CHAPTER 2

Outcomes Measurement in the Human Services: Lessons Learned from Public and Private Sectors

Harry P. Hatry

Since the first edition of *Outcomes Measurement in the Human Services* was published in 1997, a deluge of outcomes and performance measurement activity has taken place in the human services. This trend is occurring not only in the United States, but in many other developed and developing countries. The chapters in this book's second edition are testimony to the advances in activity and thinking that have occurred in outcomes measurement. To date, the bulk of this activity has been aimed at responding to the increasing demands for accountability put forth by funders of human services. However, the quality of this measurement work and its use by human service organizations (HSOs) for improving services are questionable. This chapter will shed light on this issue and discuss major outcomes measurement activity that has occurred in the public and private sectors over the past dozen years, as well as emerging issues in outcomes measurement.

(In this chapter, *outcome measurement* generally refers to the regular tracking, at least annually but almost always more frequently, by an organization of

the outcomes of its individual services and programs. *Outcomes management* refers to the use of outcome information by organization managers to help them manage better, such as better allocated resources and better formulated and justified budgets. The term *performance measurement* is frequently used to cover outcome measurement. Performance measurement covers not only the measurement of service outcomes, but also the measurement of the amount of physical output produced by an organization, such as the number of sessions held with clients. The word *outcomes* refers to what results occurred from the outputs. Performance measurement also covers the measurement of a service's efficiency, usually expressed as the cost per unit of output or, where possible, the cost per unit of outcome.)

Major Recent Activities in Outcomes and Performance Measurement

During the past dozen years, many factors and activities have shaped the approaches to outcome measurement in HSOs.

Government Performance and Results Act

The U.S. federal government fully implemented the 1993 Government Performance and Results Act (GPRA) in fiscal year 1999. The years since then have seen an increase in the quality of the reported performance indicators. Each federal department and major independent agency annually develops a performance plan as part of the budget process and, within six months after each fiscal year, provides a Performance and Accountability Report. These reports contain numerous outcome indicators covering each of the agency's major programs. This clearly has had a major effect on performance measurement in the federal government. For example, the Department of Health and Human Services (DHHS) Administration for Children and Families, through its Adoption and Foster Care Analysis and Reporting System, has pressed the states (not always without controversy) for more valid data on the outcomes of child welfare clients (such as placement frequencies and outcomes of children who entered foster care).

The George W. Bush administration added another component to GPRA, the Performance Assessment Rating Tool (PART), which was in use from 2002 to 2008. The intention was both to encourage improved performance measurement and to use the information in the Office of Management and Budget (OMB) budget review process. By the end of 2008, the OMB had examined approximately 1,000 federal programs, including more than 100 in the DHHS. A major element of this effort was an examination of a small

number of key outcome indicators for each of the programs (an average of approximately five performance indicators per program). These run the gamut from response times, to client satisfaction ratings (for example, Administration on Aging's "percent of home-delivered meals, transportation, and family caregiver services clients rating quality of services either good or excellent"), to improved condition levels (for example, Substance Abuse Prevention program's "percentage of participants who used illicit drugs at pretest who report a decrease in use at post-test," and Child Welfare Services' percentage children in care less than 12 months who had no more than two placement settings).

The PART process focused considerable attention on the performance management process in federal programs, rating elements of the program's purpose and design, strategic planning, and program management. OMB's ratings on these process elements indicated considerable improvement over the period. However, PART was considerably less successful in assessing the overall effectiveness of the programs. Although much has been written about this controversial PART process (Accenture Institute for Public Service Value, Georgetown Public Policy Institute & OMB Watch, 2009), much less has been written about the parallel Performance and Accountability Report process that provided considerably more detailed data on program outcomes (see, for example, work reported by George Mason University's Mercatus Center).

OMB Performance Data on State and Local HSOs

The pressure on U.S. federal departments to provide national performance data to the OMB and Congress has had a substantial effect on state and local HSOs in a myriad ways. Much of the human service outcome data originate at those lower levels of government. When accepting federal funds, HSOs often are required to provide performance data on selected outcome indicators, often ones selected by the DHHS program. This has required states and local governments to implement many outcomes measurement procedures that would unlikely have been implemented without such pressure.

Foundation Reporting Requirements

A number of major foundations in the United States (such as Kellogg, Hewlett, Gates, and Robert Wood Johnson) now often require reports and other information that document the results associated with the funding they provide. Sometimes they themselves sponsor in-depth evaluations of at least some of their programs. As with the federal government, this has put considerable pressure on HSOs that receive or want to receive funding from these foundations to undertake outcomes measurement or to cooperate with evaluations done by other organizations.

Focus on Community-Level Indicators

U.S. HSOs in the private, nonprofit sector were also encouraged to begin outcomes measurement in the mid-1990s by United Way of America with the 1996 release of its bestseller (more than 100,000 copies sold) report *Measuring Program Outcomes: A Practical Approach*. Other major national associations such as the Boys and Girls Clubs of America and the American Red Cross have shown leadership to their affiliates by providing support for local outcomes and performance measurement efforts.

The United Way movement has shifted to a focus on communitywide needs and community indicators. To do this effectively, partnering is required among many community groups and organizations. Other national groups, such as the Community Indicators Consortium, have emerged to help identify the communitywide outcome indicators needed to help communities tackle their problems in a more coordinated way.

Evidence-based Practice

Around the beginning of the 21st century, considerable attention began to be focused on implementing evidence-based practices in various domains (for example, health, mental health, child welfare) in an attempt is to identify which polices and procedures work well. This focus will likely continue into the future. However, what is meant by *evidence-based practice* has not been well defined. Which evaluation and measurement methods are sufficient to meet this criterion? Initially, the U.S. federal government pressed for the evidence to come from randomized controlled trials. However, it became apparent that this would greatly limit the applicability of such evaluations for many—probably most—public services, including human services. The pressure for rigorous studies has continued, but use of a larger range of evaluation methods is now acceptable. This pressure will likely considerably increase the use and volume of program evaluation in human services (Orszag, 2009).

Technology

Tremendous technological developments have made the collection and processing of substantial amounts of outcome data considerably more practical and affordable for many HSOs. This trend has added substantially to HSOs' ability to process, analyze, and report data in more sophisticated, attractive, and useful ways. For example, it is now technologically feasible to link case management records that identify an individual client's demographic characteristics and the types and amounts of services received to outcome information on that client. This can be done across services if confidentiality and privacy considerations

are resolved. Such information made available to service managers can provide them a wealth of information on what is working well (or poorly) for which groups of clients and under which conditions. However, as of this writing, such opportunities have not been widely, if at all, realized.

International Focus on Outcomes Measurement and Evaluation

Several countries in addition to the United States have introduced outcomes measurement (and evaluation) systems covering human services with a national focus, including Australia, Canada, New Zealand, and the United Kingdom. This has had ripple effects, affecting these countries' state/provincial governments as well, requiring their human service agencies to provide outcome data and encouraging implementation of outcomes measurement and evaluation.

Internationally, major organizations such as The World Bank, Inter-American Development Bank, and other multilateral funders (nongovernmental organizations that use funds obtained from many countries to help developing countries), as well as bilateral funders (countries that use their own funds to directly aid developing counties) such as the United States (through its Agency for International Development), Canada, Germany, Japan, and the United Kingdom, have pressed for outcome and evaluation information,

Numerous international HSOs that provide services to citizens in needy countries, using funds that have been obtained at least in part from the above organizations and countries, have also been encouraged to introduce outcomes measurement and evaluation practices. (These HSOs include, for example, such organizations as Heifer International, Women for Women International, Partners in Health, and the International Rescue Committee.)

Emerging Issues and Developments

Considering the expansion in interest, requirements, and need for outcome measurement, and evaluation in the United States and abroad, several questions arise:

- Are the outcome data that are being collected sufficiently valid and comprehensive? Do the data cover enough of the outcome dimensions important to HSOs and others, and especially the clients they serve? Are the data being collected and reported in a reasonably sound, valid way?
- How is outcomes and performance measurement information being used? Is it being collected mainly for accountability purposes? To what

extent are the data being used to improve services and programs, and thus improve outcomes for the clients of HSOs?

The remainder of this section discusses a number of the key issues related to answering these questions and other emerging issues and developments in outcome measurement for human services in the United States.

Need to Intertwine and Balance Outcomes Measurement and Program Evaluation

To be of most use by human service program managers, outcomes data are likely to be needed frequently, such as quarterly, if not monthly, at least for some indicators. This greater frequency lets managers obtain feedback in a more timely way, providing opportunities to undertake midcourse corrections. Outcomes measurement is analogous to information commonly available to the manager of any sports team: the running score. Managers keep track of the score to tell whether their teams are winning or losing. However, the scores do not provide information about why the teams are winning or losing.

A major misunderstanding among public officials, and probably the media and public, is that the government agencies and nongovernmental organizations that measure and report the outcomes information have complete control over these results. But these outcomes only tell what the score is. They do not indicate who or what is responsible for why the outcome occurred as it did. Many factors inevitably affect all outcome measures—the weather, the international economy, other levels of government, the parents, and so on.

Outcomes measurement, not in-depth program evaluation, has been the major focus of both public and private HSOs, both because it provides considerably more timely information and because it does not require the more sophisticated methods and statistical tools usually needed for program evaluations. Outcomes measurement, however, provides little if any evidence as to why the outcomes occurred.

In-depth, ad hoc program evaluations attempt to identify not only the outcomes for a program, but also the causes of the outcomes. HSO program evaluations are essentially intended to focus on the question, "Are the program's clients better off than they would have been without the program?" These evaluations require a substantial amount of time, both of people to do the evaluations and calendar time. Program evaluation also generally requires evaluators who have had special training, training seldom taken by HSO staffs. Program evaluations can provide HSOs with information that adds to information that is solely focused on outcomes.

Both these approaches—outcomes measurement and program evaluation—have advantages and disadvantages. However, because outcomes measurement is an ongoing process, it is likely to be considerably more useful to HSOs and their managers. Outcomes measurement is less expensive than in-depth program evaluations, requires less specialized and sophisticated technical skills, and can be applied on a continuing basis to many programs. In-depth evaluations typically require special technical skills, often require many months if not years to complete, and are costly. Generally, program evaluations need to be contracted out, or a university may be coaxed into doing them for free.

Overall, the feasibility of undertaking full in-depth program evaluations in more than a very small percentage of HSOs is unlikely. Few private, nonprofit HSOs are able to afford in-depth evaluation of their own programs, at least not without major funding from a foundation or government agency. Even in government, only a few human service programs are given a full evaluation in any given year. This is also mainly due to cost and the limited amount of time government staff have to conduct or oversee such activities. The great growth in computer technology has made outcomes measurement much more feasible for HSOs. Today's technology is also making the design and implementation of outcomes measurement systems more useful, with capabilities that can integrate multiple human service sectors and client bases and move us toward workable electronic medical and human service records.

Need to Address Disparities (Inequities)

A major issue in human services is addressing the significant disparities in service access, delivery, and quality between advantaged and disadvantaged populations. In past decades, closing the gap in funding allocation disparities has been the focus. With the heightened emphasis on outcomes measurement, a second major dimension can be added to this strategy: The outcomes for populations of disadvantaged populations can be compared more definitively with those of more advantaged populations, as well as within and across their own groups. The latter comparison is of special concern because outcomes for disadvantaged populations that are served by many public programs vary according to client characteristics, including gender, age, and geographical location (such as rural versus urban).

With the considerably more powerful computers and software technology now available, comparisons between (and among) disadvantaged and advantaged populations have started to become a relatively easy and innovative process. For example, geographic information systems and other technology make

mapping readily feasible and inexpensive, helping HSOs not only to identify geographical differences, but also to report the information considerably more effectively, and more dramatically, to HSO officials and the public.

Emergence of the "How Are We Doing"/STAT Movement

Recent years have seen the emergence of a potentially terrific management tool designed to take advantage of newly available outcome information. It is based on the simple idea that periodically, an upper-level manager meets with his or her key employees to discuss the latest performance report that addresses such questions as, "Where are we doing well? Can it be transferred elsewhere? Where are we doing poorly? Why is that? What can we do to improve this?" Then, in later meetings, the group would be asked to address the question, "Did our changes made previously have the results that we hoped for?"

This process is similar to that of regular meetings agency managers traditionally hold with their key staff. However, in this new approach, a major focus of the meeting is on examining outcomes. This approach is based on the STAT ("statistics") movement that was started in the 1990s by the New York Police Department (called CompStat; see also chapter 22 in this volume). A number of other New York City agencies adapted the CompStat approach to their own human service focus. Likewise, a number of local and state (for example, Washington and Maryland) governments have adopted the approach, and some federal agencies (for example, the Veterans Health Administration) are also experimenting with it. For example, the city of Baltimore's CitiStat process has included staff from each city agency in its biweekly mayor's meeting to discuss the agency's performance.

Most of these STAT programs have used a formal process for obtaining the outcome and related data in advance from the organization unit being reviewed, assigning dedicated staff to review the agencies' data, and establishing formal meeting rooms where the data are displayed on large screens for use by high-level officials. This type of approach to the development and use of outcome and performance measures—whether or not full STAT process has been employed—has considerable potential for the improvement of public or private HSOs. Yet, it is not likely to be necessary for a HSO to have the extensive infrastructure that has been characteristic of most STAT programs. In other words, modeling as many aspects of STAT as possible can still likely help an HSO improve its outcomes or performance measurement process, and in turn, its outcomes.

Improving Strategies that Integrate Outcomes across Programs and Agencies

A related concern is how HSOs can more effectively and efficiently track the outcomes of clients who receive multiple services and receive services in more than one agency. The concern about how to handle information flow among public services and public and private nongovernmental organization programs is certainly not new to the human services community.

In addition, we have had in the last decade experiments in "performance partnerships." These partnerships recognize that key outcomes are produced jointly. Each partner has a role in producing those outcomes and in how the desired outcomes are to be achieved. For example, in 2009, the New York State Department of Education introduced its Literacy Zone Program, which focused on partnerships in a number of communities throughout the state. These partnerships were created to improve a number of key outcomes that involved not only multiple state and local government agencies but also many public and private educational, health, and social services organizations in the community. The program used literacy improvement as a starting point for outcomes, and later added outcomes related to employment and health.

An earlier example of such partnerships in HSOs is the Harlem Children's Zone in New York City. Beginning in the 1990s, that program undertook a major effort to provide a wide variety of child development programs for disadvantaged children. "The Harlem Children's Zone is a program designed to address the entire range of community needs with a focus on changing the outcomes for children growing up in poverty" (Page & Stone, 2010). The very special element of this partnership, and the most ambitious, is a linked set of services that continues to support the enrolled children until they become of age. A major thrust has been a focus on results and the continual measurement of progress. This highly publicized program has led to the federal government's "Promise Neighborhoods" program.

Such a comprehensive program is likely to be very difficult to successfully accomplish. It requires sustained resources being available over a long period of time for a wide range of services. This set of substantial resources has not typically been available to HSOs.

Increasing the Use of Outcome Information to Motivate Service Providers

The push for performance contracting began under the Nixon administration in the 1980s, when the federal government began encouraging more private sector organizations to help deliver public services. Over the years, a substantial

number of attempts have been made to introduce performance contracting incentives for public services. For example, financial payments to service providers from public agencies have been made for such outcomes as placing the unemployed in jobs (for example, in the 1990s, the Oklahoma Department of Rehabilitation's innovative employment incentive program paid a portion of the contractors' payment based on placing clients in jobs, both for initial placements and, more unusually, for clients who were still employed several months later). Another example is the North Carolina Department of Health and Human Services' program. It paid adoption agencies on the basis of the number of placements and intact placements as of one year after the decree of adoption (Liner et al., 2001).

The federal government has used outcome program incentives with states for particular services. For example, for state efforts to seek child support payments, financial incentives are provided on the basis of each state's success in payment collections. Payment to states are based on factors such as establishment of paternity, support orders, current child support payments, and amount of payments overdue, with annual actual values compared with targeted values such as those outlined in the 1998 Child Support Performance and Incentive Act. Another example of federal program incentives involves the No Child Left Behind Act program. This incentive program has primarily used negative incentives, including takeover of schools for not meeting progress in test scores. In another example, for child care, states have used intermediate, proxy outcomes that sanction individual child care programs for not including specific service characteristics that evidence indicates are related to more successful child development.

Such outcome-based incentives can be a powerful tool for use in the human services when appropriate outcome measurements are available. They require careful planning and oversight to ensure that they are not easy to defraud. It is likely that the country will continue to increase the uses of such incentives as outcome measurement continues to grow.

Improving the Quality and Comparability of Outcome Measurement Data

The quality of outcome indicators has improved considerably since the early days of GPRA, at all levels of government and to a more limited extent in private, nonprofit organizations. Initial efforts in many organizations focused on using such indicators as response times to calls for services. Many federal agencies have also attempted to introduce more standardization of outcome indicators across states, to be able to both better compare outcomes across

states and to validly aggregate data to provide national figures. However, differences among states make complete standardization very difficult for outcomes for which a federal agency depends on data from states. For example, the Department of Agriculture's Food Stamp program seeks data on the amount of incorrect under- and overpayments and depends on the auditing procedures of individual states.

With regard to improving the quality of the outcome indicators, an example is DHHS's Adoption and Foster Care Analysis and Reporting System. Although it is now yielding improved state comparison data, a debate continues about how to best calculate some outcome indicators, such as the percentage of children placed in a permanent home via adoption or foster care (Wulczyn, 1997). In recent years, considerably more use of outcome information obtained from clients (most often by surveys) has occurred, in an attempt to better understand their conditions and their perceptions regarding the helpfulness of the services they have received. Probably one of the biggest gaps today in human services outcome measurement is the lack of follow-up data on clients after they have completed services. For regular outcome measurement, the follow-up period should probably not exceed 12 months. Assessment of longer term impacts on clients is likely to require considerable added resources and would likely be better done by in-depth program evaluation approaches. Although service providers may not believe that such follow-ups are feasible or possible because of cost, such information can be valuable to the planning, monitoring, continuous quality improvement, and sustainability of HSO services and programs. Client follow-up seems highly desirable for those programs for which the major intended benefits to the client cannot be expected at the time he or she completes the service, such as drug, alcohol, smoking cessation, family support, employment, and child placement programs. Such client follow-ups have been done for employment programs. There is a history of federal government requirements for such follow-ups to identify the number of persons who have remained employed for at least several months after initially placement.

More effective marketing of the usefulness of such follow-up information for improving human services and the identification of more practical data collection procedures is needed. A major way to incorporate such postservice follow-ups is to treat them as part of an aftercare activity, with some funding provided for the activity (Nayyar-Stone & Hatry, 2003). Overall, the pressure from funders for outcome information that enables them to compare outcomes of similar services across service delivery organizations appears likely to continue. Considerable improvement in outcome measurement has occurred, but major gaps remain.

Increasing Outcome Management: Use of Outcome Information to Improve Services

Little evidence is currently available that outcome information has or is being used by HSOs to help improve services and make them more effective for their clients. Nor does the outcome information seem to have been used for formulating and justifying budgets. A primary reason for this may be that most outcome measurement systems have been introduced from the top down and with the primary purpose of accountability. Even for budgeting, the perception appears to exist that most agencies have produced outcome information primarily for reporting as part of the budget process, as required by Congress or state legislatures, the OMB, or by foundations. The use of outcome information to formulate the budget and then justify the budget appears to have occurred to a much lesser extent.

Even the more extensively researched federal PART process used by OMB in the eight years of the George W. Bush administration has been controversial as to how the information has been used. Critics have noted that outcome information, when used at all, has been used primarily to cut or delete programs that the administration did not like. However the outcome information available in the PART process, and how it is used to develop effectiveness ratings on individual federal programs, is quite limited. Making major program decisions based to any significant extent on the PART effectiveness ratings would be highly questionable. In addition, it has been widely documented that Congress has paid little attention to the PART data, only a small proportion of which is actually devoted to outcome information (Accenture Institute for Public Service Value, Georgetown Public Policy Institute, & OMB Watch, 2009).

Although human services are only a small portion of the federal budget, the lack of real use of outcomes and performance measurement in budget formulation and service improvement is a problem common to other levels of government and to nonprofit organizations. This problem remains a significant issue for outcome measurement.

Summary

It should be clear by examining the chapters in this book that the use of outcomes measurement information has increased greatly in human services in the past dozen years. The outcome information available, however, while substantially improved in quality as well as quantity, nevertheless still has substantial limitations. Advances in technology and comprehensive approaches to integrating outcomes and performance measurement into clinical and

administrative practice, such as STAT, are making it possible to see the emerging potential for considerably more enriched data that can link client and service characteristics to outcomes. It will likely take at least the next dozen years for the human services to realize this potential.

References

Accenture Institute for Public Service Value, Georgetown Public Policy Institute, & OMB Watch. (2009). *Building a better government performance system: Recommendations to the Obama administration*. Retrieved from http://www.ombwatch.org/files/performance/buildingabettergovernmentperformancesystem.pdf

Child Support Performance and Incentive Act of 1998, 42 U.S.C. § 1305 (1998).

Liner, B., Hatry, H. P., Vinson, E., Allen, R., Dusenbury, P., Bryant, S., & Snell, R. (2001). *Making results-based state government work*. Washington, DC: Urban Institute Press. Retrieved from http://www.urban.org/uploadedpdf/310069_results-based-stategovt.pdf

Nayyar-Stone, R., & Hatry, H. (2003). *Finding out what happens to former clients*. Washington DC: Urban Institute Press.

Orszag, P. R. (2009, October 7). *Memorandum for the heads of executive departments and agencies: Increased emphasis on program evaluation* (M-10-01). Washington, DC: U.S. Office of Management and Budget.

Page, E. E., & Stone, A. M. (2010). *From Harlem children's zone to promise neighborhoods: Creating the tipping point for success*. Washington, DC: Georgetown Public Policy Institute.

Wulczyn, F. W. (1997). Methodological considerations in outcomes measurement in family and child welfare. In E. J. Mullen & J. L. Magnabosco (Eds.), *Outcomes measurement in the human services: Cross-cutting issues and methods* (pp. 181–183). Washington DC: NASW Press.