Preventing Bots and Fraudulent Data in Web Surveys: A Practical Guide for Researchers

Introduction

As survey research increasingly moves online, researchers face growing challenges from bots, fraudulent respondents, professional survey takers, and duplicate submissions. These threats can compromise data quality, inflate costs, and undermine the validity of research findings.
This guide outlines practical strategies researchers can use before, during, and after data collection to prevent, detect, and report fraudulent survey responses.

---

1. Reduce Risk During Recruitment

Avoid Open Survey Links Whenever Possible

One of the most effective ways to prevent bot activity is to carefully control how survey invitations are distributed.

Recommended approaches:

- Send survey invitations directly to individual email addresses.
- Use personalized survey links when feasible.
- Distribute through known panels or established participant databases.
- Recruit participants through trusted organizations or institutions.

Use caution with:

- Open survey links shared publicly.
- Social media posts visible to the general public.
- Public websites that may be indexed by search engines.

If social media recruitment is necessary, consider restricting distribution to:

- Closed Facebook groups
- Private LinkedIn groups
- Password-protected communities
- Membership-based organizations

---

2. Build Fraud Prevention into Survey Design

Using multiple detection methods is significantly more effective than relying on a single security feature.

a. Captcha Verification

CAPTCHAs help prevent automated programs from accessing surveys.

Best practice:

- Include a CAPTCHA near the beginning of the survey.
- If participants complete a separate incentive or contact information form, include another CAPTCHA on that form as well.

Examples include:

- Google reCAPTCHA
- Qualtrics Bot Detection and CAPTCHA functions

b. Attention Check Questions

Attention checks help identify respondents who are not carefully reading questions.

Example:

Please type the following letters into the box below:
Wzyoti
Do not type any additional letters or characters.

Additional attention checks can be placed later in the survey to confirm participant engagement. However, excessive use should be avoided, as too many attention checks can frustrate legitimate respondents.

c. **Referral Website Validation**

Use a referral website URL in the survey flow to restrict survey access to respondents who arrive through an authorized webpage. This helps prevent unauthorized access by individuals who discover or share the survey link elsewhere.

Benefits include:

- Restricting survey access to participants recruited through approved channels.
- Preventing access through direct links that have been shared publicly.
- Reducing the risk of survey links being distributed on websites, forums, or social media platforms beyond the intended audience.
- Helping protect against automated traffic originating from unauthorized sources.

d. **Consistency Checks**

Ask the same concept in different ways at different points throughout the survey.

Example:

Earlier question:

How old are you?

Later question:

In what year were you born?

Responses can be compared to identify contradictions or implausible answers.

These checks are particularly useful for:

- Demographic variables
- Household composition
- Education level
- Employment status

e. **Current Year Verification**

Including a simple question such as:

What is the current year?

This can help identify automated or inattentive respondents. Although not sufficient on its own, it can be valuable when combined with other indicators.

f. **Honeypot Questions**

A honeypot is a question or field hidden from legitimate respondents but visible to many automated programs. However, since real participants never see the item, any completed response can be treated as suspicious.

Honeypots are particularly useful on:

- Contact information forms
- Incentive registration forms
- Supplemental survey forms

---

3. Utilize Available Security Features

Most modern survey platforms offer built-in tools to reduce fraudulent activity. Researchers should enable security settings whenever available.

Useful features include:

a. **Bot Detection**

Automated systems identify suspicious or non-human behavior patterns.

b. **Fraud Scoring**

Some platforms assign risk scores based on respondent behavior.

c. **Security Scan Monitoring**
   Monitors submissions for characteristics associated with fraudulent activity.

d. **Prevent Multiple Submissions**
   Helps limit repeat participation through:
   - Browser cookies
   - Device fingerprinting
   - Authentication requirements

e. **Prevent Search Engine Indexing**
   Survey links should generally not be indexed by Google or other search engines. This helps reduce exposure to bots and individuals actively searching for surveys.

---

4. Monitor Metadata

Metadata often provides some of the strongest indications of fraudulent activity.
Researchers should review:

a. **IP Addresses**
   Look for:
   - Duplicate IP addresses
   - Large numbers of submissions from the same location
   - Clusters of responses within short time periods
   - IP addresses originating from countries outside the target population or sampling frame
     Researchers should note that some legitimate respondents may share IP addresses (e.g., workplace or university networks), so IPs should not be used as the sole exclusion criterion.

b. **Completion Time**
   Extremely short completion times may indicate:
   - Bots
   - Random responding
   - Survey farms

   Compare completion times against pilot testing results to establish reasonable thresholds.

c. **Response Patterns**
   Watch for:
   - Straight-lining (selecting the same answer repeatedly)
   - Repetitive open-ended responses
   - Nonsensical text
   - Copy-and-paste responses
   - Identical response patterns across multiple cases
   - Ai generated or “too perfect” open end responses

---

5. Use a Multi-Indicator Review Process

No single flag should automatically result in respondent removal. Instead, researchers should assess multiple indicators together.
For example, a response may be considered fraudulent if it exhibits several characteristics simultaneously:

- Failed CAPTCHA
- Failed attention check
- Duplicate IP address

- Extremely short completion time
- Contradictory responses
- Suspicious open-ended responses

Using multiple indicators reduces the risk of incorrectly removing legitimate participants.

---

6. Responding to a Bot Attack

If a bot attack occurs:

Immediately

- Pause data collection if necessary.
- Investigate suspicious submissions.
- Review recent traffic patterns.
- Implement additional security measures.

Before Relaunching

Consider adding:

- Additional CAPTCHAs
- New attention checks
- Additional validation questions
- Invite-only survey access
- Enhanced monitoring procedures

After Relaunching

- Clearly separate valid and fraudulent records.
- Maintain documentation of all actions taken.
- Record the dates affected by fraudulent activity.

---

7. Reporting Bots and Fraudulent Data Transparently

Transparency is essential when reporting any research project affected by bots or fraudulent responses.

Researchers should clearly document:

What Happened

Describe:

- When the attack or suspicious activity occurred.
- How the issue was identified.
- The period potentially affected.

How Many Cases Were Removed

Report:

- Total completed cases
- Number identified as fraudulent
- Number removed from analysis

Detection Criteria

Specify the methods used to identify suspicious records, such as:

- Failed CAPTCHA checks
- Failed attention checks
- Duplicate IP addresses
- Inconsistent responses
- Extremely short completion times
- Honeypot triggers

Corrective Actions
Document any actions taken, including:

- Survey suspension
- Additional screening measures
- Returning to the field
- Additional recruitment efforts

---

8. Follow AAPOR Reporting Guidelines

When calculating response rates, researchers should follow the standards established by the **American Association for Public Opinion Research (AAPOR)**.
For studies affected by bot activity:

- Include fraudulent submissions in disposition reporting.
- Document the number of identified bot cases.
- Under AAPOR disposition codes, bot and fraudulent cases are commonly reported under **"Other" (Disposition 4.90)**.
- Clearly explain the rationale for classification.

Providing detailed disposition tables allows readers to fully evaluate survey quality and fieldwork outcomes.

Recommended Layered Approach

Researchers should not rely on any single protection method. A layered approach is most effective:

- Controlled recruitment and distribution
- CAPTCHA verification
- Attention checks
- Consistency checks
- Honeypot questions
- Metadata review
- Duplicate submission monitoring
- Built-in survey platform security features
- Transparent reporting and documentation

By combining these strategies, researchers can significantly reduce the risk of bot activity and fraudulent responses while maintaining the integrity and credibility of their survey findings.

Please cite this document as:

Statistical Consulting and Survey Research Unit (2026) Preventing Bots and Fraudulent Data in Web Surveys: A Practical Guide for Researchers. University of Waterloo.

More resources are available at our website:
https://uwaterloo.ca/statistical-consulting-survey-research-unit/