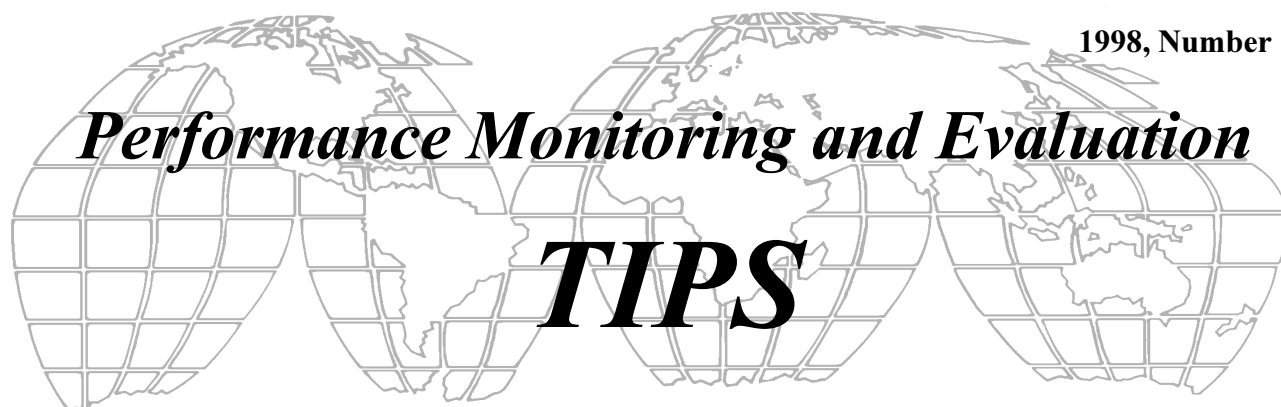




USAID Center for Development Information and Evaluation

GUIDELINES FOR INDICATOR AND DATA QUALITY

What Is the Purpose of This TIPS?

This TIPS summarizes the key references on performance measurement quality found in various parts of the Automated Directives System (ADS). It clarifies statistical, measurement, and evaluation concepts outlined in the ADS or introduced in previous TIPS and supplementary references.* In so doing, this TIPS identifies the key criteria for operating units to use in assessing the quality of their performance indicators and data.

This TIPS also describes USAID's review and approval procedures for indicators and related data. This process can be summarized as follows: *Operating units propose performance indicators and data sources in either their Country Strategic Plan or initial R4 submission. Once accepted by the relevant USAID/Washington bureau, the indicators and related data sources are judged to be satisfactory with respect to the criteria outlined in the ADS and this TIPS.*

Operating units and USAID/W bureaus are expected to apply the criteria and guidelines described in this TIPS for approximately one year when proposing a new country strategic plan, amending an existing plan, or submitting an initial R4. On the basis of experience and feedback from managers and technical officers in Washington and the field, we will revise and refine this guidance next year to ensure that it is as useful and unbureaucratic as possible. If appropriate, we will revise relevant ADS policies and essential procedures as well.

Why Do We Care About the Quality Of Indicators and Data?

USAID has moved substantially from planning our performance measurement systems to actually using performance data in managing for results. This results-oriented management approach relies on both field and Washington managers to

Because performance data have become more important to Agency decisions, clear criteria for judging the quality of these data have become increasingly crucial.

This TIPS describes Agency criteria and procedures for ensuring the quality of indicators and data in operating units' performance monitoring systems.

*Refer to TIPS #6, *Selecting Performance Indicators*, and TIPS #7, *Preparing a Performance Monitoring Plan*.

inform their decisions with performance information. But sound decisions require accurate, current, and reliable information, and the benefits of this results-oriented approach depend substantially on the quality of the performance information available. More specifically, quality performance indicators and data (combined with other information) will help

- Ensure that Agency program and budget decisions—both in the field and at USAID/W—are as well informed as practically possible
- Support efficient use of Agency resources, including those dedicated to performance measurement itself
- Meet requirements of federal legislation
- Address the information needs of the Agency’s internal and external stakeholders, including senior management, OMB, and the Congress

Over the past three years, USAID has made substantial progress in measuring performance and managing for results, including development and use of better performance measures for operational programs as the result of both field and USAID/W efforts. Through the R4 process, the Agency has also learned important lessons about how to collect, analyze, and use results information. Much has been accomplished, but given the increased use of performance information in many of the Agency’s management processes, further improvements and refinements are needed.

The key elements that determine the quality of USAID’s performance measurement systems are discussed in the remainder of this TIPS. The material is organized in three sections. First, we discuss the *quality of performance indicators* themselves. Next, we consider how to ensure the *quality of the data* that are collected in relation to those indicators. Finally, we discuss appropriate standards for *documentation*, *USAID/W review*, and *periodic reassessment* of indicators and data. While it is often useful to keep these elements distinct when thinking about performance measurement, it is also important to recognize that all three elements are crucial to measuring performance effectively and, therefore, to managing for results.

Criteria for Selecting Quality Performance Indicators

Criteria for quality performance indicators—and data—must be keyed to our intended use of the data. That is, the level of accuracy, currency, and reliability of performance information should be consistent with the requirements of good management. While we generally need something substantially better than “back of the envelope” estimates, the precision of laboratory experiments is rarely required. Determining appropriate or adequate thresholds of indicator and data quality is not an exact science. This task is made even more difficult by the complicated and often data-poor development settings in which USAID operates.

All of this underscores the need to apply the Agency’s criteria for selecting performance indicators—that is, that they be direct, objective, practical, and adequate—with a healthy measure of common sense and reasonableness. Importantly, this includes the recognition that although we always want the “best” indicators, there are inevitably trade-offs among various aspects of indicator quality. For example, data for the most direct or objective indicators of a given result might be very expensive to collect or might be available only at too great a lag or at too long an interval.

The criteria discussed in this section address the two overarching factors that determine the extent to which performance indicators function as useful tools for managers and decision-makers:

- The degree to which performance indicators and their related data accurately reflect the process or phenomenon they are being used to measure
- The level of comparability of performance indicators and data over various measurement contexts. That is, can we measure results in a consistent and comparable manner over time and across settings?

It is helpful to keep these considerations in mind when reviewing and using the following criteria of indicator quality.

Direct

A performance indicator is direct (or valid) if it closely tracks the result it is intended to measure.

(For purpose of this TIPS, the terms direct and valid are being treated almost interchangeably, although we recognize that the concepts of indicator directness and validity are complex and multifaceted.) Indicators should either be widely accepted for use by specialists in a relevant subject area, exhibit readily understandable face validity (i.e., be heuristically and intuitively valid measures), or be supported by a specific body of technical research. We suggest using indirect or proxy indicators when other considerations, such as cost or timeliness, make it necessary. When proxies are used, particularly those that are not generally accepted or widely used, the relationship between the indicator and the result it is intended to measure should be researched and well understood.

The directness of an indicator is one of the most important criteria for identifying a quality performance indicator. More direct indicators are by their nature more valid and more intuitively understandable—that is, they have high “face validity.” Child mortality and morbidity rates, for example, are very direct measures of the result *improved child health*. By contrast, wage rate is not a direct measure of *increased household food security*. The more closely an indicator reflects the result it is measuring, the more direct the indicator. Managers should place greater confidence in decisions based on direct measures.

Some objectives are conceptually simple and have relatively direct and straightforward indicators (e.g., fertility or mortality reduction, or increased school enrollment). Others are more complicated but have a long history of application in the field and are supported by a specific body of research (e.g., household expenditures as a measure of household income). In still other areas, particularly democracy and environment, identifying relatively direct measures remains a complex undertaking. In these sectors we are trying to develop new, and better indicators, but often must use proxy, or less direct, measures, which are linked to the result by one or more assumptions. As noted above, research or experience should indicate that such assumptions are sound.

Objective

An indicator is objective if it is unambiguous about 1) what is being measured and 2) what data are being collected. Objective indicators have clear operational definitions that are independent of the

person conducting the measurement—that is, different individuals would collect data for an objective indicator using the same indicator definition.

Performance indicators should be framed and defined in clear terms so as not to be open to broad and varied interpretation by sector specialists. Particularly in the case of qualitative indicators, clear and comprehensive definitions help ensure a reasonable level of objectivity and comparability over time. Multidimensional indicators (e.g., indices), must include clear definitions for each element of the indicator and specify the method of aggregation.

Objectivity of performance indicators is critical to collecting comparable data over time. If indicators are not clearly and consistently defined, the data collected are unlikely to permit a useful assessment of progress toward the relevant result over time. For example, an indicator of “number of successful firms,” without a clear and precise definition of both *successful* and *firm*, could easily lead to the collection of data that vary over time, reflecting subjective interpretations of the measure, independent of any actual changes taking place.

The ADS states that *quantitative* performance indicators are preferred in most cases. When qualitative indicators are used, they should be clearly defined so as to permit regular, systematic and relatively objective judgment regarding the change in the “value” or status of the indicator (ADS, E203.5.5a). The point here is not that quantitative indicators are in some way “better” than qualitative measures. Rather, performance measurement requires comparability of data over time. That, in turn, demands objective measurement. Although indicators that “count” objectively verifiable phenomena are easier to operationalize, they are not necessarily the most useful or meaningful measures. Good judgment and common sense are essential in weighing these factors and deciding which indicators to use.

Comparability of data also requires that our measures remain relatively stable over time. While operating units can and should modify initial indicators if better or more practical alternatives become readily available (and are acceptable to Washington), definitions should not vary across time unless clearly referenced and justified. Once monitoring begins, changes in indicator definitions may compromise our ability to understand what results are being achieved.

Quantitative Versus Qualitative Indicators

More tends to be made of the distinction between qualitative and quantitative data than is warranted. Not everything that is important can be counted, and much that can be counted is not worth knowing. Many of our most meaningful measures of development involve subjective judgments (for example, the capacity of organizations or the transparency of political institutions). Often, these judgments can be objectively operationalized, using surveys or expert panels or well-defined rating systems, and translated into rank-ordered categories, a form of quantifying judgments.

It might be more helpful to think of quantitative and qualitative as a matter of degree. On the quantitative end of the continuum are measures that involve continual, equal-interval scales with true zero points (such as GDP per capita, infant mortality rates, school enrollment rates). In the middle are data for which the frequency of various occurrences can be counted, and categorized, and perhaps even rank-ordered. At the qualitative extreme are data that can be captured only by descriptive narrative. Though USAID collects and uses data from throughout this continuum, most of the performance information collected by operating units probably falls somewhere in the middle. For example, most of the data the Agency collects on policy reform, institutional strengthening, and customer feedback are measured on some type of ranked (ordinal) scale. Such scales, when clearly operationalized, provide a good illustration of how more subjective data can be usefully and effectively quantified.

The quantitative-versus-qualitative debate is not an either-or question with some predefined best answer. Within the context of USAID's performance-based management systems, the choice of more quantitative or qualitative indicators involves trade-offs among practicality and cost, objectivity and comparability, and the directness or validity (meaningfulness) of the measure. Managers need to pick indicators that provide useful information and an adequate level of comparability over time. More meaningful and comparable indicators are the goal, regardless of how "quantitative" or "qualitative" they are.

Objective indicators are helped by being *unidimensional*—that is, when they measure a single characteristic, variable, or phenomenon (ADS, E203.5.5a). Indicators that encompass multiple dimensions (e.g., indicators incorporating both access to and use of a given service or technology) can and often do confuse procedures for data collection, collation, and analysis. Multidimensional indicators can complicate assessments of progress toward results and the consideration of relevant management options. However, multidimensional indicators are permitted (ADS, 201.5.10a), and indeed are sometimes even desirable. But when multidimensional indicators (such as indices) are used, care should be taken to ensure their objectivity. This is not usually an issue when an operating unit uses a preexisting index that is already well defined and operationalized. However, any operating unit using a new or unique index should clearly define each component or dimension of the indicator and, when relevant, the methodology for aggregating the components into a single score.

Practical

A practical indicator is one for which data can be collected on a timely basis and at a reasonable cost. Performance indicators should 1) provide data to managers at a cost that is reasonable and appropriate, as compared with the management utility of the data; 2) have data available on a frequent enough basis to inform regular program management decisions (usually on an annual basis, though for some indicators annual data collection will not be practical [see ADS, E203.5.5d]); and 3) have data available that are current enough to be useful in decision-making (generally, no more than three years in lag).

Cost considerations. Cost of data collection, in terms of both human and financial resources, is an important consideration when identifying indicators. Though difficult to assess precisely, the cost of collecting data for an indicator should not exceed the management utility of the data collected. While operating units should not incur exorbitant data collection costs, they should expect to incur reasonable—sometimes substantial—costs to collect useful

performance data. The rule of thumb provided in the ADS is that costs to an operating unit for performance monitoring and evaluations should normally range between 3 and 10 percent of the total budget for the objective activities (ADS, E203.5.4). This is a rough guideline that will not apply in all cases. For example, if the USAID program in Egypt spent 10 percent of its budget assessing performance (e.g., more than \$80 million per year), that would likely be considered “excessive.”

The costs incurred by an operating unit will relate largely to the data collection methods required by the chosen indicators and the accessibility of reliable secondary data sources. If data of adequate quality are already collected routinely by a secondary source, costs may be minimal. If primary data must be collected at the operating unit’s expense, costs will be higher—how much higher depends on the scope, method, and frequency of data collection required. A sample survey, for example, may cost several hundred thousand dollars, whereas a rapid appraisal would be much less expensive. Whatever method selected, it should provide data that are sufficiently accurate or representative.

Similarly, operating units should not expect their development partners (e.g., government departments, implementing agencies, international agencies, or other secondary sources) to bear unreasonable costs or time or paperwork in providing data specific to USAID needs (the ADS articulates a different standard for USAID’s partners who are responsible for results or assumptions upon which the Agency’s results are dependent [see ADS, E203.5.5d2]). USAID has not provided a standard or rule of thumb in this case, but recommends that USAID and its partners regularly review the costs and the utility to USAID and its partners of such information.

Timeliness considerations. When identifying indicators, operating units should also consider the frequency at which data for the indicator will be available, as well as the currency of the data (refer to the more detailed discussion of data timeliness in the data quality section of this TIPS).

Adequate

The number of indicators tracked for a given result should be the minimum necessary to ensure that progress toward the result is sufficiently captured.

There is no “correct” number of indicators. Rather, the number of indicators required to adequately measure a result depends on 1) the complexity of the result being measured, 2) the amount of information needed to make reasonably confident decisions, and 3) the level of resources available for monitoring performance. An objective focusing on improved maternal health, for example, may require two or three indicators to capture the various and constituent aspects of maternal health. It is uncommon to need more than three indicators to effectively track a result.

If a large number of indicators have been identified for a specific result, this may indicate that the result is too complex or not well enough understood to measure adequately, signaling the need for additional evaluation or investigation. It may also be that more data than necessary are being collected to *indicate* whether progress is meeting, failing to meet, or exceeding expectations.

“We must be prudent about how much and what information we collect and use for decisions. . . . More is not always better. . . . Information collected should be demonstrably useful. . . . If it is not, one should question why it is being collected.” (UNCLASS STATE 057091)

“We are collecting more data than is necessary to monitor the effectiveness of our programs and to inform budget decisions.” (Message from the Administrator, USAID General Notice 7/28/98)

Criteria for Collecting Quality Performance Data

Measuring performance effectively begins with selecting quality performance indicators, but that is not enough. We must also pay attention to the data collection process to ensure that quality data are collected and available to inform management decisions. This section discusses key criteria for assessing the quality of performance data—validity, reliability, and timeliness.

Performance Monitoring Systems And Data

“The Agency and its operating units shall establish and maintain performance monitoring systems that regularly collect data which enable the assessment of progress towards achieving results. Operating unit performance monitoring systems shall track performance at both the results framework level and the activity level.” (ADS, 203.5.5)

As with performance indicators, we sometimes have to make trade-offs, or informed judgments, when applying the criteria for data quality. This is especially true if, as is often the case, USAID relies on others to provide data for one or more indicators. For example, if our only existing source of data for a critical economic growth indicator is the Ministry of Finance, and we know that the ministry’s data collection methods are less than perfect, we may have to weigh the alternatives of relying on less-than-ideal data, having no data at all, or conducting a potentially costly USAID-funded primary data collection effort. A decision must be made as to whether the ministry’s data would allow the strategic objective team to have confidence when assessing program performance or whether they are so flawed as to be useless in reporting and managing for results. We must be careful not to let the ideal drive out the good.

Validity

Indicators are valid (direct) to the extent that they clearly and directly measure the result they are intended to measure. But even valid indicators have little value, if the data collected does not correctly measure the variable or characteristic encompassed by the indicator. It is quite possible, in other words, to identify valid indicators but to then collect inaccurate, unrepresentative, or incomplete data. In such cases, the quality of the indicator is moot. (Equally undesirable would be collecting good data for an invalid indicator).

Validity of data is affected by many factors, the most important of which—measurement errors, unrepresentative sampling, and simple transcription errors—are discussed below.

Validity and Reliability

The concepts of validity and reliability are important to both performance indicators and data. In brief, validity refers to the extent to which a measure actually represents what we intend to measure. Though simple in principle, validity can be difficult to assess in practice, particularly when measuring social phenomena. What does IQ really measure, for example? Is the poverty gap a good measure of the extent of a country’s poverty?

Reliability refers to the stability of a measurement process. That is, assuming there was no real change in the variable being measured, would the same measurement process provide the same results or findings if the procedure were repeated over and over? If we use a thermometer to measure a child’s temperature repeatedly and the results vary from 95 to 105 degrees, even though we know the child’s temperature hasn’t changed, the thermometer is not a reliable instrument for gathering data on the child’s fever.

Measurement Error

Measurement error results primarily from the poor design or management of a data collection process. The level of measurement error associated with all performance data collected or used by operating units should not be so large as to 1) call into question either the direction or degree of indicator change reflected by the data or 2) overwhelm the amount of anticipated change in an indicator (making it impossible for managers to determine whether “progress” reflected in the data is a result of actual change or of measurement error).

To ensure that data are valid, we must pay attention to a number of possible sources of measurement error. These sources of error are often grouped into two

general categories—*sampling error* (unrepresentative samples) and *nonsampling error*, including poor design of the data-collection instrument, poorly trained or partisan enumerators, or the use of questions (often related to sensitive subjects) that elicit incomplete or untruthful answers from respondents. Regardless of the source, if too much error is introduced into the data collection process, the resulting data will be invalid. (For additional information refer to “Common Problems/Issues With Using Secondary Data” in the *CDIE Resource Book on Strategic Planning and Performance Monitoring Under Re-engineering*, April 1997.)

USAID staff and partners should expect some error in any data collection effort that focuses on social and economic change. Our challenge is to determine the level of measurement error that we are willing to accept. In defining an acceptable level of error, we should remember that, at some point, reducing measurement error further can become a very expensive, if not an impossible, undertaking. Additionally, the *management utility* of data is usually not greatly enhanced by reducing measurement error beyond some acceptable threshold.

What is an acceptable level of error? There is no simple standard that can be applied across all of the data collected for USAID’s varied programs and results. As performance monitoring plans are constructed, teams should 1) assess the types and sources of error for each indicator, 2) estimate the approximate levels of error that are likely, 3) assess how this error compares with the magnitude of expected change, and 4) decide whether alternative data sources (or indicators) need to be explored. Judgments should be based on the nature of the data being collected and the intended use of the data.

For some indicators, for which the magnitude of expected change is large, even relatively large measurement errors may be perfectly tolerable; for other indicators, small amounts of change will be important and even moderate levels of measurement error will be unacceptable. Suppose, for example, that our indicator for strengthening civil society is the number of politically active NGOs. If our baseline was 900 NGOs and our preliminary data showed that after a few years this had grown to 30,000 NGOs, a 10 percent measurement error ($\pm 3,000$ NGOs) would be essentially irrelevant. If, however, our baseline was 900 NGOs and our second data point was 1,000

NGOs, a 10 percent measurement error (± 100) would be completely unacceptable because it would represent all of the apparent change in the data.

In summary, judgments about acceptable measurement error should reflect technical assessments about what level of measurement is possible, practical considerations such as cost, and, most important, management judgments about what level of accuracy is needed for decisions. Keep in mind that USAID is primarily concerned with learning, with reasonable confidence, that anticipated improvements have occurred, not with reducing error below some arbitrary level.

Judging the Quality of Data From Secondary Sources

USAID’s performance monitoring systems often rely on data from existing secondary sources, which can vary considerably in quality. In some cases a data source is sufficiently reliable so that independent data checks are necessary only at rare intervals. In other instances data may need to be spot-checked. In still others, a record-by-record check is needed. Realism, as well as technical acuity, is necessary to select the type of validation that is appropriate. It is too simplistic—and wrong—to assume whole categories of sources (e.g., NGOs, government agencies) are invalid or unreliable. Ideally, each source needs to be evaluated individually, in terms of the adequacy of its data quality assurance systems. Such an undertaking, though, is no small task and may require considerable resources.

For more on this subject and a checklist of questions for judging secondary sources, see “Common Problems/Issues with Using Secondary Data” in the *CDIE Resource Book on Strategic Planning and Performance Monitoring under Re-engineering*, April 1997.

Representativeness

Data are said to be representative if they accurately reflect the population they are intended to describe. Sample survey data collected or used by operating units should be sufficiently representative to confidently inform programs directed at the larger population to which they refer.

The representativeness of survey data is a function of the process used to select a sample of the population from which data will be collected. Drawing a sample that will allow managers to confidently generalize data/findings to the population requires that two basic criteria be met: 1) that all units of a population (e.g., households, schools, enterprises) have an equal chance of being selected for the sample and 2) that the sample is of adequate size. The sample size necessary to ensure that resulting data are representative to any specified degree can vary substantially, depending on the unit of analysis, the size of the population, the variance of the characteristics being tracked, and the number of characteristics that we need to analyze.

We will not provide an extended discussion of sampling theory here—lengthy technical tomes are widely available on the topic. But it is important to distinguish between the “completeness” of data and the representativeness of data. In short, it is not necessary or even desirable to collect data from all units of a population. If we are interested in describing the characteristics of a country’s primary schools, for example, we would not need to examine every school in the country—depending on our focus, a sample of a hundred schools might well be enough. Sample surveys, moreover, rarely obtain data for every member of an initially chosen sample. Rather, there are established techniques for determining acceptable levels of nonresponse or for substituting new respondents.

In any case, most USAID surveys involve nonrandom samples in which the criteria for inclusion are clearly defined and the problems of nonresponse are easily assessed. While there are circumstances—particularly for financial management or input/output monitoring—for which data “completeness” is an important consideration, for most results-level performance indicators this is not a critical issue.

Transcription Error

Transcription error refers to simple data entry errors made when transcribing data from one document (electronic or paper) or database to another. Transcription error is avoidable, and operating units should seek to eliminate any such error when producing internal or external reports and other documents.

When the data presented in a document produced by an operating unit are different from the data (for the same indicator and time frame) presented in the original source simply because of data entry or copying mistakes, a transcription error has occurred. Such differences (unless due to rounding) can be easily avoided by careful cross-checking of data against the original source. Rounding may result in a slight difference from the source data but may be readily justified when the underlying data do not support such specificity, or when the use of the data does not benefit materially from the originally reported level of detail. (For example, when making cost or budget projections, we typically round numbers. When we make payments to vendors, we do not round the amount paid in the accounting ledger. Different purposes can tolerate different levels of specificity.)

The use of preliminary or partial data should not be confused with transcription error. It will at times make sense to use such data (clearly identified as preliminary or partial) to inform management decisions or to report on performance because these are the best data currently available. When preliminary or partial data are updated by the original source, USAID should quickly follow suit, and note that it has done so. Any discrepancy between preliminary data that are included in a dated USAID document and data that were subsequently updated in an original source does not constitute transcription error.

Reliability

Data reliability refers to the stability or consistency of the data collection process. If we know an actual result hasn’t changed, and we collect data repeatedly, against the same indicator and for the same population, the data can be considered reliable if the findings on each occasion are approximately the same. Performance data collected or used by operating units should be reasonably reliable—that is, they should reflect a consistent data collection process from year to year such that managers can be confident that progress toward indicator targets reflects real changes rather than variations in data collection methods. If elements of the data collection process vary from year to year, operating units must assess the degree to which the resulting data can be usefully compared to understand performance over time.

“Whenever possible, reasonable standards for statistical reliability and validity should be applied, although in many cases it will not be appropriate or possible to meet these standards.” (ADS, E203.5.5e)

Ensuring that data are reliable requires not only that an indicator be objectively and clearly defined, but also that the data collection process be consistent from year to year. That is, a consistent sampling method and the same or comparable data collection instruments and data collection procedures are used. If, for example, the data collection instrument for a given survey is substantially changed between year 1 and year 2, both sets of data might be valid, but they might very well be neither reliable nor comparable.

As is the case with data validity, measurement error can compromise the reliability of data. The sampling and nonsampling errors presented in the discussion of data accuracy also affect data reliability. There are yet additional complications. If measurement error results in a *consistent* bias (e.g., due to a sampling method that consistently excludes the same segment of a given population), then data will be reliable but may well be invalid. But if managers are aware of the nature and magnitude of the bias, they may still be able to effectively use the resulting data.

Timeliness

Timeliness refers to two elements—frequency and currency. Concerning frequency, performance data collected or used by operating units should be available on a frequent enough basis to regularly inform program management decisions. As for currency, data should be sufficiently up to date to be useful in decision-making; so, as a general guideline, data should lag no more than three years. It is important also to remember that annual collection of performance data for USAID-funded intermediate results is not required until the point at which progress is expected to begin.

Data collected only once every five or six years (as is frequently the case with national-level surveys) may have limited utility for management decisions that must be made more frequently. That is, in order to “manage for results,” managers must have information regarding performance on a regular periodic basis, preferably annually. However, for some development results (such as reduced fertility) that occur

slowly over relatively long periods, it may not make sense to collect data annually, because changes are unlikely to be significant at such short intervals. Often, these are the very indicators that require relatively expensive sample surveys to collect good data. In these cases, data may be collected at several-year intervals, supplemented by proxy or indirect indicators (e.g., contraceptive distribution and sales data that track intermediate results) to get an indication of progress toward the longer term objective.

“For performance indicators for which annual data collection is not practical, operating units will collect data regularly, but at longer time intervals. . . .” (ADS, E203.5.5d1)

The second aspect of data timeliness relates to how *current* the data are. Certainly, decisionmaking should be informed by the most current data that are practically available. Frequently, though, data obtained from a secondary source, and at times even USAID-funded primary data collection, will reflect substantial time lags between initial data collection and final analysis and publication. Many of these time lags are unavoidable, even if considerable additional resources were to be expended. Sometimes preliminary estimates may be obtainable, but they should be clearly flagged as such and replaced as soon as possible as the final data become available from the source.

Although the ADS states that the results review sections of the R4 must address the operating unit’s performance for the immediate past fiscal year (ADS, 203.5.9a), we recognize that data may come from preceding calendar or fiscal years.

Moreover, data often measure results for the specific point in time that the data were collected, not from September to September, or December to December. Often the realities of the recipient country context will dictate the appropriate timing of the data collection effort, not the U.S. fiscal year. For example, if agricultural yields are at their peak in July, then data collection efforts to measure yields should be conducted in July of each year. Moreover, to the extent that USAID relies on secondary data sources and partners for data collection, we may not be able to dictate exact timing. PPC is modifying the ADS accordingly to reflect this reality.

Documentation, USAID/W Review, And Periodic Reassessment Of Indicator and Data Quality

Proper documentation, appropriate USAID/W review and periodic quality reassessments are processes that facilitate the maintenance of quality performance indicators and data. These processes should take place in a transparent and open manner. To the extent possible, they should provide opportunities for independent checks concerning the quality of USAID's performance measurement systems.

Documentation

By *documentation*, we mean recording

1. The important considerations and assumptions that went into deciding on specific performance indicators, when this is not readily self-evident
2. The detailed specifications for each indicator (e.g., a comprehensive, operational definition of the indicator and the precise unit of measurement)
3. The specifications concerning data and data collection (e.g., a detailed description of the data source, the methods used, and the frequency and timing of data collection)
4. The assessments of the quality of performance indicators and data, in relation to specific Agency criteria
5. The agreements reached during USAID/W reviews of indicator and data quality

Adequate documentation facilitates the collection of comparable performance data from one measurement period to the next (it may also allow others to independently replicate collection of the data). Proper documentation is especially important in an organization like USAID, where there is considerable staff turnover in operating units and objective teams. Documentation also ensures the availability of information critical to the analysis of performance data (i.e., the specifics of how data were collected). Finally, documenting the specific characteristics of indicators and data allows staff to explain their procedures to those who are seeking assurance that quality standards are being maintained in the collection and reporting of performance data.

Careful development, use, and maintenance of the Performance Monitoring Plan (PMP) by operating units will go a long way toward ensuring adequate documentation. The ADS requires that these plans be prepared, and periodically updated, to provide details on their performance monitoring system's indicators and data-collection efforts (ADS, 203.5.5a). TIPS #7, *Preparing a Performance Monitoring Plan* (1996), elaborates further on the ADS guidance.

For performance monitoring plans to be useful, they need to be kept current. Annual updating, timed to coincide with the R4 process, is suggested.

Performance Monitoring Plans

"Performance Monitoring Plans shall provide a detailed definition of the performance indicators to be tracked; specify the source, method of data collection and schedule of collection for all required data; and assign responsibility for collection to a specific office, team, or individual.

"Performance Monitoring Plans are one element of a performance monitoring system and function as critical tools for managing and documenting the data collection process." (ADS, E203.5.5b)

USAID/W Reviews

Operating units will propose performance indicators and data (sources) via either their Country Strategic Plan or initial R4 submission. Once accepted by the relevant USAID/W bureau, the indicators and related data (sources) will be judged to be satisfactory with respect to the criteria outlined in the ADS and this TIPS. (If with use, however, it becomes apparent that either the indicators or data are not adequately or accurately capturing progress toward the relevant result, both USAID/W and the operating unit will be responsible for revisiting the indicators and data.)

Strategic Plan, R4 or related review processes provide regular occasions for joint discussion, agreement, or reaffirmation between an operating unit and USAID/W concerning issues related to the quality of performance indicators and data. This USAID/W review

process is meant to function as a validation of operating units' judgment regarding the quality of their data and indicators. The discussion will center on outstanding or exceptional indicator or data issues. Operating units are encouraged to take advantage of the R4 performance data table (comments section), which provides space for summary information on performance indicator definitions, sources, data collection methods, and quality issues. Special qualifications and limitations concerning quality should be noted for both indicators (e.g., when proxy indicators are used, when definitions are modified) and data (e.g., that data are preliminary estimates, vary from initial definitions, refer to calendar or fiscal years or to specific dates).

The criteria outlined in this TIPS will serve as a guide for USAID/W reviews, including new country strategic plans, amendments to plans, or initial R4 submissions. Agreements reached during these review meetings regarding indicator and data quality (e.g., changes in indicator definitions or sources and actions to be taken to improve quality) should be documented. Early and ongoing participation by regional and central bureau technical and performance measurement specialists (e.g., as virtual team members) could serve as an independent check on the quality of performance monitoring systems and result in modifications and improvements to indicators and data prior to USAID/W review.

USAID/Washington Review of Performance Monitoring

“Reviewers [of strategies] will focus upon. . . the ability of the operating unit to monitor and demonstrate performance.” (ADS, 201.5.11b.4)

“The R4 will be reviewed by the parent bureau . . . this may include adjustments in indicators and targets.” (ADS, E201.5.16c)

While the ADS does not require the review of operating units' Performance Monitoring Plans by central or regional bureaus (ADS, E203.5.5b), in practice a number of regional bureaus have already encouraged operating units to share them. As concerns over the quality of operating unit performance monitoring systems grow, the need for USAID/W review of PMPs may need to be revisited (and ADS guidance revised, as appropriate).

Periodic Reassessment

In-depth reassessments of the quality of an operating unit's performance monitoring system should be undertaken periodically—at least every three years.

By *reassessment*, we mean making sure that performance indicators and data are at a level of quality that permits both confident decision-making by managers and effective reporting to those outside the program. Taking proper care when initially identifying indicators and data sources will help to ensure long-term quality of a performance measurement system. However, it is important to take a critical look at our performance measurement systems and data sources from time to time to make sure the indicators are still measuring what we think they are measuring and that data are being collected in the way that we intend them to be collected. Agency directives (ADS, E203.5.5e) call for this critical look once every three years at a minimum.

Quality Reassessments

“Data quality will be assessed as part of the process of establishing performance indicators and choosing data collection sources and methods. Data quality will be reassessed as is necessary, but at intervals of no greater than three years.” (ADS, E203.5.5e).

Reassessments should be systematic and should be driven by the criteria outlined in this TIPS. Any reassessment should include a review of all performance indicators (at both objective and intermediate results levels) and should cover each data source. Further, reassessments should be documented, as reviewers, including the OIG, will want to determine whether and to what degree such an assessment has been conducted.

These periodic indicator and data quality reviews need to be included in operating unit work plans and budgets. Operating units might want to consider using a qualified, independent individual or team—with appropriate social science research, performance measurement, and data collection expertise—to conduct indicator and data reassessments.

Questions? Comments?

If operating units, reviewers, USAID partners, or others have questions, comments, or suggestions regarding these standards, please send them to PPC/CDIE and your bureau strategic planning office. Working with your bureau, PPC commits to answering questions, comments, and suggestions; making technical assistance available, both USDH and contract (although funding may have to come from sources other than PPC); and making this TIPS more user-friendly. For information or comments on PPC's broader responsibilities for Agencywide planning, performance measurement, and reporting under the GPRA, please contact DAA/PPC, Dirk Dijkerman.

CDIE's TIPS series provide guidelines, advice and suggestions to USAID managers on how to plan and conduct performance monitoring and evaluation activities effectively. They are supplemental references to the reengineering automated directives (ADS, chapter 203). For more information on the TIPS series, contact Annette Binnendijk, CDIE senior evaluation adviser, by phone (202) 712-4459, fax (202) 216-3124, or e-mail (abinnendijk@usaid.gov). Copies of TIPS can be ordered from the Development Experience Clearinghouse by calling (703) 351-4006 or by faxing (703) 351-4039. Please refer to the PN number. To order via the Internet, address requests to: docorder@dec.cdie.org.

If you have access to the USAID internal Web site, you can access the TIPS series directly by clicking on "Information Services," then on "CDIE." From the CDIE Online Web page, click on Performance Monitoring and Evaluation TIPS. Others may view the TIPS on USAID's external Web site: http://www.dec.org/usaid_eval/#004.