

ASSESSING THE RELIABILITY OF COMPUTER-PROCESSED DATA

YELLOW-BOOK

- 6.66 Auditors should assess the sufficiency and appropriateness of computer-processed information regardless of whether this information is provided to auditors or auditors independently extract it. The nature, timing, and extent of audit procedures to assess sufficiency and appropriateness is affected by the effectiveness of the audited entity's internal controls over the information, including information systems controls, and the significance of the information and the level of detail presented in the auditors' findings and conclusions in light of the audit objectives. The assessment of the sufficiency and appropriateness of computer-processed information includes considerations regarding the completeness and accuracy of the data for the intended purposes.

GAO-09-680G

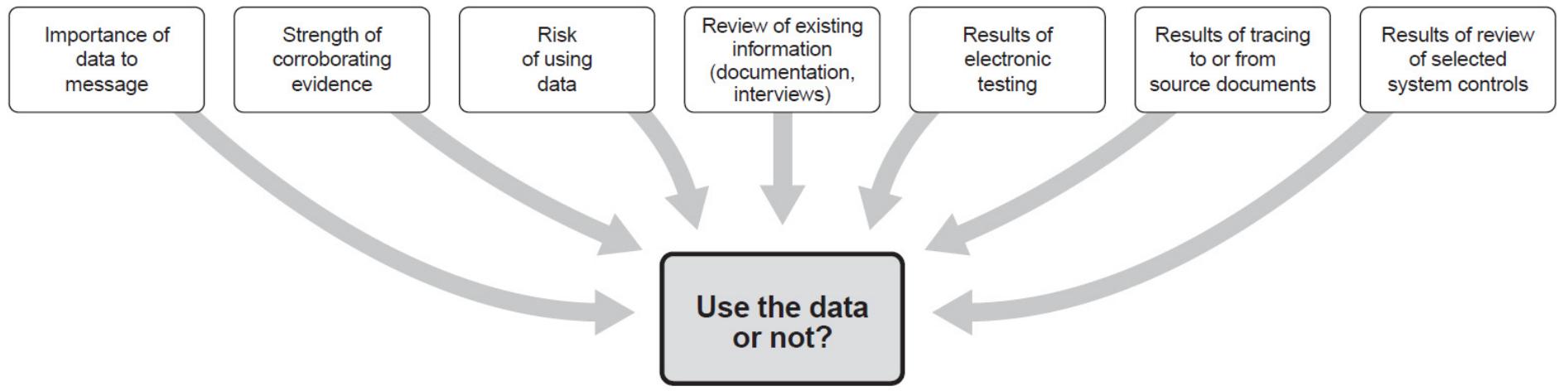
United States Government Accountability Office
GAO Applied Research and Methods

- July 2009

July 2009
External Version 1 Assessing the Reliability of
Computer-Processed Data

GAO-09-680G





KEY DEFINITIONS

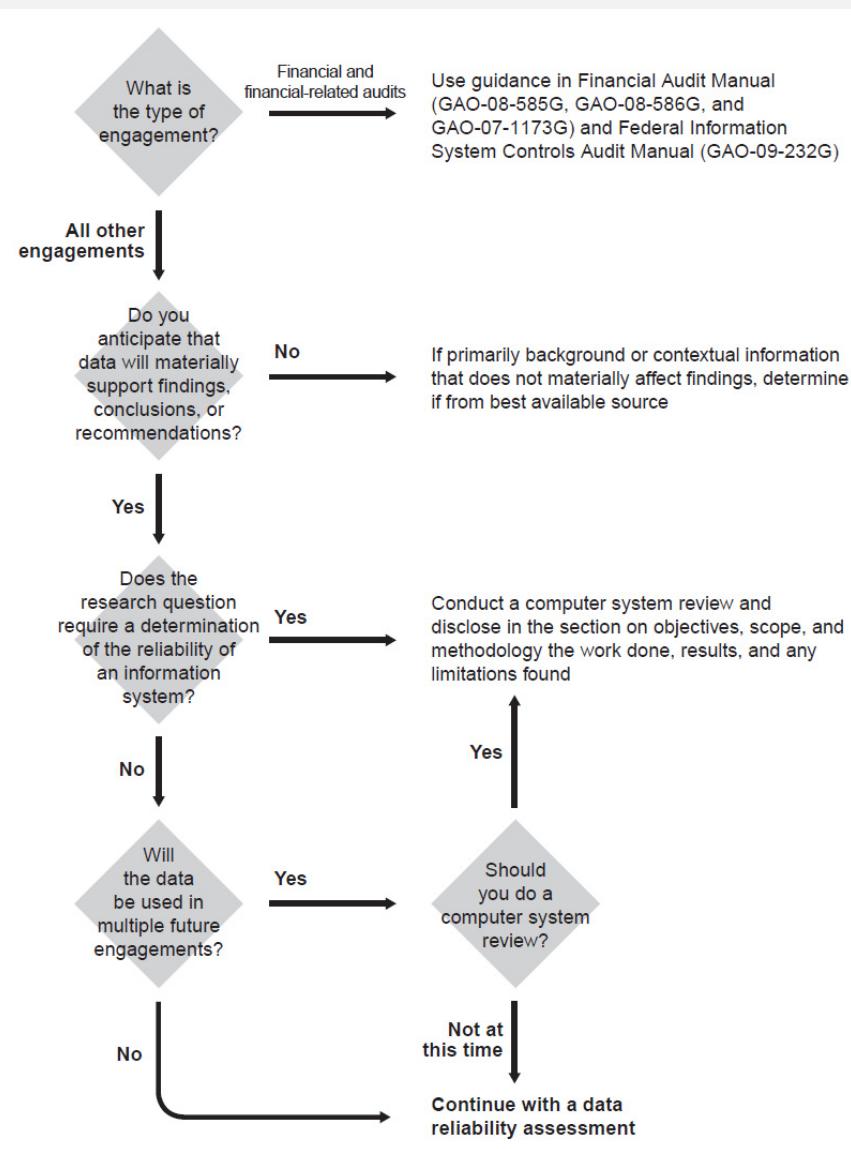
- Data reliability refers to the accuracy and completeness of computer-processed data
- Computer-processed data may be data
 - entered into a computer system or
 - resulting from computer processing

SOME SPECIFIC EXAMPLES OF COMPUTER-PROCESSED DATA

- Data extracts from databases, data warehouses, or data repositories;
- Data maintained in Microsoft excel or access or similar commercial products;
- Data extracts from enterprise software applications supported by information technology departments or contractors;
- Public use data or other replicated detail or summary-level databases accessible through an application other than the original source system;
- Data collected from forms and surveys on web portals; and
- Data summarized in a report or copied from a table in a document.

COMPLETE AND ACCURATE

- Completeness refers to the extent that relevant records are present and the fields in each record are populated appropriately.
- Accuracy refers to the extent that recorded data reflect the actual underlying information.
- Consistency, a subcategory of accuracy, refers to the need to obtain and use data that are clear and well defined enough to yield similar results in similar analyses. For example, if data are entered at multiple sites, inconsistent interpretation of data entry rules can lead to data that, taken as a whole, are unreliable.



CONDITIONS REQUIRING DATA RELIABILITY ASSESSMENT

- You should assess reliability if the data to be analyzed are intended to materially support your findings, conclusions, or recommendations.

CONDITIONS NOT REQUIRING DATA RELIABILITY ASSESSMENT

- You do not need to assess the reliability of data if their use in the report does not materially affect findings, conclusions, or recommendations.

QUESTION

- You have a finding that includes the number of uninsured Americans. To put that number in context, you report the overall U.S. population.
- Is an assessment needed for:
 - The estimate of the number of Americans who are uninsured
Yes/No
 - The estimate of the U.S. population as determined by the U.S. Census
Yes/No

FINANCIAL AUDITS

- You should not follow this guidance in assessing data reliability.
- For financial audits, which include financial statements and financial-related audits, you should follow the *Financial Audit Manual and the Federal Information System Controls Audit Manual*

GOAL

- Can you use the data to answer the research questions
 - Depends on
 - Expected importance of the data to the final report,
 - Strength or weakness of any corroborating evidence, and
 - Anticipated level of risk in using the data.

EXPECTED IMPORTANCE OF THE DATA TO THE FINAL REPORT

- Will the project team depend on the data alone to answer a research question?
- Will the data be summarized or will detailed information be reported?
- Is it important to have precise data?

CORROBORATING EVIDENCE

- Consistent with yellow book standards of evidence - sufficiency and appropriateness;
- Able to provide crucial support;
- Drawn from multiple sources;
- Drawn from multiple types of evidence, such as testimonial, documentary, and physical; and
- Independent of other sources.

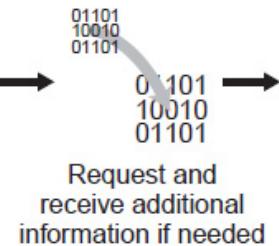
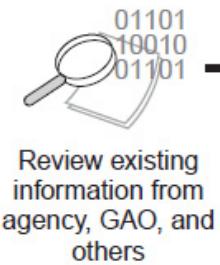
RISK LEVEL IN USING THE DATA

- Could be used to inform legislation, policy, or a program that could have substantial effect;
- Could be used to inform important decisions by individuals or organizations with an interest in the subject;
- Will be the basis for numbers that are likely to be widely quoted
- Are relevant to a sensitive or controversial subject;
- Have been judged for their quality by experts or external stakeholders who have taken positions on the information

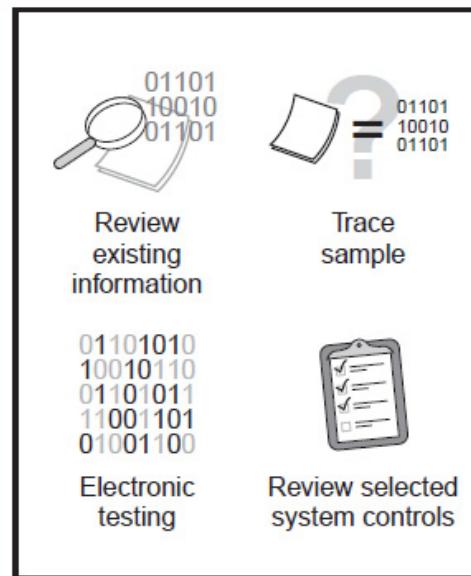
All phases of assessment are influenced by

- importance of data to message,
- strength of corroborating evidence, and
- risk of using data

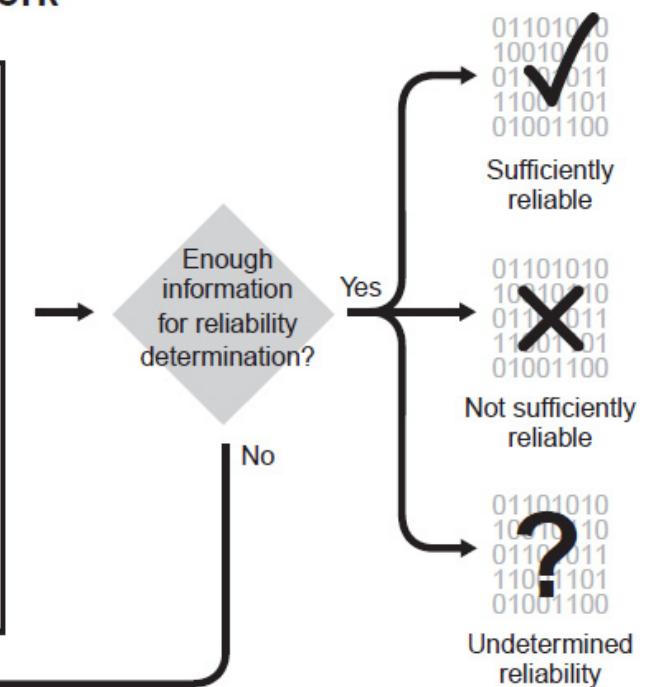
Plan the assessment



Perform data assessment with appropriate mix of work



Make determination



PLANNING A DATA RELIABILITY ASSESSMENT

- Timing an Assessment
 - Early as possible
- Level of Detail of the Data
 - Record-level data
 - Summary-level data
- Documenting the Assessment

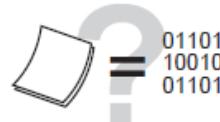
STEPS IN THE ASSESSMENT



Review existing
information

01101
10010
01101
01101010
10010110
01101011
11001101
01001100

Electronic
testing



Trace
sample



Review selected
system controls

REVIEWING EXISTING INFORMATION

- Helps you determine what is already known about the data and the computer processing
- Can indicate both the accuracy and completeness of the entry and processing of the data, as well as how data integrity is maintained
- Can be in the form of reports, studies, or interviews with individuals who are knowledgeable about the data

REVIEWING EXISTING INFORMATION

- GAO, IG, State Auditors, Single Audit Reports
- Agency under review
- Other sources

PERFORMING DATA TESTS

- Data testing can be done by applying logical tests to electronic data files or paper copies of reports. For record-level electronic data, you can use computer programs to test all entries of key data elements in an entire data file.
- For paper copy or summarized data—provided by the agency or retrieved from the Internet—ask for the electronic data file that was used to create them. If you are unable to obtain electronic data, use the paper copy or summarized data and, to the extent possible, manually apply the tests to all instances of key data elements or, if the report or summary is voluminous, to a sample of them.

PERFORMING DATA TESTS

- Checking total number of records provided against agency totals
- Testing for missing data, either entire missing records or missing values in key data elements
- Looking for duplicate records
- Looking for invalid or duplicate identifiers

PERFORMING DATA TESTS

- Testing for values outside a designated range
- Looking for dates outside valid time periods or in an illogical progression
- Following up on troubling aspects of the data—such as extremely high values associated with a certain geographic location—found while analyzing the data
- Testing relationships between data elements (sometimes by merely doing a cross tabulation), such as whether data elements follow a skip pattern from a questionnaire
- Verifying that computer processing is accurate and complete, such as testing a formula used in generating specific data elements, or testing to ensure that edit checks are working correctly.

QUESTION

- You obtained a spreadsheet from the auditee. You confirm record counts and dollar amounts to information provided by the auditee. Are you done testing?

QUESTION

- You obtained a spreadsheet from the auditee. One of the key fields is Receipt Date. What sort of test would you perform on this field?
- After performing your test, you note that there are dates outside the scope of your audit. What do you do?

TRACING TO AND FROM SOURCE DOCUMENTS

- Tracing a sample of data records to source documents helps you determine whether the computer data accurately and completely reflect these documents
- Consider the relative risks of overstating or understating conclusions drawn from the data

TRACING TO AND FROM SOURCE DOCUMENTS

- If you are particularly concerned that questionable cases might not have been entered into the computer system and that, as a result, the degree of compliance may be overstated, consider tracing from source documents to the database.
- If you are more concerned that ineligible cases have been included in the database and that, as a result, the potential problems may be understated, consider tracing from the database back to source documents.

TRACING TO AND FROM SOURCE DOCUMENTS

- Trace only a sample to save time and cost
- Random sample
- Large enough to estimate the error rate within reasonable levels of precision

TRACING TO SOURCE DOCUMENTS

- Consider tracing to source documents when
 - They are available relatively easily or
 - The possible magnitude of error is especially critical.

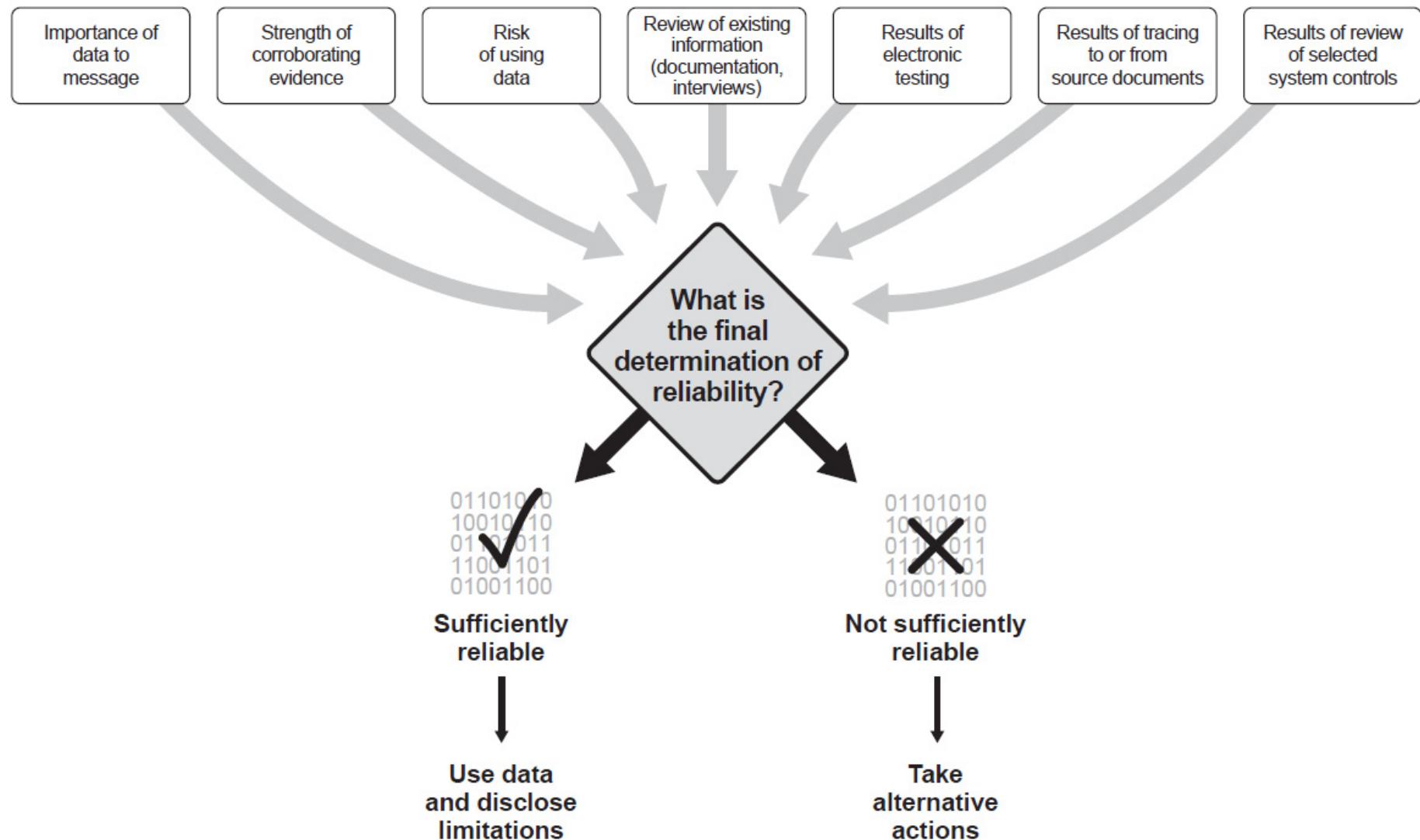
QUESTION

- Same spreadsheet. You performed all tests possible on data. You take a sample of transactions and trace them to source documents.
 - You can't find the source document for a couple transactions. Continue/Stop?
 - Some of the transactions have been amended? Continue/Stop?

REVIEWING SELECTED SYSTEM CONTROLS

- Access Controls
- Application Controls





MAKING THE DATA RELIABILITY DETERMINATION

- Considering the Results of Your Assessment Work
- Outcomes to Consider in the Assessment
 - Sufficiently reliable
 - Not sufficiently reliable
 - Undetermined reliability
- Professional judgement

MAKING THE DATA RELIABILITY DETERMINATION

- Sufficiently reliable
 - The likelihood of significant errors or incompleteness is minimal and
 - The use of the data would not lead to an incorrect or unintentional message

MAKING THE DATA RELIABILITY DETERMINATION

- Not sufficiently reliable
 - Significant errors or incompleteness in some of or all the key data elements and
 - That using the data would probably lead to an incorrect or unintentional message, given the research questions and intended use of the data

MAKING THE DATA RELIABILITY DETERMINATION

- The source is the only source.

INCLUDING APPROPRIATE LANGUAGE IN THE REPORT

- Should include in the report's methodology section a statement about having conformed to generally accepted government auditing standards.
- These standards include the appropriateness of the data being used.
- You conform to GAGAS by discussing in the report what you did to assess the data, disclose any data concerns, and make a judgment about the reliability of the data used in the report.

INCLUDING APPROPRIATE LANGUAGE IN THE REPORT

- Sufficiently Reliable Data
 - noting the kind of assessment you relied on,
 - explaining the steps in the assessment,
 - describing any corrections made to the data, and
 - disclosing any data limitations.

INCLUDING APPROPRIATE LANGUAGE IN THE REPORT

- Not Sufficiently Reliable Data
 - describe the problems with the data, as well as why using them would probably lead to an incorrect or unintentional message, and
 - state that the data problems are significant or potentially significant

INCLUDING APPROPRIATE LANGUAGE IN THE REPORT

- Data of Undetermined Reliability
 - Include such factors as the deletion of original computer files,
 - data limitations that prevent an adequate assessment, short time periods,
 - and the lack of access to the data source or to needed documents.
 - Explain the reasonableness of using the data

INCLUDING APPROPRIATE LANGUAGE IN THE REPORT

- We assessed the reliability of _____ data by (1) performing electronic testing of required data elements, (2) reviewing existing information about the data and the system that produced them, and (3) interviewing agency officials knowledgeable about the data. We determined that the data were sufficiently reliable for the purposes of this report.
- We assessed the reliability of _____ data by (1) performing electronic testing of required data elements, (2) reviewing existing information about the data and the system that produced them, and (3) interviewing agency officials knowledgeable about the data. In addition, we traced a statistically random sample of data to source documents (see appendix x for details). We determined that the data were sufficiently reliable for the purposes of this report.

INCLUDING APPROPRIATE LANGUAGE IN THE REPORT

- To assess the reliability of the data elements needed to answer the engagement objectives, we (1) performed electronic testing of required data elements, (2) reviewed related documentation, and (3) interviewed agency officials knowledgeable about the data. The results of our electronic testing showed that data elements key to our review contained high percentages of missing data. (See appendix x for further details.) Therefore, we determined that the data were not sufficiently reliable for the purposes of this report..

SSA MASTER DEATH FILE

SSA MASTER DEATH FILE

- Contains the complete and official SSA database extract, as well as updates to the full file of persons reported to SSA as being deceased.
- The Social Security Death Master file contains the following information:
 - Social Security Number, Name, Date of Birth, Date of Death, Proof/Verified Code, State (or country) of last known residence, and ZIP code of last lump sum payment.

SSA MASTER DEATH FILE

- SSA authorizes the use of this database as an identity verification tool, but notes that the Death Master File may contain inaccuracies.
- SSA cannot guarantee the accuracy of the Death Master File. Therefore, the absence of a particular person on this file is not proof that the individual is alive. Further, in rare instances, it is possible for the records of a person who is not deceased to be included erroneously in the Death Master File.

IMPORTANCE

- Sources of Date of Death
 - Match to claims with:
 - Discharge Status Code = Death
 - Medicare Enrollment Database
 - SSA Master Death File

CORROBORATING EVIDENCE

- Per SSA DOD = 07/15/2007

ENTER "D" (DETAIL), "S" (SUMMARY) OR "H" (HIGHLIGHT) IN REC TO SELECT
F2=MENU F3=RETURN F4=BENA F5=BENB F7=BWD F8=FWD F12=EXIT 08/28/09

HOSH	HOSPICE CLAIM HISTORY	PAGE 1 OF 6
HIC xxxxxxxxxx	NORTHEAST	DOB 06/30/1916 SEX F
CORR xxxxxxxxxx	NAME xxxxxxxxxxxx	DOD 07/11/07 SOURCE 3
BOI 1 CMN 3 HHB 1 HHE 1 HOS 1 MSP 1 REP 2 SUR 1	INP 0008 OUT 0000 HOS 0002 PTB 0045 DME 0001 HHA 0026 BLK N	
ICN 20721900053002	PROV 331529	INTER 00450NYR FROM 07/01/07
CANCEL ADJ CODE BILL 824	PAT STAT 41	THRU 07/11/07
PMT DIST PROV 8,048.91	PTIMB AMT 8,048.91	TOT CHRG 8,048.91
RECEIVED DTE 08/06/07	APPROVED DTE 10/24/07	SCHED PMT DTE 10/26/07
NOPAY CD: DOA 10/25/07 FORM: UB92	CANC	EXCP/NONEXCP IND 0
DEMO NUMBER UNSOL IND PRICER CD 00 NPI 1770669038	NPI PLCHDR	
ADJ: MASS CLM LIAB IND SANC-IND RAC FAC ZIP		WC-IND

RISK

- Risk in using it
- Risk of not using it
 - Overpayments not identified
 - Potential fraud continues

QUICK RESPONSE EVALUATION

*Sources of Erroneous
Death Entries Input
into the Death Master File*

A-06-09-29095



January 2009

EXISTING EVIDENCE

Sources of 7,597 Erroneous Death Reports

Description	Transactions	Percent
SSA Staff Death Entry Input	6,754	88.9
First-Party Death Report to SSA	311	4.1
Funeral Home Death Report to SSA	257	3.4
EDR State	113	1.5
Non-EDR State	53	0.7
Veterans Affairs	54	0.7
Other	55	0.7
Total	7,597	100.0

OFFICE OF
THE INSPECTOR GENERAL

SOCIAL SECURITY ADMINISTRATION

PAYMENTS TO INDIVIDUALS
WHOSE NUMIDENT RECORD
CONTAINS A DEATH ENTRY

June 2009 A-06-08-18095

AUDIT REPORT





United States Government Accountability Office
Report to Congressional Requesters

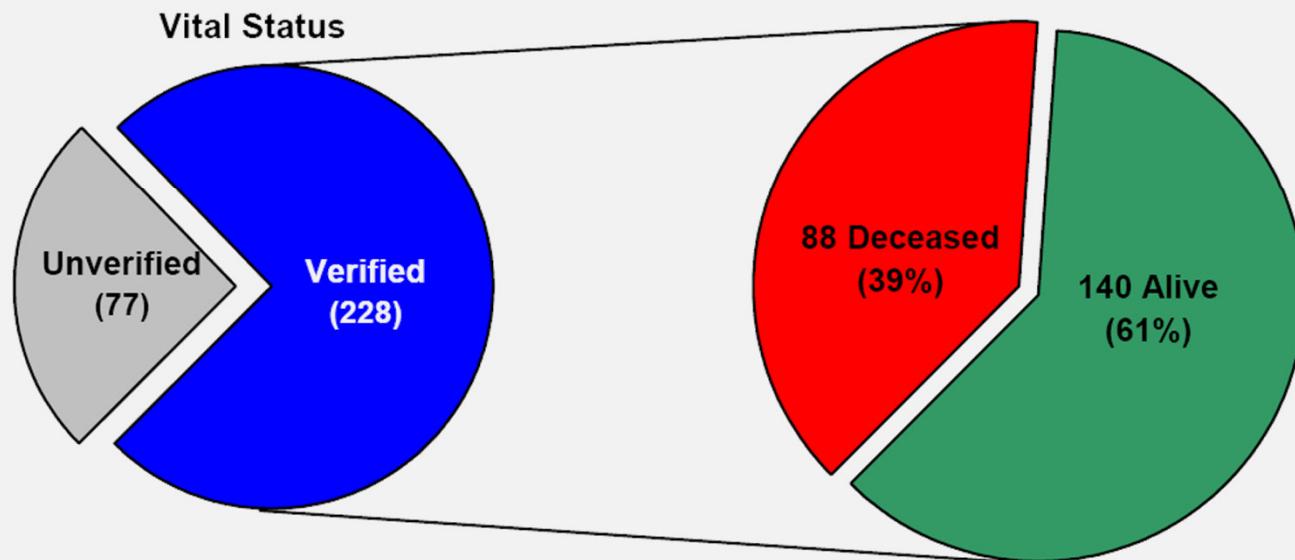
November 2013

SOCIAL SECURITY DEATH DATA

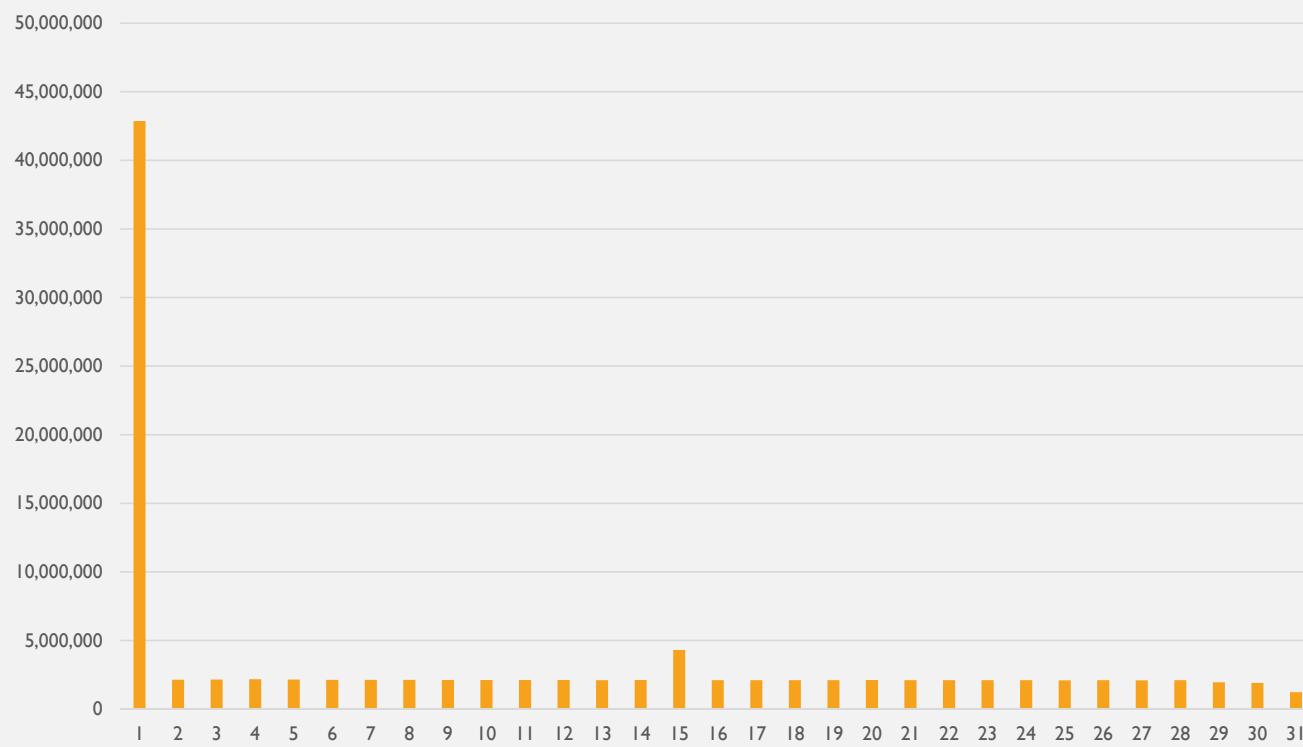
Additional Action
Needed to Address
Data Errors and
Federal Agency
Access

EXISTING EVIDENCE

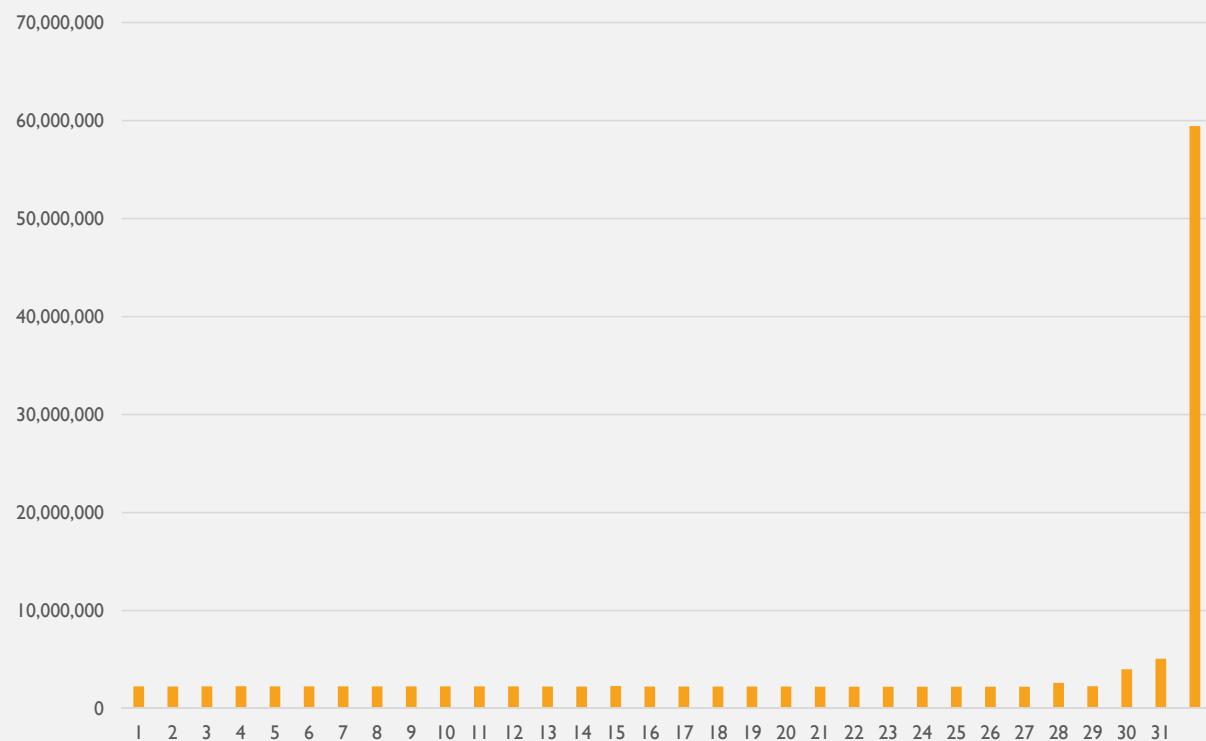
305 Beneficiaries in Current Pay Status
Whose Numident Contained a Death Entry
As of January 2008



ELECTRONIC TESTING



ELECTRONIC TESTING



***FEDERAL INFORMATION SECURITY
MANAGEMENT ACT REPORT***

**Fiscal Year 2008 Evaluation of the Social Security
Administration's Compliance with the
Federal Information Security Management Act**



September 2008 A-14-08-18063

Patrick P. O'Carroll, Jr. – Inspector General

MEDICARE PART A AND B DEATH MATCH

- Objective:
 - To determine the amount of Medicare fee-for-service payments where
 - From DOS > Date of Death

MEDICARE PART A AND B DEATH MATCH

- Matched Medicare claims data to CMS Enrollment database
 - \$67 million
- Matched Medicare claims data to SSA Master Death File
 - \$172 million

MEDICARE PART A AND B DEATH MATCH

	Amount	Claims
DME	\$10.1 million	67,789
HHA	\$3.3 million	1,441
Hospice	\$8.1 million	3,949
Inpatient	\$96.3 million	8,052
Outpatient	\$9.1 million	23,906
Physician Supplier	\$32.9 million	241,535
SNF	\$12.1 million	2,633