

Data Quality 202: Data Quality Standards



Data help us to understand the effectiveness and impact of programs. However, we can draw accurate conclusions and make appropriate decisions using data only if the data meet quality standards. In the **Data Quality 101 infographic**, we addressed how to identify and address data quality issues related to inaccuracies, inconsistencies, missing data, and outliers. This data graphic highlights examples specific to the Administration for Community Living (ACL) as well as hypothetical examples to address additional data quality issues related to misleading, inappropriate, and nonrepresentative data and ways they might be addressed.



ACL EXAMPLE | National Survey of Older Americans Act Participants

BACKGROUND

In 2003, the Administration on Aging (AoA) began the National Survey of Older Americans Act Participants (NSOAAP), a national sample survey funded by Title III of the Older Americans Act (OAA), to assess the effectiveness of the Title III programs, maintain accountability, and demonstrate the program's success in achieving legislative goals. The NSOAAP is comprised of six surveys of recipients of selected Title III services (i.e., case management, transportation, congregate meals, home-delivered meals, homemaker, caregiver). The survey instruments focus on the consumers' assessment of service quality and outcomes. The instruments also measure client characteristics, such as demographics and physical and social functioning. The surveys are conducted annually through computer-assisted telephone interviewing. NSOAAP data are available to the public and can be accessed on the [ACL's AGing, Independence, and Disability \(AGID\) Program Data Portal](#).

DATA QUALITY ISSUE

After the eighth survey ([2013 survey instrument](#)) was completed, a question was added for the subsequent survey years (e.g., [2014 survey instrument](#)) under the demographic section for gender (DE1). The new question (DE1a.) asked, "Which of the following best represents how you think of yourself? Lesbian or gay; straight, that is, not lesbian or gay; bisexual; something else; refused; or don't know." Respondents who selected the response "something else" were asked a further question: "What do you mean by something else? You are not straight, but identify with another label such as queer, trisexual, omni-sexual, or pansexual; you are transgender, transexual or gender variant; etc." Because that further question included response categories that conflated sexual orientation with gender identity, it may have resulted in misleading data.

SOLUTION

To improve the quality of the survey data, ACL deleted the follow-up question and is now working with the National Institutes of Health to improve their survey questions.



EXAMPLE

A local nonprofit group is funded to provide meals for families below the poverty line. Continued funding is contingent on providing evidence, after a year, that their process of providing meals is more efficient than a competing model being used in an adjacent community. After a year, the grantee provides data to show that they have provided meals to more families than their competitor at the same total cost. The grantee's project officer is satisfied and recommends that the grantee's funding be continued.

DATA QUALITY ISSUE

The grantee's report neglected to state that the community in which their competitor operates has larger families in the grantee's catchment area. The project officer reanalyzes the grantee's data by number of people fed, rather than number of families fed, and finds the two programs to be about equally effective.

SOLUTION

Project officers should make sure to have the full context for data reported, such as average family sizes described above, to ensure the use of the appropriate method, including the unit of analysis. When a grantee reports on improvements, project officers will also need to have sufficient data related to the amount of change in order to determine if the improvement reflects a meaningful change (e.g., is statistically significant).



EXAMPLE

A grantee agency was funded to provide services to at least 10% of eligible people in a catchment area. The grantee agency's annual report states that services have been provided to approximately 5,000 people. The grantee's project officer reads the report and is pleased by the large number who were served.

DATA QUALITY ISSUE

Of the eligible people in the catchment area, 10% is approximately 10,000 people—twice as many as the number of people to whom the grantee provided services. The project officer was mistaken to have been satisfied with the grantee's performance since the number reported is far below the number for whom the grantee is responsible. The project officer was misled by the grantee's reporting a number rather than a percent, as was required by the original application for funding. Reporting the wrong kind of data regarding outcomes can misrepresent the success of a program.

SOLUTION

When assessing grantee and program performance and outcomes, project officers should reread the original request for proposals to ensure data are reported correctly and meet indicated performance targets and outcome variables. In addition, project officers should determine whether the data reported are appropriate indicators of success, consulting with ACL evaluation experts if needed. Project officers should also make sure to have the full context for data reported such as the denominator as described above, to allow for calculation of the needed percentage. When a grantee reports program improvements, project officers will also need to have sufficient data related to the amount of change in order to determine if the improvement reflects a meaningful change (e.g., statistically significant).



EXAMPLE

A researcher developed a promising intervention to improve the problem-solving ability of individuals with moderately severe brain injuries. Administering the intervention requires specialized training and is expensive. The researcher obtained funding to test the intervention but had underestimated the cost of administering it, so was unable to test it on as many people as planned. The results of the test were quite positive, with each participant showing a noticeable improvement on the outcome measure. However, the sample size was too small for the results to be statistically significant.

DATA QUALITY ISSUE

The researcher reported to the funding agency that the test was not successful because the hypothesized effects were not statistically significant. This report was misleading because in fact the intervention's effect was substantial to the participants.

SOLUTION

The researcher could recommend that the intervention seems very promising and should be tested using the larger sample, as was originally planned, but state that until that is done, the results found here do not provide sufficient evidence to implement the intervention.

In general, when assessing the effectiveness of interventions, reviewers should request and consider a range of statistical measures, such as margins of error, confidence intervals, and effect sizes. In addition, reviewers should carefully consider the data collection and analysis methods to ensure they support and are appropriate for testing the research hypotheses. Project officers should carefully scrutinize any proposed changes to the methodology and consider additional project changes, such as increasing the length of time for the study to allow for more people to be included.



HYPOTHETICAL EXAMPLE | Fabricated Data

EXAMPLE

A grantee is running up against a deadline for completing a phone survey. Two of her three data collectors are out sick, and quite a few surveys still need to be completed. She tells the remaining data collector that it is very important to complete the surveys. She notes that if the data collector gets them done, he will get a bonus, but that if he does not get them done, he will get a negative performance evaluation. This makes the data collector very motivated to complete the surveys. He tries to get them all done but soon realizes that it will not be possible because some of the people on the outreach list are turning out to be very hard to reach. Having already done quite a few of the interviews, he believes that he has a pretty good sense of how different kinds of people answer the questions. He decides that, for the hard-to-contact people, he will create and enter data that he believes are much like the answers that similar participants have provided. That is, he fabricates the data.

DATA QUALITY ISSUE

Results of analyses of these data cannot be trusted because an unknown number of them do not reflect actual respondents but rather were concocted by the data collector.

SOLUTION

Determining whether data have been fabricated is challenging, and there is no best way to do so. Even when repeating the initial research, discrepancies in results may not be due to data fabrication. Reviewers should obtain access to the raw data collected for the study and should carefully review the data and results. The delivery of the raw data files should be made a condition of the grant/contract. Reviewers may look for red flags, such as whether the data are too perfect or whether methods and statistical analyses are vague or seem inappropriate. If there are concerns about the data and results, reviewers may also contact the researcher to probe and discuss concerns or require submission of contemporaneous notes or other documentation. Researchers can increase the likelihood of identifying data fabrication by calling a random sample of people for whom data collectors have provided data in order to confirm that they were in fact contacted by the data collector. Researchers can reduce the likelihood of data fabrication by requiring that data collectors sign a pledge to follow a list of good data collection practices (including not making up any data), not overly motivating data collectors to reach difficult-to-achieve numbers and/or types of respondents, letting data collectors know that the researcher will be contacting some of the people for whom the data are provided to confirm that they were in fact contacted and interviewed, and making clear to data collectors that data fabrication will not be tolerated.



EXAMPLE

An agency hired a contractor to determine how many exposures to their training protocol are needed in order to have a substantial effect, which is operationalized as an increase of 2 or more units on the outcome measure. Participants receive one dose of the intervention each month for a year and complete the outcome measure each month and 1 year after the last training. Partway through the study, after the training and assessment have been underway for a few months, a better outcome measure is published and is immediately accepted by the field as the gold standard. Wanting to obtain the best data possible, the contractor replaces the original outcome measure with this new measure. Data collection is eventually completed and the data are entered into a database. A data analyst proceeds to analyze the data to examine the effect of the training, but the analyst is not informed of the change in outcome measure. The analysis finds that the intervention exceeds expectations because it resulted in an average increase of 3 units on the outcome measure. The agency's director was pleased with this result and planned to provide more funding for scale-up.

DATA QUALITY ISSUE

Whereas the original outcome measure was a 5-point scale with an average typical score of around 3, the new measure is a 15-point scale with an average typical score of around 8. Outcomes assessed after the use of the new measure almost always showed a higher score because of the measure's greater scale. Thus, results reflected this artifactual increase in scale rather than an actual increase in participants' performance.

SOLUTION

Reviewers should carefully evaluate the research methods to ensure their appropriateness. Once the program has been implemented, reviewers should obtain as much detail as possible about how the interventions were administered and how outcomes were measured, then meticulously compare methods employed to the initial research plan. It is important that reviewers carefully document both the intended methodology and any change to the methodology, and that they carefully review the information before accepting the final report. Any deviations from the original plan should be carefully discussed with the researcher. Also, the contractor should not have changed to the new measure unless a crosswalk was available for converting the scores on the old measure into scores on the new measure (or vice versa), and they should have included notification of this change in the report. The contractor could also have used both measures and reported two separate variables, making it obvious that the measure changed.



EXAMPLE

A researcher is funded to conduct a project to determine if a new mobility intervention helps people who have a walking impairment following a stroke. The researcher works at a very large rehabilitation center and has access to a large number of clients who meet the inclusion criteria for her study. She includes as many of the clients as possible in her study. The researcher reports to the funding agency that the intervention is a success because the group who received the new intervention had a statistically significant higher score on a walking test than the group who received the usual intervention. The grant's project officer is satisfied that the new intervention works well.

DATA QUALITY ISSUE

Having so many people in the study resulted in this study having a statistically significant result even though its effect was very small, so small that its impact would not make much of a difference to people in their day-to-day lives

SOLUTION

For studies that employ very large sample sizes, researchers should provide justification for their N and explicitly examine to what extent the results are meaningful beyond the statistical significance (i.e., the likelihood that the results are meaningful and translate to practice). As much as possible, researchers should be objective and clear in helping reviewers understand the meaning of the results by reporting effect sizes using confidence intervals. Also, researchers and reviewers should keep in mind that “statistically significant” does not necessarily equal actual significance to individuals in terms of the magnitude of the change.

Misleading Data and Inappropriate Data Comparisons

ACL EXAMPLE | Older Americans Act Title III and VII State Performance Report

BACKGROUND

In compliance with the Government Performance and Results Act (GPRA), the OAA and ACL's AoA, 50 states, the District of Columbia, Guam, Puerto Rico, American Samoa, the Northern Mariana Islands, and the U.S. Virgin Islands submit an annual State Performance Report (SPR) on the activities carried out under the OAA Title III grant program. The data collected by state agencies, area agencies on aging, and service providers include client demographic information to assess the effectiveness of OAA Title III programs in reaching older individuals with the greatest economic and social needs, with particular attention to low-income, minority, rural, and frail older adults (including individuals with physical and/or cognitive impairments). Data from SPRs from 2005 on are available to the public and can be accessed on the [ACL's AGID Program Data Portal](#). These data can be aggregated up from the state level for analysis at the regional and national level.

DATA QUALITY ISSUE

Historically, the SPR data collection measured client race based on mutually exclusive race categories that captured unduplicated counts (i.e., White alone, Black or African American alone, Asian alone, American Indian or Alaskan Native alone, Native Hawaiian or Other Pacific Islander alone, other race alone, multiple races). This approach can be misleading. For example, if a data user wanted to know how many Black or African American clients received certain OAA Title III services, using the "Black or African American alone" total would likely be an underestimate since clients who identify as Black or African American and other race(s) would be counted in the "multiple races" category. The "multiple races" category does not allow a user to determine the specific combination of races that are represented among clients counted in this category.

Additionally, data users may want to compare an OAA Title III services population to the general population of older adults to determine if the services are reaching older adults with the greatest need as identified in the OAA. For example, a data user might want to compare the percentage of older adults who identify as Black or African American in the OAA Title III home-delivered meals population versus the general U.S. Black or African American older adult population. However, the current census data available on the AGID data portal (i.e., American Community Survey 1-Year Public Use 5% Microdata Sample [PUMS]) follow a different measurement approach to race. These census data capture race as a duplicated count (e.g., Black or African American alone OR in combination with other race or races). Given the difference in measurement approach for SPR and the census data on the AGID data portal, it would be inappropriate to make a direct comparison of these data.

SOLUTION

In 2018, ACL initiated the transition to a new State Program Report (SPR). To improve data collection around race, the new SPR measures race based on duplicated person count, thereby addressing the previous data issues around race. Clients are now counted in each race they identify with. For example, if a client identifies as both White and Asian, a state submitting data would mark both identities when collecting the demographic data, duplicating the race distributions and more accurately describing its client population.

This solution meets ACL's need related to the SPR. Other solutions may be more appropriate for other reporting and evaluation tools based on their needs.



ACL EXAMPLE | National Survey of Older Americans Act Participants

BACKGROUND

In 2003, the Administration on Aging (AoA) began the National Survey of Older Americans Act Participants (NSOAAP), a national sample survey funded by Title III of the Older Americans Act (OAA), to assess the effectiveness of the Title III programs, maintain accountability, and demonstrate the program's success in achieving legislative goals. The NSOAAP is comprised of six surveys of recipients of selected Title III services (i.e., case management, transportation, congregate meals, home-delivered meals, homemaker, caregiver). The survey instruments focus on the consumers' assessment of service quality and outcomes. The instruments also measure client characteristics, such as demographics and physical and social functioning. The surveys are conducted annually through computer-assisted telephone interviewing. NSOAAP data are available to the public and can be accessed on the [ACL's AGing, Independence, and Disability \(AGID\) Program Data Portal](#).

DATA QUALITY ISSUE

For some of the variables included in NSOAAP (e.g., sexual orientation), the cell size for the reporting is too small, resulting in a large statistical measure of standard error (>30%) which does not meet ACL standards for data quality. In general, a standard error above 30% suggests that the estimate for the variable has low accuracy and could lead to interpretations about NSOAAP participants that are not accurate.

SOLUTION

ACL has restricted access to these data at this time and is working with other experts in the field to identify ways to collect more reliable data. When examining the quality of data, reviewers should not only look at the results but also consider measures of data quality such as margins of error, confidence intervals, and standard errors.



EXAMPLE

A federal agency funded two different job training programs to test their effectiveness in reducing unemployment rates. The programs were implemented in a town with a high unemployment rate. Both programs used the same inclusion/exclusion criteria to identify eligible participants, and participants were enrolled in the same way—by posting announcements around town and on the town’s website. The agency had already determined that the town had a sufficient, if limited, number of potential eligible participants for evaluating the programs. For a number of reasons, one of the training programs was not able to start enrollment until a few months after the other program. The first program to start had no difficulties reaching its target sample size because many people who saw the announcements were eager to participate and signed up for the first opportunity to participate. Fewer people signed up for the second program because it began later. In order to reach its sample size, the second program had to reach out to eligible individuals and convince them to participate, rather than simply accept volunteers. After both programs were completed, their respective employment outcomes for enrollees were compared. Results showed that the first program was quite successful, whereas the second program was less successful. Based on these results, the funding agency is considering providing scale-up funding to only the first program.

DATA QUALITY ISSUE

Because the first program had enrolled a large proportion of the eligible townspeople who were eager to volunteer, the population of townspeople available for the second program was somewhat different. Those available for the second program tended to be less eager, less confident that they would benefit from the program, less willing to try something new, and often somewhat doubtful and suspicious of the researchers. It may well have been that this difference, rather than anything about the program itself, accounted for the lesser success of the second program.

SOLUTION

Reviewers should carefully evaluate the sampling plan before research is conducted to ensure its appropriateness. Once the program has been implemented, reviewers should obtain as much detail as possible about the reasons participants enrolled (to determine potential effects of motivation among participants), the ways they were enrolled, and the methods used to implement the program. Reviewers should compare the details to the initial plan and consider any ways in which the methods may account for the observed difference in program outcomes. Reviewers should also request and examine data that provide information about the participants in all groups and allow reviewers to draw conclusions about the comparability of the groups. If there are concerns about the methods and results, reviewers may also contact the researcher to probe and discuss concerns.

For additional information on data quality issues, see the [Data Quality 101 Infographic](#) and [Data Quality 201: Data Visualization](#).

For questions, contact ACL’s Office of Performance and Evaluation.