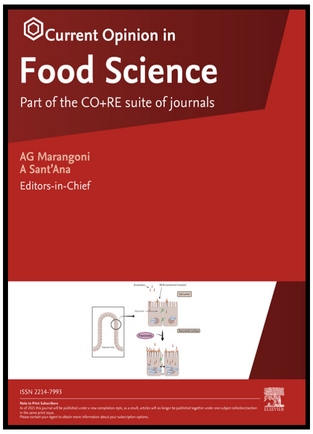

Online consumer research: More attention needs to be given to data qualityShort Title: Data quality in online consumer research

Daniele Asioli, Sara R. Jaeger

PII: S2214-7993(25)00037-2

DOI: https://doi.org/10.1016/j.cofs.2025.101307

Reference: COFS101307

To appear in: *Current Opinion in Food Science*

Manuscript for submission to: *Current Opinion in Food Science*

Online consumer research: More attention needs to be given to data quality

Short Title: Data quality in online consumer research

Daniele Asioli[1] and Sara R. Jaeger[2,*]

[1]Department of Agri-Food Economics and Marketing, School of Agriculture, Policy, and Development, University of Reading, Reading, United Kingdom.

[2]Department of Food Science, Aarhus University, Aarhus N, Denmark.

* Corresponding author: sara.jaeger@food.au.dk

ORCID IDs

D Asioli: 0000-0003-2274-8450

SR Jaeger: 0000-0002-4960-5233

Reviewers' suggestions: Cardello, Castura and Berry

HIGHLIGHTS

- More attention is needed to data quality in online food-related consumer research
- Data quality extends beyond intrinsic accuracy and is a multidimensional construct
- Intrinsic data quality can be improved using *ex-ante* and *post hoc* strategies
- Intrinsic points: attention, honesty, comprehension, reliability, and replicability

- Professionals should overcome issues about sampling, intrinsic quality, and ethics

ABSTRACT

This manuscript aims to raise awareness of the need for Sensory & Consumer Science professionals to be more diligent about data quality control when conducting consumer surveys and experiments online. This aim is achieved by i) summarising recent research on data quality by Jaeger and Cardello (2022) and Castura et al. (2023), ii) contributing to a broader understanding of data quality, and iii) recommending a more systematic adoption of practices aimed to enhance data quality in online consumer research. Various suggestions are put forward to support Sensory & Consumer Science professionals who wish to pay greater attention to the quality of online data collection. We advocate for making online consumer data quality an integral part of the research process to improve the validity and reliability of research outcomes, ultimately benefiting the final users, including science, industry, and policymakers.

Graphical abstract

KEYWORDS: Data quality; Experiments; Online data collection; Sensory & Consumer Science; Suggestions; Surveys.

1. INTRODUCTION

In the last decade, Sensory & Consumer Science has continued to experience more research using surveys[1] and experiments[2], where the participants are recruited via online platforms [1,2]. This trend, mirrored in psychology, sociology, economics, marketing, and other disciplines, [3,4] has been driven by multiple factors. These include i) the growing availability of online platforms that offer low costs, speedy data collection, access to large consumer panels to more geographically dispersed countries, and ease of use [5,6], ii) the need to test new product concepts and consumer preferences of foods not yet fully developed or approved for market commercialisation [7], iii) the need for improved external validity of research outcomes which necessitates the use of large, and more representative consumer samples [8], iv) the ability to use/adopt more flexible experimental designs [8], and v) the high costs and logistics management challenges related to central location tests, and fields experiments [8]. Assuming that the trend toward greater reliance on online participant recruitment and research implementation continues, data quality merits greater attention [9,10]since reduced or low data quality can significantly affect the informational value for science, industry, and policymakers that the data give rise to [11–13].

[1] Survey is a method typically employed to gather quantitative data about respondents' attitudes, opinions, preferences, behaviours, or characteristics. A questionnaire is the typical tool used to collect survey data with limited control, and no manipulations on the variables of interest.
[2] Experiment is a method typically employed to respondents to determine causality by examining the effects of manipulating one or more independent variables on dependent variables. Thus, data collected using experiments typically have high degree of control, and manipulations on the variables of interest.

At the outset, we wish to clarify that it is not our intent to suggest that professionals in Sensory & Consumer Science *per se* are unappreciative of the importance of data quality. A notable example to the contrary is the descriptive sensory analysis, a cornerstone methodology in sensory science where panellists are carefully screened, selected, and trained, as detailed in a core textbook [14], and their performance is closely monitored, for example, relating to discrimination, repeatability, reproducibility and scale use [15,16]. However, our field has no similar body of literature regarding consumer research. This may be linked to Food Science (and Technology) being the dominant educational background in Sensory & Consumer Science [17], which likely means that training in social sciences (e.g., marketing, psychology, and economics), where surveys and experiments are popular methods for collecting consumer data, is minimal. In turn, some Sensory & Consumer Science professionals may lack knowledge of and experience with survey design and incomplete awareness of the factors that can reduce data quality.

Against this background, we seek to strengthen awareness of the need to be more diligent with data quality control when conducting online consumer surveys and experiments (Figure 1). To this end, we briefly summarise i) Jaeger & Cardello [2], who reviewed the literature on factors affecting the data quality of online questionnaires and identified issues and metrics of relevance for Sensory & Consumer Science and ii) Castura et al. [1], who addressed respondent screening and the impact of doing so on data quality in consumer product testing. We build on these contributions and offer a broader understanding of data quality, recommend a more systematic adoption of practices to enhance data quality in online consumer research and provide several recommendations for future research. We hope to help Sensory & Consumer Science professionals address data quality issues and adapt these to individual studies as needed.

2. CURRENT RECOMMENDATIONS ABOUT DATA QUALITY: BENEFITS AND SHORTCOMINGS

Jaeger & Cardello [2] identified 16 factors (Table 1) linked to data quality, and from these, they developed ten recommendations (Table 2) covering both *ex-ante* (before data collection) and *post hoc* (after data collection) strategies to follow when conducting consumer surveys and experiments online. The use of *ex-ante* strategies, including clear and concise questions to minimise misinterpretation, attention checks [18] and instructional manipulation checks [13,19] progress indicators to keep participants engaged, designing surveys that are user-friendly and not overly long or complex, providing clear instructions [20], and emphasising the importance of honest, and thoughtful responses (e.g. cheap talk scripts, etc.) should be considered during the design and data collection phases of online consumer surveys and experiments because they can help reduce the need for extensive *post hoc* participant elimination ahead of data analysis which is wasteful and time-consuming. Elimination is typically based on imposed criteria for survey and experiment completion rates and time, item response times, response attention and accuracy, and straight-lining response behaviour. Jaeger & Cardello [2] urged that authors include written data quality statements in their articles and that these be explicit about how the final sample was achieved and data cleaning undertaken. Interested readers can find an example data quality statement in Figure 3 of Jaeger et al. [21] but are reminded that the form of the data quality statement is unimportant *per se*. The content matters most, and the information needs not be given in a single statement but could be integrated into an article or report at different relevant places.

The study by Castura et al. [1] on respondent screening as a means to increase data quality in consumer product tests is directly relevant for Sensory & Consumer Science professionals because of the popularity of such tests for understanding consumer opinions and perceptions of food products. The authors found that excluding consumers who gave low-quality screener responses (e.g., those (i) declaring to consume some fictive food brands, (ii) providing a low-effort description, (iii) responding in a contradictory manner, (iv) and completing the test faster than high quality screener respondents) affected product test results. Specifically, screener-passing respondents better differentiated between product concepts in an online consumer test. In alignment with the generally recognised importance of data quality, the authors' overarching conclusion was that "data-quality screening has the potential to yield the same results with fewer consumers, which realizes cost savings, or to obtain better quality data from a same-sized panel, which gives more insight and greater confidence in decision making" (p. 7).

The current recommendations for professionals in Sensory & Consumer Science about data quality are valuable because they are specific and can be used as-is without digging deeply into the data quality literature. Considering the high demands often placed on academic and industry professionals, we believe the recommendations' tangibility increases their likelihood of uptake. Although knowing what "tools" are available to improve and monitor data quality in online surveys and experiments, such knowledge has limited value without a shared understanding of the hallmarks of higher vs lower data quality. Progress on this is contingent on a definition of data quality, which in its simplest form reduces to fitness of use from the data users' perspective [22,23].

Table 1 about here.

Table 2 about here.

3. DATA QUALITY IS BEST UNDERSTOOD AS A MULTIDIMENSIONAL CONTINUUM

"Fitness for use" is intuitive as a definition of data quality but lacks operational details and does not clarify that data quality is multidimensional and exists on a continuum. However, the earlier sections show that data quality has multiple dimensions or properties [24] . If these, in the case of bad data, include being inaccurate, inaccessible, out-of-date, redundant, inconsistent, incomplete, incomprehensible, unrepresentative, and unethical [22,25], good data possess the opposite dimensions and properties. That a data quality continuum exists is also evident since data, for example, can be consistent, inconsistent or partly inconsistent, representative, and ethical. However, the criteria that distinguish between different levels of data quality and on what metrics are not so clear. In addition, there is the challenge that the key aspects of data quality vary by discipline. For example, data quality in the context of behavioural research [6] differs from data quality in organisational information systems [26], and big data introduces more criteria and dimensions to evaluate data quality than have previously been considered [27]. Even in the context of survey research, researchers from statistics, psychology, political science, behavioural science and consumer science approach the topic of data quality differently [28], which can hinder communication and has led to disagreements about the importance of various components of error and quality. It is beyond the scope of this manuscript to perform a systematic review of the data quality literature, but interested readers are referred to Ehrlinger & Wöß [12]. Of particular relevance for Sensory & Consumer Science professionals, are the studies on data quality and its dimensions by Peer et al. [6] and Arndt et al. [3].

4. WHEN IS DATA QUALITY GOOD ENOUGH?

Considering our experiences, we expect many Sensory & Consumer Science professionals to face difficulty deciding when data quality is good enough. A few simple and widely agreed-upon criteria would be nice, but the understanding of data quality as a discipline- and context-specific multidimensional continuum means that such criteria are unlikely to exist. Instead, it may be up to individual professionals to decide what degree of quality is needed in each study and then, with full transparency, indicate what quality criteria were implemented and why. This elevates data quality to an explicit part of the research process that needs to be incorporated from the planning stage onwards. Moreover, it places the responsibility of data quality on professionals, not respondents or online platforms, a point also made by Cuskley et al [29].

For help in deciding when data quality is good enough, the literature is surprisingly unhelpful. Tentatively, this has less to do with a lack of methodological studies and more to do with a reluctance to acknowledge an inextricable link between quality and cost. In other words, there is a cost implication to improving data quality, a point initially made by Groves [28] when writing about a lack of errors as an indicator of quality in survey research. While the cost-quality relationship may not apply to every aspect of data quality, it does apply to some. Yet acknowledging this link and referring to resource constraints as a determinant of empirical decisions (e.g., online platform selection, participant eligibility criteria, sample size and representativeness) is, in our experience, frowned upon. Nonetheless, resource constraint is a fact, with unequal consequences for professionals from different continents [30]. Whether or not data collection budgets are tight, "cost efficiency", which drawing on Groves [28], means maximising quality given the available resources, can be a way to prioritise different data quality dimensions and metrics. Arndt et al. [3] and Douglas et al. [31] are examples of research that compares different online platforms on selected quality metrics and reports participant incentives (US$). Studies like these can help explore cost efficiency and understand how overall data quality increases as additional metrics are specified and applied.

In this regard, professionals should also remember that interdependencies can exist between different quality measures. Groves [28] noted that improving response rates through heavy persuasion can increase sample size but lead to data with more measurement error, for example, due to respondent inattention. To illustrate, it may be possible to achieve 500 completions of an online experiment in 10 hours or less, but speed is only valuable for sound decision-making if responses are obtained from target participants and not prioritised over eligibility criteria (i.e., obtaining responses from participants not in the target group to achieve n=500 as quickly as possible). The question facing professionals is which aspect of data quality matters most, which may depend on the specific study context.

5. EXPLICIT AND TRANSPARENT INTEGRATION OF DATA QUALITY INTO ONLINE CONSUMER RESEARCH

Drawing on the above, we recommend that the prevalent mindset among Sensory & Consumer Science professionals regarding data quality in online consumer research should be "engineered and checked". This builds on the perspective that data quality is their responsibility [29]. The value of understanding data quality as a continuum is that it allows professionals in Sensory & Consumer Science to evaluate online platforms, identify ways to improve data collection to mitigate sources of bias, appropriately analyse data, and transparently report their research in line with the current trends [32]. The principle of "engineered and checked" elevates data quality to an explicit part of the research process and provides a framework for evaluating empirical decisions. We propose this mindset be used in addition to the recommendations by Jaeger & Cardello (2022), hereby serving as a "next step" in the quest for higher data quality in online consumer research. Thus, we provide several suggestions for professionals on how to explicitly and transparently integrate data quality considerations into online consumer research (Figure 1).

Figure 1 about here.

6. DISCUSSION AND CONCLUSIONS

We advocate that Sensory & Consumer Science professionals should be explicit about how their data perform on key dimensions of quality and document any steps taken to improve quality through cleaning, validation, etc. However, we acknowledge several limitations. For instance, using survey duration as a proxy for data quality is not always an accurate measure [13], or when consumers make inattentive real food purchasing choices; however, this "non-attention" does not automatically disqualify their responses as biased since in reality, consumers can make inattentive purchase decisions [13], or that in several cases the use of online panels produces “professional respondents” whose behaviour may be influenced by their experience as regular survey participants, potentially rendering them unrepresentative of the population of interest [33].

Future research avenues are recommended in this field to improve data quality, for example, to compare different *ex-ante* and *ex-post* approaches or how to combine them, how many and where to position trap questions to be more effective, to compare the outcomes of attentive vs. inattentive respondents (e.g. speeders vs. non-speeders), etc. Since data quality issues are also receiving considerable attention in other disciplines, Sensory and Consumer Science professionals are encouraged to stay informed about such developments. We can likely learn from their efforts and adopt or amend them as appropriate.

The rapid proliferation of artificial intelligence (AI) is likely to influence data quality issues. Integrity aspects relate to whether responses are provided by humans or AI. AI tools can also be used to help design surveys and experiments and analyse data. These options need to be used with caution to mitigate bias and prevent misuse.

In conclusion, the key message of this manuscript is that greater effort, training, and investment are needed to enhance data quality in consumer science. We advocate for integrating data quality as an important component of the online research process, which should also be documented in published articles. Overall, the main aim is to improve the validity and reliability of research outcomes, ultimately benefiting end users, including scientists, industry professionals, and policymakers.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Daniele Asioli: Conceptualization, Writing – original draft, Writing – review & editing. **Sara R. Jaeger:** Conceptualization, Writing – original draft, Writing – review & editing.

CONFLICT OF INTEREST

Both authors declare no conflicts of interest.

REFERENCES

1. Castura JC, Pohjanheimo T, Laaksonen O, McEwan JA, Varela P, Næs T: **Screening respondents to increase data quality in consumer tests**. *Food Quality and Preference* 2023, **112**:105030.

** **Research on how filtering consumers who gave low-quality screener responses affects test results.**2. Jaeger SR, Cardello AV: **Factors affecting data quality of online questionnaires: Issues and metrics for sensory and consumer research**. *Food Quality and Preference* 2022, **102**:104676.

** **Article on raising awareness of the factors affecting online data quality.**3. Arndt AD, Ford JB, Babin BJ, Luong V: **Collecting samples from online services: How to use screeners to improve data quality**. *International Journal of Research in Marketing* 2022, **39**:117–133.

4. Gao Z, House L, Bi X: **Impact of satisficing behavior in online surveys on consumer preference and welfare estimates**. *Food Policy* 2016, **64**:26–36.
5. Paas LJ, Morren M: **PLease do not answer if you are reading this: respondent attention in online panels**. *Mark Lett* 2018, **29**:13–21.
6. Peer E, Rothschild D, Gordon A, Evernden Z, Damer E: **Data quality of platforms and panels for online behavioral research**. *Behav Res* 2022, **54**:1643–1662.
7. Hess S, Daly A: *Handbook of Choice Modelling*. Edward Elgar Publishing; 2024.
8. Ahles A, Palma MA, Drichoutis AC: **Testing the effectiveness of lottery incentives in online experiments**. *American Journal of Agricultural Economics* 2024, **106**:1435–1453.

9. Kramer M: **Enhancing data quality to improve online surveys**. 2024. All Things Insights. Accessed on March 3rd 2025 from https://allthingsinsights.com/content/enhancing-data-quality-to-improve-online-surveys/

10. Kay-McClean L: **Raising the bar: How to boost data quality**. 2025. ResearchLive. Accessed on March 3rd 2025 from https://www.research-live.com/article/features/raising-the-bar-how-to-boost-data-quality/id/5135953

11. Côrte-Real N, Ruivo P, Oliveira T: **Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value?** *Information & Management* 2020, **57**:103141.

12. Ehrlinger L, Wöß W: **A Survey of Data Quality Measurement and Monitoring Tools**. *Front Big Data* 2022, **5**.

13. Malone T, Lusk JL: **Consequences of Participant Inattention with an Application to Carbon Taxes for Meat Products**. *Ecological Economics* 2018, **145**:218–230.

14. Lawless HT, Heymann H: *Sensory Evaluation of Food: Principles and Practices*. Springer New York Springer e-books Imprint: Springer; 2010.

15. Tabary C, Miège M, Brémaud D, Carvalho L, Vincenzi F: **A tool to help the panel leader to best monitor a sensory panel performance**. *Journal of Sensory Studies* 2021, **36**:e12613.

16. Sipos L, Nyitrai Á, Hitka G, Friedrich LF, Kókai Z: **Sensory Panel Performance Evaluation—Comprehensive Review of Practical Approaches**. *Applied Sciences* 2021, **11**:11977.

17. Jaeger SR, Meiselman HL, Giacalone D: **Sensory and consumer science: A complex, expanding, and interdisciplinary field of science**. *Food Quality and Preference* 2025, **122**:105298.

18. Vecchio R, Caso G, Cembalo L, Borrello M: **Is respondents' inattention in online surveys a major issue for research?** *Econ Agro-Aliment* 2020, **22**.

19. Kearney PE, Curran N, Nugent FJ: **Manipulation Checks in Focus of Attention Research: A Methodological Systematic Review**. *Journal of Motor Learning and Development* 2025, doi:10.1123/jmld.2024-0069.

20. Canavari M, Drichoutis AC, Lusk JL, Nayga Jr Rodolfo M: **How to run an experimental auction: a review of recent advances**. *European Review of Agricultural Economics* 2019, **46**:862–922.

21. Jaeger SR, Rasmussen MA, Cardello AV: **Factors affecting data quality of online questionnaires with consumers in sensory and consumer research: Metrics from the literature and empirical insights**. *Science Talks* 2024, **9**:100307.

** **Presentation raising awareness of the factors that influence online data quality.**22. Loshin D: **3 - Data quality in practice**. In *Enterprise Knowledge Management*. Edited by Loshin D. Academic Press; 2001:47–72.

23. Wang RY, Strong DM: **Beyond Accuracy: What Data Quality Means to Data Consumers**. *Journal of Management Information Systems* 1996, **12**:5–33.

24. Meier R, Gwerder N: **Dirty Data in Online Surveys. How to improve data quality before and after the field phase**. 2022,

25. Newman A, Bavik YL, Mount M, Shao B: **Data Collection via Online Platforms: Challenges and Recommendations for Future Research**. *Applied Psychology* 2021, **70**:1380–1402.

26. Cichy C, Rass S: **An Overview of Data Quality Frameworks**. *IEEE Access* 2019, **7**:24634–24648.

27. Ridzuan F, Zainon WMNW: **A Review on Data Quality Dimensions for Big Data**. *Procedia Computer Science* 2024, **234**:341–348.

28. Groves RM: **Research on Survey Data Quality**. *The Public Opinion Quarterly* 1987, **51**:S156–S172.

29. Cuskley C, Sulik J: **The Burden for High-Quality Online Data Collection Lies With Researchers, Not Recruitment Platforms.** *Perspect Psychol Sci* 2024, **19**:891–899.

** **Discussion about the link between high-data quality, online data collection platform and researchers.**

30. Otterbring T, Folwarczny M: **From WEIRD to worldwide: Auditing authors' affiliation countries over time across three leading journals publishing food-related research**. *Food Quality and Preference* 2024, **117**:105175.

31. Douglas BD, Ewell PJ, Brauer M: **Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA**. *PLOS ONE* 2023, **18**:e0279720.

* **Research that compare the quality of the data collected through five commonly used platforms.**

32. Miguel E, Camerer C, Casey K, Cohen J, Esterling KM, Gerber A, Glennerster R, Green DP, Humphreys M, Imbens G, et al.: **Promoting Transparency in Social Science Research**. *Science* 2014, **343**:30–31.

33. Jerit J, Barabas J: **Are Nonprobability Surveys Fit for Purpose?** *Public Opinion Quarterly* 2023, **87**:816–840.

FIGURE 1. Suggestions for Sensory & Consumer Science professionals on explicitly and transparently integrating data quality considerations into online consumer research. Grouped by the stages of research – before, during and after data collection.

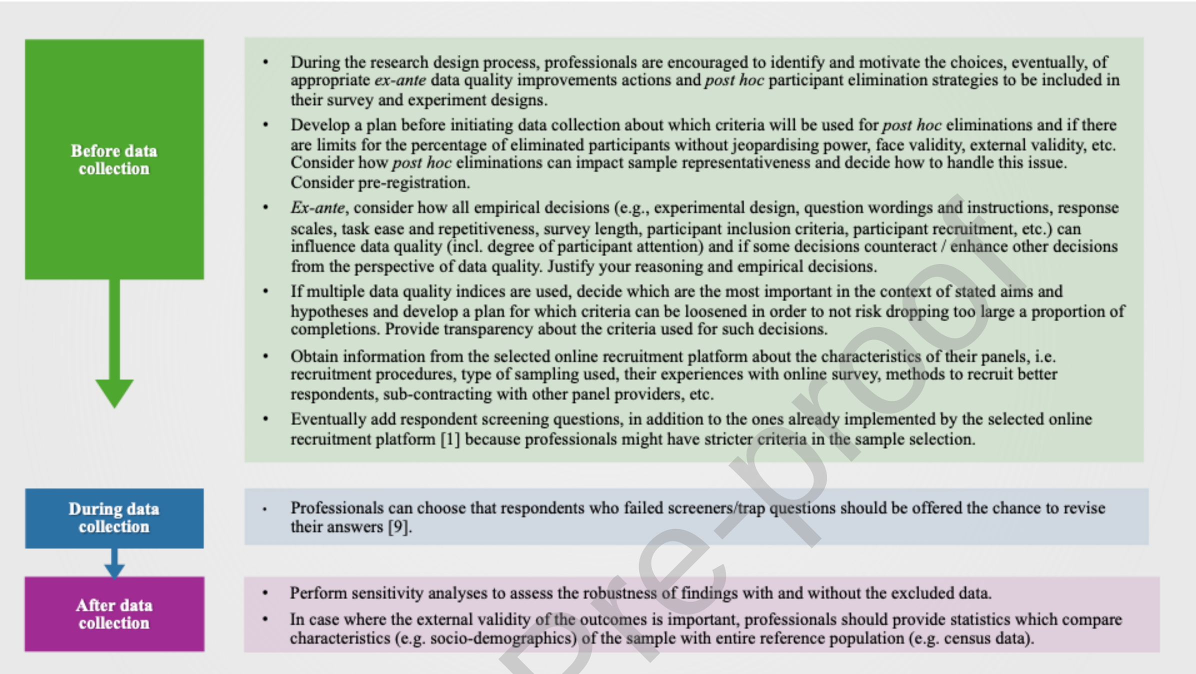

TABLE 1. List of 16 factors affecting data quality in online consumer surveys and experiments grouped into three categories. Based on Jaeger and Cardello [2].

FACTORS AFFECTING DATA QUALITY BY CATEGORY
Study design and administration
1. Mode of administration and electronic device used
2. Completion incentives
3. Respondent interest in the topic
4. Study length
5. Question difficulty and clarity of instructions
6. Question response modes and scales
Respondent demographic and psychographic characteristics
7. Respondent identity: human vs robot
8. Age and gender
9. Respondent education and cognitive ability
10. Respondent familiarity/literacy with data collection devices
11. Respondent survey/experiment-taking experience
12. Respondent personality characteristics

Distractions, carelessness and maladaptive attitudes and behaviours
13. Distractions in the survey/experiment-taking environment
14. Respondent attitudes toward surveys/experiments
15. Careless responding
16. Satisficing response behaviour

TABLE 2. List of 10 recommendations regarding study information and data quality metrics to provide when conducting surveys and experiments online and reporting the findings in scientific journals. Ordering does not indicate relative importance. Based on Jaeger and Cardello [2].

DESCRIPTIONS OF THE RECOMMENDATIONS
1. State software platform for questionnaire administration (incl. any ISO certification) and database for participant recruitment. Explain incentive structure, if any.
2. Describe the questionnaire-taking experience of the participants (e.g., average number of questionnaires per month), if applicable.
3. Calculate rates of completion, drop-out, and post-hoc elimination, ideally relative to participants invited.
4. Report proportion of skipped or missing responses by question if applicable. Indicate when forced answers were a requirement to proceed to next question in the survey/experiment.
5. Report task and/or survey completion times. If eliminating speedsters, state criteria for doing so and percent eliminated.
6. Describe protocols used to identify and eliminate non-human (bot) responses. Include percent eliminated, if available.
7. Describe post-hoc measures used to identify careless or inattentive participants (e.g., straight-line or random responders). Give details of how these are used to eliminate participants and give percentages.
8. Describe ex-ante measures used to assess data quality (e.g., trap questions, instructed manipulation checks, cheap talk scripts). Give details of how these are used to eliminate participants and give percentages.
9. Describe measures used to assess participant engagement, topic interest or questionnaire enjoyment or satisfaction. Give details of how these are used to eliminate participants and give percentages.
10. Describe proactive or embedded manipulations to improve attention or engagement. Give details of how these are used to eliminate participants and give percentages.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: