Bad data causes bad decisions:

How to solve the problem of data quality in your online research

4/5/22

Prolific

Meet the Prolific team

**Phelim Bradley**
Founder & CEO

Prolific

What is 'data quality'?

[1] Data Management Association (2021)
[2] Behrend et al. (2011)
[3] Peer et al. (2017)

- Good quality data are **data that are fit for their intended purpose**

- You need to decide which factors you put most weight on depending on your specific research goals
  - e.g., Attrition, Comprehension, Naivety, Reliability, Replicability, Speed, Thoroughness [1]

- Significant concerns persist about the quality of data from online samples [2,3]

Prolific
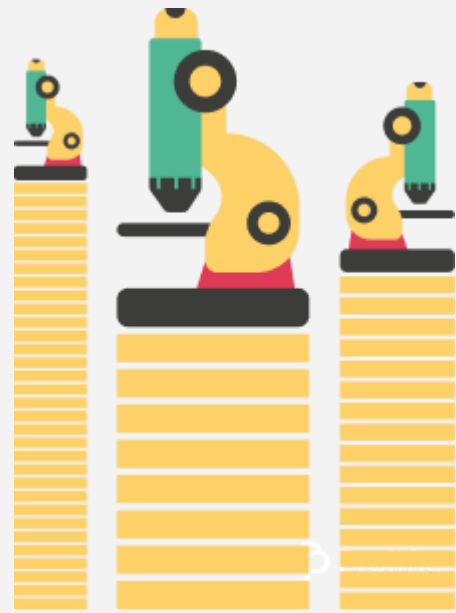
Why does it matter?

[4] IBM (2016)
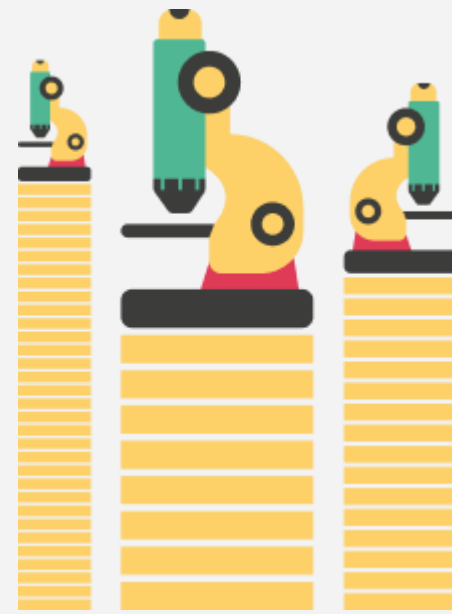[5] Gartner research (2018)

- Impact on your bottom line:

  The annual cost of data quality issues in the U.S. amounts to **$3.1 trillion** [4]

  Average financial impact of poor data quality on organizations is **$9.7M** per year [5]

- **Garbage In, Garbage Out (GIGO):** Low quality data creates a false view of the world

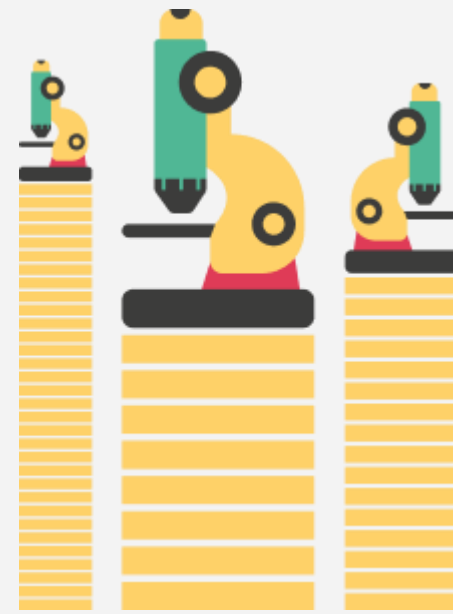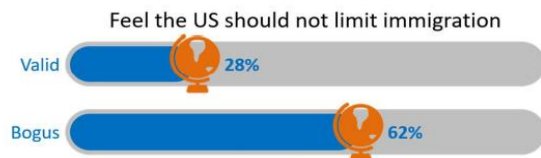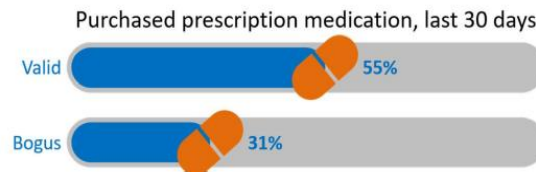 False positives (Type I errors)

  False negatives (Type II errors)

Prolific

Type I vs Type II

- **Type I**: Seeing relationships where none truly exist

  Risk of directing resources towards a dead-end

- **Type II**: Missing significant relationships that **do** exist

  Risk of ignoring potential opportunities

Brand awareness for Charity Navigator

Valid $ 15%

Bogus $ 58%

Purchased prescription medication, last 30 days

Valid 55%

Bogus 31%

Feel the US should not limit immigration

Valid 28%

Bogus 62%

Prolific

Where does low data quality come from?

- Low quality data is often due to low quality inputs

- **Who** are you testing?

  Up to 46% of all opt-in panel respondents are bogus [6,7]

- **How** are you testing them?

  95% of researchers eliminate speeders and 90% use CAPTCHAs to filter out survey bots [6] - but that's it!

Prolific

How to improve your data quality

1. Pick the **platform** with the strongest pool

1. **Validate** your participants yourself

1. Choose the **right sample** to answer your question

1. Design your study to **eliminate bad actors**

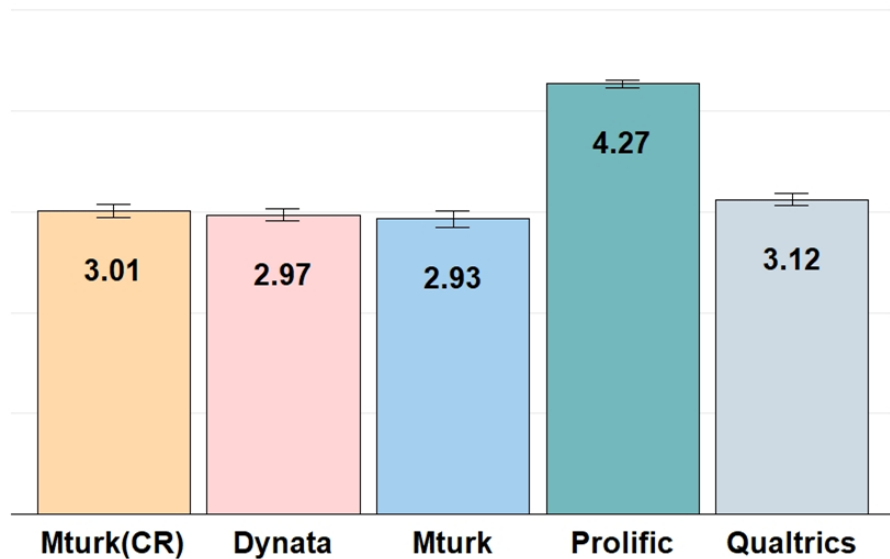1. Treat your participants **fairly**

Prolific

Pick the platform with the strongest pool

- Not all research platforms were created equal. Things to check are:
  - Participant **verification**
  - Level of **quality control** on the pool
  - **Amount of data** collected about participants
  - **Active size** of the pool

- Underlying 'pool' quality has been shown to vary greatly between the major providers, meaning that overall data quality also differs significantly
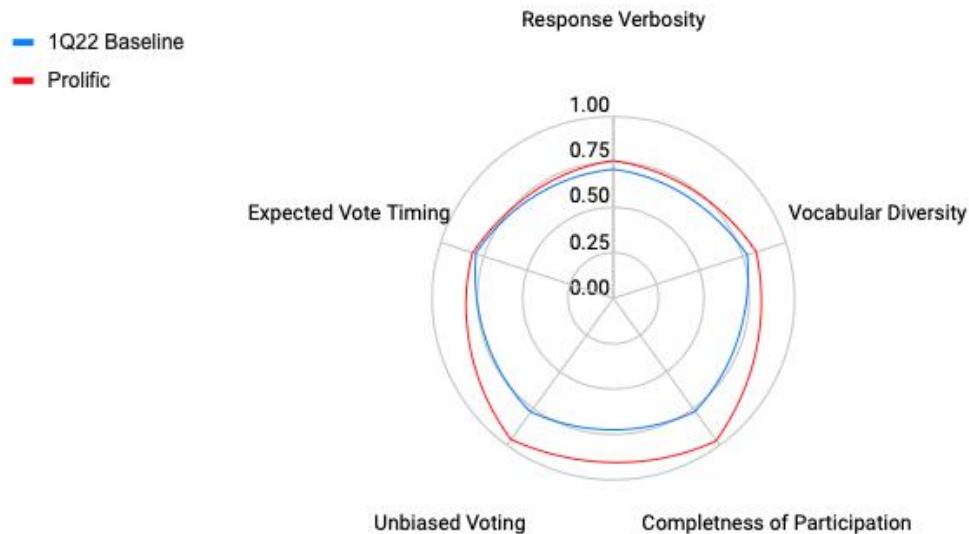
Prolific

Pick the platform with the strongest pool

Pick the platform with the strongest pool

Response Verbosity

- 1Q22 Baseline
- Prolific

1.00
0.75
0.50
0.25
0.00

Expected Vote Timing

Vocabular Diversity

Unbiased Voting

Completness of Participation

Prolific

remesh

Validate your participants yourself

- Within your study, check that your participants **are who they say they are**

- **Re-ask your screening requirements**. This will allow you to confirm that participants' prescreening responses are accurate and up-to-date

- **Never reveal your required demographics** in your task description

Prolific

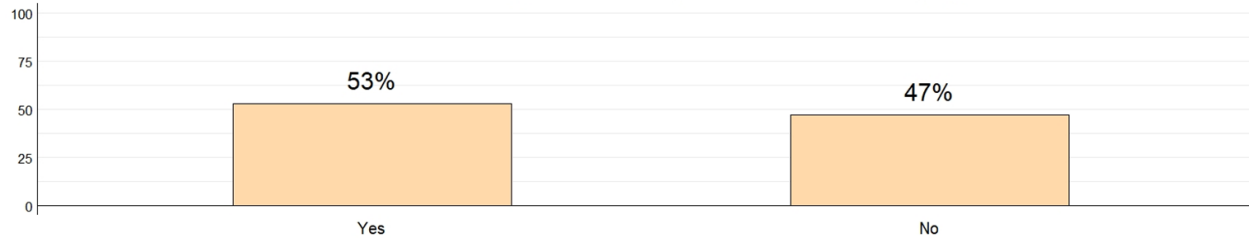# Choose the right sample to answer your question

- Not all samples can answer all research questions

- Your choice of respondents will have a huge impact on the quality of the data you receive - **What is the aim of your research?**

- Generalisability vs nicheness - **which one is more important to you**?

Prolific

Choose the right sample to answer your question

Has the COVID-19 pandemic made you feel less safe when flying?

Prolific

# Choose the right sample to answer your question

Prolific
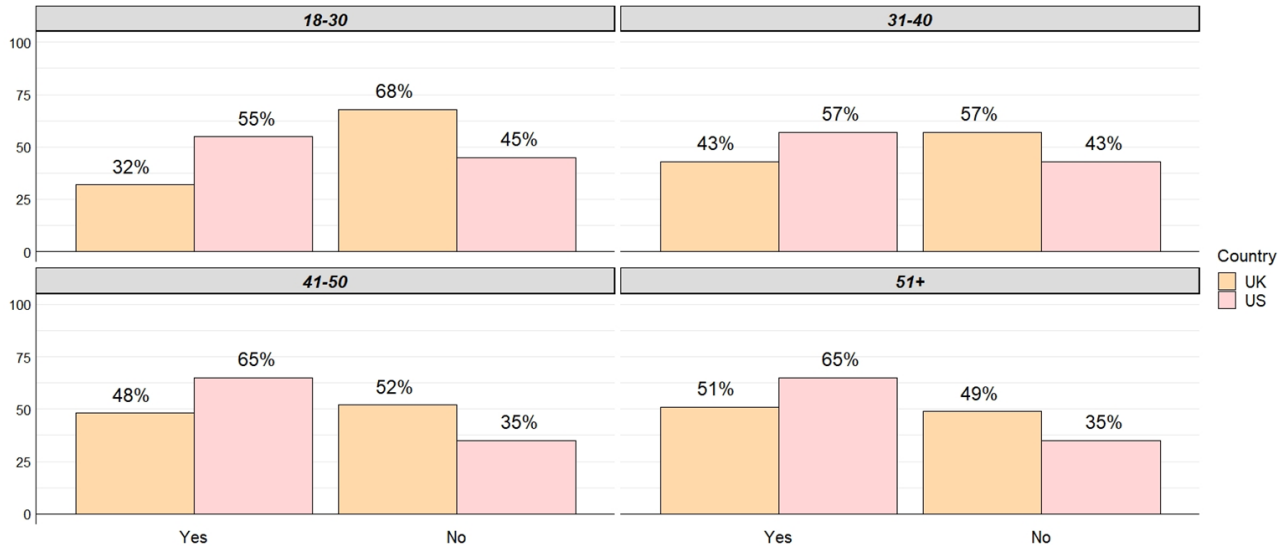
Design your study to eliminate bad actors

- Include free-text responses

- Include CAPTCHAs

- Include open-ended and duplicate questions

- Check your data for random answering patterns or careless responding

- Implement effective attention checks

Prolific

Treat your participants fairly

- Participants who are treated fairly are significantly more likely to engage with your research

- One of the best ways to do this is by paying them a fair rate for their time

- Treat your survey takers as people not machines

- **Try to sign up as a participant to your own research!**

Prolific
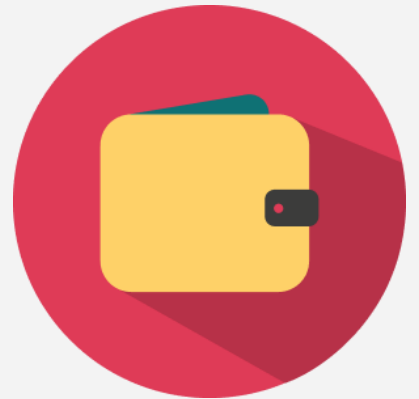
Summary

- Data quality is a function of the **participants** and your **study design.**

- Be skeptical about your sources!

- Add free text responses, attention checks, repeated demographic questions to your surveys.

- Empathise with your participants – try to take some studies yourself!

Prolific

Thanks for listening! Any questions?

- Come and see us at stall 704 to learn about how Prolific is solving the data quality issue

- You could also win £1500 in representative sample credits, or the services of our in-house research consultant for two projects simply by grabbing a bag of pick n mix!

Learn more:

www.prolific.co

@Prolific
@Phelimb

Prolific