Deciphering Medicaid Data: Issues and Needs

by Leighton Ku, Marilyn Rymer Ellwood, and John Klemm

The wide range of data bases that can be used for Medicaid analyses and research are reviewed in this article. The Health Care Financing Administration, State Medicaid agencies, and other groups have developed useful data bases and made them available to the public. Efforts could be made to obtain better quality national data, including annual reports on State participation, expenditures and program characteristics, and person-based data bases about Medicaid clients and services. State-level analyses and research could be enhanced and disseminated more widely. More complex data collection and analysis efforts are an inevitable tradeoff for the flexibility of the Federal-State structure of Medicaid.

Introduction

Folklore speaks of the difficulties four blind men had trying to describe an elephant. The blind man who felt a leg declared, "It's a tree trunk." "No," said the one touching the elephant's ear, "This is a fan." The man grasping the tail disagreed, "I think it's a rope." "You are all wrong," announced the fourth man, running his hand along the body of the elephant, "This is something without beginning or end." (Adapted from Taylor, 1951).

If there are such perplexing problems in describing a mere elephant, what can we say about this odd beast, Medicaid? Medicaid is a peculiar creature, with 1 Federal and 57 State heads and a 3-part body of acute medical, preventive, and long-term care. In its relatively short life, it has undergone continuous growth and cutbacks. With a mission to meet the health care needs of the poor, the disabled, and the elderly, Medicaid is among the most complex and least understood social programs in the Nation.

At both Federal and State levels, problems as diverse as the uninsured, acquired immunodeficiency syndrome (AIDS), infant mortality, long-term care, hospital and physician reimbursement, and deficit reduction have led to repeated policy proposals and debates over the future of Medicaid. Caught between burgeoning needs and limited public resources, policymakers have demanded careful analyses of the scope of problems, alternative solutions, and their costs and benefits. Heightened policy interest has engendered the need for better, more comprehensive data.

This diversity and constant change have made Medicaid a troubling program for policy analysts and health services researchers. There are few national data bases, and analyses of one State's data may not be applicable elsewhere. In contrast to Medicare, which has large, relatively consistent national data bases (Health Care Financing Administration, 1989), the data bases for Medicaid are more limited and less well known. This may have caused some analysts to despair and decide that Medicaid data are hopeless. The theme of this article is: Medicaid data are more available and better than most realize, but there are important gaps.

Medicaid data bases

Types and sources

Table 1 contains seven general types of data bases available for Medicaid research: State program characteristics surveys, aggregate State-reported data, Medicaid claims data bases, the Quality Control Sample, broad national sample surveys, microsimulation data bases, and miscellaneous data bases. Five of the data bases are discussed below; and the other two, the broad national surveys and microsimulation data bases, will be discussed later in this article. Readers may pursue references to obtain further explanation.

State program characteristics surveys

Given the numerous State options for Medicaid, as well as for Aid to Families with Dependent Children (AFDC) and Supplemental Security Income (SSI), one of the first issues encountered is identifying State policies—e.g., which States have medically needy programs or home and community-based service waivers. The Health Care Financing Administration (HCFA) periodically provides some description of State policies in its Medicaid and Medicaid Data Book series, but its last detailed report was Analysis of State Medicaid Program Characteristics, 1986 (Ruther et al., 1987; Health Care Financing Administration, 1987). The National Governors Association (1990), the Intergovernmental Health Policy Project (1990), and other private groups provide annual or special focus summaries of State health policies. Unfortunately, State policies have not been described consistently over time, and there have been frequent errors or gaps. In addition, some policies, e.g., reimbursement policies, may be so complex that they are almost impossible to describe succinctly or comparatively. Some analysts have used reviews of State policies to qualitatively critique the overall condition of Medicaid (Erdman and Wolfe, 1987; Joe, Meltzer, and Yu, 1985).

Aggregate State-reported data

The most commonly used Medicaid data sources are three reports submitted by States to HCFA. One is the annual HCFA Form-2082, Statistical Report on Medical Care: Eligibles, Recipients, Payments, and Services, the
| Type of data source                                      | Examples                                                                 | Strengths and weaknesses                                                                                                                                 |
|--------------------------------------------------------|--------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|
| State program characteristics surveys—Descriptions   | Health Care Financing Administration: Medicaid and Medicaid Data Book, 1986; Intergovernmental Health Policy Project: Major Changes in Medicaid and Indigent Care Programs, periodic; Congressional Research Service: Medicaid Source Book: Background Data and Analysis, 1988. | Vital to understand program variation. However, these have not been conducted consistently over time, so one cannot reconstruct full program histories for some time periods. Occasional errors and frequent gaps. Complex program rules or payment systems sometimes defy simple analytic descriptions. |
| of State program options, spending, number of         |                                                                          |                                                                                                                                                    |
| recipients, etc. Generally based on State plans or    |                                                                          |                                                                                                                                                    |
| surveys of State Medicaid officials.                  |                                                                          |                                                                                                                                                    |
| Aggregated State-reported data—National and State     | HCFA Form-2082; HCFA Form-64; HCFA Form-25 (see text for definitions and description.) | Generally the best time series data for State trends. Certain categories may be defined inconsistently across States and times, so that comparisons must be made judiciously. Counts of Medicaid clients are usually based on the number of persons receiving services, so that enrolled non-users may be excluded. |
| periodic reports to the Health Care Financing         |                                                                          |                                                                                                                                                    |
| Administration on program spending, participation,    |                                                                          |                                                                                                                                                    |
| utilization, etc.                                     |                                                                          |                                                                                                                                                    |
| Medicaid claims data bases—Voluminous data on all     | Tape-to-Tape, Medicaid Statistical Information System, other MMIS data     | Vast detail and sample sizes on enrollees, claims, services used, payments, etc. for particular States. Can be linked to other data sources using Social Security numbers, names, etc. Can be used to look at small subpopulations, e.g., patients with acquired immunodeficiency syndrome. Limited to particular States. Can be difficult to use because of the massive volume of the data. Limited demographic data on clients. No data on people not on Medicaid or on Medicaid clients while they are off Medicaid. Inter-State comparisons are not possible unless definitions are made uniform. |
| Medicaid claims in States, based on automated         |                                                                          |                                                                                                                                                    |
| Medicaid Management Information Systems (MMIS) tapes.  |                                                                          |                                                                                                                                                    |
| Include data on basis of eligibility, types of        |                                                                          |                                                                                                                                                    |
| services, diagnoses and procedures, vendor payments,  |                                                                          |                                                                                                                                                    |
| etc. Can be edited to become person based, rather than |                                                                          |                                                                                                                                                    |
| claims based.                                         |                                                                          |                                                                                                                                                    |
| Quality control sample—An ongoing national sample     | Medicaid, Food Stamps, and AFDC.                                         | Substantial verified eligibility data and total expenditures for a month. A nationally representative sample with rich sociodemographic and eligibility information. Sampled by cases, not persons. Excludes some Supplemental Security Income and other cases. |
| survey conducted by States to verify the eligibility  |                                                                          |                                                                                                                                                    |
| of cases participating in Medicaid, Aid to Families    |                                                                          |                                                                                                                                                    |
| with Dependent Children (AFDC) and Food Stamps        |                                                                          |                                                                                                                                                    |
| programs and error rates.                             |                                                                          |                                                                                                                                                    |
| Broad national sample surveys—Various national       | National Center for Health Statistics: National Health and Nutrition     | Usually includes demographic data, such as race, income, and health status on Medicaid and non-Medicaid populations. Sample size of Medicaid clients or of certain groups (e.g., disabled children) may be small. Data on health care utilization or expenditures may be limited by respondents' knowledge and recall. Comparisons with administrative data often suggest that Medicaid participation is underreported. |
| surveys which include information on Medicaid         | Examination Survey, National Maternal and Infant Health Survey, National  |                                                                                                                                                    |
| participation or utilization. Can be used to compare   | Survey of Family Growth, Longitudinal Survey on Aging; Agency for Health  |                                                                                                                                                    |
| Medicaid and non-Medicaid populations.                | Care Policy and Research (formerly NCHSR): National Medical Expenditures  |                                                                                                                                                    |
|                                                            | Survey; U.S. Bureau of the Census: Current Population Survey, Survey of   |                                                                                                                                                    |
|                                                            | Income and Program Participation; Duke University: National Long-Term      |                                                                                                                                                    |
|                                                            | Care Survey.                                                             |                                                                                                                                                    |
| Microsimulation data bases—Specially developed models | Urban Institute: TRIM2; Brookings/ICF: Long-Term Care Model; Lewin/ICF:    | Versatile, powerful tools for assessing the effects of policy changes, especially tax and income transfer policy. These are usually proprietary models, owned by particular firms. May be expensive to develop and operate. Limited by the survey data and economic and behavioral assumptions built into the models. Specification and prediction of medical needs and utilization may be a problem in these economic and demographic models. |
| that seek to simulate the effect of changes in         | Health Benefits Simulation Model.                                        |                                                                                                                                                    |
| government policies and programs, based on demographic |                                                                          |                                                                                                                                                    |
| data from surveys, information on program rules, and  |                                                                          |                                                                                                                                                    |
| coverage, etc.                                        |                                                                          |                                                                                                                                                    |
| Miscellaneous data bases—Various health care data     | Health Resources and Service Administration, Area Resource File; State or  | Depends on data source.                                                                                                                                    |
| bases that include information on Medicaid.           | local hospital or nursing home discharge data; clinical abstracts; all-payer discharge data bases. |                                                                                                                                                    |

SOURCE: Ku, L., The Urban Institute, Washington, D.C., 1990.
These problems can often be resolved using Medicaid claims data bases that contain each claim paid by a State. Medicaid claims data bases standard reports or may need measures of variability. The development of cost estimates of program changes, race and sex) in the back sections of the form. Reporting artifacts may mask true changes in the program. Historically, recipients have been defined as users of service. Counts of eligibles (or enrollees) have become available only recently and in certain States. Thus, recipient counts (i.e., counts of people actually using medical services as opposed to all those enrolled in Medicaid) are the only universal and consistent measures of Medicaid populations over time.

The other two sources are the major financial reports submitted by States to HCFA. HCFA Form-64, the Quarterly Medicaid Statement of Expenditures for the Medical Assistance Program, is the primary Medicaid financial report containing actual medical assistance funds expended quarterly by States and is the basis for apportionment of Federal Medicaid matching funds or Federal financial participation. This report contains breakouts of expenditures by type of service (more types than the HCFA Form-2082, with some types defined differently), but it does not include tabulations by eligibility group or enrollee counts. Moreover, the report contains several “bottom lines,” depending on whether certain adjustments, such as overpayments or disallowances, are included.

The other, HCFA Form-25, the Medicaid Program Budget, is submitted quarterly in February, May, August, and November and contains projections of Medicaid spending for the current and following two fiscal years. Expenditures are forecast by the same types of service reported on HCFA Form-64. A separate section contains projected recipient counts, using the eligibility groups of HCFA Form-2082. Finally, States are asked to indicate the sources of increases in Medicaid spending, whether changes in prices, covered populations, or utilization of services, including planned program changes. HCFA Form-25 projections are used by HCFA to develop the national Medicaid portion of the President’s budget.

Together, HCFA Form-64 and HCFA Form-25 constitute the best data sources for actual State spending and future funding needs.

Medicaid claims data bases

In many cases, these aggregate HCFA data are not adequate. Analysts may need statistics unavailable from standard reports or may need measures of variability. These problems can often be resolved using Medicaid claims data bases that contain each claim paid by a State, usually including eligibility group for the patient, type of provider, charges and payments, diagnoses, and procedures. The claims data are based on computerized Medicaid Management Information Systems (MMISs). Because each Medicaid client in most States has a unique identification number, these can be reconstructed to person-based records, including all of a person’s medical claims in a given time period, such as a year. Problems may occur when people have more than one identification number, although cross-checks can make them less severe. Because the sample sizes and level of detail are generally high, these voluminous data bases can be cumbersome to analyze. Further, administrative data often have errors or missing fields, especially for variables not required for claims payment.

A serious constraint of these data bases is that they are State-specific and not nationally representative. HCFA has sponsored the Tape-to-Tape project, in which MMIS data from California, Georgia, Tennessee, Michigan, and New York from 1980 to the present (years of coverage vary somewhat from State to State) are collected and edited. Uniform data files are created to make State coding comparable—that is, State-specific eligibility, procedure, and other codes are used to create consistent definitions that can be applied to each State (Dodds et al., 1988). Equally important, individual claims are linked to create a person-based record that reflects all Medicaid services used in a given year.

More recently, HCFA began the Medicaid Statistical Information System (MSIS, formerly called MEDSTAT), which encompasses a much larger number of States submitting their MMIS data. States recode their MMIS data to be person-based and to meet uniform Federal specifications. The level of detail is somewhat less than that in the Tape-to-Tape project. Insofar as MSIS is newer and is subjected to less editing than Tape-to-Tape data, its research validity, consistency, and completeness have been examined less fully. However, because 25 States have agreed to provide MSIS data, the breadth and potential value of this data base are great (Roberts, 1990). After passing HCFA editing tests, MSIS data are used to generate substitute HCFA Form-2082 reports. In addition, MSIS data have been used by Federal analysts for several recent research projects, although none has been published to date. Currently, these data are only available to Federal analysts or contractors working on approved projects.

Broad national sample surveys

Many federally sponsored surveys are used to collect data relevant to Medicaid, such as the Current Population Survey, National Medical Expenditure Survey, and Survey of Income and Program Participation. These are relatively well-known and will be discussed later.

Miscellaneous data bases

Certain specialized health care data bases have also been useful for Medicaid analysts. Hospital and nursing home discharge data bases are assembled for various purposes and may indicate which patients had care financed by Medicaid or other payers. Some States (such as California, New Jersey, and Maryland) collect these
routinely for analysis, and the National Center for Health Statistics uses its National Hospital Discharge Survey (Graves, 1990). On a similar basis, clinical abstract data bases exist that provide more information on diagnoses, procedures, and quality of care. The Area Resource File lists numerous health care indicators on a county-by-county basis.

**Special surveys**

Despite the host of data bases described in Table 1, there are times when no available data bases are appropriate, and special surveys must be conducted. Examples include the American Academy of Pediatrics' survey of pediatricians, in which factors affecting Medicaid participation are examined (Perloff, Kletke, and Neckerman, 1987), or the small survey of the health status of people who lost Medicaid eligibility because of the 1982 budget cutbacks and a comparison group that stayed on Medicaid (Lurie et al., 1984).

**Applied uses**

There are many questions we might ask about Medicaid. Sometimes we want to portray the whole "creature," and sometimes we want to scrutinize one portion of its physiology or its behavior. In this section, we summarize several examples of Medicaid analyses or research that illustrate various research issues and methods. We discuss four broad uses of Medicaid data: budget estimates and program development, program analysis and evaluation, comparisons of Medicaid clients and the rest of the population, and interactions of Medicaid with other public programs. We illustrate these with examples relevant to current policy issues, such as coverage of the uninsured, expansions of service to pregnant women and infants, and alternative payment systems.

**Budget estimates and program development**

One of the most immediate demands for Medicaid data is State and Federal budgeting. As a large and growing fraction of most States' budgets, the forecasting of Medicaid costs is a vital issue for State and Federal analysts. Unexpected growth in Medicaid spending can throw budgets out of balance. Most States require balanced budgets, and the Federal budget is regulated by the Gramm-Rudman-Hollings legislation. As a large entitlement program, Medicaid is often a tempting target for budget reductions. At the same time, there is usually pressure to expand health care for the needy and to avoid cutbacks in these services.

Budget analysis typically has two major components. First, there are current service baselines that forecast the costs of current program services, assuming no policy changes in the program. These depend on economic projections of inflation in prices and changes in the number of recipients, case mix, and service intensity. Second, there are estimates of the incremental costs or savings resulting from particular policy changes, e.g., eligibility or reimbursement rules. These require more specific estimates of the impact of a particular policy and its future costs. Budget analysts routinely estimate the costs of legislative proposals and may develop "menus" of possible program changes and their costs (or savings) for review by policymakers. Budget estimates are critical in helping policymakers align program policies and needs with the availability of government funds.

State Medicaid programs generally submit their budget estimates to their Governor's office and the State legislature for State approval and appropriations and also to HCFA via HCFA Form-25 reports for Federal budget development. At every point, the budgets are subject to review and revision. In turn, HCFA budget analysts and actuaries use these reports to generate national baseline estimates of Medicaid costs.

HCFA's Office of the Actuary combines data from HCFA Form-64 and HCFA Form-2082, after adjusting to make types of service comparable, to produce a data set of historical Medicaid spending that serves as a base for projections of future expenditures. This projection data base includes net expenditures reported in HCFA Form-64 by type of service and expenditures allocated to eligibility groups using distributions from HCFA Form-2082.

Although aggregate Medicaid program data are useful for generating budget projections, policy proposals frequently require specific data not routinely reported, such as the costs for medical care for pregnant women or persons with AIDS. As an example, Georgia Medicaid staff used diverse data and approaches to estimate the costs of serving more pregnant women (Skellic, 1990). In the early eighties, Georgia had no special programs for pregnant women and provided services only to AFDC clients with incomes up to about 33 percent of the poverty level. However, concerns about infant mortality led the State to institute special medically needy provisions for pregnant women that increased the income eligibility to, roughly, 50 percent of the poverty level, effective in 1985. Staff used MMIS data to examine the prior number of pregnant enrollees and their medical costs to estimate the baseline cost of services. Available data from the U.S. Bureau of the Census helped define the number of women in the 34 to 50 percent poverty range. Finally, staff examined the experience of nearby States with medically needy programs to anticipate participation rates for this eligibility group. These analyses guided their estimates of the costs of expanding Medicaid for pregnant women.

This Sixth Omnibus Budget Reconciliation Act required raising the income eligibility level for pregnant women to 100 percent of the poverty level and eased access in other ways, begun by Georgia in 1989. Georgia extended its prior analyses of the current costs of services and reviewed impact estimates from various national experts (Newachek, 1987; Fraser, 1987; Torres and Kenney, 1989). In 1990, Georgia staff are using MMIS data to review the actual trends in participation and costs and comparing them with their projections. Participation appears to have grown somewhat more than expected. However, they can use this experience in estimating the costs of serving women up to 133 percent of the poverty level, as required by the Omnibus Budget Reconciliation Act of 1989.
Another example of State use of MMIS data in program development and cost containment is Maryland's diabetes initiative (Stuart, 1990). MMIS data showed that diabetic patients were both frequent and expensive. Because prior research had indicated the reductions in hospitalization possible through diabetes education and nutrition counseling efforts, the Maryland Medicaid Program proposed a managed-care initiative for diabetics. Upon hospital discharge, diabetics will obtain a primary care physician or a gatekeeper, intensive outpatient education, nutrition counselling, and prescriptions shoes (because of circulatory problems common among diabetics). The MMIS data were used to indicate the large potential savings that preventive efforts might achieve. These data were also used in obtaining initial State legislative approval and in applying for a Federal waiver for the case-management services.

Federal current service baselines, projected Federal costs in the absence of legislative changes, may be based on State reports, such as the HCFA Form-25, and on prior national expenditures, such as those reported in HCFA Form-64. For cost estimates for specific policy changes, the Federal situation is more complicated. Although a State can use its own MMIS data to examine a particular service or group, there are no comparable national data, except those in the State reports. Further, because many States may already have a particular policy covered under State-only funding and because many policy proposals permit State options, Medicaid budget estimates try to reflect variation in current State policies and in State plans to implement optional policies.

In principle, some of these analyses could be conducted using MSIS data or Tape-to-Tape data, but these can be limiting because of their State specificity and cumbersome because of the volume of the data. More recently, microsimulation models have been used to estimate the costs and impacts of Medicaid expansions. Microsimulation models are complex economic or demographic models that statistically link numerous data bases (such as the Current Population Survey), program policies, and macroeconomic information to simulate policies at household or individual levels and aggregate them back to the national level (Lewis and Michel, 1990). Although microsimulation is powerful and versatile, it is complex (and therefore expensive to develop) and may provide unreliable estimates when underlying model assumptions or data are incorrect.

Used somewhat more in examining the effects of changes in tax policy or income transfer programs (e.g., Food Stamp Program), microsimulation has recently been used to examine health financing. Lewin/ICF used its Health Benefits Simulation Model to examine the costs and distributional effects of Medicaid expansions, employer mandated insurance, and other policies to reduce the number of uninsured persons (Needleman et al., 1990). The model statistically matched data from the 1988 Current Population Survey, the 1980 National Medical Care Utilization and Expenditures Survey, and the Small Business Administration Employer Health Plan Data Base. A similar project was undertaken by the Urban Institute using the TRIM2 model (Holahan and Zedlewski, 1990). In the TRIM2 model, analysts used a combination of Current Population Survey, HCFA Form-2082, and Tape-to-Tape data to estimate the probability of participation in Medicaid under alternative policies. Then they used Tape-to-Tape data to estimate regression models of the relation of household characteristics to Medicaid expenditures to estimate the costs of making new groups of people eligible. Another recent microsimulation model is the Brookings-ICF Long-Term Care Financing Model (Rivlin and Wiener, 1988).

**Medicaid program analysis and evaluation**

Budget and program development analyses often have an inherent underlying policy agenda; there is a wide range of health services research that is more analytical and less partisan in nature. We group these into a broad category of program analysis and evaluation. The broadest analyses are those describing the size and scope of program participation, services, and cost for the Nation over time. Typically, these rely on State-level aggregate data, especially the HCFA Form-2082 reports. Reilly, Clauser, and Baugh (1990) describe the patterns and trends in Medicaid participation, utilization, and cost over time. Similar analyses of HCFA-2082 data have been conducted by Burwell and Rymer (1987), Holahan and Cohen (1986), and Gornick et al. (1985).

Sometimes, aggregate HCFA Form-2082 data are not detailed enough to make a particular problem or population understandable. In these cases, researchers may use more detailed MMIS data. In many of these analyses, the focus is certain groups with high medical needs and high costs. For example, Adams et al. (1989) examined diagnoses, utilization, and expenditures for SSI disabled Medicaid recipients using 1984 Tape-to-Tape data. Similar analyses were conducted by Burwell et al. (1987) for Medicaid recipients in intermediate care facilities for the mentally retarded. These focused analyses enable us to see what disabling conditions caused people to be eligible for Medicaid and how these conditions related to services used and medical expenditures. Another virtue of person-level data is that longitudinal person records can be developed to examine patterns of enrollment and expenditures over several years (Howell, Andrews, and Gornick, 1988).

HCFA’s Bureau of Data Management and Strategy recently conducted a study to assess the effects of the skilled nursing facility and the intermediate care facility portions of the Medicare Catastrophic Act on persons dually enrolled in Medicaid and Medicare using MSIS. For 12 States, Medicaid and Medicare files were linked by social security number to examine utilization patterns. Research has also been undertaken on the effect of alternative Medicaid reimbursement mechanisms. The Nationwide Evaluation of Medicaid Competition Demonstrations, conducted in six States, examined various schemes of capitation and case management in Medicaid (Freund et al., 1989; Health Care Financing Administration, 1986). This wide-ranging project illuminates multiple approaches and data sources used in program evaluation. In conjunction with MMIS data from the States, the evaluators conducted a survey of Medicaid consumers on services, quality of care, and access; abstracted medical records; and conducted case studies, administrative cost studies, and ratesetting studies. The
effect of the demonstration projects on utilization and cost of care and selection bias (largely using MMIS data) as well as on consumer satisfaction and access (based on the consumer survey) and on quality of care (based on the consumer survey and medical records abstraction) could be determined from these data.

A recent priority in health services is outcome research, the relation of medical and surgical treatment modalities to subsequent patient outcomes. MMIS is a rich source of information for certain topics that have implications beyond Medicaid alone. Because Medicaid covers many services not covered in other data bases (such as outpatient drugs and long-term care) and many clients not well covered by other programs (such as poor pregnant women and children and AIDS patients), longitudinal Medicaid data can provide unique insights into the consequences of certain types of care. Pioneers in using Medicaid data for this area of research are Ray and Griffin (1989). Ray et al. (1987) found that elderly Medicaid patients using psychotropic drugs, such as Valium, were more likely to eventually fall and suffer hip fractures, which are serious, expensive injuries. This suggested that physicians needed to be more cautious in prescribing for elderly patients.

Comparisons with non-Medicaid recipients

An original goal of Medicaid was to give poor Americans access to health care similar to that of those who are financially better off and to improve their health status. A key gap in many of the data bases described earlier is that they lack information about people not on Medicaid or about experiences of Medicaid clients before and after participation in Medicaid. Such analyses require data bases that include both Medicaid and non-Medicaid recipients, such as broad population surveys or hospital discharge data. Linkage of Medicaid and other data bases, such as birth certificates, is another way to bring in comparison populations.

Two key data bases for participation in social assistance programs in general are the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP). CPS is an annual cross-sectional survey that, among other purposes, has been used to estimate the number of uninsured Americans (Moyer, 1989; Swartz and Purcell, 1989). SIPP is a longitudinal survey that follows families for about 2½ years. An important use of SIPP is to examine transitions into and out of Medicaid; Short, Cantor, and Monheit (1989) found that more than one-half of the persons entering Medicaid had no insurance before entry, and more than one-half of those exiting Medicaid had no private insurance afterward.

For health policy analysts, the most important issues may be health care utilization, expenditures, and health status. Kasper (1986) used the 1980 National Medical Care Utilization and Expenditures Survey (NMCUES) to compare poor and nonpoor persons with and without Medicaid; she found that Medicaid enrollees had poorer health status than those not on Medicaid and that Medicaid enrollees in poor health used health services comparable to those who were not poor and not on Medicaid. Rosenbach (1989) also used NMCUES to compare children's access to ambulatory health care and found that Medicaid children were more likely than children who were privately insured or uninsured to visit an office-based physician. The 1987 National Medical Expenditures Survey (NMES) will certainly be an important source for further understanding Medicaid. To verify utilization data in NMES and to provide expenditure data that recipients were rarely able to provide, HCFA provided Medicaid claims data from MSIS, based on matches with the social security number.

Aside from these sample survey data, all payer hospital or nursing home discharge data bases offer potential for comparing Medicaid and non-Medicaid groups. Robinson and Phibbs (1989) used 1982-86 California hospital discharge data, supplemented with information from the Area Resource File, to conduct an econometric time-series analysis of the effect of the State policy of selective contracting in Medicaid on hospital expenditures and attributed significant savings to selective contracting.

A final approach to comparing Medicaid and non-Medicaid groups is linkage of MMIS data to other broad data bases, such as birth certificates. Data base linkage requires common unique identifiers in the data sources, such as social security numbers, AFDC client numbers, names (more difficult to use), and auxiliary identifiers to help resolve difficult cases, such as date of birth and sex. If these are available and confidentiality and privacy concerns can be satisfied, data base linkage can increase the analytic potential of existing data bases for a modest investment. Recent work by Howell et al. (1989) has linked California Tape-to-Tape records with birth certificates to compare prenatal care and birth outcomes among Medicaid and non-Medicaid women, including non-Medicaid women living in low-income areas.

Interaction with other public programs

Medicaid is intertwined with a host of other public programs, such as AFDC, SSI, Medicare, the Maternal and Child Health Services Block Grant, refugee assistance, and the Women, Infant, and Children (WIC) Nutrition Program, which jointly form a social safety net for poor, disabled, and elderly Americans. Changes in one program may affect the other. Budget analysts often seek to estimate the effects of changes in eligibility for AFDC or SSI on Medicaid spending. The enactment and subsequent repeal of the Medicare Catastrophic Coverage Act of 1988 had substantial impacts on Medicaid budgets. In recent years, researchers have begun to explore interactions between Medicaid and other programs.

A longstanding concern has been that fear of losing Medicaid benefits may deter poor families from leaving AFDC. A slight increase in income may result in complete loss of Medicaid benefits (the Medicaid "cliff" effect). The expected value of Medicaid on welfare dependency, using different data bases and approaches, has been examined in three recent studies: Ellwood and Adams (1990) used Tape-to-Tape data; Moffitt and Wolfe (1989) used the SIPP and NMCUES data, and Blank (1989) used NMCUES data. These authors showed the importance of health insurance and Medicaid in employment behavior of AFDC clients and the relevance of policies offering transitional Medicaid.
benefits to people leaving welfare to work. In a similar vein, Andrews and her colleagues (1988) examined Medicaid utilization by disabled persons permitted to retain Medicaid in an SSI work incentive program and showed that Medicaid expenditures were equivalent to a large portion of their earnings.

Other researchers have examined the effects of the WIC Nutrition Program on Medicaid spending. Schramm (1985; 1989) linked records from Medicaid, WIC, and birth certificates to study the benefit-cost ratio for prenatal WIC participation in Missouri. For 1985-86, he found that a dollar spent on prenatal WIC was associated with a reduction of $0.79 in Medicaid newborn payments and one of $1.97 in Medicaid charges.

Gaps in the data

The discussion above is a brief review of the types of Medicaid data available and of selected research applications. Even so, the scarcity and uncertain quality of Medicaid data continue to be important problems. Based on our review of data sources and informal discussions with analysts and researchers at State and national levels, certain themes were recurrent. National-level researchers were concerned about the lack of and/or quality of Medicaid data across the Nation, and State-level analysts tended to be more satisfied with their MMIS data, but were concerned about a lack of support to do analyses.

From the national perspective, the critical shortcomings mentioned were lack of data to represent Medicaid program experience across the Nation and over time. We identified three specific concerns: mixed quality of HCFA Form-2082 reports, lack of consistent historical documentation of State program characteristics, and lack of a national person-based sample data base with information on Medicaid eligibility, utilization, and expenditures. These data generally exist in State MMISs or in procedure manuals; they just have not been centrally collected or edited as effectively as they could be.

Improving HCFA Form-2082 quality

HCFA Form-2082 reports are widely used, relatively convenient, and fairly detailed. They have been, and will remain, the mainstay for analysis of Medicaid trends. However, inconsistent reporting patterns cause problems. Time trends and interstate comparisons are generally uncertain, because it is not possible to separate true differences and reporting artifacts. Working together, HCFA and State Medicaid agencies have improved the reporting process over the past several years. In 1984, HCFA Form-2082 was expanded from 20 to 47 pages and States had to revise reporting programs to generate the new data. From 1984 to 1986, HCFA conducted Data Validation Reviews as part of the State Performance Review Process. Federal staff reviewed each State’s service and eligibility crosswalks to ensure consistency. Errors were reported to the States, and penalty points were assessed under the State Performance Review process if timely corrections were not made.

Under the current State reporting system, two mechanisms might improve future data quality. First, at the very least, State and Federal footnotes and explanations of how programs or reporting changed and how certain groups were defined would enable analysts to determine possible reasons for discrepancies. Second, expert reviews that ensure consistency of eligibility and service crosswalks, compare a State’s current and prior year’s submissions, and compare reports from similar States may provide better editing capability. Perhaps a shift to federally generated reports using MSIS will alleviate these problems in a different way. Correcting historical data would be quite difficult, but some explanatory notes could be collected for recent years.

Reporting State program characteristics annually

The Congressional Research Service (CRS) (1988) recently completed its detailed Medicaid Source Book: Background Data and Analyses and hopes to update it in the future. However, CRS had to commission a special survey of States by the National Governors’ Association because the data they wanted were not available. Relatively complete, good quality reviews of State program characteristics are annually reported for the AFDC and SSI programs (Family Support Administration, 1989; Committee on Ways and Means, 1989; Social Security Administration, 1989). In fact, these are important references for understanding Medicaid eligibility. However, it must also be acknowledged that the Medicaid program is much more complicated than these programs, because there are complex program policies on covered services and provider payment mechanisms, as well as eligibility policies. Some characteristics can be easily summarized and even made into a data base, but some policies, such as many aspects of reimbursement policies, are complex and not easily compared across States or over time. HCFA, which has sponsored such surveys in the past on an ad hoc basis, could institute annual surveys of State Medicaid program characteristics.

Developing national data bases

We can use examples of data bases from other programs to illustrate the types of data desired. The principal analogies that we might draw are from Medicare and from AFDC and the Food Stamps Program. Because Medicare is federally administered, HCFA maintains national data bases and makes many of these publicly available, including the Expanded Modified MEDPAR and Quality Control/MEDPAR files (containing data on utilization and institutional providers), the Part B Medicare Annual Data files (containing data on physician services, ambulatory surgical centers, and suppliers), and the Medicare Automated Data Retrieval System (which links Part A and B files) (HCFA, 1988). These data are available in various forms, organized by providers, beneficiaries, and procedures, in 100 percent or sample files.

In principle, comparable data are available from the State Medicaid computer systems. However, there is no requirement that any State submit its MMIS data to HCFA, and the MMIS data are coded differently in every State, based on State policy options and computer
facilities. This makes the process of establishing a national data base very difficult. Both the Tape-to-Tape project and MSIS have made significant strides in this direction, but they are limited in geographic coverage. Tape-to-Tape makes substantial efforts to edit data and to map State codes to uniform codes, but this process is time consuming. MSIS relies on States to process data to HCFA-specified coding, but reliability is still uncertain.

Standard diagnosis and procedure coding systems (i.e., International Classification of Diseases, 9th Revision, Clinical Modification and Current Procedural Terminology, 4th Revision) are used by States, but eligibility codes vary greatly across the Nation. Some States use 40 or more detailed eligibility codes, and others use just 6 or 7 codes. Thus, eligibility maps are less consistent across the Nation. HCFA should consider working with States to create a detailed national uniform eligibility coding scheme that can be gradually incorporated into State MMISs, permitting some variance for State-only Medicaid eligibility groups.

Some interviewees suggested that, if a national MMIS-type data base can ever be established, public use tapes should include a smaller research sample, e.g., a 5-percent file, that reduces the cost and effort of processing hundreds of millions of claims. Although Tape-to-Tape and MSIS files are for single States, they are extremely large and difficult to use. Even if it is not possible to obtain data from all States, it may be possible to develop a stratified sampling plan or a sentinel-site system that essentially emulates national data. A national person-based claims data base could vastly increase the capability to conduct analyses of the health utilization and expenditures of Medicaid recipients.

The other comparisons for Medicaid data are the AFDC and Food Stamp Programs, which use their quality control (QC) samples to develop data bases and reports on the sociodemographic characteristics of participants. The Medicaid QC system is designed as a management tool to verify eligibility of samples of recipients to estimate the dollar value of benefits paid in error and is performed in conjunction with AFDC and the Food Stamp Program as part of the Integrated Quality Control System. The Medicaid QC sample is an ongoing, nationally representative cross-sectional sample of enrolled Medicaid cases (usually households) and their Medicaid expenditures for the review month. About 28,000 AFDC cases and 20,000 non-AFDC cases are sampled by States every 6 months, and a large fraction are re-reviewed by Federal staff. Certain groups, such as SSI recipients with eligibility determined by Federal staff, 100-percent federally funded cases (such as refugees), and certain foster care and adopted child cases, are excluded from the sampling frame (Federal Register, 1990). Because SSI recipients tend to have much higher medical expenses and utilization, the lack of many SSI recipients is a problem. In one research project, this gap was overcome by incorporating SSI quality assurance data from the Social Security Administration, which was similar to the QC data (Adler and Adler, 1984).

Analysis of QC data can provide detailed sociodemographic data on Medicaid cases nationally, including eligibility group. Because Medicaid claims data bases lack family income, assets, or other detailed demographic traits, this is a sound complement to MMIS-type data bases. It may be possible, through some modification of the QC process, to include more detailed health care utilization and expenditure data to develop a data base that includes demographic, eligibility, utilization, and expenditure patterns for a nationally representative sample. However, because health care utilization varies greatly from month to month, the usefulness of a data base based on a single month of claims is limited.

Supporting State research and analysis

From the State perspective, Medicaid data needs are quite different. State analysts already have comprehensive MMIS data for their States. The problems they face pertain to populations or services not currently covered under their programs, which require such information as Census data or data on the cost or utilization of new services. These are problematic because decennial census data are often out of date, and national survey data, such as NMES and CPS, usually have too few cases within a State for statistical validity. These can be important for estimating the effects of program expansions. An additional problem in generating budget estimates is that States may be uncertain of future Federal policy in light of changing legislation or delays in issuing regulations.

However, the major problem cited was a lack of support for analysis and research within the State because of a shortage of available qualified staff, funding, and other resources (e.g., computer access). It is often the case for Medicaid, as well as other State programs, that State resources are concentrated on program management, and few or no resources are allocated to program analysis or evaluation. These tendencies may be shortsighted because policy analysis and evaluation staff could serve important roles in improving program management and containing costs. For example, research efforts examining small-area variation in practice patterns could be converted into routine reports to use in identifying high-cost providers using the Surveillance/Utilization Review Subsystems of MMIS. Efforts to improve prenatal care services could lead to reduced Medicaid neonatal costs.

Given the size and importance of Medicaid to State budgets, State officials might benefit by designating a fraction (e.g., one-quarter or one-half of 1 percent) of Medicaid funds for research and development, including improving budget estimate capabilities and cost-containment strategies. They could use researchers from public health departments or State universities to augment State Medicaid research capabilities. States also need to be encouraged to disseminate their findings more widely; usually, State administrative reports are available only to a few people. Medicaid's flexible structure permits one or a group of States to serve as "laboratories" to pilot test new systems that might be appropriate for other States or the Nation. However, the lack of research and dissemination of findings inhibits this ability to learn through experience. Through regional and national level efforts, HCFA could stimulate States to
conduct analyses and share the findings. This may include workshops or meetings for technical staff who have to work with the data.

Conclusions

As long as Medicaid is a program with both Federal and State management and funding, data and analysis needs for the program will be complex. The dual Federal-State responsibility of Medicaid permits substantial flexibility and discretion. This structure inherently makes data collection and analysis more difficult at the Federal level and more important at the State level. In light of the diversity of State programs and reporting formats, it is reasonable to expect that collecting and editing data for Medicaid will be proportionately more expensive than similar efforts for Medicare, which can impose uniform policies and data collection procedures.

At the national level, efforts to adequately understand the variations in State policies, recipient and provider participation, and service trends require complex, adequately funded data collection and analysis efforts. We mentioned three main areas for improvement: upgrading quality of the HCFA Form-2082 reports, conducting annual State program characteristics surveys and developing national person-based data bases about Medicaid clients.

At the State level, State administrators should consider how to establish and support analysis, research, and evaluation units that will provide useful, policy relevant information to improve program design and management. As a large, complex program with a dominant role in State budgets, managers need to adequately plan for research and development as an adjunct to program management and policy. Finally, State analysts could do a better job of disseminating research on their programs to other interested parties outside the State.

These concerns are all the more important given the numerous public policy issues, such as uninsurance, long-term care, AIDS, and infant mortality, that hinge on Medicaid policies. Careful management of limited government resources requires that policy analysts be able to evaluate current policies, identify problem areas, suggest possible policy solutions, and anticipate the effects of potential policy changes. Good analyses require good data.

Returning to the story of the four blind men and the elephant, we might say that the current state of affairs of Medicaid data is equivalent to several men and women, some very farsighted and the rest very nearsighted, trying to decipher a much stranger creature. Between their efforts, we can get a fuzzy, sometimes distorted picture of the Medicaid program, its clients, providers, and services. At the national level, we noted some key data gaps and proposed possible enhancements. To make significant improvements will require some additional funding and effort. However, virtually all the Medicaid data already exist at State levels; the effort to collect, edit, and disseminate them at a national level is, in that regard, incremental. At the State level, more can be done with existing data to explore the effects of program innovations and to guide future policy and program development.

A final question is: Can improved research and analysis, based on good data, make Medicaid a better program? Let us answer it this way: You would not want to start leading an elephant by tugging on its tail. Research and data analysis can help by pointing the way for future program policy and management.

Acknowledgments

We gratefully acknowledge numerous Federal, State, and private Medicaid analysts and researchers who provided their thoughts and experiences about data issues and the Medicaid Program. Two reviewers provided insights into Health Care Financing Administration’s data bases and helped us focus some policy issues. A special thanks is given to Kipp Gabriel for literary assistance.

References

Adams, E. K., Ellwood, M., and Pine, P.: Utilization and expenditures under Medicaid for Supplemental Security Income disabled. *Health Care Financing Review* Vol. 1, No. 1. HCFA Pub. No. 03286. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Oct. 1989.

Adler, M., and Adler, G.: Uses of Quality Control Data for Medicaid Policy Research. Working Paper No. 84-8. Office of Research and Demonstrations, Health Care Financing Administration. Baltimore, Md. Aug. 1984.

Andrews, R., Ruther, M., Baugh, D., et al.: Medicaid expenditures for the disabled under a work incentive program. *Health Care Financing Review*, Vol. 9, No. 3. HCFA Pub. No. 03263. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Apr. 1988.

Blank, R.: The effect of medical need and Medicaid on AFDC participation. *Journal of Human Resources* 24(1):54-87, 1989.

Burwell, B., Clauer, S., Hall, M. J., and Simon, J.: Medicaid recipients in intermediate care facilities for the mentally retarded. *Health Care Financing Review*, Vol. 8, No. 3. HCFA Pub. No. 03237. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Apr. 1987.

Burwell, B., and Rymer, M.: Trends in Medicaid eligibility. *Health Affairs* 6(4):30-45, 1987.

Congressional Research Service: *Medicaid Source Book: Background Data and Analysis*. Washington. U.S. Government Printing Office, Nov. 1988.

Dodds, S., Blake, J., Miller, K., et al.: Accessing and Using the Tape-to-Tape Data Base. Santa Barbara, Calif. SysteMetrics/McGraw-Hill, Jan. 1988.

Ellwood, D., and Adams, K.: Medicaid mysteries: Transitional benefits, Medicaid coverage, and welfare exits. *Health Care Financing Review* 1990 Annual Supplement. HCFA Pub. No. 03311. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Dec. 1990.
Robinson, J., and Phibbs, C.: An evaluation of Medicaid selective contracting in California. *Journal of Health Economics* 8:437-55, 1989.

Rosenbach, M.: The impact of Medicaid on physician use by low-income children. *American Journal of Public Health* 79(9):1220-1226, Sept. 1989.

Ruther, M., Pagan-Berlucchi, A., Wivell, K.; and Howe, C.: *Medicare and Medicaid Data Book, 1986*. Health Care Financing Program Statistics. HCFA Pub. No. 03247. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Sept. 1987.

Schramm, W.: WIC prenatal participation and its relationship to newborn Medicaid costs in Missouri: A cost/benefit analysis. *American Journal of Public Health* 75:851-57, Aug. 1985.

Schramm, W.: WIC Prenatal Participation: Cost/Benefit Analysis for Missouri 1985-1986. Paper presented at the Annual meeting of the American Public Health Association. Oct. 23, 1989.

Short, P., Cantor, J., and Monheit, A.: The dynamics of Medicaid enrollment. *Inquiry* 25:504-516, Winter 1988.

Skellie, B., Georgia Department of Medical Assistance. Personal Communications, 1990.

Social Security Administration: *Characteristics of State Assistance Programs for SSI Recipients, January 1989*. Office of Supplemental Security Income. SSA Pub. No. 17-002. Washington. U.S. Government Printing Office, 1989.

Stuart, M.: Maryland Department of Health and Mental Hygiene. Personal Communication, 1990.

Swartz, K. and Purcell, P.: Letter: Uninsured Americans. *Health Affairs* 8(4):193-97, 1989.

Taylor, A.: *English Riddles in Oral Tradition*. Berkeley, Calif., University of California Press. 1951.

Torres, A., and Kenney, A.M.: Expanding Medicaid coverage for pregnant women: Estimates of the impact and cost. *Family Planning Perspectives* 21(1):19-24, 1989.

U.S. Congress: House Committee on Ways and Means. *Background Material and Data on Programs Within the Jurisdiction of the Committee on Ways and Means*, Washington. U.S. Government Printing Office. Mar. 15, 1989.