



University of
Nottingham

UK | CHINA | MALAYSIA

A large, high-resolution image of the Earth as seen from space, showing the continents and oceans. The Earth is centered in the frame, with a white rectangular border around it. The background is a dark, starry space.

Fraudulent participants and issues with online studies

Blandine French



Agenda

- Personal experience
- Challenges in online studies
- What to look out for
- Tips and strategies
- Things to keep in mind
- UoN Guidelines



How it started



Table 1. University of Nottingham case studies.

Case study	Methodology used	Category of issues experienced	Participants impacted, n (%)	Impact on study
1	Qualitative: interviews	Nongenuine Participants Repeat responders	8/12 (66.67) were nongenuine data.	Loss of funds, inability to finish project or publish, time wasted by PI ^a and researcher to identify data.
2	Randomized controlled trial	Repeat responders	100/1123 (8.9) accounts were suspended due to repeat registration.	Human resources to develop and administer repeat registration protocol and suspend accounts suspected of repeat registration. Additional burden and distress for participants who repeatedly registered in error.
3	Randomized controlled trial	Nongenuine participants Repeat responders	Approximately 482/483 (99.79) were suspected as nongenuine participants.	Impact on study researchers in terms of increased workload but it did not impact on recruitment figures.
4	Quantitative: survey	Nongenuine participants Repeat responders	349/391 (89.26) were suspected as nongenuine participants.	The process of cleaning the data was difficult in decision-making, time-consuming, and led to delays in completing the project. Reluctance to widely advertise afterward.
5	Qualitative: interviews	Nongenuine participants	54/54 (100) were suspected as nongenuine; 7 were identified as nongenuine. No genuine participants were recruited.	Loss of time, had to widen recruitment to the general population to deliver the project.
6	Qualitative: focus groups	Nongenuine participants; repeat responders	Approximately 115 suspected nongenuine participants applied. 83% of selected participants were nongenuine.	Loss of funds, loss of time, and loss of data integrity.
7	Quantitative: administration of online task	Nongenuine participants	20/31 (64.52) nonautistic participants consented to the study were nongenuine.	Loss of comparison group, unable to fully deliver funder expectations, time, and stress.



Publications and presentations

French, B., Babbage, C., Bird, K., Marsh, L., Pelton, M., Patel, S., & Rennick-Egglestone, S. (2024). Data Integrity Issues With Web-Based Studies: An Institutional Example of a Widespread Challenge. *JMIR Mental Health*, 11(1), e58432.



French B., Babbage, C., Rennick-Egglestone S. & Cassidy, S. (2024). Misrepresentation by online study participants – a threat to data integrity. *The Lancet Psychiatry*





The impact

- Days/weeks wasted
- Whole projects cancelled
- Funds wasted
- Potential harm to other participants
- Unreliable data
- Lost of trust from participants
- Impact on researchers



Ysidron et al. (2022)

- Upload an image of medication
- A comparison of the proportions of images deemed authentic differed as a function of recruitment platform, $\chi^2 (2, N = 307) = 33.2, p < .001$, with the percentage of authentic images being:
 - 37.8% for Mturk
 - 69.2% for Prolific
 - 73.6% for Research Match.

	Authentic (n=157)	Inauthentic (n=150)
	M (SD)	M (SD)
Attention Check	2.9 (0.3)	* 2.6 (0.8)
Mental Health Quality of Life	13.8 (3.4)	13.4 (3.1)
Physical Health Quality of Life	13.9 (2.7)	* 12.6 (2.5)
Pain Intensity Rating	2.1 (2.0)	* 5.0 (3.0)
Diabetes Distress	2.4 (0.9)	* 3.3 (1.2)
Diabetes Fatalism	28.9 (8.4)	* 36.9 (10.8)
Hemoglobin A1C	7.1 (1.1)	6.9 (1.6)
Comorbidities	0.5 (0.8)	* 1.2 (1.2)

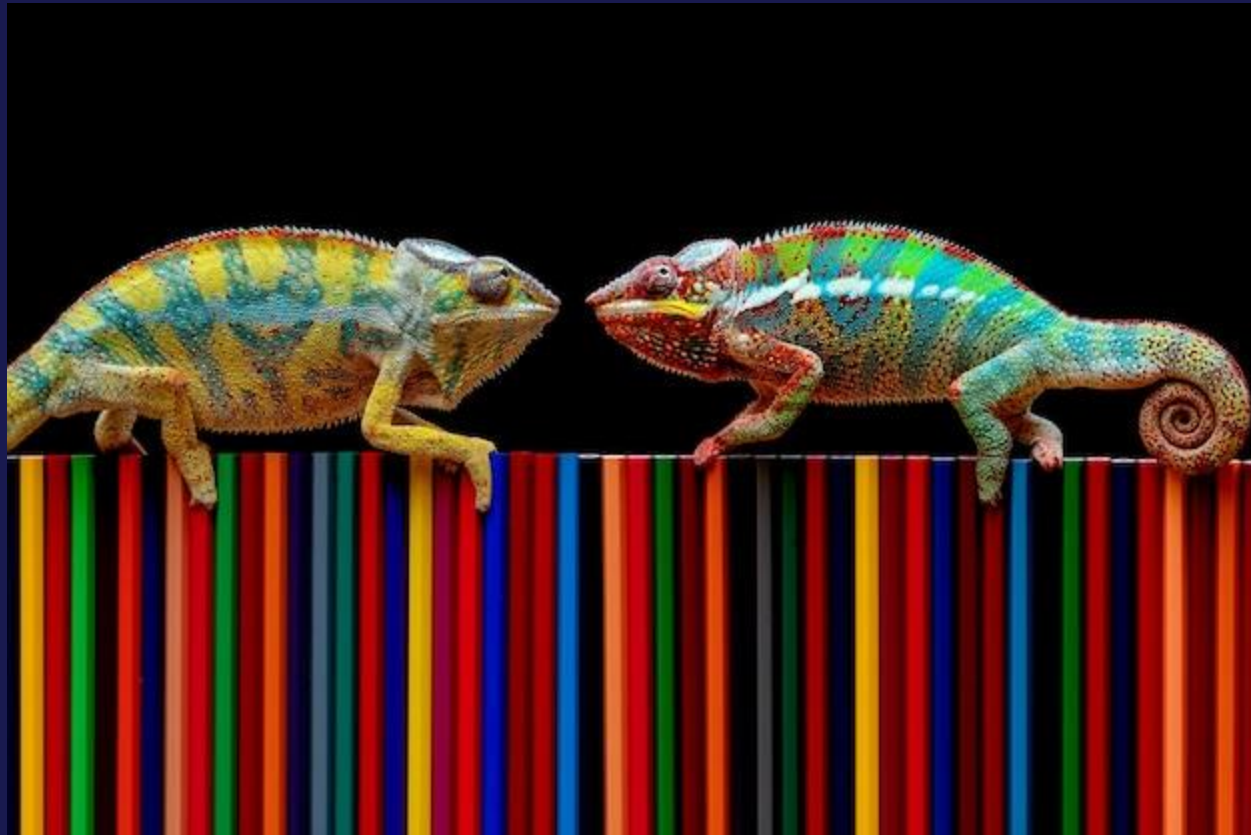
*

Significant difference between groups.



Different population, different problems

Health studies vs general population





Challenges

Issues

- Participants faking their lived-experience
- Participants taking part more than once
- Misrepresentation (exaggeration)
- Lack of engagement
- Bots





How bad is it?



RCTs

Surveys

Qualitative

Focus groups

Interviews

Clinical
recruitment

PPIE groups

Most online
studies





Salinas et al. (2022) – n=738

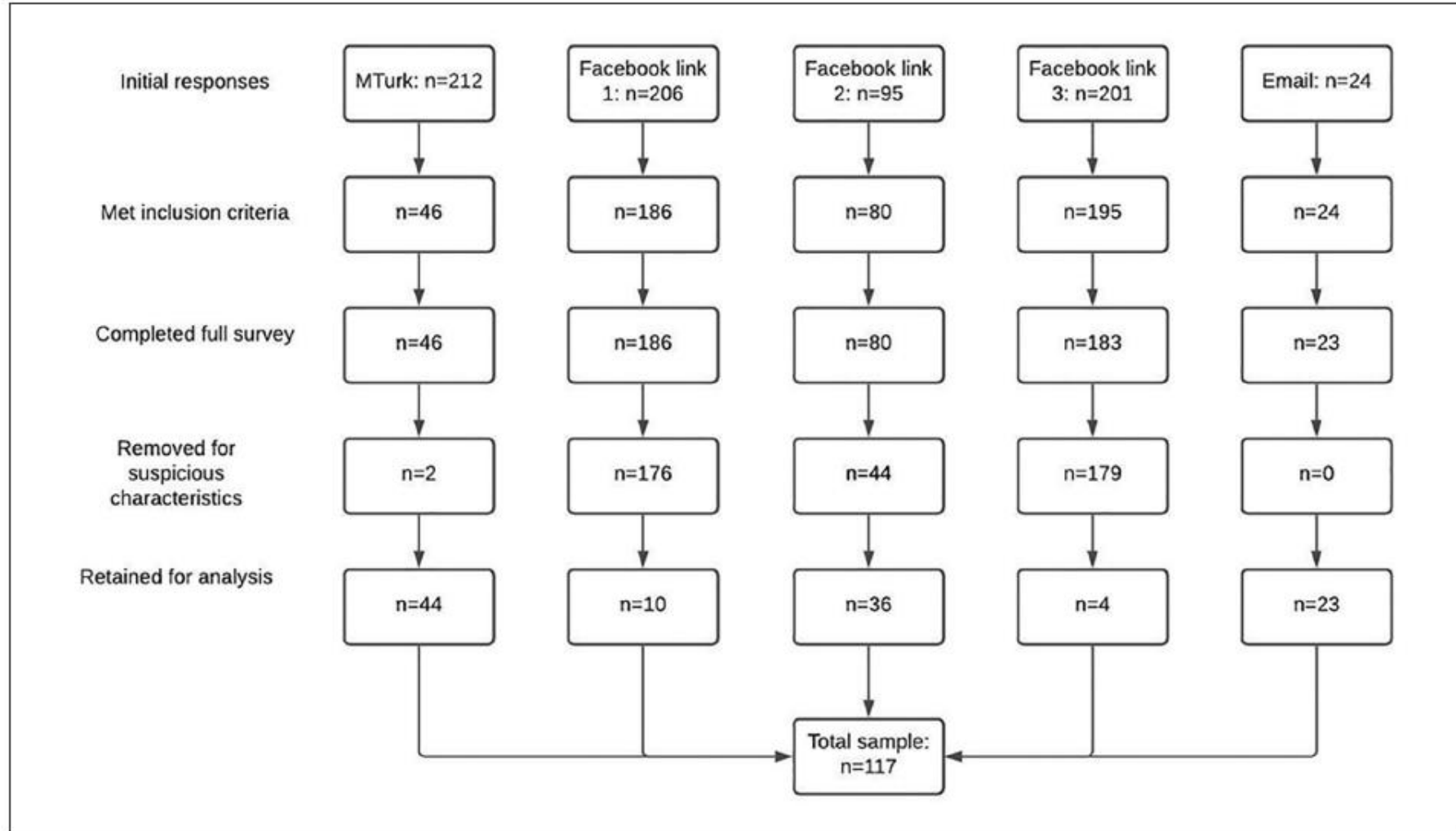


Figure 1. Flow Diagram of Participant Responses.



Bowen et al. (2008) n=1900

Table 2 Categorization of repeat responders by number of submissions and number of identical detection variables

Participant typology	Number of submissions	Percent of all multiple submissions (<i>N</i> = 627)	Number of all duplicate responders (<i>N</i> = 132)	Percent identical IP, %	Percent personal identifiers duplicated				
					0 ^a , %	1 ^a , %	2 ^a , %	3 ^a , %	4 ^a , %
Infrequent ^b	2–5	36	86.3	68.4	20.2	36.8	25.4	16.7	.9
Persistent ^b	6–10	13	7.7	80.0	50.0	10.0	20.0	20.0	0
Repeater ^b	11–30	17	3.9	100	40.0	0	0	20.0	40.0
Hackers ^b	45–67	34	2.4	100	66.7	0	0	33.3	0

^a Sum of identical detection variables (password, username, telephone, e-mail) after IP address within each fraudulent account. For example, “0” means none of the personal identifiers matched and “2” means there were two personal identifier that were the same within a multiple submission

^b Rows represent independent analyses for each participant typology



Examples

Qualitative

- Youth with HIV (Flicker et al., 2004)
 - One boy story did not fit
- Fronek and Briggs (2018)
 - Purposeful interference from competing interest organisation
- Roehl and Harland (2022)
 - 5 out of 14 participants

Quantitative

- Chandler et al. (2017, 2020)
 - “Trusted” Mturk participants misrepresenting their experiences 8-17%
- Conrique et al. (2020), students survey
 - short response time, straight-line responses, repeat entries
 - 25% of responses
- Ballard et al. (2019),
 - 28% as fraud, 10% as potential
- Dewitt et al. (2018)
 - 60% of responses were spam



What to look out for?





What to look for?

- Email configuration: Name.surname111@gmail.com
- Mismatched details: Different name/D.o.b and details registered
- Same IP address
- Cite recruitment mentioned not use
- Quick response to adverts
- Concerns about payment
- Short emails, short/empty free text responses
- Decline giving professional details
- Different country than disclosed
- Inconsistencies
- Unexpected responses



What to look for?

Interviews

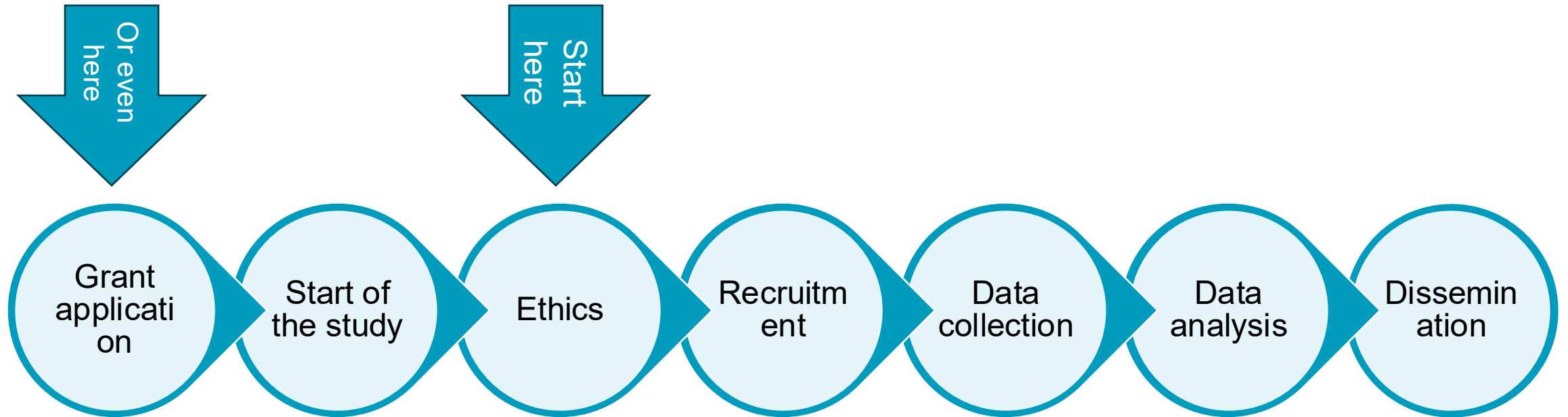
- Same stories
- Won't put camera on
- Vague answers
- Can't elaborate through prompts
- Reticence to telephone interviews
- Responses "don't feel right"
- Similarity in stories but also in participants voices
- Much shorter time

Survey

- Straight line answers
- Short response times
- Contradicting responses
- Empty free text boxes
- High level of non-response



Timeline



What to do at each stages?



Ballard et al. (2019)

Table 1. The number of surveys that violated the fraud detection measure by category.

Fraud detection measure	Total (n=414), n (%)	Missing, n (%)	Fraud, (n=119) ^a , n (%)	Potential fraud, (n=42) ^a , n (%)	Valid (n=253) ^a , n (%)
Geolocation outside of study area based on IP ^b address and unable to be confirmed through mailing address or participating at a local event	164 (39.6)	0 (0.0)	109 (91.6)	34 (81)	21 (8)
Phone number was a local business or organization phone number	94 (22.7)	93 (22.5) ^c	94 (79.0)	0 (0.0)	0 (0.0)
Mismatching names within entry	43 (10.4)	71 (17.1) ^c	37 (31.1)	2 (4.8)	4 (1.6)
Unusual email address	37 (8.9)	98 (23.7) ^c	34 (28.6)	1 (2.4)	2 (0.8)
First and last name AND one or more other personal items match other previous entry	13 (3.1)	16 (3.9) ^c	13 (10.9)	0 (0.0)	0 (0.0)
Two or more personal items match other previous entry	6 (1.4)	67 (16.2) ^c	3 (2.5)	3 (7.1)	0 (0.0)
Date of birth AND one or more other personal items match other previous entry	5 (1.2)	5 (1.2) ^c	5 (4.2)	0 (0.0)	0 (0.0)
Survey duration <5 minutes	3 (0.7)	0 (0.0)	0 (0.0)	2 (4.8)	1 (0.4)

^aSample size represents final categorizations (ie, reclassification of some that were initially classified as potential fraud as valid or fraudulent based on the verification with participants).

^bIP: internet protocol.

^cParticipants were not required to give personal information (ie, phone number, name, email address) if they did not want to be contacted about future research opportunities, declined their incentive, or if they did not complete the section of the survey describing the referral process. If a personal item was missing needed for a measure, the variable was considered missing.



Strategies





But you have to plan ahead!





Papers with strategies

Surveys

- Kramer, J., Rubin, A., Coster, W., Helmuth, E., Hermos, J., Rosenbloom, D., ... & Lachowicz, M. (2014). Strategies to address participant misrepresentation for eligibility in Web-based research. *International journal of methods in psychiatric research*, 23(1), 120-129.
- Teitcher, J. E., Bockting, W. O., Bauermeister, J. A., Hofer, C. J., Miner, M. H., & Klitzman, R. L. (2015). Detecting, preventing, and responding to “fraudsters” in internet research: ethics and tradeoffs. *Journal of Law, Medicine & Ethics*, 43(1), 116-133.
- Lawlor, J., Thomas, C., Guhin, A. T., Kenyon, K., Lerner, M. D., Ucas Consortium, & Drahota, A. (2021). Suspicious and fraudulent online survey participation: **Introducing the REAL framework**. *Methodological Innovations*, 14(3), 20597991211050467.

Interviews

- Pellicano, E., Adams, D., Crane, L., Hollingue, C., Allen, C., Almendinger, K., ... & Wheeley, E. (2023). A possible threat to data integrity for online qualitative autism research. *Autism*, 13623613231174543.



Key strategies

- All stages – pre, post and during
- Do a protocol
- Include steps to be taken for fraudulent participants in PIS
- Single use link
- Have inclusion/exclusion criteria
 - Recruitment stage
 - Data collection
 - Data analysis
- Do not advertise on social media
- Video calls before remuneration/participation
- Consistently check
 - Recruitment stage
 - Data collection
 - Data analysis

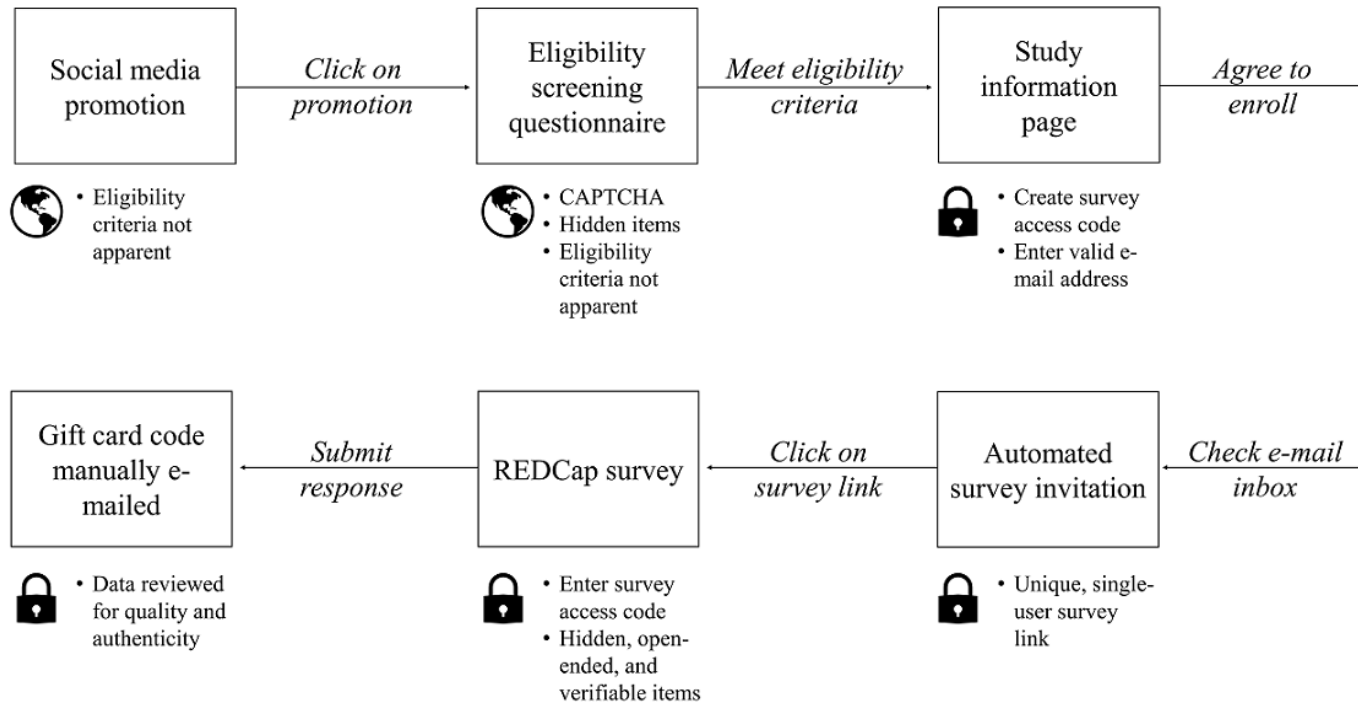


To keep in mind





Example (Pozzar, 2020)



- Social media advertisement:
 - 7h on twitter
 - 576 response
 - 271 eligible
 - 94.5% were fraudulent



Nothing is full proof

- 1650 started the survey
- 1124 passed screening
- 851 consented
- 744 participants

- But 25 from same respondent
 - Phone number formatting
 - “Handwriting”
 - Gmail accounts
 - Completion time



Additionally

- Technology evolves all the time
 - More clever way of tricking, bypassing strategies
- Each tool/strategies will have pros and cons
- Not friendly to all groups
 - Especially groups with lived experiences of certain conditions
- Country specific?
- How to report (or not)?
 - It's own section?
- How to recover if loss of funds
- What do you do if you haven't planned for it or if it happens?



Moving forward

UoN guidelines





University of Nottingham working group

- UoN guidelines
- Working group from 01/25-06/25
- Endorsed by UKRIO
- Currently being reviewed by ethics committee

Blandine French, Stefan Rennick-Egglestone, Camilla Babbage, Sarah Cassidy, Ali Alshukry, Daniel Simpkins
Jessica Jackson, James Tangen
UKRIO: Josephine Woodhams, James Parry

Non-genuine Participation in Online Research
Guidance on Handling Potential Non-Genuine Participants in Online Research

Summary	2
Objectives of the Guidelines	2
Background	2
Research Timeline	3
Potential solutions	3
Before the research starts/planning/grant applications	3
During the research/data collection	4
After the data collection	4
Caution	4
Key message	5
Consideration for Research Ethics Committees	5
Ensuring Data Integrity	5
Transparent Documentation	6
Definitions	6
Annex 1- Working group on non-genuine participation in online studies	7
Annex 2- Lessons learnt and tips	7
Annex 3- Examples	8
Example A (Survey)	9
Example B (Intervention with interviews and focus groups)	10
General solutions applicable to both examples	10
Annex 4 -Resources and Further Reading for Strategies	11
.....	12



Content

1. Summary
 2. Objectives of the Guidelines
 3. Background
 4. Research Timeline
 5. Potential solutions
 6. Before the research starts/planning/grant applications
 7. During the research/data collection
 8. After the data collection
 9. Caution
 10. Key message
 - 11. Consideration for Research Ethics Committees**
 - Transparent Documentation
 - Ensuring Data Integrity
- Annex 1- Working group on non-genuine participation in online studies
 - Annex 2- Lessons learnt and tips
 - Annex 3- Examples
 - Example A (Survey)
 - Example B (Intervention with interviews and focus groups)
 - General solutions applicable to both examples
 - Annex 4 -Resources and Further Reading for Strategies



Consideration for Research Ethics Committee

- Ethics Committees have a responsibility to support both researchers and research participants. They must **consider the ethical implications of disruptions** caused by non-genuine participants and balance these with the impact of additional countermeasures on genuine participants.
- Ethics Committees should consider the **implementation of tailored countermeasures** for specific target participant groups whilst ensuring the integrity of the research as an ethical implication.
- This includes considering approval for:
 - **The collection of personal details**, such as date of birth or emails for the purpose of verifications
 - Instances where **non-reimbursement might be acceptable** (for example, providing proof that one person responded multiple times)
 - Ensuring the integrity of the research data through **implementing extra steps to participation**, protecting public funds and where appropriate suggest alternative acceptable approaches.



Ensuring Data Integrity

- Ethics committees will want to ensure that any verification process will not exploit participants. Any methods used to verify participant identities should respect participants' privacy.
- **Validating Data Without Distorting the Experience:** Researchers need to justify how their approach is necessary and proportionate for ensuring high-quality data without distorting the research process.
- For example, if the process involves additional validation, they should ensure:
 - **Proportionality:** Ensure methods are proportional to the involvement and incentive proposed by the study. Ensure methods are proportional to the sensitivity of the data being collected.
 - **Bias Consideration:** Consider potential biases introduced by validation methods. Develop solutions to mitigate these biases.
 - **Building Trust:** Build trust in the process for participants who are genuine.
 - **Participant Experience:** Ensure the experience of genuine participants is not negatively impacted.



Transparent Documentation

- The researcher must provide a clear, detailed explanation of how they plan to authenticate participants. The ethics committee will assess whether the study design includes transparency in data handling and participant recruitment.
- Transparency in reporting in publications:
 - **Pre-defined criteria:** If exclusion criteria are pre-determined, researchers should clearly state these criteria in their study protocol and report on the number of participants excluded based on these criteria.
 - **Post-hoc exclusions:** If exclusions occur after recruitment, the reasons for these exclusions should be clearly and comprehensively reported, including the number of participants excluded for each reason.
 - **Justification for exclusions:** Researchers should provide a justifiable explanation for each exclusion, whether it relates to safety concerns, participant characteristics, data quality issues, or other relevant factor





Blandine French

Twitter: [Blandine.french](#) 

Blandine.french@Nottingham.ac.uk 

www.ndlab.org.uk 