



---

# Data Scrubbing

# Data Scrubbing

---

b3's bad responder identification (AKA Data Scrubbing) solution utilizes multiple techniques stemming from the basics of human psychology. The Basic premise being that humans are habitual and will do certain things repeatedly. Our scrubbing offer uses this trait coupled with out of the box thinking to identify bad responders.

## What is Data Scrubbing?

*scrub*

*verb (used with object), scrubbed, scrub-bing*

- to rub hard with a brush, cloth, etc., or against a rough surface in washing.
- to subject to friction; rub.

*verb (used with object), scrubbed, scrub-bing*

- to cleanse something by hard rubbing.
- to cleanse one's hands and arms as a preparation to performing or assisting in surgery (often followed by up).

## Identification and removal of bad responders

### Responders who take the survey

- Without paying attention
- Clicking through just to get to the end
- Lying to participate
- Participating just to get incentives

### Evaluates responses to questions

- Scrubbing looks at responses to survey questions
- Observes how responders type on their keyboard when answering verbatim questions
- Checks for illogical responses, e.g., 22 years old with a high school education with income over \$200,000
- Observe patterns in responses

### Identifies duplicates without PII

- Scrubbing uses survey responses along with interview duration and machine characteristics for identification of duplicates
- No PII is used in the process because
- Responders can game the system by joining panel from multiple sources or be part of multiple panels
- Surveys where sample comes from multiple panel providers, this method of duplicate identification is extremely valuable
- Unlike machine based duplicate response identification a.k.a. RelevantID

# Why is scrubbing important?

---

## **Garbage In Garbage Out**

- Uncleaned data sets impact results
- Bad responders can swing results wildly

## **Improves client confidence**

- Helps in instilling greater confidence in clients on data
- Improves brand perception with regards to quality product
- Competitive differentiator in market

## **Data Scrubbing Techniques**

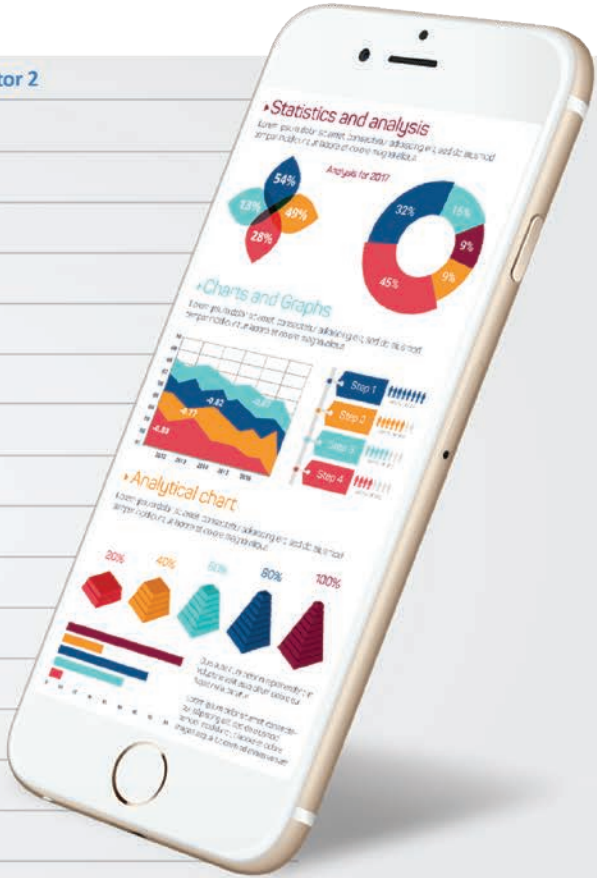
- Straightliners/Flatliner
- Bad Verbatims
- Speeders
- Inattentive
- Duplicates
- Bots
- Geolocation
- Illogical answers
- Maxdiff/Conjoint Satisficing
- Video/Audio usage
- Questionnaire Review
- Verbatim -Contextual

# Why is scrubbing important?

continued

## Competitor differentiation

Solution/Technique	b <sup>3</sup> Intelligence	Top Competitor 1	Top Competitor 2
Straightliners/Flatliner	●		
Bad Verbatims	●		●
Speeders	●		
Duplicates	●	●	
Machine response identification	●		
Use of survey responses	●		
Patterns in survey responses	●		
Click through behavior	●		
Keystroke analysis	●		
Inattentive responses	●		
Copy paste behavior	●		●
Real time results		●	●
Geolocation (Country level)	●	●	
Illogical answers	●		
Video viewing validation	●		



## Scrubbing process

Scrubbing is done parallel to fielding, to ensure there is enough time during fielding to collect replacement sample.

Round 1: Processed once 60% of data has been collected.

Round 2: Processed at 90-95% data collection mark.

Round 3 or more: Processed at field closure or based on client request.

# Turnaround times

---

## **95% of projects are turned around within 8-10 hours**

- Projects received before 3pm eastern are turned around the same day, albeit, later in the evening
- Projects received after 3pm eastern and turned around before 9 am eastern the following day

## **Expedited requests**

- Urgent requests can be turned around faster
- Projects received after 3pm can still be turned around the same day

# What impacts scrubbing cost?

---

## **Number of completes**

- Under 5,000 / Over 5,000

## **Number of files to be scrubbed**

- Each data file is processed individually
- Multiple files can constitute multiple rounds

## **Number of rounds**

- Each scrubbing project cost includes up to 2 rounds, additional rounds are charged

## **Number of non-English languages**

- Verbatim responses in non-English languages are translated via Google
- Each non-English language has an additional charge
- If no verbatims in the survey, no additional charge for non-English

# What's needed to get started?

---

## Questionnaire

Word or Excel

## Data file

- SPSS is the preferred format
- Excel files lead to longer turnaround time and higher costing
- Multiple files can constitute multiple rounds

Email request to [scrubbing@b3intelligence.com](mailto:scrubbing@b3intelligence.com)

## Case Study

### Financial Services Survey across 38 countries

Round 1: Country X went through first round of scrubbing with 320 completes. All regular checks were performed, and output was within standard removal rates barring two

Issue 1 – IP Address: 96% of all records came in from Country Y

One specific ISP was used to take the survey over 300 times, with 70% of completes coming from 8 main IP blocks  
Issue 2 – Verbatim: None of the suspect records had any responses to verbatim questions

## b<sup>3</sup> Volume & Capacity

### b<sup>3</sup> processes 420-440 scrubbing projects per month

- An average project goes through 2 rounds
- 750 to 800 files are scrubbed per month
- Team of 12 analysts

### b<sup>3</sup> can processes 550-600 scrubbing projects per month can processes 550-600 scrubbing projects per month

- At current capacity
- Scalable to a higher level due to automation and staffing