

1. Evaluation of New Statewide Health Reform Initiatives

SHERRIE H. KAPLAN and
SHELDON GREENFIELD

New England Medical Center. Boston

THE CREATION OF A SOCIAL PROGRAM LIKE THE Maryland Medicaid Diabetes Care Program (DCP) represents a major service delivery and policy effort that deserves our attention. In the context of health care reform, evaluating such a program represents a serious challenge. Program evaluations for large-scale service delivery programs, like the one described by Mary Stuart, have been plagued historically by methodological, substantive, and, sometimes, political problems. The program originators who create, develop, and oversee the implementation of their work often do not have the resources, time, or research skills to evaluate the program carefully. Moreover, a series of issues often compromises or complicates even the most optimally staffed and funded evaluation efforts.

Confronting these issues, or barriers, to effective program evaluation, before or during program implementation, improves the prospects for generating interpretable results at a moderate cost. Carefully crafted, prospective evaluations can be considerably easier to conduct and to maintain through the life of a program than retrospective evaluations

designed after the program is implemented. The payoff in advanced attention to the need for a well-designed, adequately funded evaluation of any health care reform program initiated by the states emerges in the documentation of successes and failures, so that future decisions to expand, reduce, or maintain the program can be made on a rational basis. It is unlikely that first efforts at health care reforms will be complete, uniform successes. In the evolution of such programs, prospective, formative evaluations afford the possibilities of identifying the effective elements; of targeting the ideal subgroups of patients, physicians, and health care delivery systems; and of coordinating efforts with existing health care delivery or financing programs in an optimal manner.

Various issues and barriers surface in connection with effective, prospective evaluation of statewide, health-reform-related programs for chronic disease care like the Maryland Medicaid DCP. Each of these issues, discussed below, can be addressed when evaluators participate in the initial design of the program.

1. *Difficulties in Evaluating Chronic Disease Care.* Chronic disease care is more difficult to evaluate than acute care. The process, the outcomes, and even the identification of cases are extremely difficult to document with existing databases. Comprehensive evaluation of chronic disease care therefore generally requires some level of primary data collection (Tarlov et al. 1989).

2. *Lack of Provider Coordination.* The very nature of chronic disease care means that it involves multiple providers, who may practice in varied settings, may be compensated differently, and may have different on-site services available. Even when patients are managed by a primary care provider, lack of coordination contributes to problems in identifying program targets (i.e., which patients should be selected; whether generalist or specialist physicians should be used; what role should be played by administrators) and in measuring program impact. Because few programs will be evaluated using randomized, controlled trials, evaluating program impact on specific program participants (e.g., patients of multiple providers) is difficult.

3. *Political or Practical Problems.* These also represent formidable challenges to the design of effective evaluations of large-scale, comprehensive, chronic disease programs. First, conducting large-scale, randomized, controlled trials is daunting; few, if any, states undertake experiments in health care delivery among their constituents. However, the absence of a control group seriously compromises causal assertions

about the impact of health care programs on outcomes. Even well-documented, historical prospective or time-series studies suffer from the potential for historical events, rather than the program under scrutiny, to explain observed changes in, for example, patients' health status, use of health care services, disability days, and so forth. With respect to diabetes care, the release of the results of the Diabetes Control and Complications Trial during the operational period of the Maryland Medicaid DCP might have contributed to changes in patients' use of, or referral to, specific health care services (Diabetes Control and Complications Trial Research Group 1993). The addition of a nonequivalent control group (for example, eligible participants for a later phase of the program) can contribute to the understanding of the implications of such historical events in assessing program impact.

Second, because large-scale programs are typically designed to be inclusive, they reach a broad array of patient subgroups. Even within a specific disease, patients may vary considerably with respect to sociodemographic status, extent of total disease burden, or other characteristics that may affect their health outcome prognosis. The magnitude of program impact may be over- or underestimated because a disproportionate number of severely ill patients are included in the program or comparison group. Because claims data are often a convenient source of information for large-scale program evaluations, careful case-mix adjustment represents an important methodological dilemma. Such claims databases often lack the information needed to construct adequate measures of case mix. Even when some case-mix measures can be gleaned from claims data, they are rarely adequate for use in the evaluation of programs involving ambulatory versus inpatient care. Including information on patients' sociodemographic characteristics, the severity of their target disease or condition, and the total disease burden (the sum of the severity of each comorbid disease or condition), as well as the health-habit risk profile of patients, offers a greater capacity to distinguish patient effects from program effects in identifying program impact (Greenfield et al. 1994b).

Third, for large-scale programs, the program evaluation is designed to detect differences in a small number of specific measurable outcomes, which are likely to change in specified ways in response to successful program implementation. As a result, the design is often overshadowed by the "buckshot imperative," or the perceived need to measure multiple outcomes, lest the total impact of the program be unrecognized or un-

derestimated. This tendency is costly, and it does not encourage advance formulation of specified evaluation questions or hypotheses regarding realistic program benefits that are measurable over the life of the program. Timing of the observation points for evaluating program impact is also a key methodological issue that must be addressed before the program is implemented. Considerable resources can be expended by evaluating program outcomes before it is reasonable to expect to see any measurable program impact. Rapid and progressive changes would also be difficult to document if an ongoing evaluation initiated at the time of program inception did not document early outcomes. Identifying ideal observation points, then, becomes critically important to the assessment of program impact (Greenfield and Nelson 1992).

Finally, and related to the "buckshot imperative," is the "lamppost phenomenon": the tendency to choose outcomes for assessing program impact by what is easily measurable. As the health services research community generates more well-tested outcome and process-of-care measures, it becomes easier and more tempting to use these measures without consideration of their appropriateness or relevance for evaluating specific program goals. Unfortunately, too often the specific links between program content and the likelihood that the program could affect certain outcomes (i.e., mortality, global functional status, or satisfaction with physicians' interpersonal care) is not carefully considered when these "off-the-shelf" measures are chosen to assess program impact. This does not necessarily represent an argument for developing new measures to assess each innovation in health care delivery or financing, but it does constitute a bid for evaluating outcomes, using existing or newly developed measures, as appropriate, that have an unambiguous, interpretable relation to program goals.

Some programs, like those focused on enhancing blood glucose control, may have mixed outcomes, improving, in this example, glycemic control, but possibly also compromising quality of life. Being on the alert for such potentially conflicting results will improve our prospects for an evaluation that permits midcourse corrections.

Programs like the one initiated in Maryland herald innovations that represent health care reforms occurring in advance of any national initiatives. Grappling with the considerable, but not insurmountable, methodological and substantive issues involved in designing a careful and cost-effective evaluation of these programs early on is extremely important. Imaginative use of primary data collection in conjunction with ad-

ministrative and laboratory databases for evaluating chronic disease care, as is currently being done by the Group Health Cooperative of Puget Sound and the Tufts Associated Health Plans in Boston, represents a promising avenue for future comprehensive evaluations of large-scale programs (Greenfield et al. 1994a). If evaluation becomes sufficiently streamlined, provides feedback to the health care delivery system for continuous improvement, and contributes to reform efforts to maintain quality while containing costs, then it is sensible to defray the costs of evaluation as a cost of doing business. Incentives to use such evaluation systems as the basis for designing subsequent innovations should be built into national health care reform.

References

- Diabetes Control and Complications Trial Research Group. 1993. The Effect of Intensive Treatment of Diabetes in the Development and Progression of Long-Term Complications in Insulin-Dependent Diabetes Mellitus. *New England Journal of Medicine* 329:977-1036.
- Greenfield, S., and E.C. Nelson. 1992. Recent Development and Future Uses of Health Status Measurements in Clinical Practice. *Medical Care* 30 (5, suppl.):MS23-MS41.
- Greenfield, S., S.H. Kaplan, R. Silliman, et al. 1994a. The Uses of Outcomes Research for Medical Effectiveness, Quality of Care, and Reimbursement in Type II Diabetes. *Diabetes Care* 17:32-9.
- Greenfield, S., L. Sullivan, R.A. Silliman, K. Dukes, and S.H. Kaplan. 1994b. Principles and Practice of Case Mix Adjustment: Applications to End-Stage Renal Disease. *American Journal of Kidney Diseases* 24:298-307.
- Tarlov, A.R., J.E. Ware, Jr., S. Greenfield, E.C. Nelson, E. Perrin, and M. Zubkoff. 1989. The Medical Outcomes Study: An Application of Methods for Monitoring the Results of Medical Care. *Journal of the American Medical Association* 262:925-30.

Address correspondence to: Sherrie H. Kaplan, New England Medical Center, The Health Institute, 750 Washington Street, Box 345, Boston, MA 02111.