

Outcome Reporting In Legal Services:

Caution Signs On The Road Ahead

Prepared for:

**The Legal Services Corporation Conference On
Outcomes And Program Performance
June 21, 2003
Cambridge, MA**

Melville D. Miller, Jr.
President & General Counsel
Legal Services of New Jersey

INTRODUCTION

This paper offers reflections on reporting and assessment of “outcomes” in a Legal Services context, from the perspective of a state – New Jersey – which is moving toward experimentation with such assessment as part of a comprehensive revision of Legal Services reporting. In particular, I will examine whether the national Legal Services Corporation should undertake a national initiative in outcomes assessment.

The general thesis of this paper is that experimentation with outcomes assessment is sensible only in the context of a comprehensive, strategic overall approach to performance assessment and reporting in Legal Services. Outcomes assessment should not be initiated as a new, discrete project, unrelated to overall performing reporting reform. A companion thesis is that outcome assessment, while used in some areas of social and human services, is far from science, and has numerous limitations which strongly suggest the need for caution in implementation and restraint in collecting and using the data, especially at the national level.

At the most basic level, it is natural and unsurprising to want to look at “outcomes” – what happens to clients – as a result of activities by a Legal Services provider. The hope would be find ways to improve those services by learning more about what results from them.

Several key questions, however, immediately present themselves, in the areas of (1) definition (what do we mean by the term “outcomes”, and what are the parameters of measurement), (2) purpose (which among a number of different and somewhat conflicting choices are of highest priority), and (3) how data or information will be handled (where will it be reported, who will have access to it, and how will it be used)? Before examining these issues, as well as the special challenges in trying to assess outcomes, it is best to begin with some history and context.

**Some History and Context:
A Brief Overview of Legal Services'
Experimentation With Methods
Of Program Reporting and Performance Assessment**

After inheriting the OEO management report system in 1974, the new Legal Services Corporation moved to design its own reporting system, which evolved over a few years into the Case Statistical Reports (CSR) system remarkably close to what is still in place today. Despite periodic inquiries into possible revisions, and modest changes in the late 1990's, the CSR still collects information on cases opened and closed during a period, with differentiation as to reasons for closure ("closing codes") and case type, along with accompanying demographics. Also in the late 1990's, as part of new timekeeping requirements, programs were directed to track certain types of activities by case, as well as time on certain types of "matters"; only the latter is transmitted to the LSC. These are all quantitative measures pertaining to individual legal matters; the system produces no normative assessments, no views beyond the confines of the individual case, and no definitive context for comparison across programs free from concerns about an almost infinite number of potential determinate variables.

In the same vein, there is no required recording, let alone reporting, of wins and losses or other information on case results, though an undetermined number of programs do track this information. There is no requirement that surveys of clients be conducted to determine their perspectives on representation, though many – perhaps most – programs do so in some form. A variety of other descriptive information – almost none of it case specific – is collected either as part of the grant application process or in ad hoc reports.

Three other LSC performance assessment initiatives over the years are of relevance. The first is the Delivery Systems Study (DSS), carried out pursuant to congressional mandate from 1976 to 1980. The DSS sought to compare the efficacy of various delivery models by contrasting performance assessments in four areas: quality, impact, client satisfaction and cost. Quality and impact were assessed by means of tightly structured peer review, client satisfaction through tabulations of survey results, and cost through analysis of fiscal reports. While the structured peer review measures marked some of Legal Services' most creative thinking on performance assessment, neither the methodology nor the learning was continued in any systematic way by the LSC after the conclusion of the DSS process in 1980. Of special relevance is the DSS advisory panel's effort to separate the assessment process, where possible, into four areas, looking at "inputs", "processes", "outputs", and "results" (which could be taken as another term for "outcomes"). The DSS conclusion was that, in assessing and comparing performance, it was important to look at all four areas, not just results.

A second initiative was the completion of the LSC Performance Criteria in the early 1990's. Developed as part of another LSC initiative, the Performance Criteria marked the LSC's most comprehensive and detailed effort to develop a framework for assessing program performance. In particular, the Criteria adopted a dynamic model that commenced with a community need assessment, and then moved in turn to setting goals and specific objectives in response to the needs, planning to achieve the objectives, implementation, evaluation of efforts and results (assessed in light of the initial goals and objectives and all intervening factors), and then a return to a new cycle, starting once again with community need assessment. At the heart of this process was the recognition that while a national framework was possible, community

needs, available resources and determinative contextual factors (such as applicable law and judicial culture) all were very much local in nature. Put another way, it was not possible to prescribe desirable national objectives or results, but it was possible to stipulate at a national level that this kind of strategic process should take place within each provider program. These Performance Criteria have been carried forward, essentially unchanged, as the frame of reference for the LSC's competitive bidding process.

Finally, a word on peer review in the Legal Services context. The OEO Office of Legal Services utilized external peer review teams to conduct grantee evaluations throughout its history, primarily under the auspices of its regional offices. This methodology was continued and enhanced by the LSC from its inception in 1975 through 1980. Starting in 1981, this practice of universal peer assessment effectively was terminated; since that time LSC visits to programs have been limited to audits and investigations, ad hoc limited inquiries (e.g., a study of support), and more recently, "capacity assessments" in conjunction with the competitive bidding process or special program visits related to LSC's review of a state's efforts in statewide planning.

Moving from the national LSC to state-based funders, with one exception and one footnote there is neither commonality nor significant discourse across states in relation to performance assessment and reporting. If performance assessment has been carried out at all, it is likely to be a "home-grown" version. IOLTA funders constitute the partial exception: evaluation, reporting and even assessment of outcomes frequently have been topics at twice-yearly national IOLTA seminars. While reporting and evaluation systems are solely within the purview of each state's IOLTA program, there has been ample opportunity for cross-pollination of viewpoints. One other funding source, federal Title III Older Americans Act funds distributed through local area offices on aging, typically has some home-grown aspects engrafted upon a minimal national framework which essentially mirrored the LSC's CSR system.

The IOLTA role is especially important because of the utilization, in at least a few states, of "outcome" reporting. Starting in 1993, New York, followed in later years by Virginia, Maryland, and then Texas, developed and then refined "outcome" reporting approaches. Supporters of these approaches typically point to "telling our story" – documenting and publicizing the effects of Legal Services' work – as the principal benefit from such measurement systems. In their national IOLTA presentations, however, proponents have cautioned strongly that outcomes data should not be used as a basis for comparisons across different providers, or as a basis for funding decisions.

Some view the national political history of the Legal Services Corporation, especially with its predominant emphasis on "access" to legal assistance (as distinct, say, from "law reform" or "impact"), as a nearly three decade-long push toward high-volume, brief service legal work, at the expense of representation in extended or complex matters, or cases presenting the chance for broad legal or social change. Such critics typically also implicate the CSR reporting system as reinforcing these high-volume, low-impact predilections. This critique, whether subscribed to or not, provides a useful set of questions in considering performance assessment and reporting.

Potential Purposes of Outcome Assessment

Consideration of possible purposes for looking at outcomes can help clarify some of the central choices presented. In New Jersey we have looked particularly at five main uses:

1. To compare various models or methods of delivering services.

In essence, this was the DSS, except the DSS looked at results (outcomes) as one of four subjects (also including inputs, processes, and outputs) across each of the four measurement areas (quality, impact, client satisfaction and cost). While this purpose could potentially reemerge, it has not been suggested as a motivation for the current LSC initiative, and would appear unwieldy and unnecessary to pursue.

2. To compare the relative performance of specific providers.

This purpose actually has two general subtexts, one more passive (as in making comparative performance information generally available to provide boards and staff and to other stakeholders, hoping to stimulate self-directed improvement), and one more aggressive (using such comparisons as the basis for funding decisions).

3. To hold providers “accountable”.

Akin to the aggressiveness of using comparative data to support funding decisions, this purpose suggests using the data to compare performance against a specific standard, essentially a compliance strategy, perhaps using corrective action methodology.

4. To furnish providers and stakeholders with information, not in a comparative context, concerning how particular delivery approaches or strategies hold special promise or effectiveness in maximizing benefits.

This orientation is more consistent with the limitations of the methodology. Because of the enormous number of potential explanatory variables for a particular set of results, outcome assessment is not science. Making information available, without suggesting that it is a valid basis for cross-program or cross-model comparisons, in effect acknowledges the difficulties – and perhaps indefensibility – of trying to use outcome data as a definitive or determinative measure. This approach is also consistent with the way outcome information has been used by IOLTA funders. Important variants on the same theme would be to encourage, or perhaps require, providers or states to come up with outcome assessment schemes of their own.

5. To present a compelling public case for the scope and importance of Legal Services work.

Proponents of outcomes measures offer this purpose most often, in the belief and hope that it is a persuasive way to build a case for additional resources.

Problems Of Definition, Frame Of Reference, Prioritization, Causation And Complexity

As another paper in the seminar materials describes, outcome measurement has been pushed and utilized to some degree in other human services contexts, most notably in efforts spearheaded in the 1980's and 1990's by United Way of America. Local experiences with actual implementation and utility of these efforts varied widely, notwithstanding an extensive national effort to standardize the approach. Contrasting the United Way effort with the Legal Services history concerning "outcomes", the difficulties with definition are evident. For example, to some degree IOLTA reports concerning "outcomes" appear to include items more properly considered as descriptions of the services provided – "outputs" in the lexicon of evaluation theorists.

Legal Services programs also have had some motivation to frame outputs as outcomes when reporting to a funder, because outputs are more predictable and certain, and therefore safer. When outcomes are reported to funders, the product can be distorted. Take, for example, legal representation of a tenant facing eviction for non-payment. Some reporting schemes have described "providing representation to prevent eviction" as an outcome, when it is much more like a description of an output. Moreover, a tenant might be ordered to vacate at some future point, but still might receive a right to stay six additional months, a partial rent abatement, a citing of the landlord for code violations, actual or punitive damages under a state consumer statute, an injunction to make repairs, or some combination of these results. Each of these can be considered an outcome. Which outcome do we use? If we use more than one, which are more important? Description of the "outcome" in a report can become a complex matter indeed. As pure information to the provider program, all of these results may be significant and useful to know. Their number and complexity, however, precludes meaningful comparisons across programs, states or delivery models, so they are of questionable utility at the national level.

A second issue, part definitional and part theoretical, is how far into the future results are to be assessed in order to determine "outcomes". An eviction prevented is certainly an "outcome" on the day the decision is rendered. Some outcome measurement schemes, however, seek to know whether the tenant is still in the dwelling or in other permanent housing months or even years later, a much more attenuated assessment. There are downsides to a more extended time frame (difficulty of finding clients, introduction of many more potential explanatory variables, higher costs of tracking and surveying) but some potential positives. It is not appropriate to make a national determination on these choices; providers may properly weight them differently.

Out of these definitional issues emerges the realization that outcomes are very much influenced by the frame of reference: which person or groups one sees as the intended beneficiaries. In the Legal Services context, it is possible to consider at least the following, with quite different consequences:

1. The effect on the individual (which also raises measurement issues – this effect can be determined by reference to the client's expressed wishes, or to the views of lawyers or other third party observers).
2. The effect on others "similarly situated" – others facing the same type of problem (which itself raises additional definitional questions).

3. The effect on low income people generally (as, for example, could be the case with community economic development).

And potentially even:

4. The effect on others who are not poor (as from law changes which promote greater fairness).
5. The effect on the legal and judicial system (for example, by making it more fair, or responsive to unrepresented or disadvantaged litigants).

Each of these possibilities can have importance and validity, depending on one's perspective. Some may have more force or importance in particular states or local areas. State or local funders may place different emphasis on the various possibilities. No one approach is "best" or "right". In deciding how the LSC should approach the topic of outcome measurement, it is essential important to preserve the capacity for of state and local funders, stakeholders and providers to determine their own priorities.

An additional set of challenges arises when attempting to assign normative values to an outcome. Some are suggested by the earlier discussion of the tenant facing eviction: which of the various potential outcomes should be valued most highly? There is no single correct, let alone national, answer to such a question. Similarly, a client facing eviction or some other court process may, after they receive legal advice, conclude that for any number of reasons they simply do not want to pursue or litigate the matter. The client may instead choose to leave a dangerous living situation, be unable to take time off from work, or simply be averse to judicial process. Can a client's personal, informed decision not to pursue a matter, even in the face of compelling claims or defenses, fairly be termed a "poor" outcome? How are we to define "good" and "bad"?

While the short answer surely is that such sweeping valuations are inappropriate on a national scale, there still may be benefit in assembling descriptive information which can assist a provider in assessing the effectiveness of its services. If only one-third of tenants with meritorious claims¹ choose to pursue a matter, and only half of those are successful, the information may provide important clues as to how to allocate resources and deploy staff.

A fourth set of difficulties stems from a principal assumption underlying outcome measurement – that there is an inferred causal relationship between the Legal Services provider's efforts and the outcomes observed. Of all the difficulties, this maybe the most severe, especially if the measurement takes place some time after the case or services ended. For any given outcome, there may be a host of potential independent, explanatory variables – "independent" connoting factors beyond the control or reach of the provider's conduct. There are severe practical limitations on what lawyers – even allied with others in the community – can do to change the daily lives of clients. If outcome data is retained at the provider level, many of these possible factors can be assayed and taken into account as the information is interpreted. At the national level, however, short of a controlled, random, necessarily one-time, multi-million dollar study, these possible local explanations will be completely lost in any national data aggregation.

¹ Note that even a term like "meritorious" may have definitional problems, and be the source of sharp disagreement among knowledgeable professionals.

Overall, the challenges set forth above, individually and in combination, strongly suggest that any LSC effort to use collected data at a national level – perhaps apart from the “tell the story” goal – would be seriously flawed. At the same time, there are important potential positives in encouraging providers and those at the state level to look at outcomes in a broader performance assessment and reporting context, and I turn now to some of the thinking and planning in New Jersey.

One State's Planned Approach

Leaders and planners at the state level in New Jersey have been troubled for years by the limitations inherent in the simple CSR-style, opened and closed case reporting approach. While this data is certainly viewed as relevant and meaningful information, it is not seen as sufficient to document the importance and scope of Legal Services' work, nor the effectiveness of individual providers. Several years of consideration, tabled intermittently to deal with other pressing concerns, has finally led to a more comprehensive approach.

Supplementing internal and external evaluation of program work, New Jersey will pilot a revised reporting system in the fourth quarter of 2003. If successful, it will look to full implementation for state and IOLTA funds in either the first or third quarters of 2004, depending upon the extent of any necessary refinements. While the final test system is not completed, current plans include the following key components:

1. **Inputs.** Tracking the input of staff hours, by position type (attorney, paralegal, administrator, administrative staff, social worker and other) and functional category.
2. **Processes.** Capturing increased information on certain processes, some now reported as "closing codes", especially relating to casehandling (e.g., one-time advice, advice and brief follow up services, continuing advice, complementary dispute resolution, and several more).
3. **Outputs.** Further revising (adding more detail to) the received LSC closing codes (aggregatable up to the LSC categories to avoid dual reporting systems).
4. **Outcomes.** While there is much skepticism about the utility and feasibility of outcome assessment, a new system will be instituted and examined, testing what may be possible within the five possible frames of reference discussed earlier. A starting point will be the client's own expressed goals, perhaps amplified by a third party view of individual client results, (probably using a framework centered on what was "reasonably attainable"), an attempt to assess impact on those "similarly situated", a further assessment of the effect on other low income people, and an indication of any perceived broader impacts on the public or the justice system. Categories of outcomes will be differentiated by type of case. To the extent there is any inquiry about a client's situation at some point subsequent to the end of legal representation (and there almost certainly will be), that will be carried out by means of a special or periodic study, probably at the state level, not by regular performance reporting. Such special studies are seen as critical to comparisons and judgments concerning alternative delivery methodologies – *pro se* assistance, interactive web pages and kiosks, a central hotline and other brief advice, and forms of extended representation – now in use in New Jersey.

Among other things, the pilot period will seek to assess the relative costs and benefits of such an approach, the true costs of record-keeping, and viewpoints of the data-enterers as to reliability. Such reporting will have multiple applications. Some will be primarily for the use and benefit of the provider, to assist self-evaluation and improvement. Some will be aggregated at the state level, to assist external evaluation and accountability, as well as in presenting Legal Services' performance record, to build greater public and financial support.

Recommendations For the LSC

Given the preceding observations, and based on New Jersey's experience, I would offer the following recommendations for potential LSC directions:

1. General performance reporting. LSC still needs to examine the potential and use for additional refinements in performance reporting, especially in closing codes, to make them more accurate and useful. In addition, there should be some open examination of matter reporting, beyond the report to Congress, to determine whether the data is helpful and worth the cost, and if so, whether these should be revisions. Results of matter reports to date should be made available to states and grantees.
2. Outcomes.
 - A. Starting in 2004, LSC should encourage states and programs to experiment with outcome assessment. Modest one-time grants for necessary programming if resources permit. LSC could recommend a broad organizational framework, approximating that suggested earlier in this paper, to help structure and analyze at a minimum, the information from these state or program experiments. The LSC's accompanying guidance could recommend that the states or programs the aggregation should seek to identify: (1) the outcome categories chosen, with their definitions; (2) critiques of the methodology, from the designers and users; and (3) the data produced by the state or program – based systems. This information should be made available to all designated state planning bodies and grantees. Analysis and conclusions could be worked through with an expert panel in 2005, after at least one full year of data and experience. Two full years would be even more preferable.
 - B. LSC should not: (1) try to define and impose a predetermined outcomes in a deductive fashion, even for experimental purposes; (2) require reporters to look at outcomes at some designated amount of time after representation ceases (because the effort may not be worth the cost)
 - C. Long ranges in considering any further actions in regard to program reporting and outcomes, LSC should be very skeptical of moving toward singular national frameworks, and mindful of the sensitivities and risks inherent in national aggregation of data which must be made publicly available.
3. Other performance assessment. Although resumption of national peer review visits under LSC auspices may be cost prohibitive, the LSC should consider ways of ensuring, or at least encouraging, that such reviews at least go on at a state level on some periodic basis. In addition, the Performance Criteria should be given a more explicit and renewed role in the consciousness of grantees and DSPBs.