SURVEYS · NONRESPONSE, WEIGHTING AND DATA QUALITY

# Fraud Detection in Online Panels: Duplicate and Fake Respondents

Online panels and open survey links attract people who take surveys more than once, pretend to qualify or use automated tools. This article describes the main forms of survey fraud, how it affects results and the layered checks that panel providers and researchers use to detect and prevent it.

ModQu research article · 2,150 words · 22 references (APA 7th edition)

Online panels have become one of the main sources of survey data for market research, opinion polling and academic studies. Panel members are invited to complete surveys in return for points, cash or other rewards. The same rewards that make panels work also attract fraud. Some people register several times under different identities. Some claim characteristics they do not have in order to qualify for better-paid surveys. Some use software to complete surveys automatically. Others take part in so many surveys that they learn how to pass screening questions whatever their real circumstances (Hillygus et al., 2014; Teitcher et al., 2015).

Fraudulent responses do more than add noise. Because fraudsters tend to answer in particular ways, such as claiming to own many products or agreeing with whatever is asked, they can bias estimates systematically. Kennedy, Hatley et al. (2020) found that bogus respondents in online opt-in polls tended to give positive answers to almost any question and were concentrated among people claiming to be young or Hispanic, which distorted estimates for those groups. This article describes the main forms of fraud in online panels, their effects and the methods used to detect and prevent them.

## Why online panels are vulnerable

Several features of online research make fraud attractive and possible. Respondents are not seen or heard, so identity is hard to confirm. Rewards are paid per completed survey, so completing more surveys earns more. Screening questions decide who qualifies for a survey, and surveys for rare or valuable groups often pay more, which creates an incentive to claim membership of those groups. Many panels recruit members through online advertising and affiliate websites, which can attract people interested mainly in the rewards. And demand for sample is often met by combining panels from several providers, which makes it harder to spot the same person arriving through different routes (Baker et al., 2013; Callegaro et al., 2014).

These features do not make online panels unusable. Most panel members are genuine, and well-managed panels invest heavily in quality control. They do mean that fraud is a constant pressure that must be managed rather than a rare event that can be ignored.

## Forms of fraud

Figure 1 summarises six common forms.

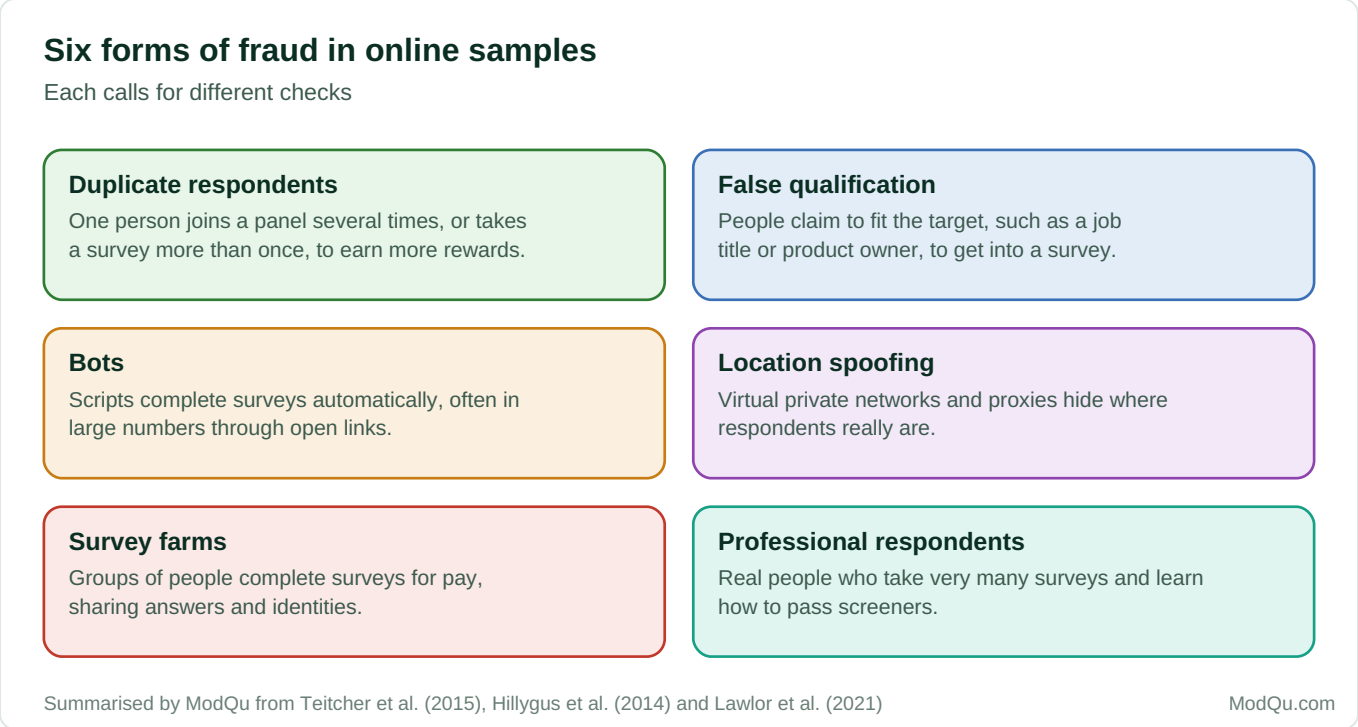



Figure 1. Six forms of fraud in online samples. Summary by ModQu.

**Duplicate respondents** take part more than once. They may hold several panel accounts, join the same survey through different panels or reuse an open survey link. Konstan et al. (2005) described an early case in which one person submitted many responses to an internet health survey, and Bowen et al. (2008) found multiple submissions in another online study and proposed methods for identifying them.

**False qualification** occurs when respondents misrepresent themselves to pass a screener. A survey seeking owners of a particular car, or managers responsible for buying software, offers rewards only to those who qualify, giving others an incentive to claim they do. Respondents who take many surveys learn which answers lead to qualification (Hillygus et al., 2014).

**Bots** are automated programs that complete surveys. They are a particular risk for surveys with open links posted online. Pozzar et al. (2020) reported that most responses to a survey promoted through social media came from bots or other bad actors, and Storozuk et al. (2020) described how bots can overwhelm an unprotected survey within hours.

**Location spoofing** uses virtual private networks or proxy servers to disguise where a respondent is. It allows people outside the target country to take part in surveys restricted to it. Dennis et al. (2020) showed that screening based on internet addresses can be defeated in this way.

**Survey farms** are organised groups that complete surveys for payment, sometimes sharing accounts and answers. Their responses may look like those of different people but show similar patterns.

**Professional respondents** are real people who belong to several panels and take very many surveys. They are not necessarily dishonest, but experience may change how they answer, and some learn to game screeners. Matthijsse et al. (2015) found that professional respondents differed from others in some respects but that their data quality was not consistently worse, while Hillygus et al. (2014) found that frequent survey-takers could differ in attitudes and behaviour.

## Effects on survey knowledge

Fraud affects results in several ways. It inflates the apparent size of rare groups, because people falsely claiming to belong to them are added to genuine members. It biases estimates within groups, because fraudulent members answer differently from genuine ones. Kennedy, Hatley et al. (2020) found that bogus respondents in opt-in samples tended to say "yes" to almost anything, inflating estimates of rare behaviours and producing strange results for small subgroups. Bell and Gift (2023) found that a substantial share of respondents in a nonprobability sample targeting a specific subpopulation were likely fraudulent and that removing them changed substantive conclusions.

Fraud also undermines trust in online research. Chmielewski and Kucker (2020) documented a marked decline in data quality on an online labour platform over a short period, which they attributed partly to fraudulent and low-effort participants, and Kennedy, Clifford et al. (2020) traced quality problems to respondents using tools to hide their location.

## Layers of defence

No single check stops all fraud. Panel providers and researchers therefore use several layers, each catching some fraudulent attempts that earlier layers missed (Lawlor et al., 2021; ESOMAR, 2021). Figure 2 illustrates the idea.

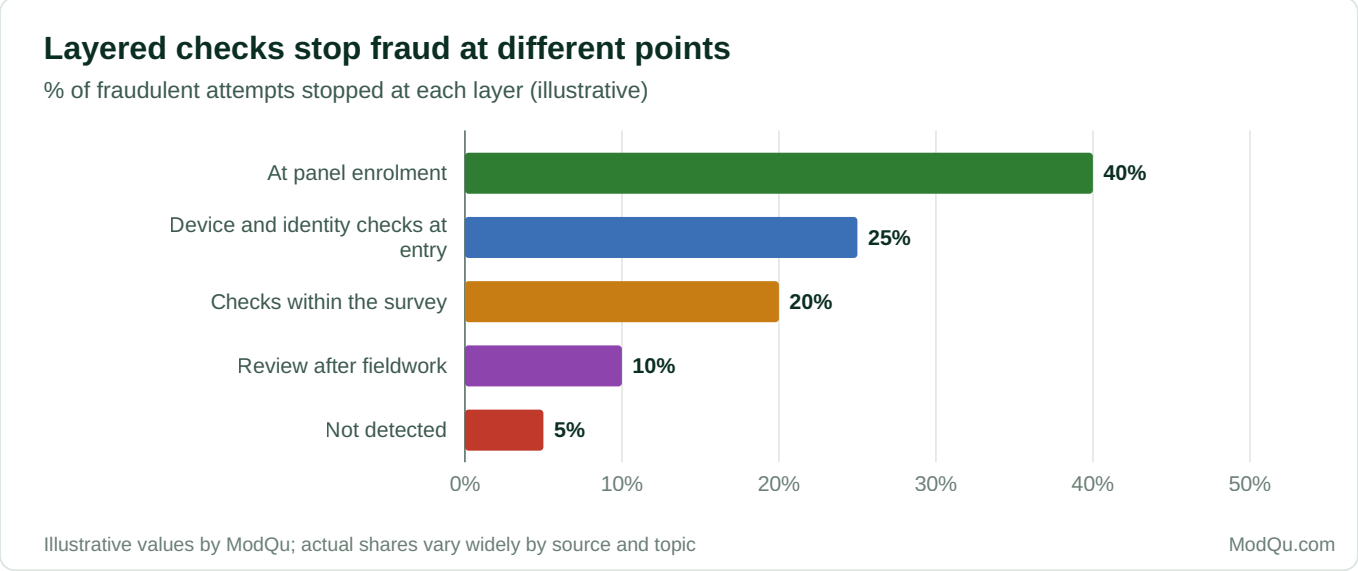



Figure 2. Illustrative shares of fraudulent attempts stopped at different layers of checking. Illustrative values by ModQu.

## At panel enrolment

Panel providers verify new members before admitting them. Common checks include confirming email addresses and telephone numbers, checking postal addresses against address databases, comparing details with existing members to find duplicates and requiring a profile questionnaire whose answers can later be compared with survey answers. Some providers verify identity against third-party records. International standards for market research require providers to describe how they validate panel members and manage duplicates (International Organization for Standardization, 2019; ESOMAR, 2021).

## At survey entry

When a respondent starts a survey, further checks can be applied:

- **Device fingerprinting** records characteristics of the browser and device to identify the same device used under different identities.
- **Internet address checks** identify addresses associated with data centres, proxies or locations outside the target area, though these can be evaded (Dennis et al., 2020).
- **CAPTCHA** and similar challenges distinguish people from programs (Storozuk et al., 2020).
- **Deduplication across sources** identifies respondents who arrive through more than one panel.

## Within the survey

Checks within the questionnaire identify fraudulent or careless respondents who pass earlier layers (Teitcher et al., 2015; Lawlor et al., 2021):

- **Screener design.** Screeners that do not reveal the qualifying answer, that include plausible but false options such as fictitious brands and that ask qualifying questions in more than one way make false qualification harder.
- **Knowledge checks.** People who genuinely belong to a group, such as nurses or software buyers, can answer questions that others cannot.
- **Consistency checks.** Answers can be compared with the panel profile or with earlier answers in the survey, such as age and year of birth.
- **Attention and quality checks**, including instructed response items, timing and straightlining measures (Meade & Craig, 2012).
- **Open-ended answers**, which are hard for bots to fake convincingly and which reveal copied or meaningless text.
- **Hidden fields** that are invisible to people but filled in by automated programs.

## After fieldwork

Before analysis, data are reviewed for patterns that suggest fraud: clusters of responses with similar timing and answers, implausible combinations of characteristics, identical open-ended text and unusually high rates of qualification for rare groups. Leiner (2019) described indicators of "too fast, too straight, too weird" responses that can be combined to identify meaningless data.

## Monitoring quality over time

Fraud changes as defences improve. When providers block one technique, fraudsters adopt another, so checks that worked last year may not work now. Researchers who run surveys repeatedly can track quality indicators across waves: the share of respondents removed, the share failing attention checks, median completion times and the share of fast completions (Zhang & Conrad, 2014; Curran, 2016). A sudden change in any of these, or in the apparent size of a rare group, is a warning sign worth investigating before results are published. Comparing key estimates with external benchmarks, where available, provides a further check that fraud has not shifted results (Ward & Meade, 2023).

## Fraud in business and professional samples

Surveys of business decision-makers, doctors, engineers and other professionals are especially exposed. These respondents are hard to reach, so incentives are high and samples are often drawn from specialised panels or from general panels using screeners. A general panel member who claims to be a purchasing manager may be telling the truth or may have learned that the claim leads to a well-paid survey (Hillygus et al., 2014). Several practices help. Professional status can be verified against external directories or professional registers where permission allows. Knowledge questions that a genuine professional would find easy, but an outsider would not, are effective if they are not too obscure. Asking about the respondent's firm, such as its size and industry, and checking consistency with other answers can reveal invented profiles. Because the number of genuine members of such groups in any panel is limited, a sample that fills unusually quickly is itself a reason for caution.

## Balancing detection and fairness

Fraud checks can wrongly exclude genuine respondents. Households share devices and internet addresses. People travel or use privacy tools for legitimate reasons. Some genuine respondents write short open answers or make honest mistakes on knowledge questions. Overly strict checks can therefore remove real people and bias the sample, often towards those who are older, more educated or more comfortable with technology (Teitcher et al., 2015).

Good practice is to set rules in advance, require several independent signals before excluding someone, review borderline cases, keep records of the reasons for each exclusion and report how many cases were removed. Lawlor et al. (2021) proposed a structured framework for reviewing suspicious participation that emphasises clear criteria and documentation. Ethical considerations also apply: researchers should state in consent materials that responses may be checked for quality and that rewards may be withheld for fraudulent participation, and should treat any personal data collected for verification securely (Teitcher et al., 2015).

## Questions for buyers of online samples

Organisations buying online sample can ask providers how they manage fraud. ESOMAR (2021) publishes a set of questions for this purpose, including:

- How are new panel members recruited and validated?
- How are duplicates identified within and across panels?

- What device, location and identity checks are applied at survey entry?
- How are professional respondents and survey frequency managed?
- What in-survey and post-survey quality checks are used, and how are flagged respondents handled?
- What share of completes is typically removed for quality reasons?

Clear answers to these questions, and willingness to share quality data, are signs of a provider that takes fraud seriously. Buyers can also protect themselves by adding their own checks to every questionnaire rather than relying entirely on the provider, by asking for respondent-level quality flags with the data and by agreeing in advance that cases removed for quality reasons will be replaced. Where several providers are used for one study, asking each to deduplicate against the others, or using an independent deduplication service, reduces the risk that the same people are counted more than once.

## An example

A survey seeks 300 information technology decision-makers in mid-sized firms. The panel provider applies its standard enrolment and entry checks. The screener asks respondents to select their role from a list that includes several genuine and two non-existent job titles, and to identify which of several technology products their firm uses, including one fictitious product. A knowledge question asks respondents to choose the correct definition of a technical term. Of 900 people who start the survey, 280 are screened out as not eligible, 130 select a fictitious option or fail the knowledge question, and 40 are removed after fieldwork because of duplicate devices, identical open-ended answers or completion times far below what reading requires. The final sample of 450 is larger than needed, so the team keeps those with the fewest flags. Comparing results with and without the 40 removed after fieldwork, the team finds that their inclusion would have raised the reported adoption of several new technologies by 6 to 10 percentage points (illustrative example).

## Conclusion

Fraud is a persistent feature of online survey research, driven by the rewards that make panels work. Duplicate respondents, false qualification, bots, location spoofing, survey farms and professional respondents each threaten the accuracy of survey knowledge, often by inflating rare groups and pushing answers in predictable directions. Layered defences, from validation at panel enrolment through entry checks, careful screener design and in-survey checks to review after fieldwork, can catch most fraudulent attempts. Applying these checks with clear rules, attention to fairness and transparent reporting allows researchers to protect their data without excluding genuine respondents.

## References

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J., & Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology, 1*(2), 90–143. https://doi.org/10.1093/jssam/smt008

Bell, A. M., & Gift, T. (2023). Fraud in online surveys: Evidence from a nonprobability, subpopulation sample. *Journal of Experimental Political Science, 10*(1), 148–153. https://doi.org/10.1017/XPS.2022.8

Bowen, A. M., Daniel, C. M., Williams, M. L., & Baird, G. L. (2008). Identifying multiple submissions in internet research: Preserving data integrity. *AIDS and Behavior, 12*(6), 964–973. https://doi.org/10.1007/s10461-007-9352-2

Callegaro, M., Baker, R., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (Eds.). (2014). *Online panel research: A data quality perspective*. Wiley. https://doi.org/10.1002/9781118763520

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science, 11*(4), 464–473. https://doi.org/10.1177/1948550619875149

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Dennis, S. A., Goodson, B. M., & Pearson, C. A. (2020). Online worker fraud and evolving threats to the integrity of MTurk data: A discussion of virtual private servers and the limitations of IP-based screening procedures. *Behavioral Research in Accounting, 32*(1), 119–134. https://doi.org/10.2308/bria-18-044

ESOMAR. (2021). *Questions to help buyers of online samples*. ESOMAR.

Hillygus, D. S., Jackson, N., & Young, M. (2014). Professional respondents in nonprobability online panels. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 219–237). Wiley. https://doi.org/10.1002/9781118763520.ch10

International Organization for Standardization. (2019). *Market, opinion and social research, including insights and data analytics — Vocabulary and service requirements* (ISO Standard No. 20252:2019).

Kennedy, C., Hatley, N., Lau, A., Mercer, A., Keeter, S., Ferno, J., & Asare-Marfo, D. (2020). *Assessing the risks to online polls from bogus respondents*. Pew Research Center.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. G. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods, 8*(4), 614–629. https://doi.org/10.1017/psrm.2020.6

Konstan, J. A., Simon Rosser, B. R., Ross, M. W., Stanton, J., & Edwards, W. M. (2005). The story of subject naught: A cautionary but optimistic tale of internet survey research. *Journal of Computer-Mediated Communication, 10*(2). https://doi.org/10.1111/j.1083-6101.2005.tb00248.x

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). https://doi.org/10.1177/20597991211050467

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods, 13*(3), 229–248. https://doi.org/10.18148/srm/2019.v13i3.7403

Matthijsse, S. M., de Leeuw, E. D., & Hox, J. J. (2015). Internet panels, professional respondents, and data quality. *Methodology, 11*(3), 81–88. https://doi.org/10.1027/1614-2241/a000094

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

Pozzar, R., Hammer, M. J., Underhill-Blazey, M., Wright, A. A., Tulsky, J. A., Hong, F., Gundersen, D. A., & Berry, D. L. (2020). Threats of bots and other bad actors to data quality following research participant recruitment through social media: Cross-sectional questionnaire. *Journal of Medical Internet Research, 22*(10), e23021. https://doi.org/10.2196/23021

Storozuk, A., Ashley, M., Delage, V., & Maloney, E. A. (2020). Got bots? Practical recommendations to protect online survey data from bot attacks. *The Quantitative Methods for Psychology, 16*(5), 472–481. https://doi.org/10.20982/tqmp.16.5.p472

Teitcher, J. E. F., Bockting, W. O., Bauermeister, J. A., Hoefer, 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. https://doi.org/10.1111/jlme.12200

Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology, 74*, 577–596. https://doi.org/10.1146/annurev-psych-040422-045007

Zhang, C., & Conrad, F. G. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods, 8*(2), 127–135. https://doi.org/10.18148/srm/2014.v8i2.5453