PDPglobal

Research and Development Division
Professional DynaMetric Programs, Inc.

Data Quality Control and Integrity Methodological Process

Prepared By:
Lauren Matheny, PhD, MPH.
January 2023

PDPglobal

Data Quality Control and Integrity Methodological Process

When collecting data electronically, and especially in large amounts, it is common to encounter issues in the quality and integrity of the data from some respondents. With any type of data collection, it is customary to encounter respondents who may rush through the instrument's questions without paying much attention to their responses (speeders), become distracted while completing questions (distracted respondents), or simply respond to questions with unreasonable responses that may take the form of systematic or identifiable patterns (straightliners/patterns of response integrity). Although these issues in data collection are commonplace, they are often overlooked due to numerous factors including the complex nature in which to identify some of these issues, as well as the cost and the large amount of time spent to comprehensively identify all of the potential issues.

As part of the continual process to invest in the evolution and quality of the ProScan instrument, responses provided from 29,082 responses were assessed based on multiple criteria to identify any integrity issues with the data. All data were first assessed for "speeders." For each response, completion time was recorded and assessed for times less than a previously defined threshold. In addition, responses were assessed for abnormally long completion times, which are referred to as "distracted responders." If a response time was beyond the previously defined threshold for acceptable time, the response was flagged with an indicator variable.

All responses were additionally evaluated for 33 different systematic response patterns. For example, responders who selected the same answer for every question, who are referred to as "straightliners," were flagged for "pattern response integrity" issues. All of the data were also evaluated for nonsensical responses. For example, if there was an activity level scale with an item that asked a respondent whether they could walk up 4 flights of stairs with "no difficulty," it would not make sense for them to state "extreme difficulty" walking up 1 flight of stairs. This same logic regarding nonsensical responses was applied to 16 of the ProScan items.

Responses were also assessed for re-try manipulation, which is defined as respondents who completed the ProScan more than once. All ProScans completed by the same individual within a three-day period were identified and flagged. Responses flagged as duplicates were assessed to determine whether the Basic Structure changed from the initial ProScan to the duplicated ProScan. If an individual altered their responses enough to change their Basic Structure, this was also flagged with an indicator variable.

Overall, the ProScan was evaluated for response integrity and quality control based on 50 criteria. While this type of thorough quality control is not the status quo in data collection, it is important to note that these integrity checks are being performed on the ProScan for each and every response. Flagging a response for any of these issues initiates the next step in quality assessment, indicating the need for further inspection. Performing these rigorous evaluations helps ensure that ProScan remains a trusted and reliable source for trait assessment.