GAO
Accountability * Integrity * Reliability

Assessing the Reliability of Computer-Processed Data

Greg Wilmoth
U.S. Government Accountability Office (GAO)
Washington, D.C.

Sept. 21, 2009

Outline

- Part I: Beginning to think about data reliability
- Part II: Overall framework and key factors
- Part III: How do we do the assessment?
  - Assessment work
  - Making the decision
- Part IV: Documenting and reporting

Beginning to Think About Data Reliability

- **What are computer-processed data?**
- **When do we need to do a data reliability assessment?**
- **What do we mean by data reliability?**
  - Generally, data are considered reliable if
    - the data entered into a computer system/application are reasonably accurate, complete, and consistent,
    - the data, while stored in the computer system/application, are not inappropriately modified or deleted, and
    - any computer processing performed on that data is reasonably accurate, complete, and consistent.

Why Do We Care About It?

- According to the Yellow Book, auditors should assess the sufficiency and appropriateness of computer-processed information regardless of whether this information is provided to auditors or they extract it independently.
- Bottom line:
  - ***We do not want to issue a product that contains incorrect information or the wrong message.***

How Do We Tackle This Issue?

- The guidance *Assessing the Reliability of Computer-Processed Data (GAO-09-365G)* has recently been reissued internally for use by GAO staff and an external version will also soon be issued for use by other governmental audit entities. (They are not substantively different.)

- The guidance is geared toward performance evaluations (not financial audits).

How Do We Tackle This Issue?

- The guidance provides a flexible, risk-based framework built on:
  - making use of existing information about the data,
  - conducting only the amount of work necessary to determine whether the data are reliable enough for our purposes,
  - maximizing professional judgment, and
  - bringing the right people, including management, to the table at key decision points.

Overall Framework

See Figure 3: The Framework of the Data Reliability Assessment Process

Determining the Extent of the Assessment – Key Factors to Consider

- The extent of the assessment and your decision about reliability depend on:
  - expected importance of the data to the product,
  - strength or weakness of corroborating - or conflicting – evidence,
  - anticipated level of risk in using the data.

- Considering these factors, determine the most appropriate mix of work needed to assess the reliability of the data.

How Do We Do the Assessment?
Review of Existing Information

- Early on, learn as much as possible about the data from existing sources. This will help you determine whether the data are appropriate to answer your research questions and provide information for your assessment.

- Generally, this is done by
  - examining existing audit or quality assurance reports, studies, and documentation related to the data; and
  - interviewing staff knowledgeable about the data and the system that produces them.

How Do We Do the Assessment?
Data Testing

- Data testing can be done by applying logical tests to electronic data files or paper copies of reports. For record-level data, you can use computer programs. Test only data elements you plan to use.
- Examples of such tests are:
  - range tests and frequencies, to look for out of range values or other anomalies
  - checks for
    - missing data (records or values in key data elements),
    - valid dates, and
    - erroneous duplicates (entire records or unique identifiers),
  - examine relationships between variables (cross-tabulations),
  - duplicate automated processes (testing a formula or edit checks)

How Do We Do the Assessment?
Testing: Example

State Crime and Grant Data

State	Murders	Grant Amount	Grant start date
AZ	237	150000	2/15/2000
AL	455	200000	9/1/1999
AK	9999	150000	3/25/2005
CA	780	500000	
KS	300	200000	1/1/1998
NY	1000	750000	6/6/1997
NM	154	-150000	11/13/2001
MI	356	250000	5/15/1995
PZ	0	150000	10/1/1990

How Do We Do the Assessment?
Testing: Examples

- Range Tests (Minimum and Maximum), Sums, Number Missing (Miss)

Variable	Label	Minimum	Maximum	Sum	Miss	N
X5729	Income	-1000000.0	122660000.0	11547685585	0	22210

- Outliers

Extreme Observations (X5729)

Lowest		Highest	
Value	Obs	Value	Obs
-1000000	19053	104410000	325
- 999999	14025	115990000	6494
- 876342	51	122660000	492

How Do We Do the Assessment?
Testing: Examples

- Frequencies:

Gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
	5590	26.32	5590	26.32
F	4995	23.52	10585	49.84
J	4	0.02	10589	49.86
M	10650	50.14	21239	100.00

Based on the following data dictionary entry:

Field Name	Description	Possible Values & Keys
Sexcode_desc	Gender	Male: Female: DNS (did not supply)
Ethnic_desc	Ethnicity	White - Black - Hispanic - NAT AM (Native American) -
		NAT AM (Native American) - Asian/PI (Asian or Pacific Islander) -
		DNS (did not supply)

How Do We Do the Assessment?
Tracing To/from Source Records

- **We might choose to trace a sample of data to and/or from source records when**
  - there is a need to quantify the degree of reliability of required data elements (proper random sample needed),
  - preliminary work has led us to believe that there is no assurance that internal controls are in place to assure the reliability of the data,
  - there is a need for assurance the data are precisely accurate, or
  - there are concerns about the completeness of the data.

How Do We Do the Assessment?
Reviewing Selected System Controls

The team might choose to review selected controls when they have

- identified potential system control problems, or
- learned that source documents are not readily available.

- Controls will usually include 1) general controls, such as logical access and control of changes to the data, and 2) application controls that help ensure that data are accurate, complete, and authorized.
- You may need to involve qualified information technology auditors or security specialists

How Do We Do the Assessment?
Making the Decision about Reliability

- Examine the results of your work in the context of the previously discussed factors: importance of data to final product, corroborating evidence, and level of risk in using the data. If needed, take further steps to assess the data.
- With relevant stakeholders, come to a decision:
  - The data are sufficiently reliable (e.g. to answer the researchable question).
  - The data are not sufficiently reliable (e.g. to answer researchable question).
  - The reliability of the data is undetermined.

How Do We Do the Assessment?
Decision

- The decision that the data are sufficiently reliable should be made when the assessment provides sufficient assurance that the data are reasonably complete and accurate and therefore are *sufficiently* reliable to answer researchable question. Specifically, that the likelihood of significant errors or incompleteness is minimal and the use of the data would not lead to an incorrect message.
    - Use the data and disclose work done along with any limitations of the data.

How Do We Do the Assessment? *Decision*

- The assessment should result in a decision that the data are not sufficiently reliable when the work shows that the data are unacceptably incomplete and/or inaccurate and could possibly lead to an incorrect message.
- Do not use the data.
- Explore other options, including the following:
  - Modify the audit question or approach.
  - Seek other sources.
  - Collect primary data.
  - Report on problems with the data and/or consider whether results should be reported or explored further in another engagement.

How Do We Do the Assessment?
Decision

- But what if we're not sure and we are unable to determine reliability?
  - There may be times when we've done all the data reliability investigative work we have time to do, yet we are still uncertain about the reliability. At these times, ask: ***Will using the data likely result in an inaccurate message?***
    - If this is the case, do not use the data.
    - If circumstances force you to, clearly lay out all limitations and how those limitations affect the interpretation of the data. Do not use the data as the sole basis for findings, conclusions, or recommendations.
    - Strongly consider reporting on the problems found.

Documenting and Reporting the Assessment

Three important parts of performing a data reliability assessment are:

- planning and performing the assessment,
- documenting the assessment and decisions made, and
- reporting the assessment steps and decision(s), in relation to the use of the data in the product.

Documenting the Assessment

- All work performed as part of the data reliability assessment should be documented and included in the engagement documentation. This includes:
  - all testing and results,
  - information review and interviews related to data reliability, and
  - decisions made during the assessment, including the final assessment of whether the data are sufficiently reliable for the purposes of the engagement.

Report Language

- Engagement reports conform to GAGAS by
  - discussing what was done to assess the data,
  - disclosing any data concerns, and
  - reporting the judgment about the reliability of the data for use in the product.
- See appendix III in the guidance for examples of report language.

Summary

- A few points to emphasize as we wrap up:
  - This guidance provides a framework that aids planning, discussion, and reporting.
  - Assessments of data reliability should be performed only on those data elements needed for the engagement.
  - We are not making a decision on the reliability of the database or system as a whole, only whether or not the data we examined are *sufficiently reliable to answer the research question.*
  - The framework is based on bringing the people with the right skills into the process when they are needed, and obtaining management buy-in on key decisions about the extent of the assessment and use of the data for the assignment.