Developing a performance management system for a Federal public health program: the Ryan White CARE ACT Titles I and II

Jennifer Katesa,*, Katherine Marconib, Thomas E. Mannle Jr. c

a The Henry J. Kaiser Family Foundation, 2400 Sand Hill Road, Menlo Park, CA 94025, USA
b HIV/AIDS Bureau, Health Resources and Services Administration, USA
c The Lewin Group, USA

Received 27 May 1999; received in revised form 24 July 2000; accepted 6 August 2000

Abstract

Both public and private sector organizations are increasingly being called upon to measure performance. In the public sector, this trend is evidenced by such legislative initiatives as the Government Performance and Results Act (GPRA). Yet performance measurement presents complex analytic and behavioral challenges and few models exist to guide policy makers and program managers in their attempts to respond to legislative mandates and market pressures for accountability through performance measurement. This article describes one approach to introducing performance measurement into a large Federal health program, the Ryan White Comprehensive AIDS Resources Emergency (CARE) Act, in response to GPRA. It places performance measurement within the broader concept of performance management. This approach is particularly applicable to large grant programs characterized by significant autonomy and variation at the local level. The article also discusses some of the characteristics of the HIV/AIDS epidemic which present unique challenges to performance measurement. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: HIV; AIDS; Ryan White CARE Act; Performance measurement; Outcome measurement; Government Performance and Results Act (GPRA)

1. Introduction

A convergence of economic and political forces is driving a systematic search for greater effectiveness, efficiency, and accountability in both the public and private sectors. In health care, the proliferation of managed care, the consolidation of delivery systems, and other marketplace changes have contributed to a growing movement towards performance measurement, quality improvement, and ‘report card’ efforts. Shrinking resources underscore the need to ‘do more with less’ and challenge programs to increase their effectiveness, develop partnerships and alliances, and demonstrate value. In the public sector, activities designed to measure program effectiveness have become known under the general description of ‘reinventing government’ and include a number of specific legislative and executive initiatives, such as the National Performance Review (NPR) (Gore, 1993), the Chief Financial Officers (CFO) Act of 1990 (P.L. 101-576), Federal Performance Partnerships with States, the Government Performance and Results Act (GPRA) of 1993 (P.L. 103-62), and the Federal Management Improvement Act (FMIA) of 1996 (P.L. 104-208). The NPR, for example, mandates that US Executive Agencies define ‘results-oriented goals’ and collect performance data. The other Acts link performance data to annual budgets and agency operations while the FMIA also directs agencies to assure that their financial management systems provide the necessary information to better invest scarce resources and reduce costs. Taken together, these initiatives instruct Agencies to measure their performance in order to demonstrate impact.

This paper describes one approach to introducing performance measurement into a large Federal health program, the Ryan White Comprehensive AIDS Resources Emergency (CARE) Act Titles I and II, in response to the Government Performance and Results Act. The Ryan White CARE Act (P.L. 101-381, P.L. 104-146) represents the largest targeted Federal spending on HIV/AIDS treatment and is one of the most important sources of funds for the development of the Ryan White CARE Act funding model.

The effort described in this paper was performed by The Lewin Group for the Health Resources and Services Administration and was designed to assist the Agency and its programs in responding to GPRA requirements. It was funded through separate delivery orders under HRSA Contract No. 240-94-0037. Ms. Kates was a consultant in The Lewin Group’s Public Policy Practice at the time of this effort.
systems of care for and delivery of services to uninsured and underinsured individuals and families with HIV. The CARE Act, passed in 1990, was first funded in 1991. It is a discretionary grant program, administered by the Health Resources and Services Administration (HRSA), that depends on annual appropriations by Congress. In FY 2000, $1.5 billion was allocated through grants to states, localities, and providers of HIV services. Title I (PHSA, S2601-08) provides formula based and supplemental grants to eligible metropolitan areas (EMAs) that are disproportionately affected by the HIV epidemic. In FY 2000, Congress appropriated $543.6 million in formula and supplemental grants to Title I grantees. Title II (PHSA, S2611-20) provides grants to states, the District of Columbia, Puerto Rico and US territories to improve the quality, availability, and organization of health care and support services for individuals and families with HIV. In FY 2000, Congress appropriated $819.6 million (including dollars for the AIDS Drug Assistance Program or ADAP, which provides HIV-related prescription drugs to under and uninsured people with HIV) to Title II grantees.

In addition to the GPRA mandate, and to the growing emphasis on the need to measure performance within the public sector more generally, the Ryan White CARE Act was reauthorized during the course of this project, charging grantees with considering ‘cost and outcome effectiveness’ in their funding allocation decisions (P.L. 104-146). Measuring performance in the context of the Ryan White CARE Act is very important given that the CARE Act, which provides the bulk of the nation’s care to uninsured and underinsured people living with HIV, is dependent on annual appropriations by Congress. It is therefore critical for the Ryan White Program to be able to articulate its activities and accomplishments—namely, its performance. Performance measurement is also important in the context of the HIV/AIDS epidemic more generally, where treatment regimens are complex and ever changing, access to care remains a problem for many (see Bozzette, Berry, Duan, Frankel, Leibowitz, Leftkowitz et al., 1998), and the need to develop, test, and utilize standard measures of performance can enhance efforts to improve the quality of HIV-related care.

The approach presented in this paper offers lessons for policymakers, program administrators, and planners concerning the complexities and the value of performance measurement and its role in the development of a performance management system. It is particularly applicable to large grant programs characterized by significant autonomy and variation at the local level, as is the case with the Ryan White CARE Act and many other health programs administered by the Federal government. At the same time, while this approach may be more broadly applicable, we recognize that the particular characteristics and history of the HIV/AIDS epidemic present unique challenges—some of which are highlighted here—to performance and outcome measurement, as well as underscore the importance of demonstrating ‘what works’. It, therefore, also offers guidance for those working to provide care to people with HIV/AIDS.

We first present a background to GPRA and the complexities of performance measurement. We then discuss our approach, including key steps, to developing a performance management system using the examples of Titles I and II of the CARE Act. Overall, we believe that our approach, and the experience of the Federal Ryan White Program, may serve as a guide to the development of a performance management system for public health more generally.

2. Background

GPRA is the Federal government’s effort to establish measurable performance goals and actual results that can be reported as part of the budgetary process, thus linking funding decisions with the achieved results and performance of public programs. Among its purposes, GPRA is intended to ‘improve Federal program effectiveness and public accountability by promoting a new focus on results, service quality, and customer satisfaction’ (P.L. 103-162). GPRA requires each Federal agency to develop performance measurement systems that include comprehensive strategic plans, annual performance plans with measurable goals and objectives, and annual reports on actual performance compared to performance goals.2

GPRA, the legislation, and performance measurement more generally, represent a paradigm shift in the operation of public health programs by explicitly linking program budgets and activities to outcome measures—the actual results of program activities. This shift positions Federal programs within largely uncharted territory (see also Newcomer, 1997). There is no single blueprint or roadmap for measuring performance; efforts to produce program report cards, outcome measures, and other initiatives are relatively new and underdeveloped within the public sector (Eddy, 1998).

In addition to the dearth of models for implementing performance measurement, responding to GPRA requirements presents both behavioral and analytic challenges to public health program managers and staff: ‘GPRA forces a shift in the focus of Federal agencies—away from such traditional concerns as staffing and activity levels and toward a single overriding issue: results … This will not be an easy transition, nor will it be quick’ (GAO, 1996, p. 1). Program staff (with some justification) may be skeptical of the value of change, in particular the type of change anticipated by GPRA, that ultimately may require increased innovation, accountability, and more explicit allocation of program resources. Staff may have

---

2 The first strategic plan was due by September 30, 1997. Strategic plans are to be updated at least every three years. Agencies’ annual performance plans are due each February, beginning in fiscal year 1999. Annual performance reports are due each March, the first of which was due by March 31, 2000, on program performance for fiscal year 1999.
experienced similar efforts with varying degrees of success, or may question, based on their actual experience, whether increased knowledge about a program’s needs, priorities, or actual performance translates into stable resources or support of program initiatives.

Performance measurement is also methodologically complex, and there are several challenges to the effective measurement of health programs. Variation in health systems infrastructure, market dynamics, epidemiology, and demographics makes it difficult to connect Federal level activities to effects at the local level. In addition, because clients of Federal programs may receive services at multiple locations, each with multiple funding sources, it is difficult to align a Federal program activity (or grant dollar) with a client-level outcome such as reduction in morbidity and mortality. Indeed, individual client characteristics and varying levels of disease severity may require sophisticated case mix adjustment to make comparisons meaningful:

The fundamental reason that establishing results-oriented goals is so difficult is that, to set such goals, agencies must move beyond what they control—that is their activities—to focus on what they merely influence—their results. (GAO, 1997, p. 5, emphasis added)

In other words, federally funded public health programs are mediated at the local level by many other factors, including the impacts of other federally, state, and locally funded programs, socio-economic and cultural factors more broadly, and individual behaviors (see also GAO, 1999). In our work with a large federal health program, we came to describe this as the ‘problem of attribution’ or the difficulty of attributing client or local level impacts to higher (state or Federal) level activities.

We also described the related ‘problem of aggregation’, or the difficulty in aggregating client or local level impacts into meaningful but substantially broader and standardized categories. This is especially important given that, while the state of the art in measuring results is at the program or activity level, the regulatory attention of GPRA is focused on Cabinet-level departments and independent agencies.

Despite the difficulty with performance measurement, there is a strong feeling on the part of many program managers and advocates that if a program such as the Ryan White CARE Act were taken away, there would be a significant negative impact on communities and individuals—measuring that impact, or, the difference between having a program and not having it, is the crux of performance measurement. Yet, this type of scientific measurement is impossible to do, since such a ‘controlled experiment’ could not, and many would argue, should not, exist.

Performance measurement is valuable for public programs to undertake; indeed, in an era of increasing resource competition, performance differences may be one of the only ways to distinguish between equally valuable and potentially worthwhile public goods. In addition, performance measurement may help program managers and operators gauge how they are doing, against both external and internal goals. For performance measurement efforts to demonstrate this value (beyond merely responding to legislative requirements such as GPRA or CARE Act Reauthorization) and be practicable, they must be grounded in and emerge from program experience. This is particularly true of many Federal public health programs, such as Ryan White, where significant planning, management and decision-making responsibilities are delegated to the state and local levels. An approach which is grounded in program experience and works to create staff engagement, support and participation is necessary for the selection of meaningful measures that support effective service delivery.

Thus, we believe a more ‘holistic’ approach, one that accounts for program context and activities and informs ongoing program planning, focuses on the development of a performance management system designed to reconcile local needs for flexibility with a centralized need for standardization. Conceived of here, performance measurement becomes one activity—albeit a critical one—in a larger effort to develop an underlying system around the management of performance that uses performance data to assess how it is doing and chart its course. Such a management system supports planning, accomplishes internal and external/ environmental surveillance, and provides data and information feedback into the system. The specific aim of such an approach is to build a performance infrastructure, or the sustainable capacity of an organization or program to produce and learn from its performance measurement data. Moreover, it enables a program to tell its unique story over time. Table 1 presents the key steps used as part of our approach.

It is important to emphasize that this approach is predicated on the use of an iterative process recognizing that the development of a performance management system is both ongoing and nonlinear—conceptual models and specific values or data inform one another throughout the process. In the context of the Ryan White CARE Act, the performance system developed was ‘poked and prodded’ to reflect ongoing staff input and to capture alternate or even competing perspectives from the different program operating units on the relationships between program operations, goals, and results. This strategy combined the concurrent activities of the design of the performance management system with data collection and analysis. As the design of the performance management system
progressed, the process informed decisions about what program information and data were still needed to move forward. Similarly, program information and the availability of data both structured the development of the system and pointed to its current limits. For example, in the case of the HIV/AIDS epidemic, the lack of reliable estimates of HIV prevalence limits the ability to identify the population in need of Ryan White services and, therefore, the ability to fully measure the CARE Act’s response to this need.

The next section of this paper elaborates upon these key steps using the example of Titles I and II of the Ryan White CARE Act.

3. Development of the performance system

3.1. Involving key stakeholders

One of the first steps in the development of a performance management system is to involve key stakeholders. Key stakeholders include program staff, grantees, and clients, each of whom represents different components and levels of the program (see also Newcomer, 1997; GAO, 1996). Involving stakeholders in the development of a performance management system is critical for making the system meaningful, operational, and sufficiently flexible, as well as creating stakeholder engagement in the system. Stakeholders from a variety of program levels have unique perspectives and different levels of knowledge about the importance of a program measure and/or the feasibility of operationalizing measurement. In addition, stakeholders can assist in specifying the expected or desired values of program indicators (e.g., the percent reduction in inpatient utilization or the number of expected ambulatory medical care visits). In the context of the CARE Act, stakeholders initially were limited to staff at the Federal level because of the immediate need for the Federal program to respond to Congressional GPRA requirements, the complexity of the endeavor, and the need to first achieve consensus among multiple program operating units at the Federal level. Plans are underway to involve grantees and clients in the development of measures that are appropriately specified and operational.

3.2. Articulating program purposes and goals

Effective performance management must identify and describe a program or organization’s mission, purpose, and objectives. These guide the development of a performance management system so that it appropriately specifies the relationships among different organizational components and the dimensions along which performance measures must be developed. A statement of program purposes(s) should answer these questions: ‘what is the program trying to accomplish?’ and ‘why is the program necessary?’

A primary source of information about a public program’s purpose is its legislative language and legislative history. The legislative purpose of the CARE Act is ‘to improve the quality and availability of care for individuals and families with HIV disease’ (P.L. 101-381). Title I provides emergency assistance to localities that are disproportionately affected by the HIV epidemic via the designation of HIV Health Services Planning Councils which establish priorities for allocating funds, develop comprehensive HIV services plans, and assess the efficiency of the administrative mechanism. The purpose of Title II is to enable states to improve the quality, availability, and organization of health care and support services for individuals with HIV disease and their families in three ways: through health services Consortia and home and community-based care services; through the continuation of health insurance coverage; and by development of or funding for AIDS Drug Assistance Programs (ADAPs) (P.L. 101-381).

In addition to legislative language, program managers have much to offer about a program’s purpose based on their own attempts to implement and administer the legislative intent. Program managers are invaluable sources for describing how legislative intent has been operationalized at various program levels. In the context of Ryan White, the Federal program’s role is to ensure that grantees use their state or local authority to meet the provisions of the CARE Act. Significant planning and funding allocation decisions are delegated to the local level. An Aggregate Reporting System (AAR), already in place, provides data on each grantee’s clients and services.

Performance measures developed as part of a Ryan White performance management system, therefore, should collectively demonstrate how the CARE Act develops systems of care for people with HIV and improves the quality and availability of care at the state and local levels through Title I and II activities. Agreement on program purposes may be difficult and require significant up front investment of staff time.
3.3. Developing a program logic model

Once the purposes and goals have been identified, it is necessary to develop a clear picture of the program's 'logic'. A program's logic is the representation of the interactions and relationships among a program's objectives and stages of activities, or performance (e.g. the flow of inputs, activities, outputs, and outcomes required to achieve the objectives). A performance management system needs to capture both the individual properties of each program component while connecting them within, at least theoretically, a coherent system. Specifying these relationships is an important step in identifying locations of control over program results—that is, what a program may be held accountable for and, therefore, the targets of performance measurement. A logic model is one tool for representing these relationships (see, for example: Cozzens, 1997; Julian, 1997; Plant & Rushworth, 1997, on identifying potential points of measurement in the context of a micro-organism). A logic model helps to map the flow of dollars, decision-making, and responsibility throughout a program's operations demonstrating the potential locations of program impact and, therefore, measurement. It also assists in identifying those points where program activities are mediated by other factors (such as system or client variation).

A sample logic model for Titles I and II is presented in Fig. 1. The model depicts the complexity of this public health program and the multiple levels of activities expected to occur in order to achieve the desired results. Federal grant funding is dispersed to multiple and diverse grantees at the state and local level. Title I and II planning and decision-making activities occur at multiple levels as well and

![Sample logic model for Titles I and II.](image-url)
diverse constituencies are involved in the funding allocation and disbursement process to numerous sub-grantees, who may include providers and community-based organizations.

3.4. Using standard terminology

In the new era of ‘performance measurement’, ‘quality assurance’, and ‘total quality management’, a plethora of definitions and concepts has emerged. These are not necessarily consistent across efforts or organizations. Therefore, it is critical to identify standardized terminology within a performance management effort. Standard terminology provides the necessary common lexicon for achieving mutual understanding across stakeholder views and grantee variation and allows each to tell a consistent yet flexible story over time. Here, we define the concepts of performance dimensions, performance measures, and performance indicators, each of which is a distinct but interrelated component of an overall performance management system. It is important to note that the terms ‘measure’ and ‘indicator’ are often used interchangeably, but we define them differently here (see Table 2).

**Performance dimensions** function as ‘themes’ used to tell a story over time. An example of a performance dimension in the context of public health is ‘ACCESS’. Enhancing ACCESS to health services is fundamental to most public health programs, such as Ryan White, designed to serve vulnerable populations (see Aday, 1993, for example). While the desired story of enhancing ACCESS (the performance dimension) will remain consistent, reliable and meaningful measures (factors or variables, such as the ratio of providers to clients) of ACCESS may become more readily available over time and/or program data systems may be enhanced to more adequately measure ACCESS. Similarly, the expected indicators (actual values, such as the numerical value of the provider client ratio) of each ACCESS measure will likely change as, in the context of HIV for example, populations affected by the epidemic change and as health insurance coverage and other system level factors change.

Performance measures may operate at different stages of program performance, along a continuum from input to outcome. These stages are presented in Table 3. It is often difficult to locate a measure along this continuum. It is important to recognize, for example, that a single performance measure may represent different stages of performance at each program level or to different stakeholders. In other words, ‘my output may be your input’. For example, if the output of the Ryan White program at the Federal level is the disbursement of Federal funds to ‘x’ number of grantees, the input at the grantee level is the amount of funds whereas the output at the grantee level may be the disbursement of funds to ‘y’ number of providers. In addition, depending upon data availability, the feasibility of measurement, and the current level of knowledge, process and/or output measures may have to serve as surrogates or proxies for outcome measures; they may also be thought of as intermediate outcomes, or those which may reasonably be assumed to lead to the ultimate outcomes. In order to use a proxy measure to represent an outcome, however, the cause and effect relationship between them must be clearly described.

Meaningful indicators of performance derive from comparisons between two values, positions, or points in time, and may be of different types (Table 4). For instance, actual program performance (e.g. number of clients receiving services in the year 2000) may be compared to baseline performance (e.g. number of clients receiving services in the year 1997), a planned level of performance (e.g. the planned number of clients to be served), an average level of performance (e.g. the average number of clients served among all service providers), or a benchmark/best practice standard of performance. Comparison, however, pre-supposes that programs know their baseline, or how they are functioning today. This is particularly challenging within the context of HIV, where it is quite difficult to determine the population in need. Therefore, performance management might initially look at comparisons of grantees to themselves (e.g. what did we do compared to what we planned to do), rather than a comparison to total need or across grantee entities. Over

---

5 This is largely because reliable estimates of the number of people living with HIV at the sub-national level are not readily available. In addition, many individuals with HIV do not know their serostatus.
time, as performance management systems become more sophisticated and use techniques which can adjust for grantee variation (e.g. risk adjustment, HIV prevalence estimation or measurement techniques), estimates of unmet need (e.g. what did we do compared to what is needed) or cross-site comparison (e.g. what did we do compared to others) may be more easily accomplished.

In sum, Table 5 presents an illustration of the overall relationship between performance dimensions, measures, and indicators (in this case, the baseline and planned values compared to the actual value), within a performance management system.

3.5. Identifying performance dimensions

As described above, performance dimensions are broad areas or themes that align with and are designed to capture program purposes and functions (e.g. ‘ACCESS’, ‘QUALITY’). Based upon an analysis of the legislative purpose of the CARE Act, discussions with program managers, and review of the program logic model, six performance dimensions, most of which are frequently referenced in the health care literature and measurement field (in addition to Aday, 1993, see also Hawe et al., 1997; Iglehart, 1997) were identified to collectively capture Title I and II activities and purpose. Taken together, these dimensions are designed to tell a story about the performance of Titles I and II over time:

1. UTILIZATION. This dimension is designed to provide data on clients, providers, service units, and costs for Titles I and II funded programs. Utilization data are important for establishing a baseline and a denominator. They provide a fundamental measure of Title I and II operations because it is difficult to measure access, unmet need, or any health outcomes unless the program knows how many clients it is currently serving. In addition, utilization trends over time can assist the program in projecting need and future capacity.

2. ACCESS. This dimension reflects the legislated goal of the CARE Act—to improve service use for individuals who otherwise would not have access to services, the uninsured and underinsured.

3. UNMET NEED. By improving access, Titles I and II are also intended to reduce unmet need for care for people with HIV.

4. QUALITY. Enhancing the quality of care delivered to individuals with HIV is another important goal of the CARE Act. Quality of care includes such components as the types of clinical services offered, their appropriateness, the types of professionals who deliver care, and related health outcomes.

---

Table 3
Measures and stage of program performance

<table>
<thead>
<tr>
<th>Stage of performance</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>Measure those program components (financial and non-financial) which comprise the ‘raw material’ of the program, e.g. budget dollars, number of FTEs</td>
</tr>
<tr>
<td>Process</td>
<td>Measure the way in which a program or intervention (planning, service delivery) occurs, e.g. number of long-range plans completed; representation on Planning Councils</td>
</tr>
<tr>
<td>Output</td>
<td>Measure tangible results of program activity, that which is produced, e.g. number of service units delivered; number of clients served</td>
</tr>
<tr>
<td>Outcome</td>
<td>Measure the end result or impact of a program or intervention, e.g. number of preventable opportunistic infections prevented; increase in the time between AIDS diagnosis and death</td>
</tr>
</tbody>
</table>

Table 4
Types of indicators

<table>
<thead>
<tr>
<th>Type of indicator for comparison</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Known value or position used to locate something; data used for comparison or as control prior to start of program (e.g. current mortality rate)</td>
</tr>
<tr>
<td>Planned</td>
<td>Goal or target against which actual performance data can be measured (e.g. ‘x’ percent reduction in mortality rate)</td>
</tr>
<tr>
<td>Average</td>
<td>Average or median position of many programs or grantees within a particular measure (e.g. the average reduction in mortality rate)</td>
</tr>
<tr>
<td>Benchmark/Best practice</td>
<td>Something that serves as a standard by which others may be measured, best practices, clinical standard (e.g. the ideal or greatest possible reduction in mortality rate)</td>
</tr>
</tbody>
</table>

Table 5
Relationship between performance dimensions, measures, and indicators

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
<td>Number of individuals served</td>
</tr>
<tr>
<td>Indicator</td>
<td>Baseline: 2000 (Baseline), Planned: 3000 (50% increase), Actual: 4000 (100% increase)</td>
</tr>
</tbody>
</table>
5. **CONTINUUM OF CARE.** This dimension encompasses both the availability of appropriate treatments and services throughout the course of illness and the availability of system and service delivery linkages to enable clients to receive the full range of services.

6. **CAPACITY BUILDING.** Finally, Titles I and II, through the use of planning bodies and the creation of service delivery systems, are intended to help create and enhance the capacity of health systems to deliver services to people with HIV and to have an ongoing impact on care.

### 3.6. Developing a universe of potential measures

Once performance dimensions have been agreed upon, measures (the factors or variables which describe dimensions) can be specified within them. In the initial stages of performance planning, it is important to identify an inclusive list (as inclusive as possible) of potential measures from which the more critical or relevant can be selected by all stakeholders involved. We conceptualized this as a ‘universe’ of potential measures, or the broad set of measures that might be important for a particular program. A review of relevant materials (program information, existing data sources, research) and development of a conceptual framework described above enabled us to create a universe of potential measures for the Ryan White CARE Act Titles I and II.

Measures may be selected from this universe for different purposes or because they measure different stages of performance or are appropriate for different audiences. For example, some measures may be appropriate for reporting performance to Congress (e.g. the total number of clients served nationally by all Ryan White Title II grantees) while others may be more useful for internal program management (e.g. the number of clients served in a particular state) and still others may be more useful at the Planning Council level. Again, what links the measures together is their connection to performance dimensions. Therefore, different measures may tell different parts of the same story.

### 3.7. Identifying data capacity, needs and gaps to specify performance indicators

Integral to performance management is the identification of current data capacity, data needs, and data gaps, a process that must occur throughout the development of a performance management system. This process informs system design, the specification and selection of measures, and, ultimately, the designation of performance indicators (or actual values) for those measures. If a potential measure is considered extremely important and meaningful but is not supported by current data collection efforts, an organization may either decide not to use the measure or to revise data collection efforts to collect appropriate data. There may be many reasons why an organization would choose not to use a measure despite its theoretical importance—it may be too expensive or cumbersome to introduce new data collection requirements or measuring the data may be too methodologically complex. The benefits of instituting new data collection efforts to support performance measurements must be weighed carefully against the burdens, including cost and time, particularly since the burden of data collection increases as one moves closer to the operational location of program activity, e.g. at the grantee, Planning Council, and provider levels. On the other hand, by deciding to collect new data, an organization might enhance its performance system by enabling the measurement of critical activities or outcomes. In the context of HIV, for example, if a new study suggests that physician experience in treating patients with HIV has an impact on patient survival (see, for example, Kitahata, Koepsell, Deyo, Maxwell, Dodge, & Wagner, 1996), Planning Councils may decide to start collecting data on individual physician experience (as measured by experience with AIDS during residency, the cumulative number of patients with HIV/AIDS cared for over time, or the number of patients with HIV/AIDS seen over the past year). Assessing the limits of current data sets is thus an important part of performance measurements that helps to guard against selecting measures based only upon readily available and easily understandable data.

Data for performance measurement should be defined broadly to include sources that provide data on services and clients, such as utilization and demographics, and sources that provide information on program functions and expectations. In addition to internal program-specific data sources, external data sources may provide important environmental surveillance and/or comparative data (e.g. for averages or benchmarks). Table 6 presents some examples of data sources available for performance management of Titles I and II.

There are several limitations to these data sources, many of which are limitations faced by other public health programs. First, the primary internal program data system (the AAR) does not provide an unduplicated client count (except for limited demonstration sites and for some state level data). Without unduplicated counts, it is difficult to have accurate baseline data on the number of clients served. Another limitation of internal data sources is their lack of standardization (e.g. grantees conduct their own needs assessments, which are not standardized).

---

6 For example, the costs of a proposed set of measures should not be greater than actual program costs. This represents a potential problem of performance measurement, not discussed in this paper, namely that measurement and accountability efforts have emerged in part due to scarce resources and the need to make hard choices. Yet, these efforts run the risk of requiring significant resources on the part of public programs including dollars, time, and staff.
External data sources also have limitations. The primary one in the context of HIV is the lack of reliable national, regional, or local estimates of the number of individuals with HIV in need of Ryan White services. In addition, available studies and research are not usually applicable beyond their study population, limiting their ability to serve as best practice or comparison studies. One study, the HIV Cost and Services Utilization Study or HCSUS (a public/private collaboration) is the first to offer nationally representative data on the population infected with HIV in care. HCSUS, however, is also limited in that it only provides data on people with HIV who are already receiving care (Bozzette et al., 1998).

### 3.8. Specifying criteria for selecting measures

Once the universe of measures is developed and data sources examined, it is important to identify the critical (see also the GAO’s ‘vital few’ in GAO, 1996) set of measures, that which is most meaningful and useful now. These may also serve as pilot measures (see next step).

The process of moving from a universe of measures to the critical subset is a complex one. There are several criteria that can assist decision-makers in the selection of appropriate performance measures as shown in Table 7. This list of criteria is not meant to be exhaustive. Stakeholder involvement in an ongoing process may produce new and important criteria for selection. Rather, this list is presented to provide some guidance on the type of questions to be considered when selecting measures. In the case of the Ryan White program, the limitations of existing electronic health data sets and the time needed for new data collection were important factors in selecting measures and related indicators. Some of the measures being considered and piloted include measures of disease progression, such as CD4 counts and HIV viral load burdens, and the number of preventable AIDS-related hospitalizations—these measures may be indicative of inadequate access to or delivery of primary care, including recommended antiretroviral therapy.

---

### Table 6
Potential data sources for Ryan White performance measurement

<table>
<thead>
<tr>
<th>Internal data sources</th>
<th>External data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual administrative reports (AAR)</td>
<td>CDC surveillance data</td>
</tr>
<tr>
<td>Client level data demonstration initiatives</td>
<td>Clinical studies/research</td>
</tr>
<tr>
<td>Grantee applications</td>
<td>Best practice studies</td>
</tr>
<tr>
<td>Grantee needs assessments</td>
<td>Hospital discharge data</td>
</tr>
<tr>
<td>Grantee studies and evaluations</td>
<td>HIV Cost and Services Utilization Study (HCSUS)</td>
</tr>
</tbody>
</table>

*a This would include such data sets as the CDC’s regular surveillance of HIV and AIDS incidence and prevalence, studies such as the Supplemental HIV/AIDS Surveillance study (SHAS), and other sentinel studies of HIV in selected populations and/or facilities including those of sexually transmitted disease clinics and Job Corps entrants.

### Table 7
Criteria for selecting measures

<table>
<thead>
<tr>
<th>Does the measure have explanatory power?</th>
<th>Are data available to support measurement and define indicators?</th>
</tr>
</thead>
<tbody>
<tr>
<td>For external sources, e.g. to Congress</td>
<td>Are data reliable?</td>
</tr>
<tr>
<td>For internal management, e.g. HRSA, grantees, Planning Councils, Consortia</td>
<td>Are data accurate?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Are different measures appropriate at different program levels?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Federal (HRSA) level indicators to monitor grantee performances and provide technical assistance</td>
</tr>
<tr>
<td>Grantee and Planning Council level indicators to plan for service delivery and monitor contracts</td>
</tr>
<tr>
<td>Provider level indicators to monitor quality of patient care and clinical pathways</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Are measures more appropriate at different times?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate or near-term use</td>
</tr>
<tr>
<td>Long-term use</td>
</tr>
<tr>
<td>Not to be used</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How much of a burden is presented by this measure?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
</tr>
<tr>
<td>Time</td>
</tr>
<tr>
<td>Need for new data collection efforts</td>
</tr>
<tr>
<td>New resource requirements</td>
</tr>
</tbody>
</table>

| Is this measure meaningful on its own? Does it need to be grouped into a more meaningful combination? |
and prophylaxis and treatment for AIDS-related opportunistic infections.

3.9. Pilot testing performance management system and measures

Once stakeholders agree on a set of measures, it is important to pilot test the set at all program levels. Even if a measure has sufficient explanatory power in theory, it may be too difficult to use or may at least result in new technical assistance needs (particularly for those operating the program, such as grantees and Planning Councils) in practice. Or, program measures may not accurately capture what they were originally intended to measure. In addition, new data or information, derived from operationalizing measurement, may shed new light on the meaning of a measure. Finally, measurement may highlight the dependence of a measure on other measures or conditions. All of these kinds of operational concerns may best be examined through pilot testing. Currently, the Ryan White Program is pilot testing the ability of seven state and local grantees to collect and measure health-related program data. The pilot test has underscored the difficulty in measuring performance. Recent data, for example, indicate that HIV-related treatment failures and treatment-induced illnesses, such as diabetes and heart disease, could increase hospitalizations or viral load burdens even though standard treatment protocols may have been followed (Henry, 2000).

3.10. Ongoing assessment and evaluation

Finally, to bring the process full-circle, it is important to continually revisit the performance management system developed to ensure that the system accurately reflects program goals and objectives and that measures remain meaningful and operational. In addition, as measures and their corresponding indicators continue to tell important information about a program (and as new data become available), the program and its performance management system should be modified accordingly. New data, often external to the program, may serve to support measurement where previous data were lacking (such as newly available estimates of the number of people infected with HIV).

4. Conclusion

In the context of the Ryan White CARE Act, the short-term success of this approach may be affirmed by the use of program measures and indicators in the 1999 and 2000 Federal budget by the Ryan White Program. However, the long-term success will depend upon the extent to which the Ryan White Program is able to maintain a performance management system whose measurement activities improve both Federal program and grantee performance and contribute to future efforts to ensure delivery of HIV services for under and uninsured individuals and families living with HIV.

The next steps facing the Ryan White Program are to formally involve grantees, Planning Councils, Consortia, and clients in performance management system development, measure specification, and measure selection (all of the steps described in this paper). A series of consultations have begun with Title I and II grantees around the six performance dimensions presented above. These consultations range from one-day discussion groups to electronic bulletin boards, to more formal reviews of revised data collection systems. Measures will need to continue to be piloted at the local level. Consideration should be given to expanding the development of common measures to the other Titles of the Ryan White CARE Act and similarly piloting these at the local level. In addition, more sophisticated models for comparing performance across grantees and providers must be developed to support long-term performance management and the identification of best practices for others to learn from. Finally, some analytic challenges will have to be addressed, including the challenge of estimating the true population in need and of adjusting for variation across client populations and health systems. These are problems faced by many public sector programs, particularly those with significant variation at and delegation of planning and activity to the local level.

Despite the analytic and behavioral challenges presented by performance management, it is clear that an emphasis on accountability, efficiency, and quality is becoming an integral part of the public sector landscape. Within the context of a large public health program, a complex, changing disease, such as HIV, and tighter resources, models which assist programs in demonstrating their value and re-assessing their activities to provide greater value, are one critical tool for improving the public health.

References


