United States General Accounting Office

GAO

Applied Research and Methods

October 2002
External Version 1

Assessing the Reliability of Computer-Processed Data

In July 2009, this product was superseded by GAO-09-680G, *Applied Research and Methods: Assessing the Reliability of Computer-Processed Data*, available at http://www.gao.gov/products/GAO-09-680G.

GAO-03-273G

Contents

Preface		iii
Section 1: Introduction		1
Section 2: Understanding Data Reliability		3
Section 3: Deciding If a Data Reliability Assessment Is Necessary		6
	Conditions Requiring a Data Reliability Assessment	8
	Conditions Not Requiring a Data Reliability Assessment	8
Section 4: Performing a Data Reliability Assessment		10
	Timing the Assessment	10
	Documenting the Assessment	10
Section 5: Viewing the Entire Assessment Process		12
Section 6: Taking the First Steps		14
	Reviewing Existing Information	14
	Performing Initial Testing	15
	Dealing with Short Time Frames	16
Section 7: Making the Preliminary Assessment		18
	Factors to Consider in the Assessment	18
	Outcomes to Consider in the Assessment	21

Section 8: Conducting Additional Work 23
Tracing to and from Source Documents 24
Using Advanced Electronic Testing 25
Reviewing Selected System Controls 26
Using Data of Undetermined Reliability 27

Section 9: Making the Final Assessment 28
Sufficiently Reliable Data 29
Not Sufficiently Reliable Data 29
Data of Undetermined Reliability 30

Section 10: Including Appropriate Language in the Report 31
Sufficiently Reliable Data 31
Not Sufficiently Reliable Data 31
Data of Undetermined Reliability 32

Glossary of Technical Terms 33

Figures
Figure 1: Factors to Consider in Making the Decision on Using the Data 1
Figure 2: Decision Process for Determining If a Data Reliability Assessment Is Required 7
Figure 3: Data Reliability Assessment Process 13
Figure 4: The First Steps of the Assessment 14
Figure 5: The Preliminary Assessment 19
Figure 6: Choosing and Conducting Additional Work 23
Figure 7: Making the Final Assessment 28

Preface

Computer-processed data, often from external sources, increasingly underpin audit reports, including evaluations (performance audits) and financial audits. Therefore, the reliability of such data has become more and more important. Historically, computer-processed data have been treated as unique evidence. However, these data are simply one form of evidence relied on, although they may require more technical assessment than other forms of evidence. In addition, the very nature of the information system creating the data allows opportunities for errors to be introduced by many people.

This guidance is intended to demystify the assessment of computer-processed data. It supplements GAO's "Yellow Book" (*Government Auditing Standards*, 1994 Revision), which defines the generally accepted government auditing standards (GAGAS), and replaces the earlier GAO guidance, *Assessing the Reliability of Computer-Processed Data* (GAO/OP-8.1.3, Sept. 1990).

For all types of evidence, various tests are used—sufficiency, competence, and relevance—to assess whether the evidence standard is met. You probably have been using these tests for years and have become quite proficient at them. But because assessing computer-processed data requires more technical tests, it may appear that such data are subject to a higher standard of testing than other evidence. That is not the case. For example, many of the same tests of sufficiency and relevance are applied to other types of evidence. But in assessing computer-processed data, the focus is on one test in the evidence standard—competence—which includes validity and reliability. Reliability, in turn, includes the completeness and accuracy of the data.

This guidance, therefore, provides a flexible, risk-based framework for data reliability assessments that can be geared to the specific circumstances of each engagement. The framework also provides a structure for planning and reporting, facilitates bringing the right mix of skills to each engagement, and ensures timely management buy-in on assessment strategies. The framework is built on

- making use of all existing information about the data,
- performing at least a minimal level of data testing,

- doing only the amount of work necessary to determine whether the data are reliable enough for our purposes,
- maximizing professional judgment, and
- bringing the appropriate people, including management, to the table at key decision points.

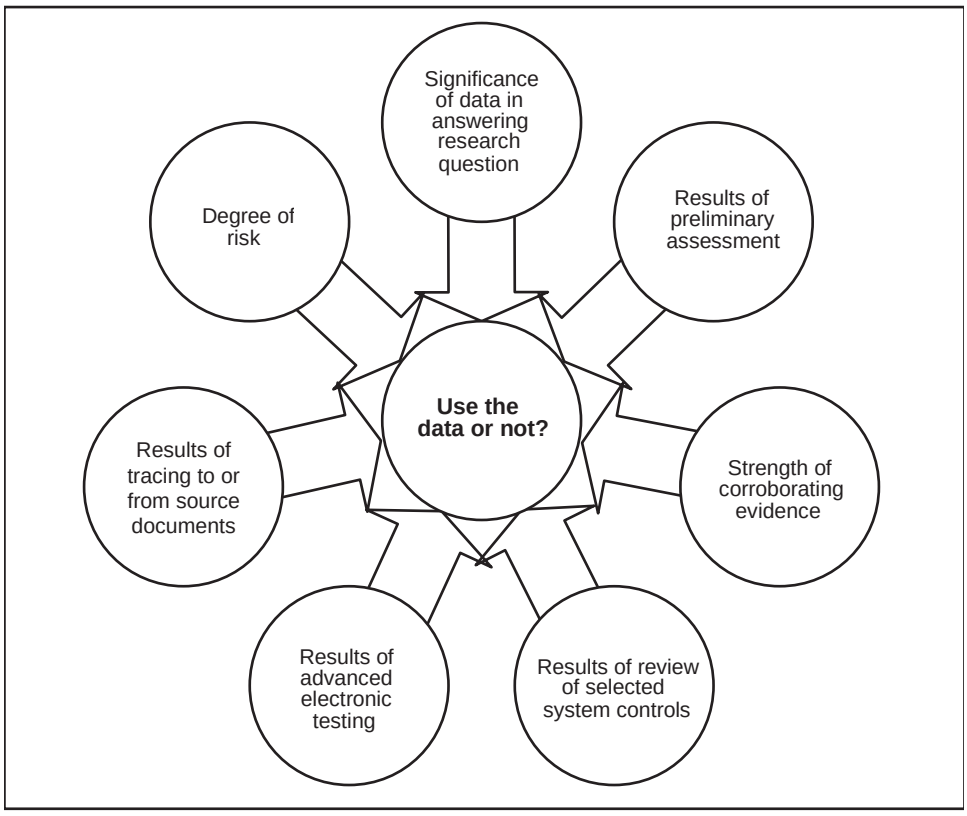
The ultimate goal of the data reliability assessment is to determine whether you can use the data for your intended purposes. This guidance is designed to help you make an appropriate, defensible assessment in the most efficient manner. With any related questions, call Barbara Johnson, focal point for data reliability issues, at (202) 512-3663, or Barry Seltser, the Acting Director of GAO's Center for Design, Methods, and Analysis, at (202) 512-3234.

Nancy Kingsbury
Managing Director, Applied Research
and Methods

Section 1: Introduction

This guidance explains what data reliability means and provides a framework for assessing the reliability of computer-processed data. It begins with the steps in a preliminary assessment, which, in many cases, may be all you need to do to assess reliability. This guidance also helps you decide whether you should follow up the preliminary assessment with additional work. If so, it explains the steps in a final assessment and the actions to take, depending on the results of your additional work. The ultimate goal in determining data reliability is to make the following decision: For our engagement, can we use the data to answer the research question? See figure 1 for an overview of the factors that help to inform that decision. Not all of these factors may be necessary for all engagements.

Figure 1: Factors to Consider in Making the Decision on Using the Data



Source: GAO.

In addition, this guidance discusses suggested language—appropriate under different circumstances—for reporting the results of your assessment. Finally, it provides detailed descriptions of all the stages of the assessment, as well as a glossary of technical terms used (see p. 33). An on-line version of this guidance, which will include tools that may help you in assessing reliability, is currently being developed. The overall process is illustrated in figures 2 (p. 7) and 3 (p. 13).

Section 2: Understanding Data Reliability

Data reliability refers to the accuracy and completeness of computer-processed data, given the intended purposes for use. Computer-processed data include data (1) entered into a computer system and (2) resulting from computer processing. Computer-processed data can vary in form—from electronic files to tables in published reports. The definition of computer-processed data is therefore broad. In this guidance, the term data always refers to computer-processed data.

The "Yellow Book" requires that a data reliability assessment be performed for all data used as support for engagement findings, conclusions, or recommendations.[1] This guidance will help you to design a data reliability assessment appropriate for the purposes of the engagement and then to evaluate the results of the assessment.

Data are reliable when they are (1) complete (they contain all of the data elements and records needed for the engagement)[2] and (2) accurate (they reflect the data entered at the source or, if available, in the source documents). A subcategory of accuracy is consistency. Consistency 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 rules can lead to data that, taken as a whole, are unreliable. Reliability also means that for any computer processing of the data elements used, the results are reasonably complete and accurate, meet your intended purposes, and are not subject to inappropriate alteration.

Assessments of reliability should be made in the broader context of the particular characteristics of the engagement and the risk associated with the possibility of using data of insufficient reliability. Reliability does not mean that computer-processed data are error-free. Errors are considered acceptable under these circumstances: You have assessed the associated risk and found the errors are not significant enough to cause a reasonable person, aware of the errors, to doubt a finding, conclusion, or recommendation based on the data.

[1]U.S. General Accounting Office, *Government Auditing Standards*, GAO/OGC-94-4 (Washington, D.C.: June 1994), pp. 62-87.

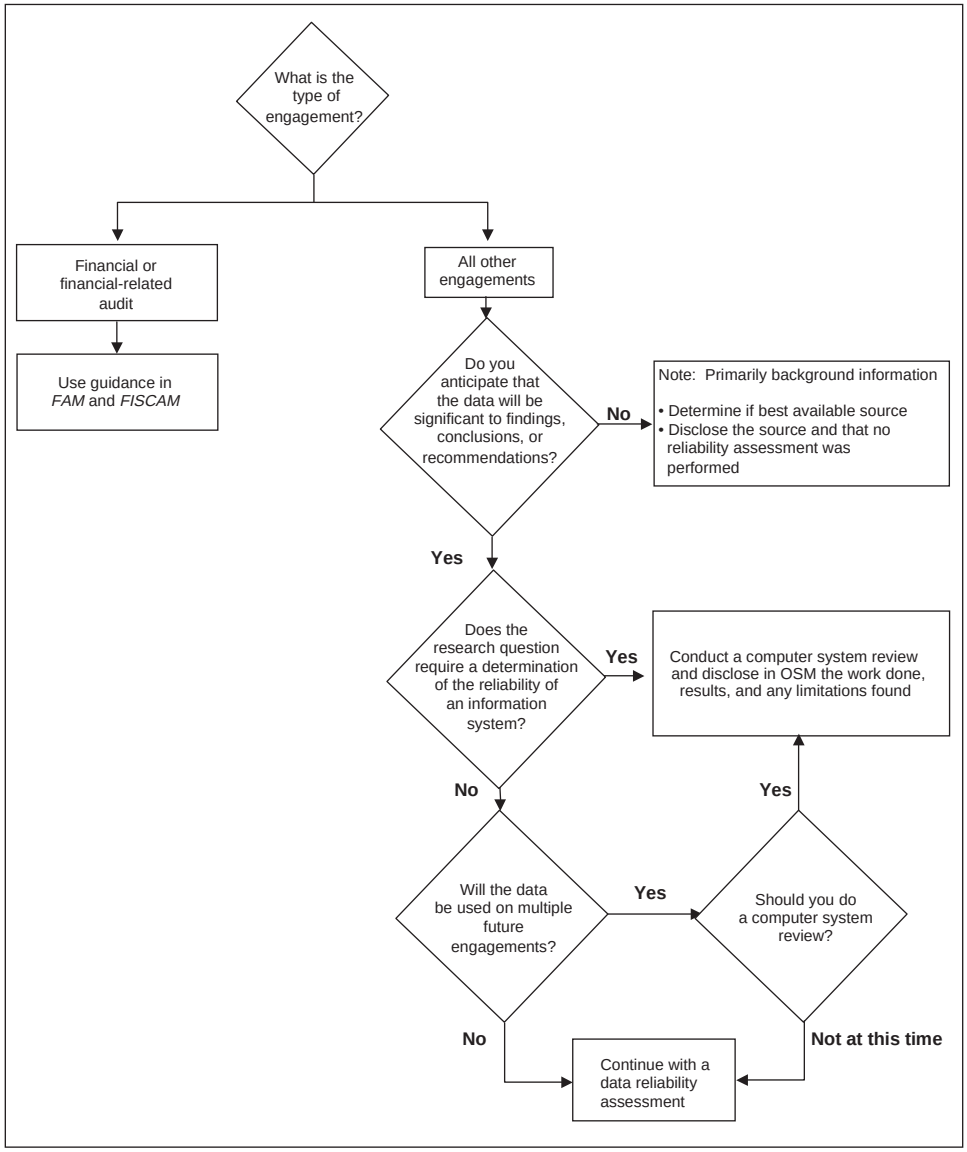
[2]A data element is a unit of information with definable parameters (for example, a Social Security number), sometimes referred to as a *data variable* or *data field*.

While this guidance focuses only on the reliability of data in terms of accuracy and completeness, other data quality considerations are just as important. In particular, you should also consider the validity of data. Validity (as used here) refers to whether the data actually represent what you think is being measured. For example, if a data field is named "annual evaluation score," is this an appropriate measure of a person's job performance? Considerations of data validity and reliability issues should be addressed early in the engagement, and appropriate technical specialists—such as data analysts, statisticians, or information technology specialists—should be consulted.

Section 3: Deciding If a Data Reliability Assessment Is Necessary

To decide if a data reliability assessment is necessary, you should consider certain conditions. The engagement type and planned use of the data help to determine when you should assess data reliability. See figure 2 for an illustration of the decision process that you should use.

Figure 2: Decision Process for Determining If a Data Reliability Assessment Is Required



Source: GAO.

Conditions Requiring a Data Reliability Assessment

You should assess reliability if the data to be analyzed are intended to support the engagement findings, conclusions, or recommendations. Keep in mind that a finding may include only a description of the condition, as in a purely descriptive report. In the audit plan for the engagement, you should include a brief discussion of how you plan to assess data reliability, as well as any limitations that may exist due to shortcomings in the data.

Conditions Not Requiring a Data Reliability Assessment

You do not need to assess reliability if the data are used (1) only as background information or (2) in documents without findings, conclusions, or recommendations. Background information generally sets the stage for reporting the results of an engagement or provides information that puts the results in proper context. Such information could be the size of the program or activity you are reviewing, for example. When you gather background or other data, ensure that they are from the best available source(s). When you present the data, cite the source(s) and state that the data were not assessed.

Sometimes, as a best practice, however, you may want to do some assessment of background data. Your judgment of the data's importance and the reliability of the source, as well as other engagement factors, can help you determine the extent of such an assessment.

Finally, for financial audits and information system reviews, you should not follow this guidance in assessing data reliability. For financial audits, which include financial statement and financial-related audits, you should follow the *GAO/PCIE Financial Audit Manual (FAM)* and the *Federal Information System Controls Audit Manual (FISCAM)*. In an information system review, all controls in a computer system, for the full range of application functions and products, are assessed and tested. Such a review includes (1) examining the general and application controls of a computer system,[3] (2) testing whether those controls are being complied with, and (3) testing data produced by the system.[4] To design such a review, appropriate to the research question, seek assistance from information technology specialists.

[3]*General controls* refers to the structure, policies, and procedures—which apply to all or a large segment of an organization's information systems—that help to ensure proper operation, data integrity, and security. *Application controls* refers to the structure, policies, and procedures that apply to individual application systems, such as inventory or payroll.

[4]Guidance for carrying out reviews of general and application controls is provided in the U.S. General Accounting Office, *Federal Information System Controls Audit Manual*, GAO/AIMD-12.19.6 (Washington, D.C.: Jan. 1999).

Section 4: Performing a Data Reliability Assessment

To perform a data reliability assessment, you need to decide on the timing—when to perform the assessment—and how to document it.

Timing the Assessment

A data reliability assessment should be performed as early as possible in the engagement process, preferably during the design phase. The audit plan should reflect data reliability issues and any additional steps that still need to be performed to assess the reliability of critical data. The engagement team generally should not finalize the audit plan or issue a commitment letter until it has done initial testing and reviewed existing information about the data and the system that produces the data. In addition, the team should not commit to making conclusions or recommendations based on the data unless the team expects to be satisfied with the data reliability.

Documenting the Assessment

All work performed as part of the data reliability assessment should be documented and included in the engagement workpapers. This includes all testing, information review, and interviews related to data reliability. In addition, decisions made during the assessment, including the final assessment of whether the data are sufficiently reliable for the purposes of the engagement, should be summarized and included with the workpapers. These workpapers should be (1) clear about what steps the team took and what conclusions they reached and (2) reviewed by staff with appropriate skills or, if needed, technical specialists.

Section 5: Viewing the Entire Assessment Process

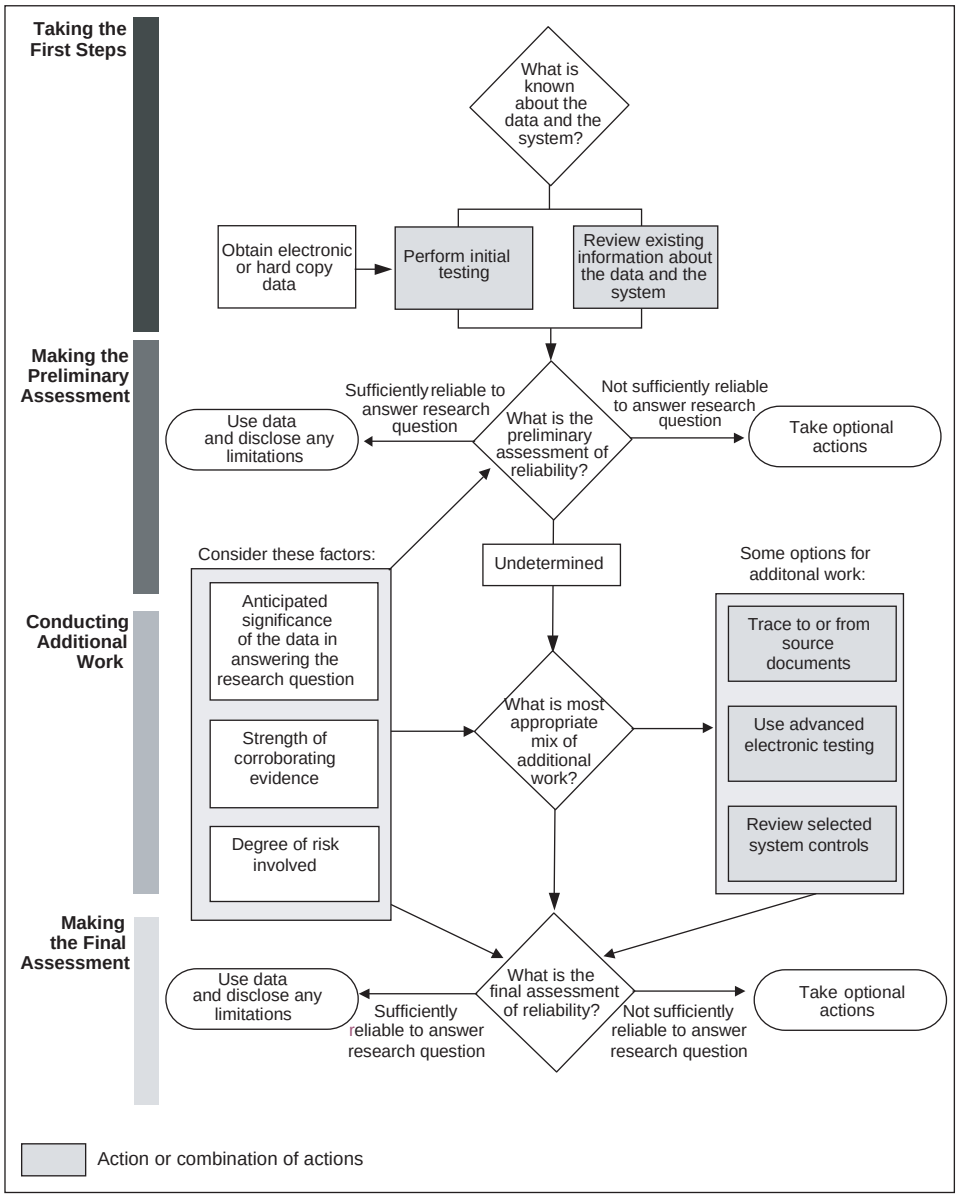
The ultimate goal of the data reliability assessment is to determine whether you can use the data to answer the research question. The assessment should be performed only for those portions of the data that are relevant to the engagement. The extensiveness of the assessment is driven by

- the expected significance of the data to the final report,
- the anticipated risk level of using the data, and
- the strength or weakness of any corroborating evidence.

Therefore, the specific assessment process should take into account these factors along with what is learned during the initial stage of the assessment. The process is likely to be different for each engagement.

The overall framework of the process for data reliability assessment is shown in figure 3. The framework identifies several key stages in the assessment, as well as actions and decisions expected as you move through the process. The framework allows you to identify the appropriate mix of assessment steps to fit the particular needs of your engagement. In most cases, all of the elements in figure 3 would not be necessary in completing the assessment. Specific actions for each stage are discussed in sections 6-10.

Figure 3: Data Reliability Assessment Process



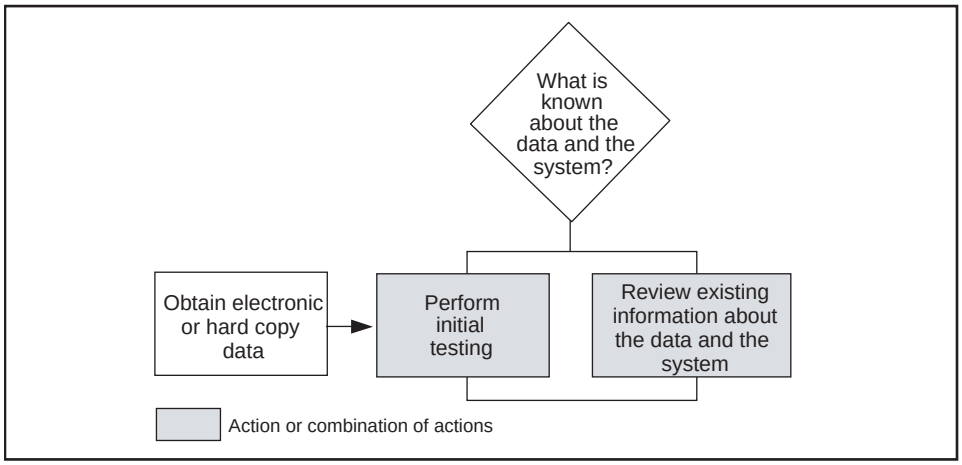
Source: GAO.

Section 6: Taking the First Steps

The data reliability process begins with two relatively simple steps. These steps provide the basis for making a preliminary assessment of data reliability: (1) a review of related information and (2) initial testing (see figure 4). In some situations, you may have an extremely short time frame for the engagement; this section also provides some advice for this situation.

The time required to review related information and perform initial testing will vary, depending on the engagement and the amount of risk involved. As discussed in section 4, these steps should take place early in the engagement and include the team members, as well as appropriate technical staff.

Figure 4: The First Steps of the Assessment



Source: GAO.

Reviewing Existing Information

The first step—a review of existing information—helps you to determine what is already known about the data and the computer processing. The related information you collect can indicate both the accuracy and completeness of the entry and processing of the data, as well as how data integrity is maintained. This information can be in the form of reports, studies, or interviews with individuals who are knowledgeable about the data and the system. Sources for related information include GAO, the agency under review, and others.

GAO

GAO may already have related information in reports. Those from fiscal year 1995 to the present are available via GAO's Internet site. This site also provides other useful information: for example, as part of the annual governmentwide consolidated financial audit, GAO's Information Technology Team is involved with reporting on the effectiveness of controls for financial information systems at 24 major federal agencies.

Agency under Review

Officials of the agency or entity under review are aware of evaluations of their computer data or systems and usually can direct you to both. However, keep in mind that information from agency officials may be biased. Consider asking appropriate technical specialists to help in evaluating this information.

Agency information includes Inspector General reports, Federal Managers' Financial Integrity Act reports, Government Performance and Results Act (GPRA) plans and reports, Clinger-Cohen Act reports, and Chief Information Officer reports. (Some of this information can be found in agency homepages on the Web.)

Others

Other organizations and users of the data may be sources of relevant information. To help you identify these sources, you can use a variety of databases and other research tools, which include the Congressional Research Service Public Policy Literature Abstracts and organizations' Web sites.

Performing Initial Testing

The second step—initial testing—can be done by applying logical tests to electronic data files or hard copy reports. For electronic data, you use computer programs to test all entries of key data elements in the entire data file.[5] Keep in mind that you only test those data elements you plan to use for the engagement. You will find that testing with computer programs often takes less than a day, depending on the complexity of the file. For

[5] Though an in-depth discussion of quality-assurance practices to be used in electronic testing and analyses is beyond the scope of this guidance, it is important to perform appropriate checks to ensure that you have obtained the correct file. All too often, analysts receive an incorrect file (an early version or an incomplete file). Appropriate steps would include counting records and comparing totals with the responsible agency or entity.

hard copy or summarized data—provided by the audited entity or retrieved from the Internet—you can ask for the electronic data file used to create the hard copy or summarized data. If you are unable to obtain electronic data, use the hard 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.

Whether you have an electronic data file or a hard copy report or summary, you apply the same types of tests to the data. These can include testing for

- missing data, either entire records or values of key data elements;
- the relationship of one data element to another;
- values outside of a designated range; and
- dates outside valid time frames or in an illogical progression.

Be sure to keep a log of your testing for inclusion in the engagement workpapers.

Dealing with Short Time Frames

In some instances, the engagement may have a time frame that is too short for a complete preliminary assessment, for example, a request for testimony in 2 weeks. However, given that all engagements are a function of time, as well as scope and resources, limitations in one require balancing the others.

Despite a short time frame, you may have time to review existing information and carry out testing of data that are critical for answering a research question, for example: You can question knowledgeable agency staff about data reliability or review existing GAO or Inspector General reports to quickly gather information about data reliability issues. In addition, electronic testing of critical data elements for obvious errors of completeness and accuracy can generally be done in a short period of time on all but the most complicated or immense files. From that review and testing, you will be able to make a more informed determination about whether the data are sufficiently reliable to use for the purposes of the engagement. (See sections 7 and 8 for the actions to take, depending on your determination.)

Section 7: Making the Preliminary Assessment

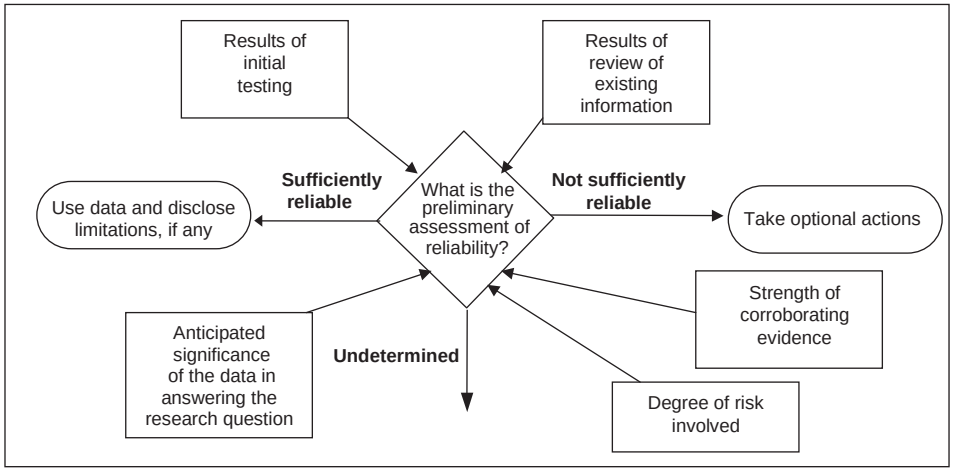
The preliminary assessment is the first decision point in the assessment process, including the consideration of multiple factors, a determination of the sufficiency of the data reliability with what is known at this point, and a decision about whether further work is required. You will decide whether the data are sufficiently reliable for the purposes of the engagement, not sufficiently reliable, or as yet undetermined. Keep in mind that you are not attesting to the overall reliability of the data or database. You are only determining the reliability of the data as needed to support the findings, conclusions, or recommendations of the engagement. As you gather information and make your judgments, consult appropriate technical specialists for assistance.

Factors to Consider in the Assessment

To make the preliminary assessment of the sufficiency of the data reliability for the engagement, you should consider all factors related to aspects of the engagement, as well as assessment work performed to this point. As shown in figure 5, these factors include

- the expected significance of the data in the final report,
- corroborating evidence,
- level of risk, and
- the results of initial assessment work.

Figure 5: The Preliminary Assessment



Source: GAO.

Expected Significance of the Data in the Final Report

In making the preliminary assessment, consider the data in the context of the final report: Will the engagement team depend on the data alone to answer a research question? Will the data be summarized or will detailed information be required? Is it important to have precise data, making magnitude of errors an issue?

Corroborating Evidence

You should consider the extent to which corroborating evidence is likely to exist and will independently support your findings, conclusions, or recommendations. Corroborating evidence is independent evidence that supports information in the database. Such evidence, if available, can be found in the form of alternative databases or expert views. It is unique to each engagement, and its strength—persuasiveness—varies.

For help in deciding the strength or weakness of corroborating evidence, consider the extent to which the corroborating evidence

- is consistent with the "Yellow Book" standards of evidence—sufficiency, competence, and relevance;
- provides crucial support;

- is drawn from different types of sources—testimonial, documentary, physical, or analytical; and
- is independent of other sources.

Level of Risk

Risk is the likelihood that using data of questionable reliability could have significant negative consequences on the decisions of policymakers and others. To do a risk assessment, consider the following risk conditions:

- The data could be used to influence legislation, policy, or a program that could have significant impact.
- The data could be used for significant decisions by individuals or organizations with an interest in the subject.
- The data will be the basis for numbers that are likely to be widely quoted, for example, "In 1999, the United States owed the United Nations about $1.3 billion for the regular and peacekeeping budgets."
- The engagement is concerned with a sensitive or controversial subject.
- The engagement has external stakeholders who have taken positions on the subject.
- The overall engagement risk is medium or high.
- The engagement has unique factors that strongly increase risk.

Bear in mind that any one of the conditions may have more importance than another, depending on the engagement.

Results of Initial Assessment Work

At this point, as shown in figure 5 (p. 19), the team will already have performed the initial stage of the data reliability assessment. They should have the results from the (1) review of all available existing information about the data and the system that produced them and (2) initial testing of the critical data elements. These results should be appropriately documented and reviewed before the team enters into the decision-making phase of the preliminary assessment. Because the results will, in whole or in part, provide the evidence that the data are sufficiently reliable—and therefore competent enough—or not sufficiently reliable for the purposes

of the engagement, the workpapers should include documentation of the process and results.

Outcomes to Consider in the Assessment

The results of your combined judgments of the strength of corroborating evidence and degree of risk suggest different assessments. If the corroborating evidence is strong and the risk is low, the data are more likely to be considered sufficiently reliable for your purposes. If the corroborating evidence is weak and the risk is high, the data are more likely to be considered not sufficiently reliable for your purposes. The overall assessment is a judgment call, which should be made in the context of discussion with team management and technical specialists.

The preliminary assessment categorizes the data as sufficiently reliable, not sufficiently reliable, or of undetermined reliability. Each category has implications for the next steps of the data reliability assessment.

When to Assess Data as Sufficiently Reliable for Engagement Purposes

You can assess the data as sufficiently reliable for engagement purposes when you conclude the following: Both the review of related information and the initial testing provide assurance that (1) the likelihood of significant errors or incompleteness is minimal and (2) the use of the data would not lead to an incorrect or unintentional message. You could have some problems or uncertainties about the data, but they would be minor, given the research question and intended use of the data. When the preliminary assessment indicates that the data are sufficiently reliable, use the data.

When to Assess Data as Not Sufficiently Reliable for Engagement Purposes

You can assess the data as not sufficiently reliable for engagement purposes when you conclude the following: The review of related information or initial testing indicates that (1) significant errors or incompleteness exist in some or all of the key data elements and (2) using the data would probably lead to an incorrect or unintentional message.

When the preliminary assessment indicates that the data are not sufficiently reliable, you should seek evidence from other sources, including (1) alternative computerized data—the reliability of which you should also assess—or (2) original data in the form of surveys, case studies, or expert interviews.

You should coordinate with the requester if seeking evidence from other sources does not result in a source of sufficiently reliable data. Inform the requester that such data, needed to respond to the request, are unavailable. Reach an agreement with the requester to

- redefine the research questions to eliminate the need to use the data,
- end the engagement, or
- use the data with appropriate disclaimers.

Remember that you—not the requester—are responsible for deciding what data to use. If you decide you must use data that you have determined are not sufficiently reliable for the purposes of the engagement, make the limitations of the data clear, so that incorrect or unintentional conclusions will not be drawn. Finally, given that the data you assessed have serious reliability weaknesses, you should include this finding in the report and recommend that the agency take corrective action.

When to Assess Data as of Undetermined Reliability and Consider Additional Work

You can assess the data as of undetermined reliability when you conclude one of the following:

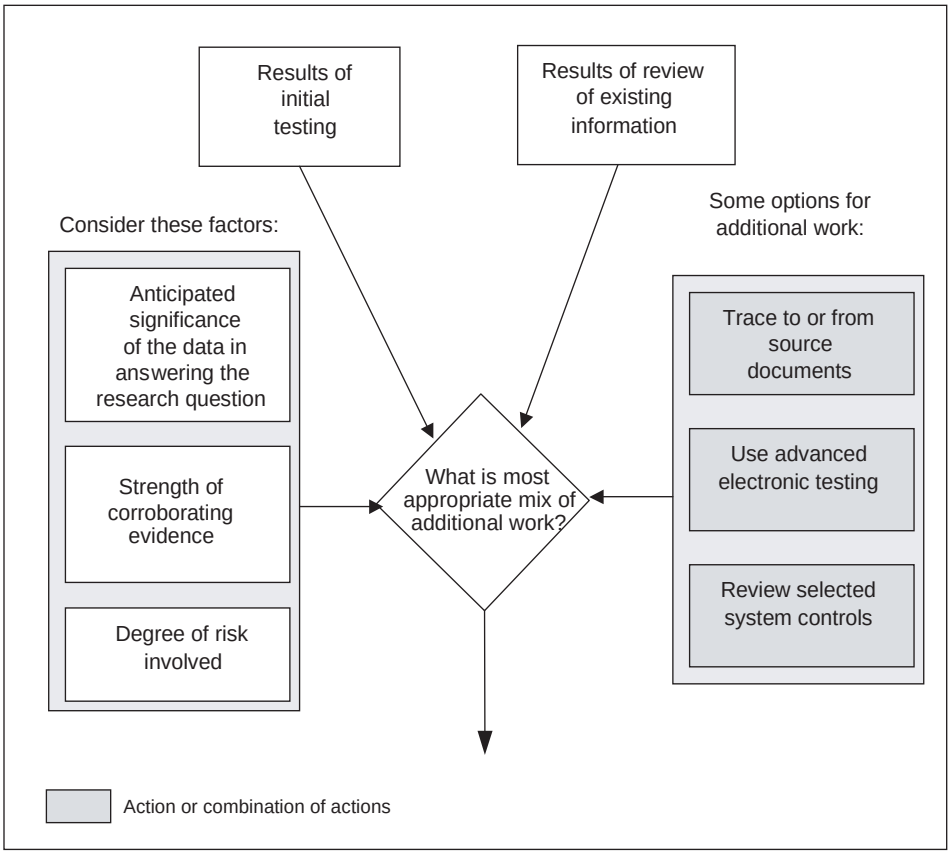
- The review of some of the related information or initial testing raises questions about the data's reliability.
- The related information or initial testing provides too little information to judge reliability.
- The time or resource constraints limit the extent of the examination of related information or initial testing.

When the preliminary assessment indicates that the reliability of the data is undetermined, consider doing additional work to determine reliability. Section 8 provides guidance on the types of additional work to consider, as well as suggestions if no additional work is feasible.

Section 8: Conducting Additional Work

When you have determined (through the preliminary assessment) that the data are of undetermined reliability, consider conducting additional work (see figure 6). A range of additional steps to further determine data reliability includes tracing to and from source documents, using advanced electronic testing, and reviewing selected system controls. The mix depends on what weaknesses you identified in the preliminary assessment and the circumstances specific to your engagement, such as risk level and corroborating evidence, as well as other factors. Focus particularly on those aspects of the data that pose the greatest potential risk for your engagement. You should get help from appropriate technical specialists to discuss whether additional work is required and to carry out any part of the additional reliability assessment.

Figure 6: Choosing and Conducting Additional Work



Source: GAO.

Tracing to and from Source Documents

Tracing a sample of data records to source documents helps you to determine whether the computer data accurately and completely reflect these documents. In deciding what and how to trace, consider the relative risks to the engagement of overstating or understating the conclusions drawn from the data, for example: On the one hand, 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, you should consider tracing from source documents to the database. On the other hand, 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, you should consider tracing from the database back to source documents.

The reason to trace only a sample is because sampling saves time and cost. To be useful, however, the sample should be random and large enough to estimate the error rate within reasonable levels of precision. Tracing a random sample will provide the error rate and the magnitude of errors for the entire data file. It is this error rate that helps you to determine the data reliability. Generally, every data file will have some degree of error (see example 1 for error rate and example 2 for magnitude of errors). Consult statisticians to assist you in selecting the sampling method most suited to the engagement.

Example 1: According to a random sample, 10 percent of the data records have incorrect dates. However, the dates may be off by an average of only 3 days. Depending on what the data are used for, 3 days may not compromise reliability.

Example 2: The value of a data element was incorrectly entered as $100,000, rather than $1,000,000. The documentation of the database shows that the acceptable range for this data element is between $100 and $5,000,000. Therefore, the electronic testing done in the initial testing phase would have confirmed that the value of $100,000 fell within that range. In this case, the error could be caught, not by electronic testing, but only by tracing the data to source documents.

Tracing to Source Documents

Consider tracing to source documents when (1) the source documents are available relatively easily or (2) the possible magnitude of errors is especially critical.

To trace a sample to source documents, match the entered data with the corresponding data in the source documents. But in attempting to trace entered data back to source documents, several problems can arise: Source documents may not be available because they were destroyed, were never created, or are not centrally located.

Several options exist if source documents are not available. For those documents never created—for example, when data may be based on electronic submissions—use interviews to obtain related information, any corroborating evidence obtained earlier, or a review of the adequacy of system controls.

Tracing from Source Documents

Consider tracing from source documents, instead of or in addition to tracing a sample to source documents, when you have concerns that the data are not complete. To trace a sample from source documents, match the source documents with the entered data. Such tracing may be appropriate to determine whether all data are completely entered. However, if source documents were never created or are now missing, you cannot identify the missing data.

Using Advanced Electronic Testing

Advanced electronic testing goes beyond the basic electronic testing that you did in initial testing (see section 5). It generally requires specialized computer programs to test for specific conditions in the data. Such testing can be particularly helpful in determining the accuracy and completeness of processing by the application system that produced the data. Consider using advanced electronic testing for

- following up on troubling aspects of the data—such as extremely high values associated with a certain geographic location—found in initial testing or while analyzing the data;

- testing relationships—cross-tabulation—between data elements, such as whether data elements follow a skip pattern from a questionnaire; and

- verifying that computer processing is accurate and complete, such as testing a formula used in generating specific data elements.

Depending on what will be tested, this testing can require a range of programming skills—from creating cross-tabulations on related data elements to duplicating an intricate automated process with more advanced programming techniques. Consult appropriate technical specialists, as needed.

Reviewing Selected System Controls

Your review of selected system controls—the underlying structures and processes of the computer in which the data are maintained—can provide some assurance that the data are sufficiently reliable. Examples of system controls are limits on access to the system and edit checks on data entered into the system. Controls can reduce, to an acceptable level, the risk that a significant mistake could occur and remain undetected and uncorrected. Limit the review to evaluating the specific controls that can most directly affect the reliability of the data in question. Choose areas for review on the basis of what is known about the system. Sometimes, you identify potential system control problems in the initial steps of the assessment. Other times, you learn during the preliminary assessment that source documents are not readily available. Therefore, a review of selected system controls is the best method to determine if data were entered reliably. If needed, consult information system auditors for help in evaluating general and application controls.

Using what you know about the system, concentrate on evaluating the controls that most directly affect the data. These controls will usually include (1) certain general controls, such as logical access and control of changes to the data, and (2) the application controls that help to ensure that the data are accurate and complete, as well as authorized.

The steps for reviewing selected system controls are

- gain a detailed understanding of the system as it relates to the data and

- identify and assess the application and general controls that are critical to ensuring the reliability of the data required for the engagement.

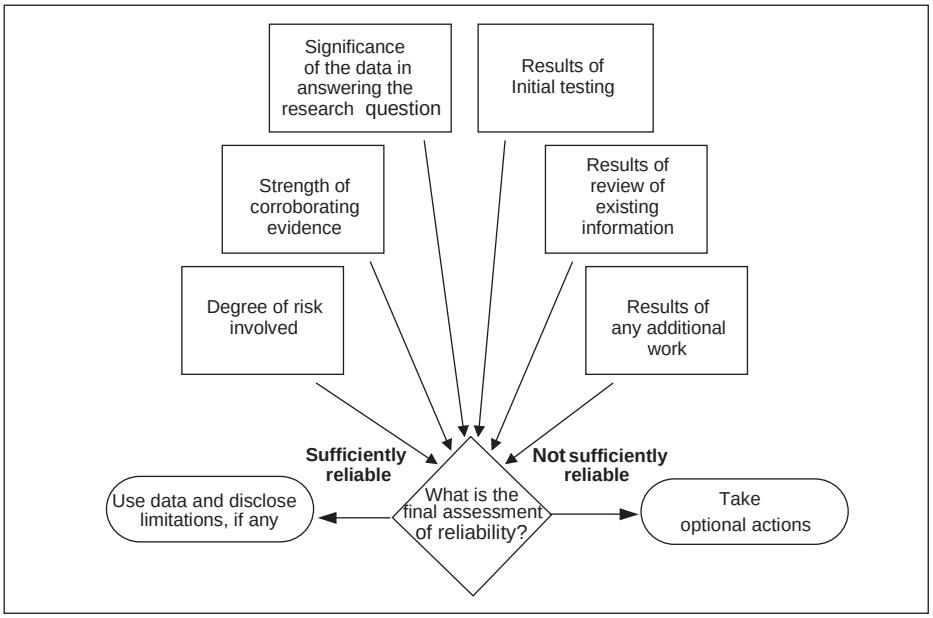
Using Data of Undetermined Reliability

In some situations, it may not be feasible to perform any additional work, for example, when (1) given a short time frame (too short for a complete assessment), (2) original computer files have been deleted, or (3) access to needed documents is unavailable. See section 9 for how to proceed.

Section 9: Making the Final Assessment

During the final assessment, you should consider the results of all your previous work to determine whether, for your intended use, the data are sufficiently reliable, not sufficiently reliable, or still undetermined. Again, remember that you are not attesting to the reliability of the data or database. You are only determining the sufficiency of the reliability of the data for your intended use. The final assessment will help you decide what actions to take (see figure 7).

Figure 7: Making the Final Assessment



Source: GAO.

The following are some considerations to help you decide whether you can use the data:

- The corroborating evidence is strong.
- The degree of risk is low.
- The results of additional assessment (1) answered issues raised in the preliminary assessment and (2) did not raise any new questions.

- The error rate, in tracing to or from source documents, did not compromise reliability.

In making this assessment, you should consult with appropriate technical specialists.

Sufficiently Reliable Data

You can consider the data sufficiently reliable when you conclude the following: On the basis of the additional work, as well as the initial assessment work, using the data would not weaken the analysis nor lead to an incorrect or unintentional message. You could have some problems or uncertainties about the data, but they would be minor, given the research question and intended use of the data. When your final assessment indicates that the data are reliable, use the data.

Not Sufficiently Reliable Data

You can consider the data to be not sufficiently reliable when you conclude the following: On the basis of information drawn from the additional assessment, as well as the preliminary assessment, (1) using the data would most likely lead to an incorrect or unintentional message and (2) the data have significant or potentially significant limitations, given the research question and intended use of the data.

When you determine that the data are not sufficiently reliable, you should inform the requester that sufficiently reliable data, needed to respond to the request, are unavailable. Remember that you—not the requester—are responsible for deciding what data to use. Although the requester may want information based on insufficiently reliable data, you are responsible for ensuring that data are used appropriately to respond to the requester. If you decide to use the data for the report, make the limitations of the data clear, so that incorrect or unintentional conclusions will not be arrived at. Appropriate team management should be consulted before you agree to use data that are not sufficiently reliable.

Finally, given that the data you assessed have serious reliability weaknesses, you should include this finding in the report and recommend that the agency take corrective action.

Data of Undetermined Reliability

You can consider the data to be of undetermined reliability when you conclude the following: On the basis of the information drawn from any additional work, as well as the preliminary assessment, (1) use of the data could lead to a incorrect or unintentional message and (2) the data have significant or potentially significant limitations, given the research question and the intended use. You can consider the data to be of undetermined reliability if specific factors—such as short time frames, the deletion of original computer files, and the lack of access to needed documents—are present. If you decide to use the data, make the limitations of the data clear, so that incorrect or unintentional conclusions will not be arrived at.

As noted above in the case of not sufficiently reliable data, when you determine that the data are of undetermined reliability, you should inform the requester—if appropriate—that sufficiently reliable data, needed to respond to the request, are unavailable. Remember that you—not the requester—are responsible for deciding what data to use. Although the requester may want information based on data of undetermined reliability, you are responsible for ensuring that appropriate data are used to respond to the requester. If you decide to use the data in your report, make the limitations clear, so that incorrect or unintentional conclusions will not be arrived at. Appropriate team management should be consulted before you agree to use data of undetermined reliability.

Section 10: Including Appropriate Language in the Report

In the report, you should include a statement in the methodology section about conformance to generally accepted government auditing standards (GAGAS). These standards refer to how you did your work, not how reliable the data are. Therefore, you are conforming to GAGAS as long as, in reporting, you discuss what you did to assess the data; disclose any data concerns; and reach a judgment about the reliability of the data for use in the report.

Furthermore, in the methodology section, include a discussion of your assessment of data reliability and the basis for this assessment. The language in this discussion will vary, depending on whether the data are sufficiently reliable, not sufficiently reliable, or of undetermined reliability. In addition, you may need to discuss the reliability of the data in other sections of the report. Whether you do so depends on the importance of the data to the message.

Sufficiently Reliable Data

Present your basis for assessing the data as sufficiently reliable, given the research questions and intended use of the data. This presentation includes (1) noting what kind of assessment you relied on, (2) explaining the steps in the assessment, and (3) disclosing any data limitations. Such disclosure includes

- telling why using the data would not lead to an incorrect or unintentional message,
- explaining how limitations could affect any expansion of the message, and
- pointing out that any data limitations are minor in the context of the engagement.

Not Sufficiently Reliable Data

Present your basis for assessing the data as not sufficiently reliable, given the research questions and intended use of the data. This presentation should include what kind of assessment you relied on, with an explanation of the steps in the assessment.

In this explanation, (1) describe the problems with the data, as well as why using the data would probably lead to an incorrect or unintentional message, and (2) state that the data problems are significant or potentially significant. In addition, if the report contains a conclusion or recommendation supported by evidence other than these data, state that fact. Finally, if the data you assessed are not sufficiently reliable, you should include this finding in the report and recommend that the audited entity take corrective action.

Data of Undetermined Reliability

Present your basis for assessing the reliability of the data as undetermined. Include such factors as short time frames, the deletion of original computer files, and the lack of access to needed documents. Explain the reasonableness of using the data, for example: These are the only available data on the subject; the data are widely used by outside experts or policymakers; or the data are supported by credible corroborating evidence. In addition, make the limitations of the data clear, so that incorrect or unintentional conclusions will not be drawn from the data. For example, indicate how the use of these data could lead to an incorrect or unintentional message. Finally, if the report contains a conclusion or recommendation supported by evidence other than these data, state that fact.

Glossary of Technical Terms

accuracy. Freedom from error in the data.

completeness. The inclusion of all necessary parts or elements.

database. A collection of related data files (for example, questionnaire responses from several different groups of people, with each group's identity maintained.)

data element. An individual piece of information that has definable parameters, sometimes referred to as variables or fields (for example, the response to any question in a questionnaire).

data file. A collection of related data records, also referred to as a data set (for example, the collected questionnaire responses from a group of people).

data record. A collection of related data elements that relate to a specific event, transaction, or occurrence (for example, questionnaire responses about one individual—such as age, sex, and marital status).

source document. Information that is the basis for entry of data into a computer.

GAO's Mission

The General Accounting Office, the investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through the Internet. GAO's Web site (www.gao.gov) contains abstracts and full-text files of current reports and testimony and an expanding archive of older products. The Web site features a search engine to help you locate documents using key words and phrases. You can print these documents in their entirety, including charts and other graphics.

Each day, GAO issues a list of newly released reports, testimony, and correspondence. GAO posts this list, known as "Today's Reports," on its Web site daily. The list contains links to the full-text document files. To have GAO e-mail this list to you every afternoon, go to www.gao.gov and select "Subscribe to daily E-mail alert for newly released products" under the GAO Reports heading.

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. General Accounting Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone: Voice: (202) 512-6000
TDD: (202) 512-2537
Fax: (202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Public Affairs

Jeff Nelligan, managing director, NelliganJ@gao.gov (202) 512-4800
U.S. General Accounting Office, 441 G Street NW, Room 7149
Washington, D.C. 20548

PRINTED ON RECYCLED PAPER

United States
General Accounting Office
Washington, D.C. 20548-0001

Official Business
Penalty for Private Use $300

Address Service Requested

Presorted Standard
Postage & Fees Paid
GAO
Permit No. GI00