Making Homelessness Programs Accountable to Consumers, Funders and the Public

by
Dennis Culhane, Ph.D.
David Eldridge, M.S.W.
Robert Rosenheck, M.D.
Carol Wilkins, M.P.P.

Abstract

This paper discusses how different types of performance measurement can be used to improve the accountability of homeless programs to consumers, funders, and to the public. A distinction is made between the kinds of data used in formal research projects and data that can be practically obtained in a practice setting. Consumer outcomes are discussed in terms of accountability to consumers, program outcomes in terms of accountability to funders, and systems outcomes in terms of accountability to the public. Cost-benefit analyses are also discussed as providing another critical dimension of accountability to funders and the public.

If performance effectiveness is determined by appropriate measures of consumer need, services delivered, and outcomes attained, policy makers and practitioners can gain important insight into what policies have the greatest impact on homelessness and what practices serve homeless people the most effectively. A reliable performance accounting system will require collaboration among policy makers, practitioners, and consumers to collect systematic consumer- and program-specific information.

Lessons for Practitioners, Policy Makers, and Researchers

- Outcome-based program evaluation uses methods that range from simple and inexpensive to complex and resource-intensive. Doing at least some basic outcome measurement provides valuable information about program effectiveness.
- Research measures and practice measures are necessarily different. For instance, there is tension between low demand clinical engagement and the intimidation of comprehensive "intake," so it may not always be possible to get a baseline measurement.
- Standardized data collection at the consumer level is a critical building block. Decisions about desired program outcomes should include consumer input and results of program evaluations should be shared with consumers.
- System-wide standards and provider information are needed to compare the relative effectiveness of program. A number of Management Information Systems (MIS) programs are available that standardize outcome.
- The homeless system must demonstrate effectiveness to compete with other public priorities. Funders are increasingly using outcome measures to evaluate programs and make choices about which programs to fund.
- The homeless system is related to performance of other systems (e.g., health, welfare, and housing) so evaluating their impact involves complex interagency data acquisition issues.

Introduction

Funding agencies at all levels of government have begun to consistently use outcome measures to evaluate the effectiveness of social services. Some local governments have been using outcome-based evaluations of their programs in their reporting and application for federal funds. The link between receipt of funding and performance has been further strengthened with the establishment of the Government Performance and Results Act which obligates federal departments to report on the performance of all funded programs in meeting their specified objectives by the year 2000.

This paper reviews the literature on homeless services outcomes and discusses how different kinds of performance measurement can be used to improve the accountability of programs to consumers, funders, and to the public. In doing so, a distinction is made between the kinds of information gathered by formal research projects and that which can be practically obtained in a practice setting. Regardless, accountability should be based on outcome measurements that provide as clear indicators as practicable that the public and private homeless services are meeting their intended objectives, for consumers, funders and the public. While each constituency may have interests that cross analytic boundaries, this paper is organized according to the primary accountability issues for each audience, and the corresponding level of analysis. Thus, consumer outcomes are discussed primarily in the section on accountability to consumers, program outcomes in the section on accountability to funders, and systems outcomes in the section on accountability to the public.

Introduction: Analysis Framework

Level of Accountability	Primary Units of Analysis
Consumers	Consumer Needs Services Received
Funders	Provider
Public	System

Cost-benefit analyses provide another critical dimension of accountability by weighing the costs of not conducting programs along with the impact of current programs. These kinds of analyses must incorporate a range of data, typically in different units of analysis, with varying degrees of reliability, and from a variety of service systems that use different data management systems. Thus, even less literature in this area is available than in homelessness outcome measurement evaluation generally, providing provisional guidance at best. However, this methodology may help to increase accountability and has been given particular attention in this paper for the potential benefits it offers to policy analysts, program administrators, and state and local officials.

When performance effectiveness is determined by outcome measures, policy makers and practitioners gain important insight into what policies have the greatest impact on homelessness and what practices serve homeless people the most effectively. It must be noted that the needs of consumers and the limited resources of providers can interfere with the types and extent of measurements that can be reasonably and reliably gathered. Thus, an active collaboration between policy makers and practitioners is required to accommodate to the clinical realities that make outcome measurement challenging. Nevertheless,

sound policy and good practice require a concerted effort to collect information on what consumers need, what they receive, and to what effect.

Accountability to Consumers

In order for programs to be accountable to consumers, they must be able to demonstrate responsiveness to consumers' needs—both as consumers' perceive them, and as may be ascertained through reliable and clinically appropriate means by service providers. Because consumers and providers often differ in their views of what consumers need, as well as the relative priority of those needs, multiple methods of assessment should be used. Correspondingly, to assess the adequacy by which consumers' needs have been served multiple methods of performance and outcome measurement should also be employed. These include measures for services provided, consumer progress in meeting service goals, and consumer satisfaction. The usefulness of these measures for achieving accountability to consumers depends in large measure on the degree to which these measures are accessible to and reviewed by program staff and program managers, as well as by consumers. The use of such measures for achieving accountability to funders and the public will be discussed in later sections; this section will review how they can be used to increase the accountability of programs to consumers.

Accountability to Consumers: Measurement Strategies

Assessment Type	Methodology
Consumer Preferences	Checklists, Likert scales (domain specific)
Standardized Assessment Tools	Standardized scales, MIS assessment instruments
Match-Mismatch	Compare consumer needs with services delivered ("appropriateness")
Outcomes	Housing stability Adequate income Social relationships/functioning Consumer satisfaction Quality of life

Consumer-Perceived Needs. Several published studies have reported attempts to gauge what consumers perceive as their needs and the relative priority of those needs. For example, in an early study of consumer-defined needs, Ball and Havassy (1984) found that homeless repeat users of psychiatric facilities in San Francisco prioritized their lack of basic resources for survival, over their lack of access to social services, as the main cause for their homelessness. Results from the 112 subjects surveyed identified the following needs: affordable housing (44%), financial entitlements (38%), alcohol treatment (9%), and counseling (7%). More recently, a needs assessment study conducted as part of the Access to Community Care and Effective Services and Supports (ACCESS) program found that among 1,482 homeless people with mental illness, 91 percent identified a need for long-term housing, including 61 percent who identified this as one of their top three needs (Rosenheck & Lam, 1997b). In each case, long term housing was the most frequently identified need. But it was clearly not the only need. Other needs, in order of importance, were for mental health services (78%), dental services (73%), medical services

(72%), financial assistance (70%), help getting a job (56%) and substance abuse treatment (28%). Other studies of consumer preferences have employed similar measures, and derived similar conclusions (see Herman, Stuening & Barrow, 1994; Moxley & Freddolino, 1991; Acosta & Toro, in press; Linn & Gelberg, 1989).

Compared to assessments of homeless consumers' psychiatric status, employment history, or substance use history, these studies of consumer-perceived needs have employed relatively simple and straightforward surveys of consumer preferences that are quick and easy to complete. Consumer-perceived needs tend to be assessed in fairly broadly conceived domains, with standard Likert-type scales, checklists, or rank ordered lists. In some cases, such as those assessing consumers' preferences for a specific type of housing placement, more detailed questions may be asked. Nevertheless, standardized, psychometrically tested (i.e., tested for reliability and validity) instruments, have not been developed in this area. This has been due to the presumption that consumers can accurately describe what they want, and how much they want it, without major threats to validity. The reliance on consumers to identify their own needs helps alter the focus of practitioners, researchers, and policy makers from measuring dysfunctions to determining survival skills and assessing consumer aspirations (Friedman, 1998).

Of course, consumers' and case managers' assessments of consumer needs are not always in concordance. One study that measured the discrepancy between these assessments is the study mentioned above from the ACCESS program (Rosenheck & Lam, 1997b). In that study assessment data gathered from consumers were matched with similar needs assessments from providers in the same domains (mental health, general health, substance abuse, public support, housing assistance/support, dental care, and employment). The greatest consumer/provider differences in perceived service needs were in dental services (identified by 73% of consumers, but only 44% of providers); medical service (identified by 72% of consumers but only 55% of providers); substance abuse services (identified by only 28% of consumers but 44% of providers) and mental health services (identified by 78% of consumers and 93% of providers). Mental health providers were thus less likely than consumers to identify needs for non-mental health services, but more likely to identify needs for mental health services. Awareness and respect for the potential discrepancies between consumer and provider needs assessments is important: it can help agencies in their self-evaluation process to reflect on their missions, and to affirm a more consumer-centered approach to policy, program design and/or advocacy. Thus, simple consumer surveys conducted periodically can offer program managers with a useful self-study tool for improving their accountability to consumers.

Standardized Assessment Tools

More traditional consumer assessment techniques usually take the form of an intake interview. Researchers, focused as they are on obtaining thorough data, have created many standardized instruments to assess a broad range of consumer characteristics. Since homeless people typically have a multiplicity of problems, consumers' needs must be assessed along multiple dimensions: housing; health status (including psychiatric illness, substance abuse, and medical and dental problems); income support; access to necessities such as food and clothing; social support; employment; involvement in the criminal justice system; and access to health care and/or social or vocational, rehabilitation services. In addition to these specific components there is also value to assessing quality of life according to a global assessment (Lehman, 1988), and general satisfaction with services (Attkisson & Greenfield, 1996; Rosenheck, Wilson & Meterko, 1997). Assessment is further complicated by the fact that each of these domains may have multiple sub-components. For example, housing status can be assessed by the number of days a consumer has been free of homelessness in the past 30 days; by the stability of their residence

(how many times they have moved in the past 6 months); by the quality of the housing (safety, state of repair, privacy, proximity to transportation etc.); or by the number and type of people with whom they reside.

In general, it is best to use standard measures of health outcome that have well-characterized validity and reliability. There are typically numerous measures to chose from. In the ACCESS program psychiatric status is assessed by self-reported symptoms of depression (Robins, Helzer, Croughan & Ratcliff, 1981), psychosis (Dohrenwend, 1976), and interviewer ratings of psychotic behavior on standardized scales. Substance use was assessed with the composite alcohol and drug indices of the Addiction Severity Index (ASI) (McLellan, Luborsky, Wood & O'Brien, 1980). Psychological distress can also be measured using the Brief Symptom Inventory, a 53-item version of the well-known Symptom Checklist-90 (Derogatis & Spencer, 1982).

One of the major dilemmas evaluators must face is the trade off between obtaining comprehensive data and consumer tolerance for participating in assessments. To paraphrase Abraham Lincoln, "You can get all of the data on some of the people, and some of the data on all of the people. But you can't get all of the data on all of the people." Only people with the least problems and the greatest willingness to cooperate will complete assessment batteries, and the data will thus not be representative of the population being served. Moreover, from a service delivery perspective, too many questions may pose a barrier to engagement of homeless mentally ill consumers. Consequently, most agencies use relatively brief intake or assessment forms to collect basic demographic information, income status, reasons for homelessness, educational status, employment status, health status, family issues, etc. These intake interviews typically do not employ standardized instruments, and their psychometric properties have not been established. They are more commonly developed to facilitate the planning of direct services for individual consumers or to verify eligibility, as opposed to answering research questions.

That said, many public agencies, including the federal government, have encouraged the development of more standardized intake or assessment interviews. For example, for nearly ten years, cities such as New York, Philadelphia, Columbus (OH), St. Louis, and Maricopa County (Phoenix), have been collecting basic demographic and psychosocial information on nearly all persons entering emergency shelters. The Department of Veterans Affairs Health Care for Homeless Veterans (HCHV) Programs and Domiciliary Care for Homeless Veterans (DCHV) programs have conducted over 150,000 standardized assessments since beginning operation in 1987, using selected items from standardized instruments (Seibyl, Rosenheck, Medak & Corwel, 1997; Kasprow, Rosenheck & Chapdelaine, 1997). In addition, many individual shelters throughout the country have also developed their own intake and assessment tools.

The federal government has also encouraged broader use and standardization of such instruments through their support for the development of the Participant Outcomes Monitoring System (POMS) (Fosburg, Locke, Peck & Finkel, 1997), and the Runaway and Homeless Youth Management Information System. Many jurisdictions are also beginning to implement local Management Information System (MIS), or to install other consumer-tracking software applications.[1] Because of their automation and consequent standardization across providers in a jurisdiction, these systems and assessment instruments are very useful for establishing accountability of programs to funders and to the public (to be discussed later).

---

[1] One of the authors (D. Culhane) was involved in the development of the ANCHoR system with PRWT Services, Inc. (Philadelphia, PA). Other Homeless Services MIS systems and their vendors include, SOPHIA by Caracole, Inc. (Cincinnati, OH), Community Link by Community Services Network (Orlando, FL), Client Track by Data Systems International (Layton, UT), Homeless Prevention Network by School of ECE, Purdue University (West Lafayette, IN), Locator 2000 by Gulf Coast Software (Groves, TX), Homeless Services Network by Paradigm Systems, Inc. (Charlotte, NC), FACTORS & HelpWorks by Peter Martin Associates, Inc. (Chicago, IL), C-STAR by St. Vincent de Paul Village, CSC (San Diego, CA), Homeless Information System by Colorado Department of Human Services (Denver, CO).

Whether or not discrete services are measured, programs can still be held accountable to consumers simply by showing whether or not consumers progress or achieve other intended outcomes. Again, researchers have been far more careful to collect exhaustive information when trying to assess program outcomes than is typically possible in a practice setting. In the research setting, assessment tools like those described previously are typically used as periodic follow-up measures. This would be the equivalent of repeatedly administering an intake/assessment interview to a consumer. Other instruments are explicitly designed to collect "outcome" data (i.e., housing stability). But, again, while researchers typically receive special funds to hire staff to track consumers, to interview them, and to compensate them for their participation, these are resources that a typical practice setting cannot afford. Still, it is important to be able to demonstrate to consumers that their participation has real effects in improving the homeless service system.

To some extent less rigorous outcome measurement has into some of the consumer tracking systems mentioned earlier. In the VA's HCHV and DCHV programs, mentioned above, a brief discharge form has been completed after each one of over 50,000 episodes of residential treatment. This form documents where each veteran will be residing after discharge, their employment status, whether they completed the program to a mutually agreed upon discharge or premature departure, and staff assessments of clinical improvement. While these measures lack the rigor and reliability of research-level measures, they provide useful and face-valid evidence of program results (Rosenheck, Leda & Gallup, 1992). Some of these consumer-tracking systems not only have assessment forms but also allow for collecting periodic measures of status in several domains. These relatively consumer-friendly techniques can be applied to a broader population and tend to be much more brief than an assessment interview. They can be used to assess whether programs are meeting their objectives in serving either consumer- or program-identified needs. In some cases, jurisdictions may elect to collect outcome measures on a sample of discharged consumers, given the difficulty of tracking consumers, especially after they have left a given program. Indeed, the major challenge in collecting outcomes information is finding former consumers. Doing so for all discharged consumers would require an unrealistic (and unjustifiable) expenditure of effort and resources in most cases. Alternatively, some agencies may choose to follow a sample (i.e., ten-percent) of their consumers for six months, measuring their progress once or twice over this time period.

The subject of follow-up raises another issue: how long after leaving a program should consumers be followed? Unfortunately, there are no established standards for this important consideration. Outcome data are most easily and comprehensively gathered during participation in or at the time of discharge from a program. However, while some program "graduates" may be tracked through aftercare programs, these typically represent the most successful exits from a given program. Again, research and demonstration projects may have the resources to conduct long-term follow-up (perhaps two years or more) with a broader cohort of service users, while service agencies will have to choose a more practical strategy. Nevertheless, care should be used to avoid selection biases, whereby only the successful consumers are followed (sometimes referred to as "creaming").

Accountability to Funders

While it is theoretically possible for consumers to hold programs accountable by reviewing their own services histories and progress reports, as well as aggregate reports on other consumers (indeed, such a practice might be encouraged among consumer advocacy groups), providers are more commonly held accountable by their funders. Funders, especially government funders, will usually enforce

accountability through annual or periodic reporting requirements, and through applications for new or renewed funding. Unfortunately, because of the range of funders that may be involved, providers have to struggle with widely variable and inconsistent reporting requirements. In general, funders tend to require data elements similar to those used to establish accountability to consumers described above. However, since funders often must regulate or oversee a large number of programs, and programs of different types, they may have more generic benchmarking practices. It is worth noting that to the extent that some large social service agencies have several programs that they directly manage, they might conduct similar analyses as described here, and could therefore be considered "funders." Some of the more common reporting formats and issues are discussed below.

Accountability to Funders: Performance Measurement

Components	Measures
Consumers served	Contract Performance (units of service per consumer)
Units of Service delivered	Effectiveness (maximize positive outcomes per consumer)
Outcomes	Efficiency (maximize ratio of positive outcomes :: services delivered per consumer)

Again, the research literature, where resources are available for tracking consumers and analyzing large quantities of data, offers only a few examples of detailed provider or program evaluation research. The vast majority of program evaluations measure either the relative effectiveness of two or three different program interventions, or the impact of a single program model, implemented by a number of providers. The first type of program evaluation, which typically uses an experimental or quasi-experimental design, compares models of housing or services in terms of how well they serve consumers, using aggregated consumer-level data over time. For example, Miesher and Galanter (1996) experimentally compared two programs that served homeless alcoholic men according to consumer retention in each program. (Interestingly, the authors' conclusions point to the importance of service integration, or systems-level concerns, rather than the program-level issues, in producing better outcomes.)

A second type of program evaluation in the literature derives system-level conclusions by aggregating program-level outcomes from a large project. For example, Matulef et al.'s (1995) National Evaluation of the Supportive Housing Demonstration Program evaluates the impact of 93 percent of the over 700 programs that participated in this demonstration project. While this strategy generates information on the need for systems-level interventions across sites, the relative effectiveness of each program is not evaluated. Alternatively, Huebner and Crosse (1991) used an innovative approach that combines experimental and quasi-experimental comparisons with site-level and systems-level analyses for a nine-site demonstration project evaluation. They conclude that inconsistent definitions of homelessness, problems with missing data and difficulty in measuring treatment effects were major obstacles to developing cross-project comparisons. This lack of standardization could similarly frustrate analysts working within, let alone across, jurisdictions. Finally, from a research perspective, the authors point out that common instrumentation is not a panacea, but that data-collection procedures and schedules, comparison groups, and selection into these groups are also important factors to consider in assessing program effectiveness. Another strategy that may help performance measurement is the institution of quality standards for providers that are developed in conjunction with consumers (Friedman, 1998).

In the Department of Veterans Affairs HCHV and DCHV programs reporting formats are uniform nationally. A total of 33 "critical monitors" (benchmark measures) are used to evaluate the comparative performance of over 100 programs across the country in the areas of program structure (6 measures); contacting the intended target population (7 measures); delivering intended services (8 measures) and outcomes (11 measures) (Kasprow et al., 1997; Seibyl et al., 1997). Less able or interested in funding research projects, more typically funders of homeless services will require basic data for reporting purposes, such as the number of people served, consumer characteristics, and, less often, the needs of those consumers. They also usually require some description of the program or services provided, and may even ask for the total number of units of service delivered by service category. Funders have less traditionally asked for outcomes, but, as stated previously, that is becoming more common as federal, state, and local governments are more conscious of performance, or employ performance-based auditing or contracting procedures.

Performance measures can be as crude as units of service per person (efficiency) to more complicated attempts to link units of services to some desired outcome (efficacy). For example, a funder of homeless programs may want to know the number of consumers served and the number of shelter days used, and – in combination – the average length of stay in a program. Given that the residential component of homeless programs is usually the most significant in terms of cost, the "average length of stay" (ALOS) is perhaps the most common proxy measure of what a consumer receives, or what the funder is "purchasing." For some funders, this might constitute performance in that it may document what amount of service was provided for the average consumer, or in that it may be used to compare the efficiency of providers in serving a given pool of consumers. However, even this level of reporting is only recently becoming more common in jurisdictions. To calculate an average length of stay per episode, a provider must know, at a minimum, the total number of consumer episodes (which may include duplicates of individuals) and the total number of days of service provided. To calculate average length of stay per consumer, agencies must furthermore be able to unduplicate consumers across episodes. To do so across a jurisdiction, or some other grouping of agencies, would undoubtedly require some automated information system.

Given the rather basic state of reporting at present, research is needed to develop more accurate and discriminating performance criteria, as well as organizational or service delivery factors that can be associated with those criteria. The average length of stay is one indicator that has been discussed; still others might include units of case management delivered, hours of counseling, frequency of outreach contacts, rate of readmission to shelter, etc. For example, a recent "provider performance" analysis (Culhane, Eldridge & Metraux, 1999) chose the rate of readmission to shelter as the critical benchmark for measuring the effectiveness of transitional housing providers. Of particular interest to the funders of the research was whether or not rate of return was associated with length of stay. In other words, do shelters with longer lengths of stay have lower rates of return? In later models, length of stay was treated as a control variable, and the effect of various social services (frequency, duration, etc.) on readmission rates was assessed (i.e., Does providing case management or a certain amount of it reduce the rate of return to homelessness?). Still further models included consumer characteristics ("case mix") and discharge type (percent receiving housing subsidies) as variables. The operationalization of these variables involved a number of challenging conceptual decisions and statistical procedures. Nevertheless, the study provides a method for systematically comparing programs while taking account of variations in consumer mix, service mix, etc., and doing so based primarily on administrative records, rather than depending exclusively on costly, time-consuming survey methods.

Accountability to Funders: Benchmarking
Example (Culhane, Eldridge, & Metraux, 1999)

Variables		Required Standardized Data
Dependent Variable:	Rate of Return	1. Consumer needs/characteristics
Independent Variables:	Specific services delivered	2. Services delivered 3. Discharge type 4. Provider characteristics
Control Variables:	Length of Stay Exit type Case-mix adjustments Provider characteristics	5. System-wide admission data

From a funder's perspective, if more days, or certain services, or more of certain services, are not associated with reductions in rate of return, one could conclude that those are either poorly performing programs or are not critically important services. Similarly, one could measure placements to housing, tenure in housing, increases in income, etc., as performance measures against which various inputs are measured. In choosing benchmarks, one must be careful that the outcome measures reflect an intended policy or programmatic objective. For example, some might argue that rate of return, while a readily available measure in some cities, is really a poor measure of how a person or family is functioning, particularly given that some persons may prefer to live on the streets instead of return to a shelter. For these persons, the lack of "return" to shelter would not indicate "success."

Also, as indicated by the preliminary study of provider performance mentioned above (Culhane et al., 1999), if funders want to be accurate in judging providers, additional sets of data are needed, beyond length of stay and rate of return. First, consumer characteristic information must be collected so that adjustments can be made for variations in case mix. Some providers may serve a more chronically disabled population, and would therefore be expected to have different performance parameters. The VA programs described previously use over ten baseline measures to risk adjust outcomes assessments that are used to compare programs seeing different types of patients across the country. Additionally, a number of provider level characteristics, which cannot be obtained through consumer-tracking systems, may be important as qualifying or control variables. These may include organizational size, auspices, funding sources, staffing levels, ownership, revenues by source, expenditures by category, etc. For example, Rosenheck and Lam (1997a) used ACCESS program data to find that geographic location of homeless services provider sites was a greater obstacle to service use than consumer characteristics among homeless persons with serious mental illness.

Thus, in addition to implementing consumer tracking information systems, funders concerned with accountability should consider the creation of provider inventories, similar to that which exists for providers of mental health services (the Inventory of Mental Health Organizations, National Institute of Mental Health, 1983 and 1986). Such periodic surveys would enable comparisons of providers' performance controlling for differences in provider characteristics. A survey developed by the U.S. Department of Commerce, Bureau of the Census (1995) on behalf of the Interagency Council on the Homeless (the National Survey of Homeless Assistance Providers and Clients) may serve as a good beginning point for such a survey.

The major problem for providers will be collecting the information necessary to measure performance, however it is measured (Huebner & Crosse, 1991). Clearly, standardization of information collection,

automation of information collection, and, in particular, explicit reporting requirements by governmen funders, will drive the collection of performance data. Though challenging, streamlining data anc achieving unduplicated counts of consumers is doable (Friedman, 1998). A great deal of work remains however, before the multitude of service and funding systems are fully coordinated. For example one administrator wondered how it was possible to report to multiple, and as many as fifteen or more, funders (Harris, 1998).

Besides these data collection difficulties, an equally challenging quandary for both providers and funders is that the homeless services system, is primarily reactive, and cannot always be appropriately accountable for patterns of homeless service utilization. People may stay longer in shelters because the supply of housing certificates or affordable housing shrinks, or more people may re-enter shelters because of welfare reform. Thus, the performance of homeless providers is affected by significant externalities that are beyond the control of the homeless providers, or their funders. This contingency makes any assessment of homeless providers' "performance" necessarily tentative and provisional.

Accountability to the Public

Because public funders essentially represent the public, requiring accountability for funding expenditures serves the public's interest. However, funders cannot just require reporting from providers, but must make available to the public information regarding how their systems, or networks of providers, are performing, what they cost, and what objectives they do or do not meet. In so doing, the public may be more able to hold policymakers accountable for sound public policymaking and efficient administration of programs.

Accountability to the Public: Research Strategies

Strategy	Methodology
Homeless System Performance	Services delivered and costs Effectiveness Efficiency
Program Innovations	Comparison of program types Costs/Effectiveness/Efficiency Pre-Post comparisons
Policy Changes	Pre-Post comparisons Multiple Systems Analysis (Data integration across systems, secondary impacts on health, welfare, housing and criminal justice systems.)

The Congress, state legislatures, and their administrative departments are institutions, which through their funding structures, can require local agents to report standardized information on services. Unfortunately, there are fewer requirements that obligate federal agencies to make easily available standardized, comprehensive measures of relative performance by jurisdiction. For example, while HUD and HHS and other federal departments currently require providers to report the number of consumers served, consumer characteristic information, and services received (i.e., through the Annual Performance Report (APR), through the Projects for Assistance in Transition from Homelessness (PATH) and Health Care for the Homeless (HCH) reporting requirements, and the Runaway and Homeless Youth (RHY) reporting system), they less consistently require outcome information or jurisdiction level information (which would require unduplication across the jurisdiction). The HUD APR does collect outcomes such as housing stability and changes in income, and the Center for Mental Health Services (CMHS) does

require outcomes data on persons served by PATH. These efforts will hopefully be further advanced by the Government Performance and Results Act. However, from a public accountability stand point (a system versus provider level of analysis), most Federal programs do not require jurisdiction-wide aggregation of those measures on an unduplicated consumer basis. Thus, it is difficult for the public to judge whether policies are having an intended effect (i.e., Are fewer people homeless? Are people homeless for shorter periods of time? Do peoples' incomes and quality of life improve as a result of program participation?). Policy can consequently be more often driven by anecdote, the exigencies of funding constraints, and the ideological perspective of policy makers, rather than a reasoned discussion and review of the evidence. One could argue that the inability to have institutionalized jurisdiction-wide performance measurement results in public policy being unaccountable to consumers as well as the general public.

This lack of performance information is particularly disconcerting given that many important and national initiatives have been undertaken by state, local, and federal governments. Despite their value, without information to substantiate efficiency and efficacy, good programs often remain demonstration projects and are not realized on an appropriate scale. Similarly, failed policies can remain in place because they suit the prerogatives of established interests, though they may have little evidentiary basis for continuing.

Alternatively, some service integration strategies and some large demonstration projects have been documented and there is much to be learned from them. Dennis, Cocozza and Steadman (1998) present findings from 10 different systems integration projects which used different levels of evaluation procedures, ranging from none at all to sophisticated quasi-experimental, outcome-oriented designs. Some, such as the HUD Shelter Plus Care Program, have demonstrated significant improvements in the lives of program participants in both service usage (e.g. engagement in needed treatment and reduced hospital and jail use), and material well-being (e.g. increased income, employment, and housing stability) (Fosburg et al., 1997). Similarly, the NIMH/CMHS McKinney Demonstrations showed a dramatic improvement in residential stability resulting from increased collaboration between the participating housing authorities and mental health centers. Only one project, the ACCESS Program, has built in outcome measures at the consumer, program, and systems levels. Some of the improvements demonstrated by these programs are very important, and because they have supporting evidence, arguments can be advanced for their continuation, and they are less vulnerable to legislative or departmental program cuts.

Cost-Benefit Analysis

Very little research has been published about the cost-effectiveness of services to homeless people. Studies that have been published fall into two major groups: research on small groups of homeless consumers (usually fewer than 100), and program evaluations which are usually completed by (or on behalf of) service providers and/or government agencies that fund programs serving homeless people. Research studies often use rigorous data collection strategies and powerful statistical tools, and results are published in academic journals several years after services are provided. Program evaluations tend to use routinely collected administrative data, have substantially larger sample sizes, and their results are often reported quickly in order to support program planning and decision-making about ongoing funding.

As with any consumer level, provider level and systems level analysis, cost-benefit analysis uses information that most homeless service providers have limited capacity to collect: consumer demographics, service utilization, and outcomes for the same group of consumers over time or across service settings or programs. Very few programs even use unique consumer identifiers, so they may not

have the capacity to determine how many intakes or bed nights are unique to individuals, or how many of the people served at one site are also served later in the year at another program site. The kind of data that providers do collect relates to the following program needs: (1) establishing consumers' eligibility for services, (2) documenting service utilization for billing purposes, and (3) to a very limited extent, documenting outcomes that are of interest to the source of funding for the program. While there is clearly a need for more comprehensive data collection by providers, cost-benefit analysis can utilize a variety of measures based on the kind of data that providers may already collect.

Ideally, cost-benefit analyses would compare the total costs and benefits of providing homelessness interventions with the total costs and benefits of providing no interventions, or some other standard set of services. This analysis requires measurement of costs in a wide range of systems, and we are not aware of any study that synthesizes data from every relevant system to measure the costs of not addressing homelessness at all. Instead, it is likely that we will need to patch together information from a variety of sources and methodological approaches to draw conclusions about what service models "work" to reduce homelessness and to identify both the costs and the savings benefits to consumers and to the public. Cost-benefit analysis includes the following methods: cost effectiveness, before and after comparisons, service utilization comparisons, waiting list comparisons, comparisons between groups, comparisons between programs, multivariate analysis, and multiple systems analysis. All of these approaches can provide valuable information about the relative costs of homelessness interventions.

Cost-Benefit Analysis

Strategy and Analysis Level	Description
Cost Effectiveness (Consumer Level)	Measures the costs for providing a unit of service to an eligible consumer.
Before and After Comparison (Consumer Level)	Compares service utilization costs by homeless consumers before entering program with costs after leaving program.
Service Utilization Comparison (Consumer Level)	Compares service utilization costs by homeless consumers with non-homeless consumers.
Waiting List Comparison (Consumer Level)	Compares service utilization costs and benefits by currently served consumers with eligible consumers on waiting list.
Programs Comparison (Program Level)	Compares service utilization costs and benefits between two or more of the same kind of programs or between programs that offer different kinds of intervention.
Multiple Systems Analysis (System Level)	Measures costs and benefits of services to consumers relative to not providing those services using data from the homeless-ness, housing, health care, mental health and substance abuse treatment, criminal, and welfare and employment systems.

Cost Effectiveness. Some studies simply measure the costs for providing a unit of services to an eligible person. If the costs appear to be "reasonable" and the projects appear to achieve desired outcomes for most consumers, the report concludes that services are "cost effective" (Matulef et al., 1995). While not

a complete picture of cost-effectiveness, this measure may be meaningful if a desired outcome is the engagement of the project's target population. This is particularly true in the case of projects serving consumers who have not consistently participated in treatment. It can also be a useful indicator of the efficiency and effectiveness of program implementation in terms of evaluating such program activities as filling vacancies, or providing outreach to maintain program capacity.

Before and After Comparison. Costs of services utilized by homeless individuals before they enter a program are compared to costs and benefits during and after program participation. For example, the evaluation of the Minnesota Supportive Housing Demonstration Program (Tilsen, 1998) found that average monthly costs for a range of services, including hospitals, jails, state institutions, and income support, were reduced from an average of $2,168 to $1,370 per month, resulting in annualized savings of $1.7 million a year for 180 units of supportive housing. Changes in residential stability, consumer functioning, and consumer satisfaction are also reported. Because some homeless people might have achieved some improvements in functioning or reduced utilization of other services without program assistance, it has been suggested that changes measured using this approach may be useful as an estimate of the upper boundary of program impacts (Rosenheck et al., 1995).

Service Utilization Comparison. In a recent study which examined hospital discharge data from New York City's public hospitals, Salit et al. (1998) found that homeless patients stayed 36 percent longer than other patients, after adjustments were made for clinical and demographic characteristics. Rosenheck and Seibyl (1998) reached similar conclusions after comparing a national study of the health service use and costs for homeless and domiciled veterans hospitalized in psychiatric and substance abuse units at VA medical centers. The authors found that 13.3 percent more money was spent serving homeless veterans than domiciled veterans. Measurements of the costs associated with homelessness are one way of identifying the savings that could be achieved by successful interventions. Ideally this information can be linked to data from programs serving homeless people to compare the costs of programs that are successfully targeting and retaining people with the same clinical and demographic characteristics. Programs to prevent homelessness can use a similar approach by creating models to project what would have happened to program participants if services had not been provided. The New York State Department of Social Services (1990) used this method to estimate the cost effectiveness of programs that intervened to prevent evictions, and estimated that the late stage eviction component of their Homeless Prevention Program saves approximately $11.6 million in averted homeless costs.

Waiting List Comparison. Another strategy that has been discussed but not fully pursued is to compare costs and benefits for program participants with those for homeless individuals who are eligible for the program but on wait lists because of limited program capacity. Where participants are selected by lottery, (as is often the case in housing for homeless people), there is *de facto* random selection into a treatment group and a control group. Comparison between the service utilization costs and program benefits between these groups may yield valuable cost-benefit data. While it would be very difficult to track homeless individuals who do not receive services, information from public data systems such as public hospitals and clinics, mental health services, and other systems of care could be used to gather longitudinal data on wait list subjects.

Program Comparison. A number of research studies compare two or more programs which use interventions that are based on different service models. For example, Wolff et al. (1997) compared costs and outcomes for three types of case management after randomly assigning consumers to each service model. The authors found that higher average costs for two assertive community treatment approaches were associated with increases in financial assistance (vocational/educational, residential, and income support) and decreases in costs for inpatient psychiatric services, when compared to less

intensive and less expensive brokered case management services. Dickey, Latimer, Powers, Gonzalez and Goldfinger (1997) compared two types of housing with services and found that treatment and case management costs did not vary by housing type, but that one housing program was significantly more costly than independent living arrangements, without producing significantly different outcomes. Rosenheck et al. (1995) used multivariate analysis to assess relationships between specific treatment elements and outcome measures in VA homeless programs. Using this approach, the researchers measured the costs for determining a standardized amount of improvement for different outcomes (housing stability, symptoms, employment), using different service strategies (e.g., case management, residential treatment, and income supports). They found that the cost of residential treatment was three to five times the cost of case management for achieving a standardized amount of improvement in outcomes, but that the interventions were associated with improvements in different areas of functioning. Residential treatment was associated in improvements in the number of days housed and days employed, while case management was associated with improvements in days of substance abuse and psychological distress. Income transfer payments were associated with improvements in days homeless but worse outcomes in paid employment.

Multiple Systems Analysis. Because homeless people often face a number of additional problems, such as mental illness and drug and alcohol addiction, it is important to compare costs from many different service systems. These data include utilization of shelter, housing, and other "homeless" services; utilization of health care, mental health and substance abuse treatment; utilization of welfare and employment services; and involvement with the police and jail systems. For families, we need to also consider costs of child protective services and foster care. We know anecdotally that effectively stabilizing the lives of homeless people can prevent or reduce the costs of services in each of these systems. A number of the program comparisons cited above (Wolff et al., 1997; Dickey et al., 1997; Tilsen, 1998) combine data from several of these systems, but to date there have been no published studies that have been successful in accessing and matching data from all of these systems.

Challenges. The limited availability of cost-benefit information is a result of a variety of challenges frequently encountered during cost-benefit analyses. These challenges include: high attrition rates, small sample sizes, imprecise measurement of service utilization costs, inconsistent data over time and across systems, and lack of provider commitment to and resources for data collection. A description of the extent of these challenges is followed by a description of strategies that providers can use to meet them and conduct effective cost-benefit studies.

Attrition rates in most studies are very high, reflecting the instability in the lives of many homeless people. In some studies plans to track participants over an extended time period are abandoned when findings are compromised by high rates of attrition. After 12 months, Rosenheck et al. (1995) were able to conduct follow-up interviews with only 37 percent of homeless veterans who had agreed to participate in the study at entry into the program. Wolff et al. (1997) found that between 33 percent and 63 percent of participants assigned to case management dropped out, and that higher rates of attrition in brokered case management compared to assertive community treatment may have resulted in unknown bias that compromised the study's conclusions. Attrition is particularly problematic if participation in data collection is dependent upon participation in services, as is the case in most data collection by programs that serve homeless people. Because our current research does a poor job of tracking outcomes for homeless people who discontinue treatment, findings can be generalized only to those who are most successfully engaged in services. Better outcomes for consumers are usually associated with more consistent participation in services, but it is hard to tell if this is cause or effect. It is possible that both participation and outcomes such as residential stability, employment, or improved quality of life, are associated with other variables that are not identified upon intake. For example, people who experience

substance abuse relapse problems are likely to discontinue participation in services and have worse outcomes that are not included in study results because they drop out of data collection.

Another challenge for cost-benefit analysis is sample size. Sample sizes that are too small will not yield statistically valid results. Consequently, it is often very difficult to identify statistically significant differences between program outcomes, even when there appear to be changes in the expected direction. Compounding this problem, observed changes are often small and progress for many homeless people is unsteady and rarely linear. In addition, service utilization after intervention is often strongly correlated with prior service utilization (Wolff et al., 1997). Large sample sizes would resolve these two problems, because the larger the sample size the easier it is to detect small or non-linear changes and the easier it is to control such variables as prior service utilization.

A third challenge to conducting cost-benefit analysis is the difficulties with using billing records data. Public data systems that have been created for billing purposes usually report charges which may be much higher than actual costs, while payment rates from Medicaid or other programs may be lower than costs. Services which are not reimbursed (e.g. jail medical care) are often not recorded in a way that is easy to assign costs. One partial solution is to use data systems to gather information about the types and quantity of services used by consumers, and then use other budget information to get average costs for each type of service. This is particularly appropriate when variability in the costs of other services, (e.g., regional variations in the cost of a hospital day), is not relevant to measuring the effectiveness of different programs (Dickey et. al., 1997).

The inconsistent reliability of billing records is related to the wider problem of "messy data." People who are homeless are often not identified consistently when they use services, and many will have more than one "unique identifier" in public data systems. If it is not possible to gather data from all of these systems, significant cost shifting may not be identified in the analysis. For example, a reduction in hospital use may be a benefit, or may reflect increased rate of incarceration for study participants. Another source of data problems is the changes in the data systems of public health departments, hospitals and mental health systems data systems that has accompanied the recent transition to managed health care. This often results in significant year-to-year differences in data quality, format, and availability, making it extremely difficult to get consistent information about service utilization over time.

Homeless service provider capacity and willingness to participate in data collection may also limit the availability of data. Direct service providers have challenging jobs, are often required to use creative strategies to engage consumers who may be resistant to the offer of services, and often deliver services in unconventional settings. This leaves little time to fill out paperwork, particularly when providers do not see analysis of the data they collect data collection as helping their front-line efforts. This problem can be mitigated to some degree if the collected data is analyzed and given back to program staff quickly and in a format that is useful to them and their consumers. Administrators can provide incentives for the collection of data by proving a link between cost-benefit analysis and improvement in consumer outcomes and/or in the acquisition of additional funding.

Because of the above challenges, powerful statistical techniques and a very significant investment of resources are needed to achieve ideal levels of scientific validity and reliability. However, this sort of research usually takes years to complete, requires resources outside of the reach of most homeless programs, and the results may not be published until three to five years after the program intervention has been tested. Homeless advocates, then, must be prepared to utilize a number of less thorough cost-benefit analyses to develop an understanding of the savings represented by various approaches to addressing homelessness. Fortunately, there are a variety of strategies advocates can use that will

increase the rigor of their cost-benefit analyses. Forging partnerships with providers in other service systems, for example, can address the problem of attrition by facilitating longitudinal tracking of thei consumers. Utilizing the MIS programs discussed above can help standardize routine data collection and increase sample size for studies that include multiple providers and multiple service systems. These MIS systems can also facilitate speedy feedback from program administrators and funders that both provider and consumers could use to measure and improve consumer outcomes.

The importance of system-wide cost-benefit analysis is evident when considering programs whos consumers have multiple problems and utilize multiple programs. For example, those people who vis public hospital emergency rooms 12 to 50 (or more) times a year and the visible homeless people in or streets and parks tend to have similar characteristics—chronically homeless, out-of-treatment addict out-of-treatment mentally ill, and often living with chronic health problems and at very high risk f HIV/AIDS. As we move toward establishing new systems of accountability which measure outcom and cost-effectiveness, we need to be sure that we do not create fiscal or other incentives for programs "cream" and to exclude those most at risk of failure. For example, programs that require 60 days sobriety before intake may achieve better outcomes related to employment and housing stability, but m be completely inaccessible to most of the homeless people who are seen in public hospital emergen rooms or city doorways.

As described earlier, to get a complete picture of cost effectiveness, we need to look across systems a funding streams (e.g., federal, state, and local, as well as "homeless," health care, criminal justice, a welfare). However, program and funding decisions are almost never made from this global perspecti For example, it is hard to convince a county mental health department to allocate funds for interventi that will create savings in other systems by reducing entry into shelters, hospitals, or jails and that n result in increased demand for the limited resources available within the mental health syst Homelessness does not occur in isolation and so must not be addressed as an isolated phenomenon. need for cost-benefit analysis that integrates data from a variety of systems, then, reflects the need policies that bring providers from a wide number of systems together to solve homelessness.

Finally, cost-benefit analyses can provide powerful tools for homeless advocates in the policy ar Any intervention designed to reduce or end homelessness requires funding, and quantitative argum that a given intervention can save money relative to not investing in that intervention can go a long to obtaining that funding. The matter of funding is perhaps the clearest aspect of accountability to public whose taxes fund the vast majority of homeless programs and who has a vested interes eliminating homelessness.

Next Steps

The assessment of consumer needs, reporting of services, and the measurement of consumer outc are where policy and practice meet. Policymakers need information to drive decisions such a allocation of resources and the design of programs, and practitioners need data to understand who are serving, how they serve them, and to what effect. Thus, information collection represen intersection of interests and around which there could be concerted effort to improve upon cu practice. Consumers, providers, funders, and the public, all stand to gain by a system of m accountability that can be enabled by reasonable and accurate data collection.

At the most basic level, providers and consumers can be the keystone for such an effort throug initiation or cooperation with the development and/or implementation of standardized inforr

collection. Organizations of providers or "service systems" (i.e., local government) can foster such standardization by establishing agreed upon protocols or through adoption of existing mechanisms for collecting information. Service systems can use the goal of accountability as one way to encourage participation in such arrangements. Local government can also demonstrate their conviction and value for such efforts by supporting the functions of assessment, tracking and outcome assessment, financially and materially through equipment and technical support.

Similarly, the federal government can support accountability for consumers, funders, and the public by funding and even mandating the establishment of minimum data collection requirements. At a minimum, homeless programs can begin to be accountable to the public if jurisdictions were required to report basic information such as: unduplicated counts of consumers served, units of service by consumer type, and direct (i.e., housing stability, reduced hospitalization, improved quality of life, even if just on a sample) and indirect (i.e., rate of readmission) measures of service utilization. The extensive data collection that has been part of the Department of Veteran Affairs national homeless program has clearly demonstrated the possibility of generating higher levels of accountability on a national scale in integrated service systems (Kasprow et al., 1997; Seibyl et al., 1997). If inducements to collect standardized information were in place, local, state and federal government as well as consumer and provider advocacy groups and other interested parties could exploit information technologies and establish monitoring systems. These systems could be used to gauge demand for emergency shelter, and measure duration of shelter stays, rate of shelter exit and return, and assessment of consumer satisfaction.

With these tools at hand, an evidentiary discussion of the merits of existing or proposed policies could ensue. Perhaps even more importantly, the system of homeless services from a policy perspective can become more proactive and apply such information technologies as an accountability check and as a critical performance measurement for the larger social welfare and insurance systems. For example, if welfare reform produces increases in shelter admission or longer shelter stays, homeless providers and consumers would be in a position to prove it. If increases in the development of support housing units yield increases in the rate of shelter exits, local governments can show it. If managed care of Medicaid programs is leading to premature discharge of hospital patients, some of whom end up in the shelters, regulators can prove it. In this way, the homeless system can become an accountability check on the larger arena of public policy related to poverty.

Recommended Research Initiatives

- Develop new instruments that measure consumer outcomes, provider performance, and systems effectiveness and efficiency. These instruments should be brief, reliable and valid so they can be used by practitioners without interfering with their ability to meet their consumers' needs and at the same time can provide useful evaluative data. Such instruments should capture a variety of data including: performance variables that measure inputs (services provided), outputs (placements to housing, tenure in housing, increases in income), ALOS (average length of stay), and organizational characteristics (size, auspice, funding sources, staffing levels, ownership, revenues by source, expenditure by category, etc.).
- Forge a closer relationship between providers and researchers so that provider-identified trends could become the source of formal research projects.
- Develop techniques and guidelines for tracking the outcomes of consumers longitudinally. Protocols should be established for sample sizes and periodicity of follow-up.
- Develop benchmarking measures for performance. These include ratios of inputs to outputs, which may include cost-benefit analyses. A crucial component to these benchmarks is the development of a

more common understanding of what services and units of services mean in different settings so that they can be more comparably measured. This common understanding should include input from consumers, providers, funders, and researchers.

- Conduct system-wide analysis that provides the "bigger" picture policy effectiveness to answer the following questions: Do expenditures yield reductions in costs in other systems? Do broad policy initiatives yield overall gains in housing stability, costs, etc.?
- Devise creative strategies (e.g., data integration strategies, interagency task forces, or case reviews) to use data from different systems that take different forms.
- Adopt automated systems (e.g., software programs, or management information systems) more widely in order to track consumers at the site of service delivery.