

# Standard Operating Procedure: Data Quality

## Purpose

The purpose of this document is to establish best practices for checking and cleaning data files for online surveys programmed by Schlesinger on Decipher. The goal is to ensure that online surveys and responses produce high quality output that can be relied on to make good business decisions.

For surveys not hosted by Schlesinger, the guidelines below should be used a reference to our standards and expectations from the hosting Company/Client.

## Scope

Schlesinger Programmers and Project Managers may be involved in discussing the questionnaire design with the Client before programming begins, and building real-time data cleaning measures during the programming phase. Project Managers will review and clean the final data file prior to delivering it to the Client.

## Policy

Schlesinger recommends employing a “multi-tier” approach to identify sub-optimal survey taking behaviors. Removing responses based on a standalone criterion may result in over-elimination and may impact the sample representation negatively. An appropriate balance between suitable responses and human behavior needs to be employed. Actions on data cleaning can be taken appropriately based on the severity of sub-optimal behavior, survey length, design, and complexity, impact on data representation, and other empirical considerations.

As a general policy, the Client will be responsible if the data cleaning results in more than 3% of respondents being removed. An analysis of responses and the entire survey will be performed.

Specific criteria and actions include:

CRITERIA	ACTION	TOOL/METHOD
Multiple Panels	(Removal)	RelevantID
Duplicate IP Addresses	(Flag/Removal)	RelevantID
Geo Location	(Flag/Removal)	RelevantID
Speeders	(Flag)	Decipher (Median Length)
Straight Liners	(Flag)	Decipher (Custom Script/Manual)
Open End Responses	(Flag/Removal)	Data Review (Manual)
Red Herring	(Flag)	Decipher (Custom script)
(By Request)		

Creating a count of the various “flags” or “strikes” found among the respondents in the data set allows the program or the person cleaning the data to set a threshold and mark those respondents who fail it as “sub-optimal.”

## Definitions

The following represents common definitions of various survey-taking offenses, as well as relevant considerations.

### Duplicate Respondents

Duplication may occur due to the following reasons:

1. Multiple Panels: A single respondent can be a member of multiple online panels.
2. Duplicate IP addresses: In most developed countries with adequate Internet penetration, IP addresses for each interview should be unique. Since an IP address takes into account a unique identifier based on the system to which it's attached, any duplication should be suspect and be marked as suspicious and should be investigated. That said, note that duplicate IP's can genuinely occur when people connect from the same company, campus, doctor's office or household, etc., so this should be considered a "soft strike," at best. In emerging countries where internet cafes and/or kiosks are more prevalent, duplicate IP addresses may be more common. Based on where the interview originated, the person who is cleaning the data should exercise some discretion over how this strike is reflected.

### RelevantID

On all surveys using the Decipher Beacon tool, Schlesinger employs RelevantID, digital fingerprinting software which gathers a large number of data points from a respondent's computer, such as operating system version, browser version, plug-in, etc. and assigns a relative weight to each data point. The data gathered is put through deterministic algorithms to create a unique "digital fingerprint" of each computer.

The digital fingerprint identifies duplicate respondents who take the same survey more than once from the same machine. RelevantID flags a computer each time a user tries to take a survey, so it is able to detect if multiple e-mail accounts are being used to take surveys from a single computer. In addition, RelevantID has the unique ability to identify multiple panel accounts from different research firms on the same computer. Suspect respondents are flagged in the system, prior to taking the first question in the survey, and based on business rules, are either allowed, redirected or completely filtered out of surveys in which they attempt to participate.

### Geo Location

RelevantID also identifies the geographic location of the user by country, state/region, city, US postal code, metro code, latitude, and longitude information and verifies that the IP address is originating from where the individual should be located. If not, the respondents are flagged for further review.

Please note sometimes due to legitimate reasons such as travels, respondents take surveys from a different country than where they belong. An individual respondent who is occasionally in another country may be considered safe. An individual respondent who is continuously flagged in another country is a suspect.

This process is invisible to the user and does not interfere with the user experience. RelevantID is consistent with global data and privacy laws.

### Speeders

A speeder is someone who completes a survey in an unreasonably short period of time. If respondents speed through a survey it is unlikely that they are paying full attention to the survey, and their data may

skew the results. Please note some surveys with complex routing logic may exhibit a wide range of survey completion times. Schlesinger follows a rule of 40% of median LOI as a threshold. Anyone completing the survey in 40% or less of the overall median survey duration should be flagged as a speeder.

Speeding behaviors should be evaluated individually and at a high level. The recent ARF FoQ2 study (across 17 panels and more than 77,000 interviews) found that certain demographic groups (younger, more educated, males, etc.) regularly exhibit speeding behaviors. By excluding the speeders, one can potentially exclude certain demographic groups and potentially bias the data. The ARF FoQ2 data also showed that by eliminating speeders, the accuracy of the data was not improved and in some instances resulted in excluding key demographic groups.

### Straight Liners

Within table-based or attribute-based questions (grids), there exists the possibility that a respondent may simply click answers in a straight line. Schlesinger employs the use of following method to identify straight liners: If a respondent straight lines 3 grids (all responses in each grid) for a survey longer than 25 minutes (or 2 grids for a survey shorter than 25 minutes) in median length then s/he will be flagged as a straight liner. These straight lining metrics should not be applied to attitudinal questions, where straight-lining of responses may be legitimate.

There are evidences that the length of the survey, number of attribute-based rating questions (grids), and scale of the grids have an impact on respondents' straight lining behavior. In order to mitigate this issue, we recommend keeping the survey to 25 minutes in length or shorter and the number of grid type questions at a minimum. Creative representation of grids, such as card sorts or slider scales, can be viable alternates and can help minimize straight lining.

### Open End Responses

A respondent who rushes through a survey is apt to take short cuts and may type in gibberish or nonsense when prompted for an open-ended response. The best way to catch this behavior is to review the open end responses.

To the best extent possible, all open ended responses should be reviewed and any "nonsense" or "inappropriate" responses should be noted as a "soft" strike. If a respondent accumulates multiple strikes of this particular type, or uses egregious or offensive language, that case should be removed (i.e. the entire respondent record should be removed). This is a more of an art than a science.

Depending on the length, design and the number of open ended questions, respondent behavior may vary from respondent to respondent and survey to survey. In certain scenarios a respondent may not have a specific comment for those questions and comments such as "n/a," "don't know" or a blank open end may be valid.

### Red Herring

Within a list of possible answer choices, a nonexistent/impossible answer choice may be programmed ("red herring" questions). The benefit to implementing such questions is that these types of questions are helpful in catching spurious respondents; the downside of using such questions is that they tend to upset legitimate respondents.

Schlesinger does not add red herring questions to Client surveys, unless requested by Clients. Instead, Schlesinger uses a combination of other measures, such as speeders, straight liners, and open ended responses to isolate poor survey taking behavior.

## Summary

In general, if any respondent exhibits unusual behavior provides inconsistent or inappropriate responses, they should be marked as suspicious and in some cases removed from the data set after review. Inconsistent or inappropriate behavior may be exhibited with such examples as respondents selecting all possible choices for a multi-choice question or respondents providing contradictory responses.

Since each survey design, complexity, length and target is unique, the person cleaning the data should be thoughtful in how these "rules" are applied. The survey responses should be looked at holistically, rather than making a decision based on a single measure. Our goal is to provide quality data!

## Process Flow and Accountability

### Pre Field

The Client will provide the final questionnaire to Schlesinger. The questionnaire should clearly show the data cleaning logic to be implemented in the survey by the Schlesinger team. The Schlesinger Project Manager (PM) and Programmer will review the questionnaire and engage the Client in the discussion if they note issues with questionnaire causing potential sub-optimal responses. The Schlesinger programming team may implement appropriate logic to identify speeders, straight liners and Red Herring checks, if any. If/when survey logic is implemented to remove respondents from data their incentive impact should be discussed with the Client.

For all surveys programmed by Schlesinger on Decipher Beacon, the Schlesinger team will implement RelevantID deduplication checks. The Schlesinger team will be responsible for testing survey link prior to launching in field.

### In Field

At times, we have found it beneficial to complete Soft Launch to discuss possible ways to implement logic to identify sub-optimal behavior, and to discuss enhancement in survey design to mitigate such occurrences. Any survey where more than 3% of responses are considered sub-optimal may indicate a poor survey design and an analysis of responses and the entire survey will be performed as a joint exercise between the Schlesinger team and Client. The Client will be responsible for paying out Recruit and Incentive costs for removing respondents in excess of 3%.

### Post Field

At the end of fielding, the Schlesinger team will perform data cleaning adhering to this policy, and send the Client a summary and raw data for all respondents flagged as sub-optimal and/or recommended for removal showing:

Length of interview (median)

Completed interviews (including sub-optimal interviews) #

Recommended responses to be cleaned # - TOTAL

1. Recommended responses to be cleaned # - Speeding
  2. Recommended responses to be cleaned # - Straight lining
  3. Recommended responses to be cleaned # - Red Herring
  4. Recommended responses to be cleaned # - Open Ended responses
- (There may be overlap in 1-4 above)

## Treatment of Questionable Respondents

Schlesinger team follows the guidelines below when any combination of sub-optimal survey taking behaviors line up to denote a potentially questionable respondent:

- o If survey logic was implemented to identify sub optimal survey takers and automatically disqualify, those respondents will get the appropriate disqualification statuses when they exit the survey.
- o For the checks which are implemented as “soft strikes” or “flags,” respondents failing those checks will be allowed to proceed through the survey as they normally would, and receive appropriate statuses at the end of the survey upon exit.
- o The Schlesinger team will take the following actions:
  - o Any respondents who are removed from the final data set, up to 3% of total completes, will be replaced at no additional charge to the Client (if feasibility and time allow for it). If, however, it is not possible to find a replacement for each such case, the Client will not be charged for the removed respondents, up to 3% of total completes. Any respondents above 3% of total completes will be charged on the bid rate for Recruit and Incentives.
  - o The Schlesinger team will provide the panel source and the unique panel IDs of the panelists deemed “unusable” to appropriate panel teams for further action, including removal from the panel.

## Survey Quality Impacts Survey Taking Behaviors

The quality of the survey will be taken into account when assessing survey taking behaviors. Long surveys (in excess of 25 minutes), complex or long grids, poorly written questions, redundant questions and other survey attributes directly impact the quality of the responses. The Client and the Schlesinger PM will review the survey proactively during the kick off phase, and assess the likelihood of quality respondents exhibiting poor behavior, so it is possible to modify the questionnaire before fielding begins.

One of the findings of the ARF FoQ2 study is that sub-optimal behavior occurs at low rates in well-designed surveys. Sub-optimal behavior rarely has major effects on aggregate estimates and speed of survey completion has far less effect on overall results than does sample provider. The study also found that panelists segments such as younger, more educated and males exhibit speeding more often and excluding them from the data set may bias the data.

The ARF FoQ2 study recommends that reducing survey length, redundancy and difficulty level will reduce sub-optimal survey taking behavior. It also recommends that 2-4% of responses can be eliminated using a “multi-hurdle” techniques (employing a combination of criteria such as straight lining, speeding, and responses to open ended questions); however, instances where more than 2-4% of the respondents are removed may indicate a poor quality survey.