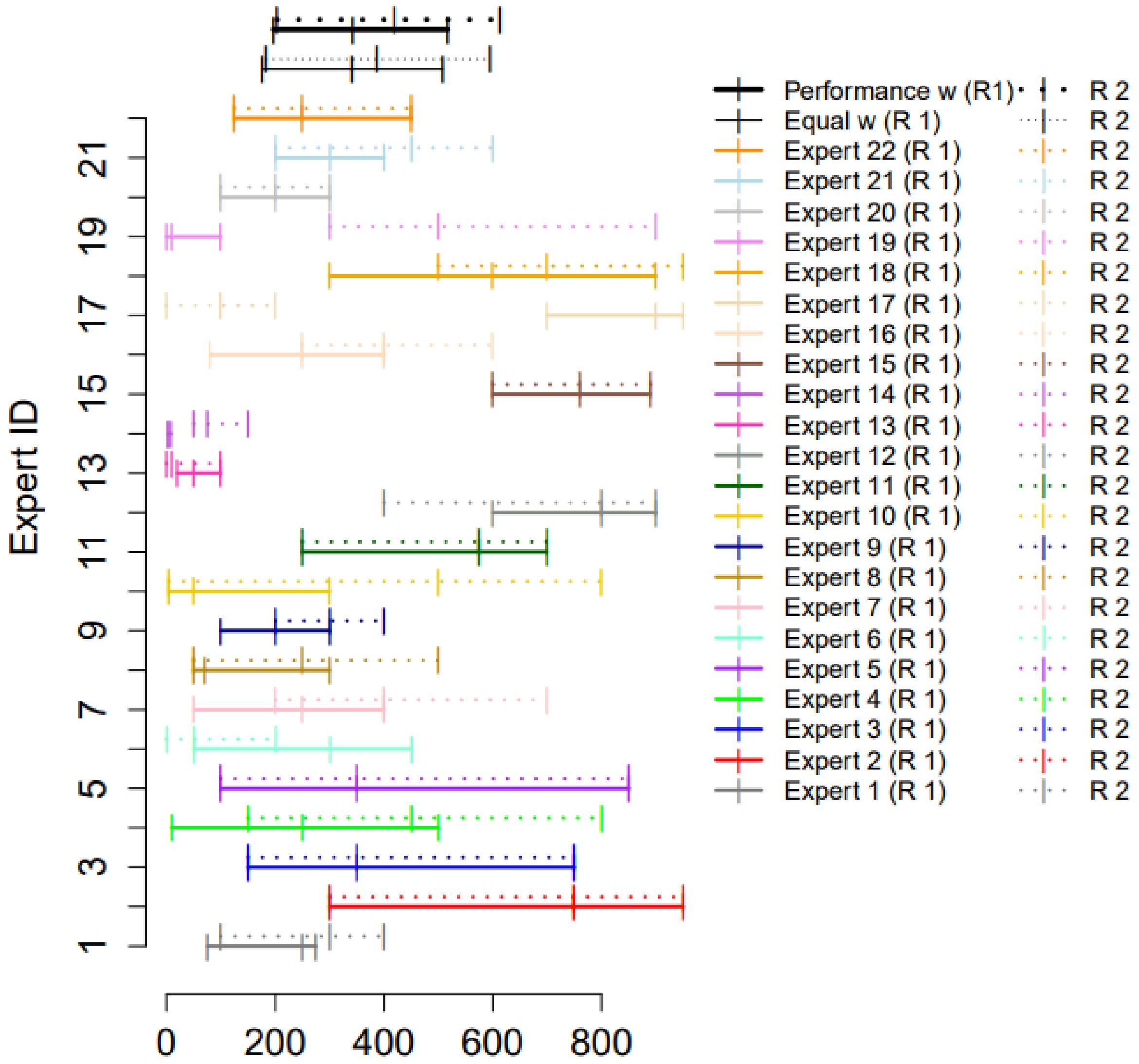
THE NATIONAL ARCHIVES

WARWICK

**Plot 15 (Question 24)**



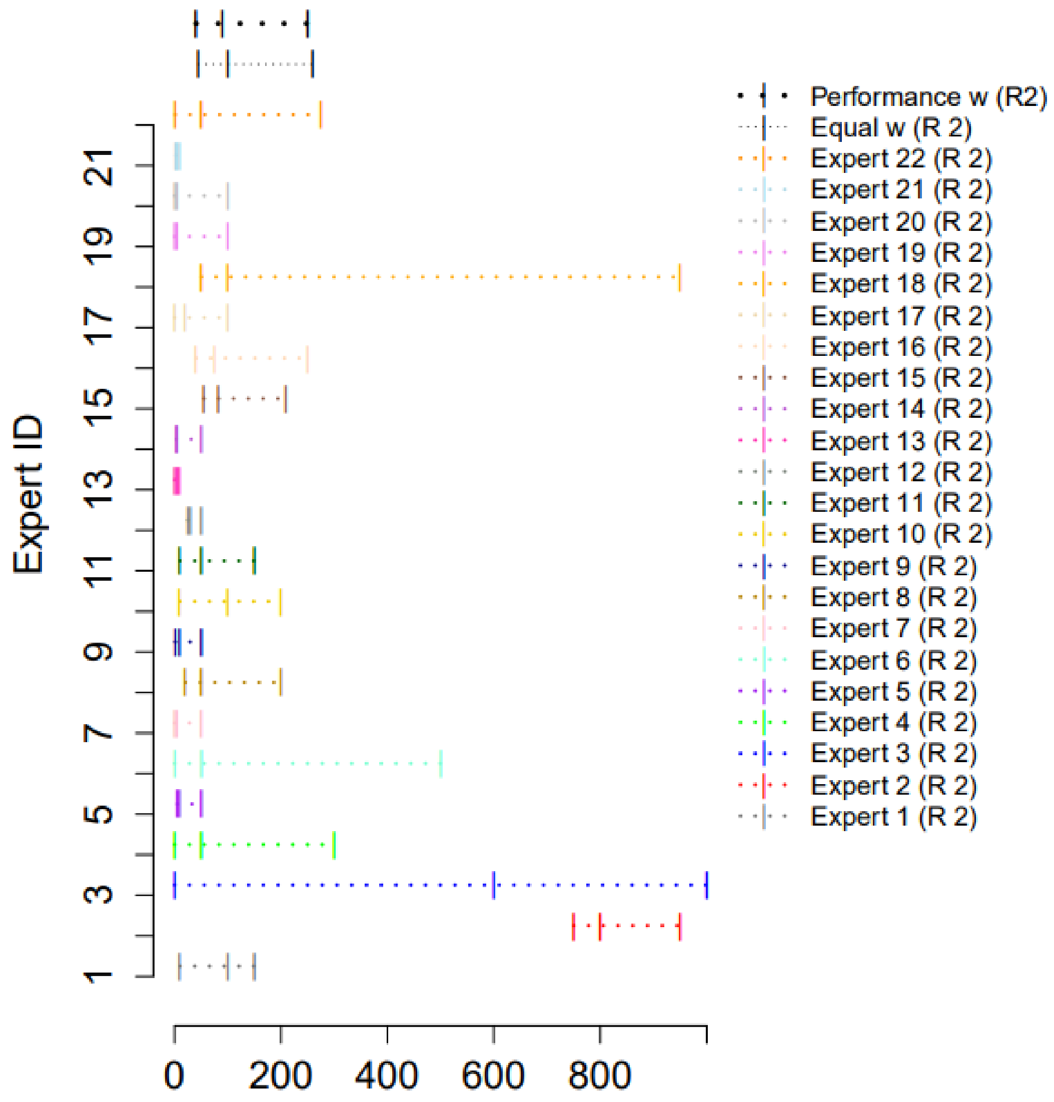
**Plot 16 (Question 25)**



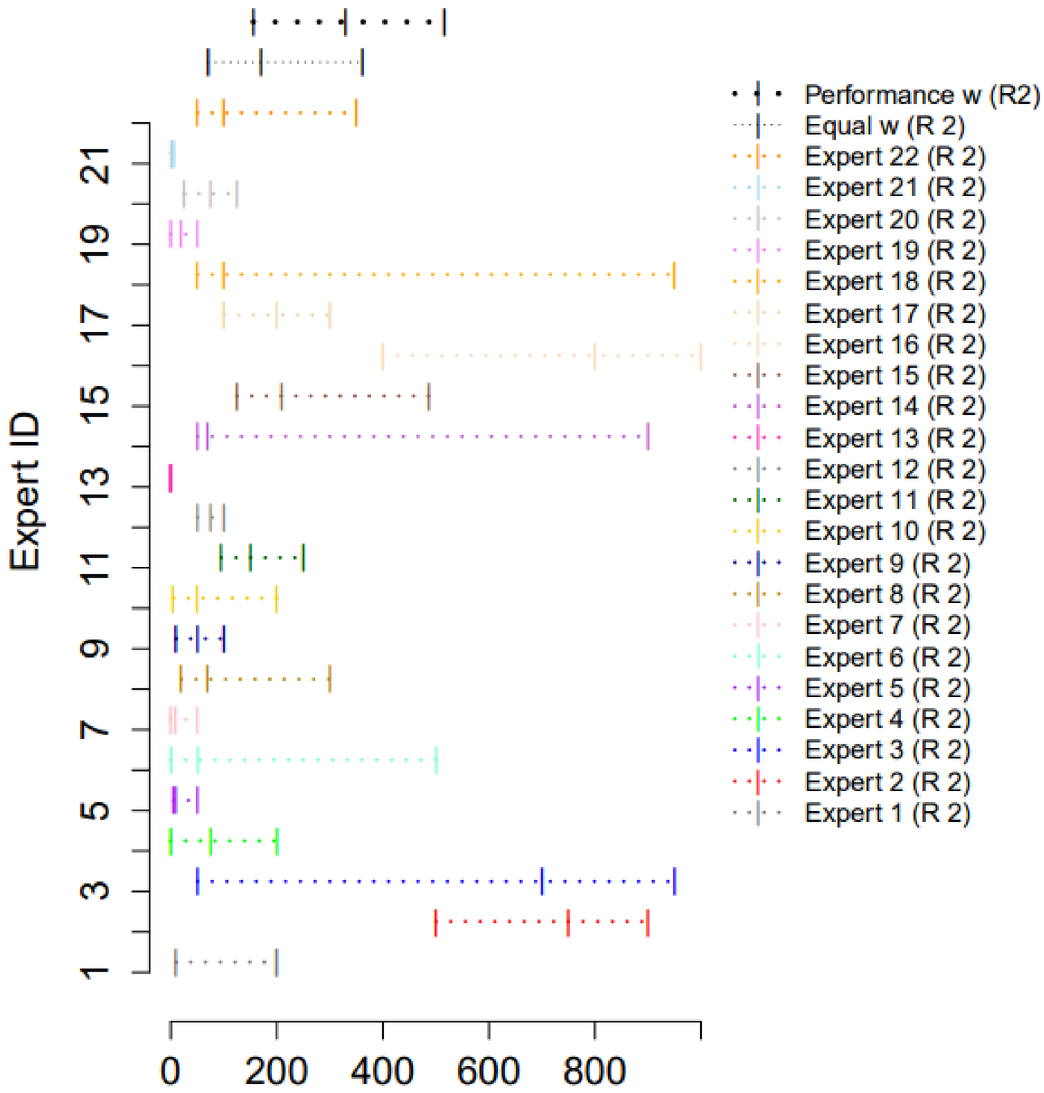
THE NATIONAL ARCHIVES

WARWICK

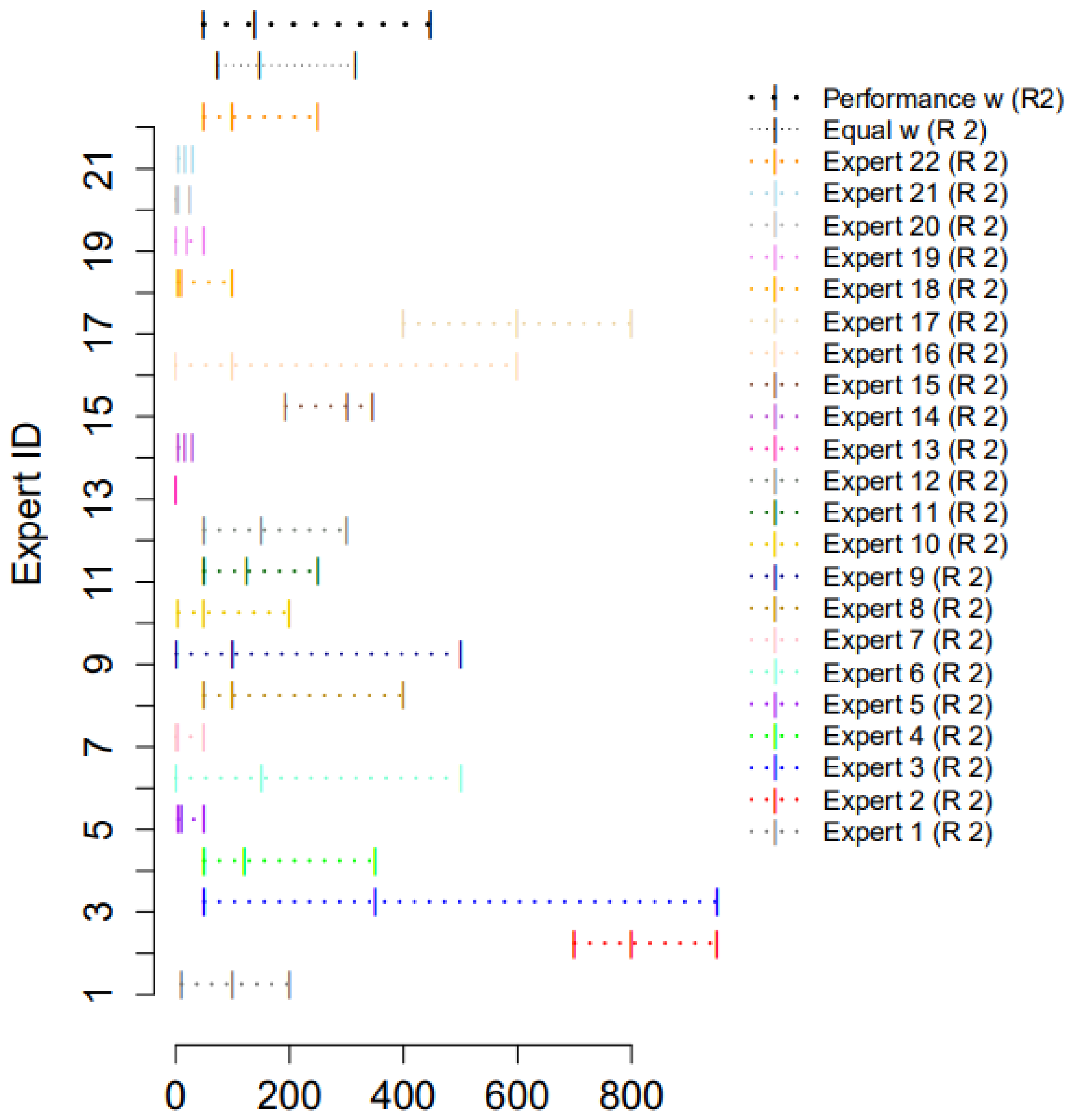
**Plot 17 (Question 31)**



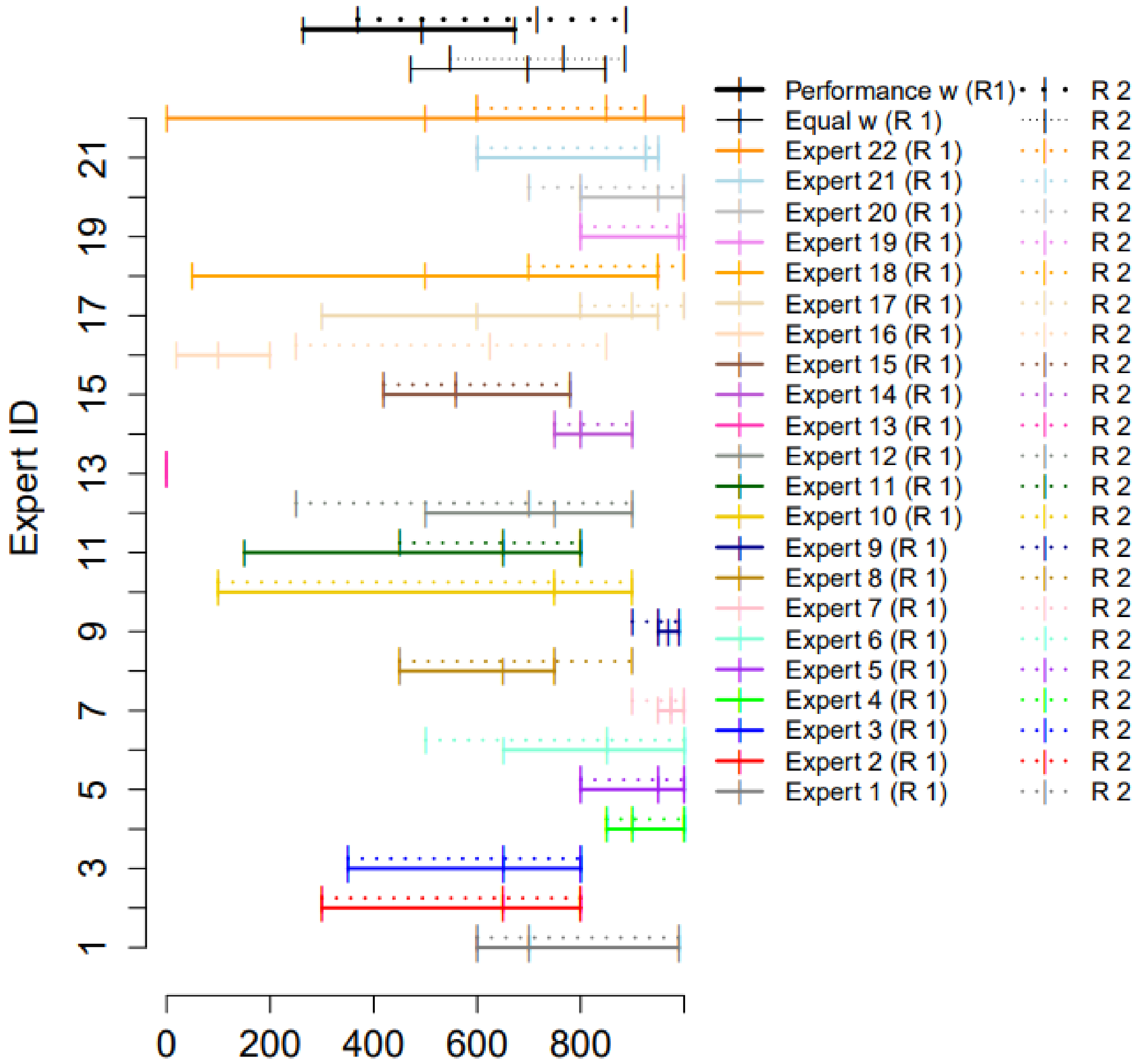
**Plot 18 (Question 32)**



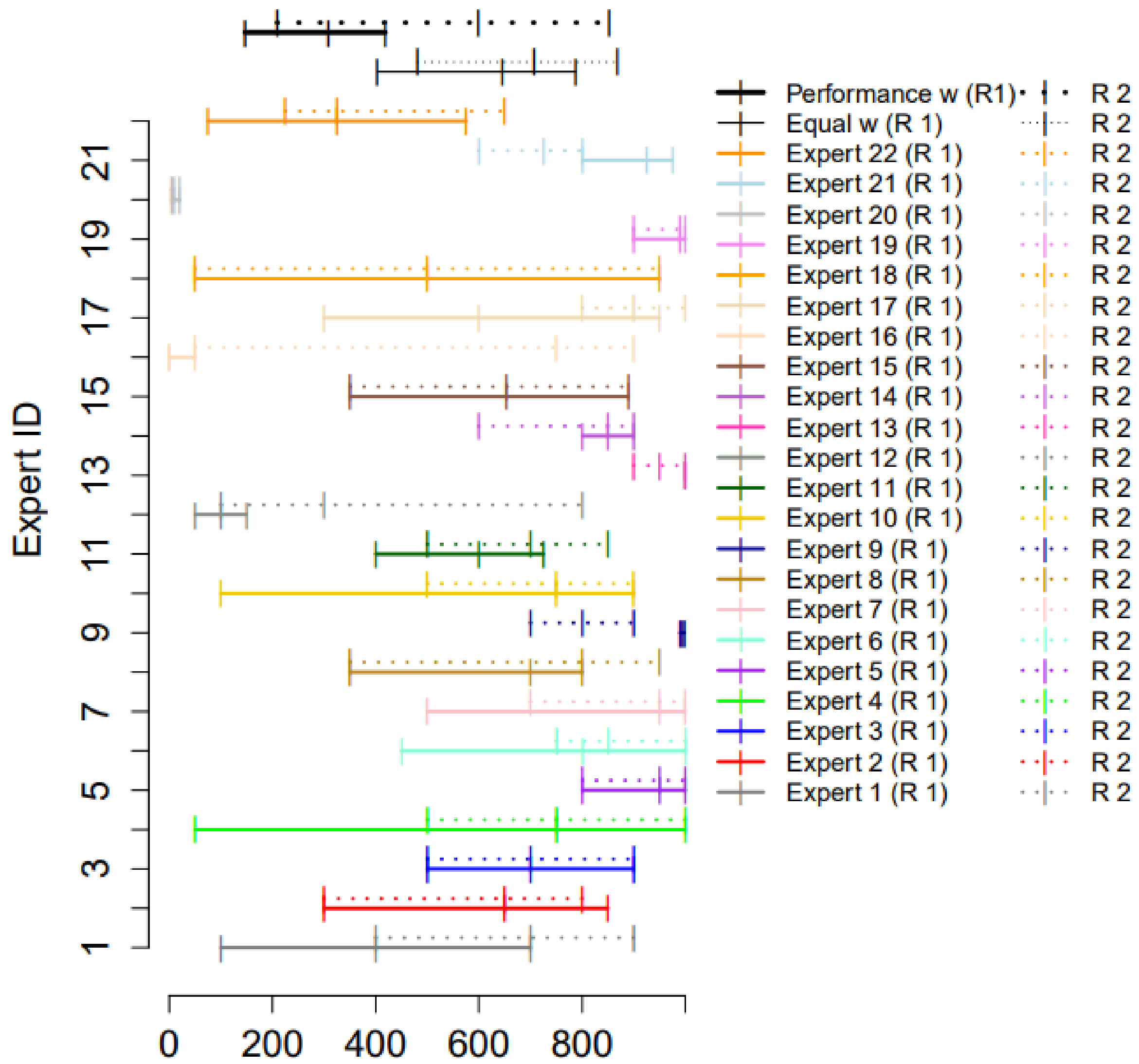
**Plot 19 (Question 33)**



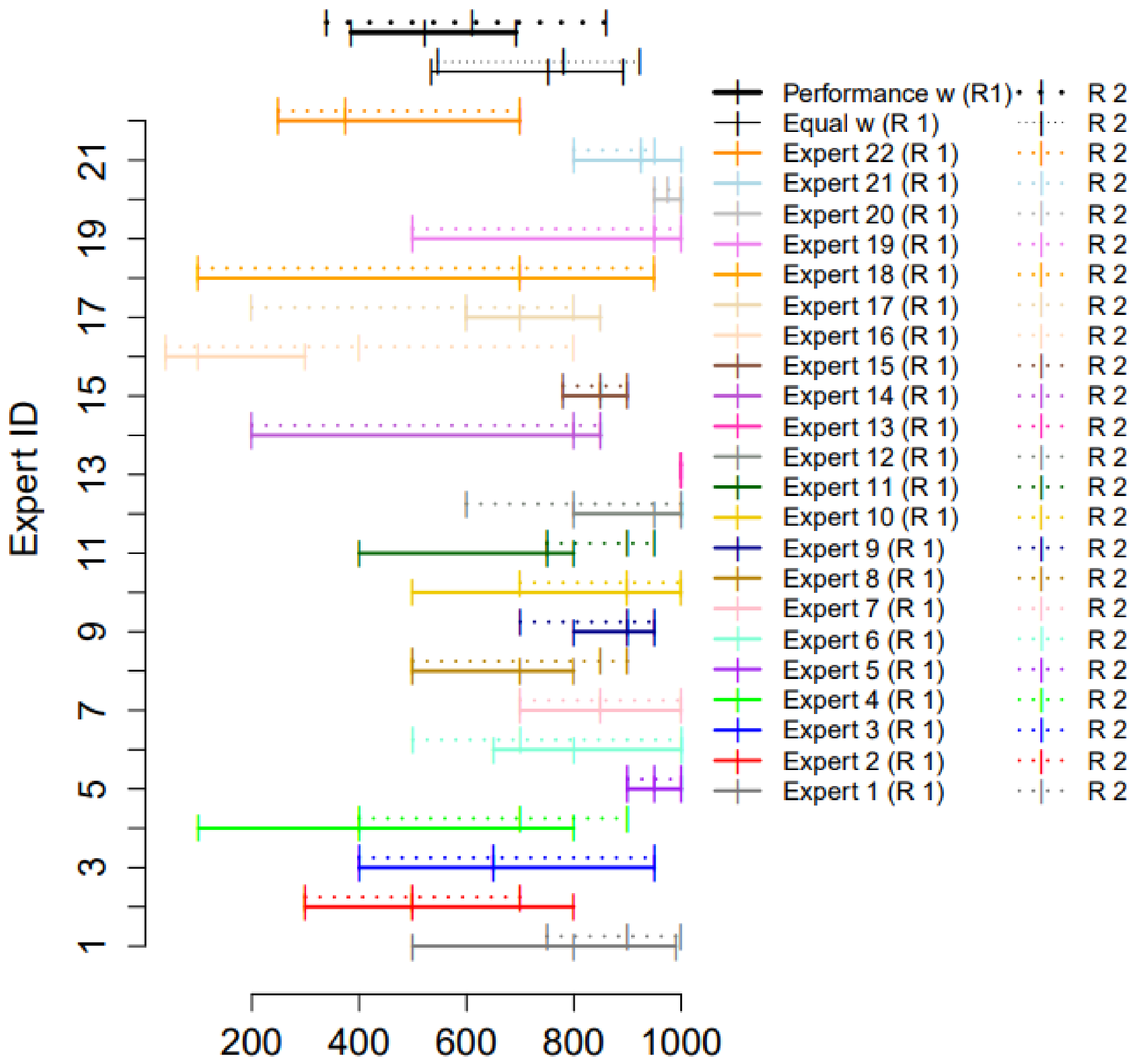
**Plot 20 (Question 34)**



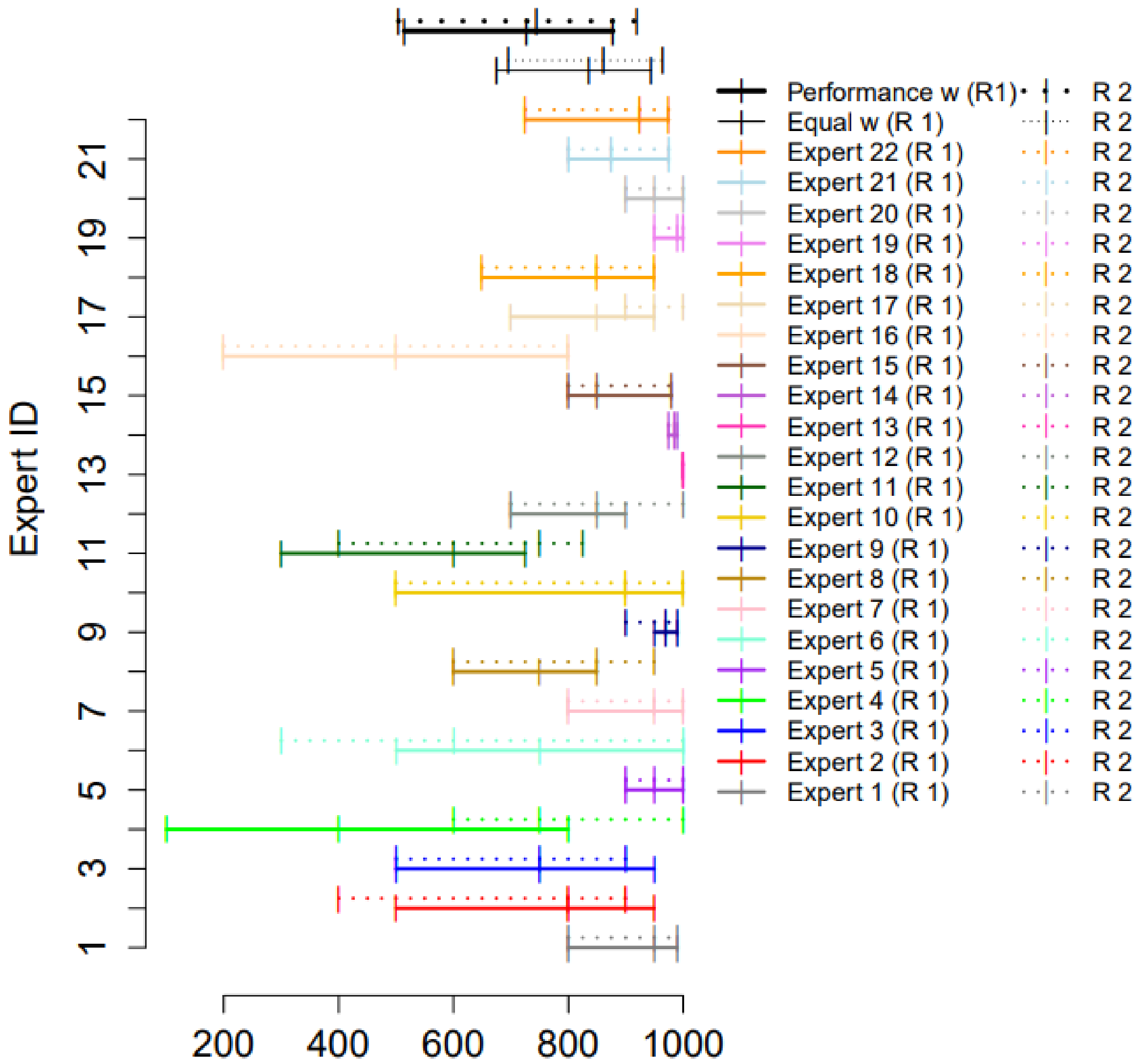
**Plot 21 (Question 35)**



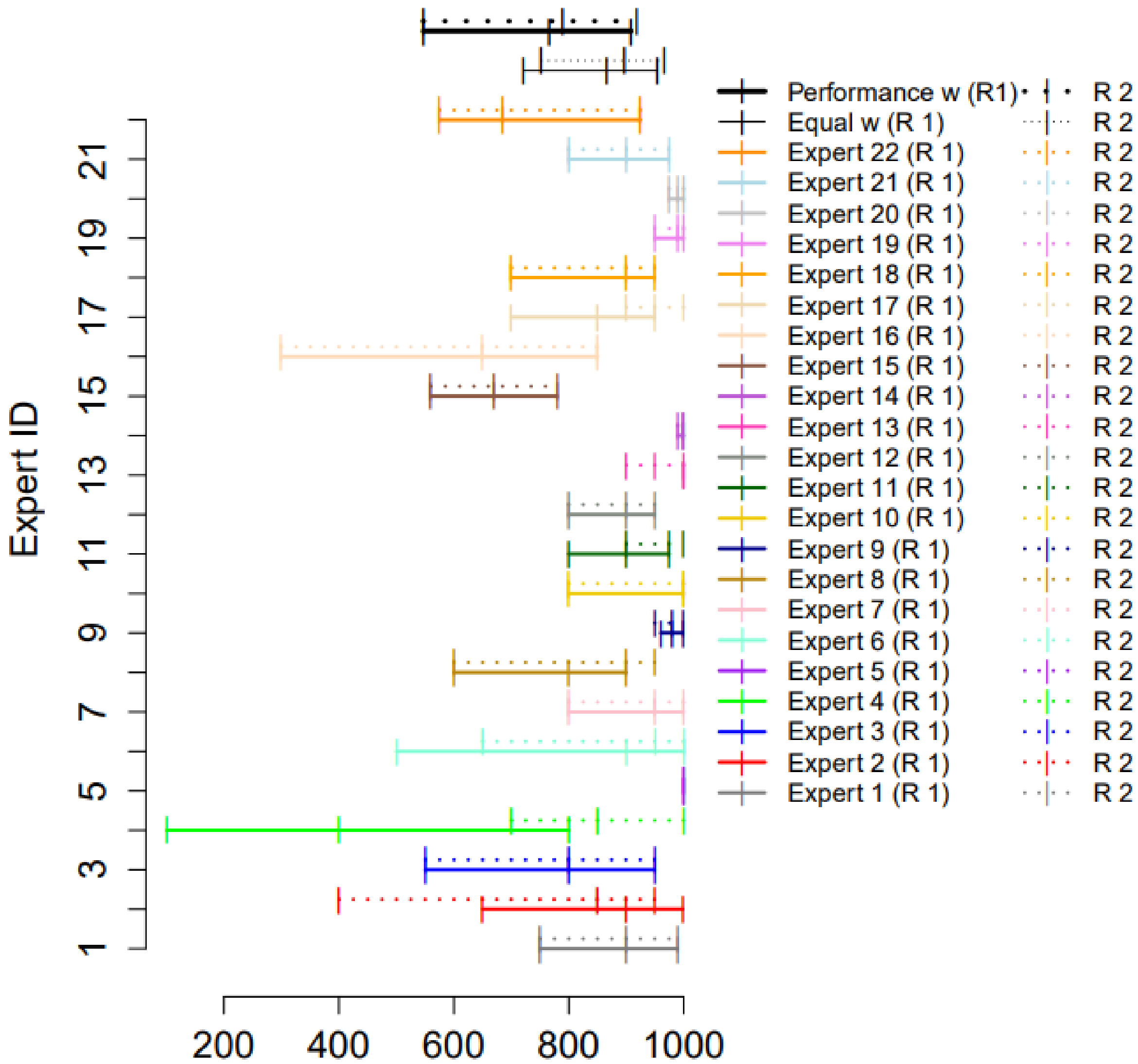
**Plot 22 (Question 36)**



**Plot 23 (Question 37)**



**Plot 24 (Question 38)**



# Appendix A

This appendix features correspondences and material sent in advance of the elicitation workshop, specifically:

- The email sent to experts on 22nd April from Alex Green
- The informed consent form
- The explanatory statement on expert elicitation

Some minor changes of formatting may have occurred as a result of copying this material from the source documents into this report.

**Email to experts**

Good morning,

I hope this email finds you well.

It's now just a week before our online Elicitation Workshop on 28th and 29th April and we have some important information for you to read ahead of it.

Please note your attendance will be required for the full two days as the second day follows on from the first.

Consent Form

It is vital that we have your informed consent before we start the elicitation, so we would be grateful if you could review the attached 'Informed Consent Form' at your earliest convenience. Once you have read and fully understood the document, you can confirm your consent to me either via an email message or by returning the completed the form (digital signatures will be accepted). If you have any questions or concerns about how we will be using your data for the model, please don't hesitate to get in touch.

Background Reading

We have invited you to participate in this for your **expertise in digital preservation and archiving -** not because you are experts in probability and statistics. We intend to do all we can to help you understand the process and make you comfortable, so that we can harness your best knowledge and consequently build a rigorous statistical model to help UK archivists better understand and manage their digital preservation risks. It is your expert input that will make this project a success!

Ahead of the workshop, we would recommend reading this blog post which introduces structured expert elicitation and the protocol we will be using. We have also attached a more detailed explanation of expert elicitation and what to expect during the workshop.

We will be explaining the processes from scratch at the start of the workshop allowing plenty of time for questions and someone will always be on hand to help throughout the session, so please don't feel you have to fully understand all these techniques beforehand. However, we hope these might be useful to introduce you to some of the ideas and terms we will be using and feel more prepared.

For more information about the project as a whole, please see our TNA webpage. This DPC blog post also gives a gentle introduction to why this project started and what we hope to cover, followed by a TNA blog post after our ideas began to crystallise some more and a webcast (the latter is available to DPC members only).

Logistics on the day

The meeting invitation has now been updated with the expected timings and agenda. **If you have not received this, please let me know.** The mornings will be used for presentations and discussions, and in the afternoons you will be free to answer the elicitation questions off the Zoom call at your own pace. We understand that some of you may not be able to follow live for the whole of the morning events, so we intend to record them to allow people to play back anything they may have missed. We will only share this privately with individuals that request it and for this purpose alone – it will not be made public or used elsewhere in the project. If you are uncomfortable about this or have any questions, please get in touch.

The workshop will be hosted online via Zoom. To get the best functionality we recommend that you have a personal Zoom account and join via the desktop application if you can. However, you can also join in a web browser or via a phone call. If you intend to dial in please let us know your telephone number in advance, so we know to accept you into the meeting. A link to the meeting and full joining instructions will follow later this week.

Many thanks,

Alex Green

**Informed consent form**

# Digital Preservation Risk Model
# Expert Judgement Workshop 28th-29th April 2020

# Informed Consent Statement

The elicitation workshop will ask experts to provide their subjective beliefs and their degree of uncertainty about various quantities required for a new decision support tool being implemented to support decision-making for digital preservation risk management. We aim to protect participants' anonymity to the fullest possible extent. When we provide feedback to the group, individual comments will not be identifiable and all results will be pseudonymised: names will be disaggregated from answers, and a code name used instead.

The purpose of the workshop is to provide evidence for a new decision support system that draws on judgements of panels of experts who themselves rely on experimental and observational evidence. The decision support system combines these expert judgements in a coherent and defensible way so that a user can use the outputs to score policy options against each other. This tool is being developed by the University of Warwick led by Dr Martine Barons, Director of the Applied Statistics and Risk Unit; the National Archives led by Alex Green, Service Owner for Digital Preservation; and other collaborating partners from five other UK archives.

Producing a quantitative decision support tool for policies designed to manage digital preservation risks has clear novelty and potential for impact. We are hoping the results will be published in a peer-review journal, presented at international conferences and shared in online reports. Your name will not be listed next to your answers; however, we will ask if we may list your name in the acknowledgements section. You can withdraw consent at any time prior to publication, in which case you will not be listed in the acknowledgements, and you are free to do so without prejudice.

We will report group results and illustrative examples only, that is, nothing that could identify you as an individual.

**Participation is totally voluntary**, there is no penalty for declining, and you can withdraw at any time without penalty. There is no financial compensation for participation.

**Funding**

This research is funded by a grant from the National Lottery Heritage Fund awarded to The National Archives and an Engineering and Physical Sciences Research Council grant awarded to the University of Warwick.

**Contact**

If you have any questions, please contact Alex Green, Service Owner: Digital Preservation, The National Archives.

If you consent to your data being used in this way, please complete the declaration below and return this by Monday 27th April 2020.

A virtual signature or email confirmation will be accepted.

---

**Declaration:**

"I (______________________________________________________)
have read and understood this sheet and any questions I have asked have been answered to my satisfaction. I agree to participate in the project, realising that I may withdraw at any time. I agree that the data provided by me during the project may be included in technical reports, conference proceedings or scientific publications on the condition that my name will not be listed next to my answers. My name will be listed in the acknowledgements section of any publication unless I have withdrawn my consent at any time prior to publication, which I am free to do without prejudice."

Signature: Date:

**Explanatory statement**

# *Expert Elicitation*

## *Introduction*

Structured expert judgment is an accepted tool in risk analysis to use when data is not available or is inadequate. It enables us to quantify uncertainty and build rational consensus. As R. M. Cooke and L. H. J. Goossens describe in their procedures guide, “Increasingly, expert judgement is recognised as just another type of scientific data, and methods are developed for treating it as such.”

In practise, the process involves a panel of experts quantifying uncertainty of variables relating to their subject area. Individual results are then carefully combined to maximize the statistical accuracy and Informativeness of the data. For more information on uncertainty, you may find this blog post helpful: Uncertainty, Error, and Confidence | Process of Science.

As outlined in the informed consent form, results are pseudonymised so that individuals cannot be associated with responses in any open documentation, but a collective list of participating expert names may be published as an acknowledgement of their input. Expert reasoning is captured during the elicitation and may become, where indicated, part of the published record.

## *The Elicitation Protocol*

The elicitation will follow the structure and steps of the IDEA protocol for quantifying uncertainty with experts. The acronym IDEA arises from the combination of the key features of the protocol: it encourages experts to **I**nvestigate and estimate individual first round responses**,** **D**iscuss, **E**stimate second round responses, following which judgements are combined using mathematical **A**ggregation.

An outline of the basic approach is as follows. First, experts provide private, individual estimates in response to the questions posed to them. This first step is meant to encourage independent thinking and avoid anchoring on other experts' estimates.

Then experts receive feedback in the form of the judgements of the other experts.

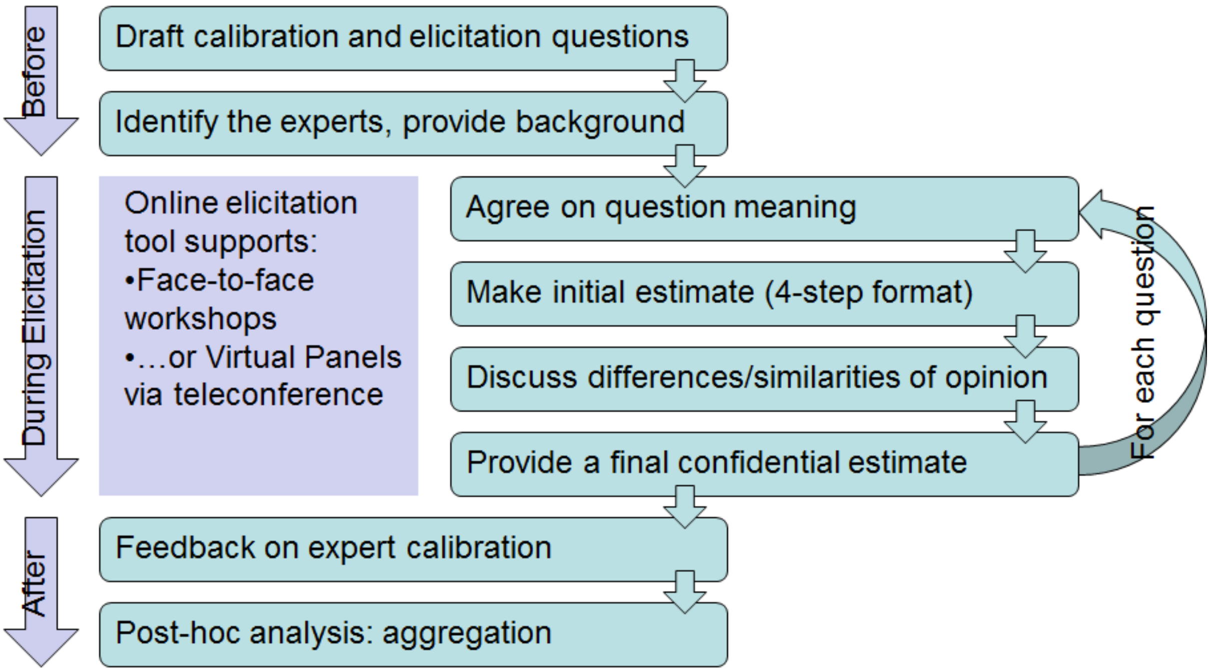
With the assistance of a facilitator, the experts discuss their initial estimates with the others, sharing information, clarifying terms, and establishing a shared understanding of the problem. This discussion stage may take place remotely or face-to-face. We recommend:

- Use discussion to clarify the terms and scope of the question. Use discussion time to make sure everyone understands the information and terms in the same way.
- Use discussion to share information, and to comment on the information shared by others.
- Use discussion to ask questions about extreme values, that is the very low or very high estimates that look different to the rest of the groups' responses. Also, look out for very narrow intervals, suggesting very confident responses (see the format of the questions section). Do those individuals have extra information? In contrast, very wide intervals suggest unconfident responses. Perhaps those individuals are aware of contradictory evidence?
- Use discussion to raise counterarguments to others' arguments. The purpose of this is not to annoy each other, but to make sure your group as a whole is considering the full range of evidence. Moreover, this guards against the availability bias which makes people "remember events that have occurred most recently and deem them to be more plausible, therefore assigning them a higher probability".

We would like to emphasize that consensus is NOT the goal.

Experts are then asked to revise their judgements in light of this discussion and make a second, private and anonymous estimate. By having this opportunity, we hope to avoid group/social biases such as deferring to the most senior, highest ranking or most charismatic member of the group. These second round estimates are finally combined mathematically. The final aggregation may combine the estimates using equal weights, or differential weights calculated based on answers to calibration questions (see the next section for details).

The full process from pre- to post-elicitation is depicted below.




## *The Questions*

There are two types of questions:

- **Calibration questions** – these are questions of which the true answers are known (with certainty) by the analyst or can be found within the time span of the study and are not known (hopefully) by experts. The performance of the experts on the calibration variables is taken as indicative for the performance on the questions of interest. Therefore, the calibration questions must resemble as much as possible the questions of interest.
  A fundamental assumption of using calibration questions to measure performance is that the future performance of experts (as uncertainty assessors) can be judged on the basis of past performance, reflected in the so-called calibration questions.

- **Questions of interest** – these are the questions for which nobody knows the answer and we are trying to quantify this uncertainty in the best way possible.

All questions will be asked in the same way/format.

*Question Format*

All questions will be about quantities surrounded by uncertainty and measured on a continuous scale.

Consider the example:

| ***Out of 1,000 adults in England, how many visited a museum or gallery between April 2018 and March 2019?*** | | |
|---|---|---|
| ______ | ______ | ______ |
| **5%** | **95%** | **50%** |

Suppose you respond as shown below:

| ***Out of 1,000 adults in England, how many visited a museum or gallery between April 2018 and March 2019?*** | | |
|---|---|---|
| **50** | **460** | **200** |
| **5%** | **95%** | **50%** |

This means that*, according to your belief,* the true value is equally likely to be above or below 200 *and* there is only a 1 in 20 chance that the true value is below 50 and there is only a 1 in 20 chance that the true value is above 460 *and* therefore there is an 18 in 20 (90%) chance that the true value lies between 50 and 460.

We ask for 3 quantities per question, and the order is important.

> First, we ask for the number for which the true value has only a 1 in 20 chance of being BELOW and a 19 in 20 chance of being ABOVE. We call this the 5%-tile. Another way of explaining this is the 5%-tile is that number for which the chance that the true value is BELOW is 0.05 and the chance that the true value is ABOVE is 0.95.
>
> Second, we ask for the number for which the true value has a 19 in 20 chance of being BELOW, and only a 1 in 20 chance of being ABOVE. We call this the 95% -tile. Another way to explain this is the 95% -tile is that number for which the chance that the true value is BELOW is 0.95, and the chance that the true value is ABOVE is 0.05.
>
> Lastly, we ask for the number for which the true value has a ½ chance of being ABOVE and a ½ chance of being BELOW. We call this the 50%-tile, and you can think of it as your best estimate (expressed as a median). Another way to explain this is the 50%-

tile is that number for which you judge the chance ½ that the true value is above or below.

**ALWAYS: 5%-tile ≤ 50%-tile ≤ 95%-tile**

Answering the questions in this order is meant to account for anchoring on the central estimate, and in this way reduce overconfidence (the temptation to give intervals which are too narrow). We can represent the uncertainty using a probability distribution, and we will use your quantiles of these distributions.

A good probability assessor is one whose assessments capture the true values with the long run correct relative frequencies (statistically accurate or calibrated), with distributions that are as narrow as possible (informative). Informativeness is gauged by 'how far apart the percentiles are' relative to an appropriate background (Shannon relative information).

Measuring statistical accuracy requires the true values for a set of assessments. The true value for the above question is 502. It falls above the 95%-tile. If the expert's assessments are statistically accurate, then in the long run, 5% of their answers should fall within this inter-percentile interval. Similarly, 90% of the answers should fall between the 5%-tile and the 95%-tile, etc.

In gauging overall performance, statistical accuracy is more important than informativeness. Non-informative but statistically accurate assessments are useful, as they sensitize us to how large the uncertainties may be; highly informative but statistically very inaccurate assessments are not useful. **Do not shy away from wide distributions if that reflects your real uncertainty.**

If you have little knowledge about an item, this fact by itself does NOT disqualify you as an uncertainty assessor. Knowing little means that your percentiles should be 'far apart'. If other experts are more informative, without sacrificing accuracy, then they will exert more influence on the decision maker. But if there are no statistically accurate experts with more informative assessments, then the uninformative assessments accurately depict the uncertainty. That in itself is VERY important information.

Finally, note that the combined score which accounts for calibration and information does not aim to score experts in the same way as grading students. The aim is to find the best combination of answers, that is, a spanning set of options which best represent the group.

# Appendix B

This appendix features material used and sent during the elicitation workshop, specifically:

- The definitions list as presented to the experts
- The questions as presented to the experts, but with calibration questions starred (*)
- The slide packs and presentations used for the virtual meeting
- A transcription of the discussions on the question semantics and round 1 answers

Some minor changes of formatting may have occurred as a result of copying this material from the source documents into this report.

**Definitions list**

| Node Name | Definition |
|---|---|
| **Tools_to_Render** | Availability of tools and software to render the digital material and the expertise to use them. |
| **Technical_Skills** | Bespoke digital preservation skills such as awareness of technological trends, detailed knowledge of storage media, hardware and software, skills to perform file format migration, skills to find emulating software etc…<br><br>"Good" is having the capability to perform actions such as file format migration, software emulation and data recovery, as well as having intermediate Excel skills (e.g. pivot tables) and/or basic programming. |
| **System_Security** | A secure system can protect data from deletion or modification from any unauthorized party, and it ensures that when an authorized person makes a change that should not have been made the damage can be reversed.<br><br>Definition from Forcepoint, https://www.forcepoint.com/cyber-edu/cia-triad<br><br>"Good" is actively managing access restrictions virtually and physically, having antivirus software, performing regular audits, etc. |

| | |
|---|---|
| **Obsolescence** | Essential equipment, hardware or software required to access the bit stream becoming out of date or unusable.<br><br>For example, hardware no longer being produced or essential software to access the media no longer supported. |
| **File_Format** | A file format is a standard way that information is encoded for storage in a computer file. It tells the computer how to display, print, and process, and save the information. It is dictated by the application program which created the file, and the operating system under which it was created and stored.<br><br>Definition from http://en.wikipedia.org/wiki/File_format<br><br>The file format can be described as ubiquitous if is widely known and used by non-specialists.<br><br>If the file format is not proprietary, is can be described as open.<br><br>The availability of tools and software to render a digital object depends on the file format. |

| | |
|---|---|
| **Checksum** | A unique numerical signature derived from a file that can be used to compare copies.<br><br>Definition from the DPC handbook: https://www.dpconline.org/handbook/glossary<br><br>This is needed to ensure integrity of the digital object. |
| **Tech_Metadata** | Technical information that describes how a file functions and that enables a computer to understand it at the bit level, so that it can be rendered in a way that is faithful to its original content.<br><br>Adapted from: https://www.nedcc.org/fundamentals-of-av-preservation-textbook/chapter-4-introduction/chapter-4-section-5 |
| **Info_Management** | Internal systems and support for coherent information management and documentation of preservation actions.<br><br>This is needed to ensure integrity and provenance of the digital object. |

| | |
|---|---|
| **Digital_Object** | Whether it is a born-digital record, digitised record or surrogate.<br><br>- Born-Digital records are digital materials which have originated from a digital source.<br><br>- Surrogates are digital materials which have been created as a result of converting analogue originals where the archive also holds the analogue original. (They are surrogate as the analogue is the official record).<br><br>- Digitised records are digital materials which have been created as a result of converting analogue originals, but the archive does not hold the analogue original. |
| **Content_Metadata** | Describes the intellectual entity through properties such as author and title, and supports discovery and delivery of digital content. It may also provide an historic context, by, for example, specifying which print-based material was the original source for a digital derivative (source provenance). It also includes provenance information of who has cared for the digital object and what preservation actions have been performed on it.<br><br>Adapted from the Digital Preservation Metadata Standards publication https://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isqv22no2.pdf |

| | |
|---|---|
| **Storage_Medium** | The type of media on which the digital material is stored.<br><br>Type A - less stable<br><br>Expected lifespan below 10 years or unknown, highly susceptible to physical damage, requires specific environmental conditions and very sensitive to changes, does not support error-detection methods, supporting technology is novel, proprietary and limited. Examples include USB flash drive, floppy disk, SD drive and CD-R discs.<br><br>Type B - more stable<br><br>A proven lifespan of at least 10 years, low susceptibility to physical damage, tolerant of a wide range of environmental conditions without data loss, supports robust error-detection methods, supporting technology is well established and widely available. Examples include LTO tapes, Blu-ray discs and CD-ROM discs.<br><br>Type C - outsourced data storage<br><br>An external company is taking responsibility for your data storage. Examples include Amazon Simple Storage Service, Microsoft Azure Archive Storage and Google Cloud Storage. |

| | |
|---|---|
| **Storage_Life** | The length of time for which the physical storage device is expected to store the digital object's bit stream.<br><br>The end of storage life can be described as the point at which you can no longer store or retrieve data due to hardware defects or malfunction. |
| **ReplicationAndRefreshment** | The archives policies on making and maintaining copies of digital materials.<br><br>"Good" is always having at least two copies of the bitstream at any one time. |
| **Op_Environment** | The conditions under which the storage media are kept, in terms of their resilience to damage from a flood. |
| **Integrity** | The assurance that the bit-stream is identical to when it was added to the archive. |
| **Bit_Preservation** | A term used to denote a very basic level of preservation of digital resource as it was submitted (literally preservation of the bits forming a digital resource).<br><br>Activities may include maintaining onsite and offsite backup copies, virus checking, fixity-checking, and periodic refreshment to new storage media.<br><br>To have bit-preservation, you need sufficient storage life, the bit-stream to be accessible (i.e. not affected by obsolescence) and have integrity. |

| | |
|---|---|
| **Renderability** | The object is a sufficiently useful representation of the original file. |
| **Intellectual_Control** | Having full knowledge of the material content, provenance and conditions of use. |
| **Identity** | Knowing what the material is and where it is from.<br><br>Specifically:<br><br>- Can you locate the file?<br><br>- Is it sufficiently described for you to know this is what you want?<br><br>- Can you understand its context within the archive?<br><br>- Can you find other versions of the file which were created by preservation actions?<br><br>- Can you find the provenance of the file? |
| **Conditions_of_Use** | Knowing of the conditions of use and any restrictions on the digital material, including the legal status, copyright, who owns the intellectual property and Freedom of Information restrictions. |
| **Physical_Disaster** | For this first version, we will only consider the risk of a flood at your safest storage location. |

| **Other useful terms** | **Definition** |
|---|---|

| | |
|---|---|
| **Provenance** | Information regarding the origins, custody, and ownership of an item or collection.<br><br>Provenance is a fundamental principle of archives, referring to the individual, family, or organization that created or received the items in a collection. The principle of provenance dictates that records of different origins be kept separate to preserve their context and the archivist should maintain the original order in which the records were created and kept.<br><br>Definition from the Society of American Archivists, for more details see here: https://www2.archivists.org/glossary/terms/p/provenance |

| | |
|---|---|
| **Migration** | A means of overcoming technological obsolescence by transferring digital resources from one hardware/software generation to the next. The purpose of migration is to preserve the intellectual content of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. Migration differs from the refreshing of storage media in that it is not always possible to make an exact digital copy or replicate original features and appearance and still maintain the compatibility of the resource with the new generation of technology.<br><br>Definition from the DPC Handbook: https://www.dpconline.org/handbook/glossary |
| **Emulation** | A means of overcoming technological obsolescence of hardware and software by developing techniques for imitating obsolete systems on future generations of computers.<br><br>Definition from the DPC Handbook: https://www.dpconline.org/handbook/glossary |
| **Redaction** | The separation of disclosable from non-disclosable information by blocking out individual words, sentences or paragraphs or the removal of whole pages or sections prior to the release of the document. |

| | |
|---|---|
| **Fixity Check** | A method for ensuring the integrity of a file and verifying it has not been altered or corrupted. During transfer, an archive may run a fixity check to ensure a transmitted file has not been altered en route. Within the archive, fixity checking is used to ensure that digital files have not been altered or corrupted. It is most often accomplished by computing checksums such as MD5, SHA1 or SHA256 for a file and comparing them to a stored value.<br><br>Definition from https://en.wikipedia.org/wiki/File_fixity |

THE NATIONAL ARCHIVES

WARWICK

**Questions**

Definition reminder

**File Format**: A file format is a standard way that information is encoded for storage in a computer file. It tells the computer how to display, print, and process, and save the information.

**Technical Skills**: Bespoke digital preservation skills such as awareness of technological trends, detailed knowledge of storage media, hardware and software, skills to perform file format migration, skills to find emulating software etc.

**Tools to render**: Availability of tools and software to render the digital material and the expertise to use them.

Out of 1,000 UK archivists, how many would you expect to say that their archive has a list of all the digital file formats collected?*

1. Out of 1,000 UK archivists, how many would you expect to say that they had at least some knowledge/skill to perform file format analysis of a digital accession?*
2. Out of 1,000 UK archivists, how many would you expect to say that there was some capability within their organisation to carry out at least one of a) file format migration, b) software emulation or c) data recover from damaged/obsolete media, even in a limited way?*
3. Out of 1,000 files with file formats that are ubiquitous and/or open, how many would you expect to have the tools to render?
4. Out of 1,000 files with file formats that are neither ubiquitous nor open, at an archive where staff have good technical skills, how many would you expect to have the tools to render?

**Ubiquitous**: The format is widely known and used by non-specialists.

**Open**: The format is not proprietary.

**Good technical skills** is having the capability to perform actions such as file format migration, software emulation and data recovery, as well as having intermediate Excel skills (e.g. pivot tables) and/or basic programming.

Definition reminder

**Content Metadata**: Describes the intellectual entity through properties such as author and title, and supports discovery and delivery of digital content. It may also provide an historic context, by, for example, specifying which print-based material was the original source for a digital derivative (source provenance). It also includes provenance information of who has cared for the digital object and what preservation actions have been performed on it.

5. Out of 1,000 born-digital files, how many would you expect to have the content metadata that meets an archive's requirements?

6. Out of 1,000 digitised files, how many would you expect to have the content metadata that meets an archive's requirements?

7. Out of 1,000 surrogate files, how many would you expect to have the content metadata that meets an archive's requirements?

**Born-Digital** records are digital materials which have originated from a digital source.

**Digitised** records are digital materials which have been created as a result of converting analogue originals, but the archive does not hold the analogue original.

**Surrogates** are digital materials which have been created as a result of converting analogue originals where the archive also holds the analogue original.

8. Out of 1,000 UK archivists, how many would you expect to say that their catalogue management system met the needs of their organisation?*

9. Out of 1,000 UK archivists, how many would you expect to say that their digital asset management system met the needs of their organisation?*

10. Out of 1,000 files with insufficient content metadata, at an archive where there is sufficient information management, how many files would you expect to be able to identify (i.e. knowing what they are and where they are from)?

11. Out of 1,000 files at an archive where staff have good technical skills, how many files would you expect to have sufficient technical metadata?

**Good technical skills** is having the capability to perform actions such as file format migration, software emulation and data recovery, as well as having intermediate Excel skills (e.g. pivot tables) and/or basic programming.

12. Out of 1,000 files at an archive where staff have poor technical skills, how many files would you expect to have sufficient technical metadata?

<u>Definition reminder</u>

**Information Management:** Internal systems and support for coherent information management and documentation of preservation actions.

**Identity**: Knowing what the material is and where it is from. Specifically:

- Can you locate the file?

- Is it sufficiently described for you to know this is what you want?

- Can you understand its context within the archive?

- Can you find other versions of the file which were created by preservation actions?

<u>Definition Reminder</u>

**Storage life**: The length of time for which the physical storage device is expected to store the digital object's bit stream. The end of storage life can be described as the point at which you can no longer store or retrieve data due to hardware defects or malfunction.

13. Out of 1,000 hard drive disks kept in a monitored commercial environment, how many drives would you expect to fail in any 12 month period?*

14. Out of 1,000 hard drive disks kept in a monitored commercial environment, how many drives would you expect to fail within their first 12 months of use?*

15. Out of 1,000 NAND solid state drives kept in a monitored commercial environment, how many drives would you expect to fail in any 12 month period?*

Definition Reminder

**Storage media - type A (less stable)**: Expected lifespan below 10 years or unknown, highly susceptible to physical damage, requires specific environmental conditions and very sensitive to changes, does not support error-detection methods, supporting technology is novel, proprietary and limited. Examples include USB flash drive, floppy disk, SD drive and CD-R discs.

**Storage media - type B (more stable)**: A proven lifespan of at least 10 years, low susceptibility to physical damage, tolerant of a wide range of environmental conditions without data loss, supports robust error-detection methods, supporting technology is well established and widely available. Examples include LTO tapes, blu ray discs and CD-ROM discs.

16. Out of 1,000 NAND solid state drives kept in a monitored commercial environment, how many drives would you expect to experience a persistent read error within any 12 month period?*

17. Out of 1,000 media of type A (less stable), how many would you expect to reach the end of their life within 12 months?

18. Out of 1,000 media of type B (more stable), how many would you expect to reach the end of their life within 12 months?

19. There is a flood at your storage location and the archive has inadequate mitigations. Out of 1,000 media of type A (less stable), how many would you expect to be destroyed?

20. There is a flood at your storage location and the archive has inadequate mitigations. Out of 1,000 media of type B (more stable), how many would you expect to be destroyed?

Note: Mitigations to protect against a flood would include considering the location of the storage media within the building, having local flood defences, sandbags on site etc.

Definition reminder

**Obsolescence**: Essential equipment, hardware or software required to access the bit stream becoming out of date or unusable. For example, hardware no longer being produced or essential software to access the media no longer supported.

**Technical Skills**: Bespoke digital preservation skills such as awareness of technological trends, detailed knowledge of storage media, hardware and software, skills to perform file format migration, skills to find emulating software etc.

21. Out of 1,000 files all stored on a storage medium of type A (less stable) at an archive where staff have good technical skills, how many files would you expect the bit-stream to be inaccessible due to obsolescence?

22. Out of 1,000 files stored on a storage medium of type A (less stable) at an archive where staff have poor technical skills, how many files would you expect the bit-stream to be inaccessible due to obsolescence?

23. Out of 1,000 files stored on a storage medium of type B (more stable) at an archive where staff have good technical skills, how many files would you expect the bit-stream to be inaccessible due to obsolescence?

24. Out of 1,000 files stored on a storage medium of type B (more stable) at an archive where staff have poor technical skills, how many files would you expect the bit-stream to be inaccessible due to obsolescence?

Definition reminder

**Checksum**: A unique numerical signature derived from a file that can be used to compare copies.

**Integrity**: The assurance that the bit-stream is identical to when it was added to the archive.

**System Security**: A secure system can protect data from deletion or modification from any unauthorized party, and it ensures that when an authorized person makes a change that should not have been made the damage can be reversed.

25. Out of 1,000 UK archivists, how many would you expect to say that they had at least some knowledge/skill to be able to generate a checksum of a digital file?*

26. Out of 1,000 UK archivists, how many would you expect to say that their IT provider supports the requirements of their organisation's archival activities to a large extent or a very great extent?*

27. Out of 1,000 UK organisations which experienced a cyber-security break or attack, how many would you expect to have experienced by viruses, spyware or malware, including ransomware attacks?*

28. Out of 1,000 global data breaches, how many would you expect to be due to system glitches?*

29. Out of 1,000 UK archivists, how many would say that their digital collections were fixity checked at regular intervals?*

30. Out of 1,000 files, how many would you expect to have become corrupted during transfer from a depositor to an archive?

**Good system security** is actively managing access restrictions virtually and physically, having antivirus software, performing regular audits, etc.

31. Out of 1,000 files, at an archive where you have poor system security, how many would you expect to have become corrupted and identified this, given you have a checksum to compare to?

32. Out of 1,000 files, at an archive where you have poor system security, how many would you expect to have become corrupted and not been able to identify this, despite having a checksum to compare to?

> Definition reminder
>
> **Bit-preservation**: A term used to denote a very basic level of preservation of digital resource as it was submitted (literally preservation of the bits forming a digital resource).

33. Out of 1,000 files with their bit-stream stored and accessible but where you cannot guarantee their integrity, how many files would you expect to be bit-preserved?

34. Out of 1,000 files which are bit-preserved at an archive which has full access to rendering tools but where the files have insufficient technical metadata, how many files would you expect to be able to render?

> Definition reminder
>
> **Technical Metadata**: Technical information that describes how a file functions and that enables a computer to understand it at the bit level, so that it can be rendered in a way that is faithful to its original content.
>
> **Renderability**: The object is a sufficiently useful representation of the original file.
>
> **Conditions of use**: Knowing of the conditions of use and any restrictions on the digital material, including the legal status, copyright, who owns the intellectual property and Freedom of Information restrictions

35. Out of 1,000 born-digital files, for how many would you expect an archive to know their conditions of use?

36. Out of 1,000 digitised files, for how many would you expect an archive to know their conditions of use?

37. Out of 1,000 surrogate files, for how many would you expect an archive to know their conditions of use?

> **Born-Digital** records are digital materials which have originated from a digital source.
>
> **Digitised** records are digital materials which have been created as a result of converting analogue originals, but the archive does not hold the analogue original.
>
> **Surrogates** are digital materials which have been created as a result of converting analogue originals where the archive also holds the analogue original.

38. Out of 1,000 UK archivists, how many would you expect to say that their digital collections were protected with access restrictions/permissions?*

39. Out of 1,000 UK archivists, how many would you expect to say that they had at least some knowledge to be able to digitally redact part of a document for web publication?*

40. Out of 1,000 UK archivists, how many would you expect to say that their organisations' born-digital materials were made available to users?*

41. Out of 1,000 UK archivists, how many would you expect to say that their organisations' digitised materials were made available to users?*

42. Out of 1,000 UK non-cyber security incidents, how many would you expect to be due to data being posted, emailed or faxed to an incorrect recipient?*

43. Out of 1,000 UK non-cyber security incidents, how many would you expect to be due to loss/theft of paperwork or data left in an insecure location?*

**Slide packs and material**

***appendix-b-dprm-elicitation-workshop.pptx*** is the main slide pack used to structure both days and includes the agenda and presentations. This was shared with participants in advance.

***appendix-b-elicitation-questions-for-discussion.pptx*** is a supplementary slide pack used for the discussion on question semantics on day 1.

***appendix-b-presented-plots-of-expert-opinion.pdf*** contains the Round 1 graphs used for discussion on day 2.

# Appendix C

This appendix links to the complete transcript and chat saved from Zoom for each session within the elicitation workshop. Participants are identified only by an expert reference which matches the identifiers used in the final range graphs (see Appendix D).

### Transcripts and chat, day 1

***appendix-c-transcript-day-1-am.pdf*** is the transcript starting from Martine's presentation until the lunch break.

***appendix-c-chat-day-1-am.txt*** is the chat for the same period.

***appendix-c-transcript-day-1-pm.pdf*** is the transcript after the lunch break where discussion on question meaning is continued. Also included is a later brief discussion between one participant and Martine and Alex to clarify how they should be answering the questions.

***appendix-c-chat-day-1-pm.txt*** is the chat for the same period.

### Transcripts and chat, day 2

***appendix-c-transcript-day-2-am.pdf*** is the transcript starting from the discussion of range graphs from round 1 until the lunch break.

***appendix-c-chat-day-2-am.txt*** is the chat for the same period.

***appendix-c-transcript-day-2-pm.pdf*** is the transcript after the lunch break where discussion on the round 1 results was continued.

***appendix-c-chat-day-2-pm.txt*** is the chat for the same period.

# Appendix D

This appendix features supporting material from the calibration exercise, specifically:

- The individual results from each expert on the calibration questions (identified by the same reference as in the transcripts and saved chats provided in Appendix C).
- Statistical summaries of the data used to create the weightings.
- An overview of the methodology and a summary.

***appendix-d-expert-performance-updated.pdf***

# Appendix E

This appendix provides details of all questions considered for inclusion as calibration questions in the elicitation exercise.

***appendix-e-working-doc-calibration-questions.xlsx***