

The Potential Use of Health Care Financing Administration Data Sets for Health Care Services Research

by Judith Lave, Allen Dobson, and Carol Walton

Administrative Record Systems may be an overlooked source of data for health services researchers. Through its administration of the Medicare and Medicaid Programs, the Health Care Financing Administration (HCFA) routinely receives data on items such as its beneficiary population, providers certified to deliver care to its beneficiary population, providers certified to deliver care to the beneficiaries, the use of services and reimbursements to providers. This article introduces the reader to the HCFA data, it describes the most important data bases that are useful for research, their relative strengths and weaknesses and the extent to which they are available to outside users.

Introduction

The decrease in Federal funding for health services research will lead to a number of changes in the nature and type of expensive primary data collection activities undertaken by investigators and an increased reliance on primary data collected for other purposes as well as on secondary data analysis. Researchers will search for readily available data, preferably already in machine-readable format.

Health researchers have relied extensively on such data in the past. Data from the Health Interview Survey, The Annual Survey of Hospitals and the Area Resource File are examples of widely used data. Other possible and perhaps overlooked sources of data are the administrative records of agencies that gather data as part of their ongoing activities.

This article describes some of the data that are routinely collected by the Health Care Financing Administration (HCFA) as well as that collected through two surveys, and it indicates the extent to which these data can be made available to outside researchers. The article begins with a very brief discussion of the two HCFA programs, Medicare and Medicaid. Next, it describes the data that make up the "Medicare Statistical System," the Medicare cost report information, the Medicaid data, and data collected in two surveys. The purpose is to introduce the HCFA data and its availability rather than to give a complete description of it. The paper concludes with a discussion of the relative strengths and weaknesses of the major HCFA data bases from a user's perspective.

The HCFA Programs

The two major programs funded by HCFA are Medicare and Medicaid, which were established by the 1965 amendments to the Social Security Act (Titles XVIII and XIX respectively).

The Medicare Program

The Medicare program covers hospital, physician and other medical services for persons age 65 and over, disabled persons entitled to Social Security cash benefits for 24 consecutive months, and most persons with end-stage renal disease. Total Medicare expenditures were nearly \$50 billion in fiscal year 1982 (Muse and Sawyer, 1981).

Medicare has two complementary but distinct parts: Hospital Insurance (HI), known as Part A, and Supplementary Medical Insurance (SMI), known as Part B. The HI program covers 90 days of inpatient hospital care in a benefit period ("spell of illness") — which begins with a hospitalization and ends when the beneficiary has not been an inpatient in a hospital or skilled nursing facility (SNF) for 60 continuous days. There is no limit to the number of benefit periods an individual may use. The program also provides a one-time ("life-time") reserve of 60 days to use if a beneficiary exhausts the 90 days available in a benefit period. In addition to inpatient hospital care, the hospital insurance program covers up to 100 post-hospital days in a SNF if the beneficiary requires such care. The program also covers home health agency (HHA) visits. About 95 percent of the nation's aged population are enrolled in the HI program.

Nearly everyone covered by the HI program voluntarily enrolls in the SMI program. Unlike the hospital insurance program, SMI coverage is contingent upon the payment of a monthly premium \$12.20 per month as of July 1982. Under "buy-in" agreements, most State Medicaid programs pay these premiums for persons who qualify for Medicaid in addition to Medicare. The SMI program provides payments for physicians as well as related services and supplies ordered by the physician. SMI also covers outpatient hospital services, rural health clinic visits, and home health visits.

Several health care services that the aged generally use on a continuing basis such as drugs, dental care, routine

eye examinations, and preventive services are not covered by Medicare. Long-term institutional services are not covered either (as opposed to SNF stays resulting from an acute care episode).

Both the HI and SMI programs require some beneficiary cost sharing. Under the hospital insurance program, the patient pays an inpatient hospital deductible in each benefit period. This deductible approximates the cost of one day of hospital care (\$260 in 1982). Coinsurance based on the inpatient hospital deductible is required for the 61st through 90th day of inpatient hospital care (always equal to 1/4 of the hospital deductible), for the 21st through 100th day of skilled nursing facility (1/6 of the deductible), and for the 60 lifetime reserve days for inpatient hospital care (1/2 of the deductible). The patient is also liable for the cost (or replacement) of the first three pints of blood.

Under SMI, in addition to paying a monthly premium, the beneficiary must meet a deductible each year (\$75.00 in 1982). On each claim for payment, physicians can accept or reject assignment. Acceptance of assignment means that the physician agrees to accept as full payment the amount Medicare allows for the service. The program reimburses 80 percent of allowed charges directly to the physician. Beneficiaries are liable for the remaining 20 percent (coinsurance) of allowed charges. On unassigned claims, the beneficiary is also responsible for the difference between the physician's charge and the allowed charge. Beneficiaries covered under Medicaid "buy-in" agreements are relieved of these cost-sharing obligations. HI is financed primarily by a tax on earnings while SMI is financed by enrollee premiums and Federal general revenues.

The Medicaid Program

Medicaid is a Federally-supported and State-administered assistance program providing medical care for certain low income individuals and families. Medicaid accounted for about \$27 billion in Federal and State expenditures in fiscal year 1981 and is the primary source of health care coverage for the poor in America. About one third of the poor are covered by Medicaid (Dobson, 1982). The program is designed to provide medical assistance to those groups or categories of people who are eligible to receive cash payments under one of the existing welfare programs established under the Social Security Act; that is, Title IV-A, the program of Aid to Families with Dependent Children (AFDC), or Title XVI, the Supplemental Security Income (SSI) program for the aged, blind and disabled. In most cases, receipt of a welfare payment under one of these programs means automatic eligibility for Medicaid. In addition, States may provide Medicaid coverage to the "medically needy", that is, to people (1) who fit into one of the categories of people covered by the cash assistance programs (aged, blind, or disabled individuals or members of families with dependent children when one parent is absent, incapacitated, or unemployed), and (2) who have enough income to pay for their basic living expenses (not welfare recipients), but not enough income to pay for their medical care.

Title XIX of the Social Security Act requires that every State Medicaid program offer certain basic services: inpa-

tient hospital services, outpatient hospital services, laboratory and X-ray services, skilled nursing facility services for individuals 21 and older, home health care services for individuals eligible for skilled nursing services, physicians' services, family planning services, rural health clinic services, and early and periodic screening, diagnosis, and treatment services for individuals under 21. In addition, States may provide private duty nursing, intermediate care facility services, inpatient psychiatric care for the aged and persons under 21, physical therapy, and dental care.

Medicaid operates as a vendor-payment program. Payments are made directly to providers of services for care rendered to eligible individuals. Providers must accept the Medicaid reimbursement level as payment in full. In mental institutions and intermediate care facilities, individuals are required to submit income in excess of their personal care needs to help meet the cost of care. States may require Medicaid recipients to cost-share on services under certain circumstances.¹ As noted above, most State Medicaid programs have buy-in agreements with Medicare. Under these agreements, Medicaid assumes responsibility for the Medicare cost-sharing obligations of persons covered under both programs.

Medicaid is financed jointly with State and Federal funds. Federal contributions vary with States' *per capita* income and currently range from 50 percent to 77.55 percent of program medical expenditures. Administration and Medicaid Management Information System (MMIS) costs are matched at other rates. States administer their Medicaid programs within broad Federal requirements and guidelines. The requirements allow States considerable discretion in determining income and other resource criteria for eligibility, covered benefits, and provider payment mechanisms. As a result, the characteristics of Medicaid programs vary considerably from State to State.

The Omnibus Budget Reconciliation Act of 1981 made major changes to the program. The act attempts to limit Federal reimbursement to States by manipulating Federal participation rate rules and it modifies coverage and service requirements for the medically needy. The act also provides for waiver of freedom of choice which allows the Medicaid program to contract for services with a limited number of providers and for a home- and community-based waiver program designed to reduce Medicaid beneficiary reliance on long-term care institutional support.

¹Previous to the tax Equity and Fiscal Responsibility Act of 1982, cost-sharing was limited to optional services. Current law now allows States to impose nominal copayments in all cases except for any services provided to categorically and medically needy children under age 18 (or up to age 21 at the State's option) or to categorically and medically needy institutionalized persons who are required to spend all their income for medical expenses except for a personal needs allowance; copayment may not be imposed on pregnancy-related services provided to categorically needy Health Maintenance Organization (HMO) enrollees.

The Medicare Statistical System

Content

The Medicare Statistical System (MSS) was designed to provide data to measure and evaluate the operation and effectiveness of the Medicare program (Goldstein, 1981). The statistical system is a by-product of three administration record systems which are centrally maintained in the operation of the Medicare program: the Health Insurance Master File, the Provider of Service File, and the Utilization File.

The Health Insurance Master File

This file contains a record on each person who is enrolled in Medicare. (Approximately 26 million aged beneficiaries and 3 million disabled beneficiaries are currently entitled to Medicare benefits.) Data elements for each individual include the Medicare claim number, age, sex, race, place of residence, and reason for entitlement (for example, end-stage renal disease). The Medicare Statistical System obtains data from this file to prepare enrollment statistics and to provide a denominator for calculating all Medicare utilization rates.

The Provider of Service File

This file contains information on every hospital, skilled nursing facility, home health agency, independent laboratory and other institutional providers that have been certified to participate in the program. Each is assigned a distinct provider number. Approximately 7,000 hospitals, 5,200 SNFs, 3,000 HHAs, 3,500 independent laboratories, and 1,600 other types of facilities that participate in Medicare or Medicaid, or both, are in this file. The available descriptive information differs by type of provider but includes geographic location, bed size (hospitals and nursing homes), type of control (nursing home, hospitals, and home health agencies), and medical school affiliation (hospitals).

The Utilization File

The third part of the Medicare Statistical System — the Utilization File — is based on the Medicare billing information which is received and processed centrally to update the Health Insurance Master File for data elements such as copayments, deductibles and spells of illness. After this administrative process is complete, 100 percent of the bills enter the statistical system. The bill information includes the amount billed and the amount of interim reimbursement. For services it includes date of admission and date of discharge. For a sample of the bills the MSS obtains more extensive information. For example, on approximately 20 percent of the hospital bills, information on the nature of the hospital episode (diagnostic and surgical procedures) is obtained. These data elements are reported by the hospital

as International Classification of Diseases, 9th edition, Clinical Modification (ICD-9-CM) codes or as narratives that are coded at HCFA. (Although diagnoses are coded for a sample of other institutional providers, they have not been really used for research purposes.)

Since each record in the utilization file contains the beneficiary's claim number and the provider's number, the utilization records can be matched to the enrollment and provider records. This then provides the basis for developing population-based statistics or provider-based statistics. These record systems are extremely large. For example, about 12 million inpatient hospital bills, 30 million outpatient hospital bills and 150 million physician payment records were received and processed in 1981. Usually HCFA staff works with samples drawn from the parent population.

Availability

Primary Data

The data in the Medicare Statistical System (MSS) are confidential and access to the files is governed by the Privacy Act. Conditions under which researchers can access these data are given in the October 2, 1981, issue of the Federal Register.

Access to these data, however, is limited because of factors other than confidentiality. The HCFA data system was designed to serve internal needs. Unlike some other government agencies, HCFA has not developed public use data tapes. No tapes are ready to be given to researchers on request and no supporting documentation is available. Thus, every approved request has to be produced individually by a staff whose primary responsibility it is to prepare data for internal purposes.

While HCFA data bases have not been widely disseminated, they have been utilized by HCFA grantees and contractors. This practice will undoubtedly continue, and in a few instances may be bolstered by the development of general purpose public use tapes. These public use files would provide summary information on Medicare enrollment and utilization for the most recent year(s) available.

Secondary Data from the Medicare Statistical System

The Health Care Financing Administration publishes a variety of reports that make available statistics on the Medicare program. These publications, including the Health Care Financing *Program Statistics* series, provide statistics on beneficiary utilization, billed charges, and reimbursement for each covered service. The statistics are given for the Medicare population as a whole and for subgroups of the population based on age, sex, race, and place of residence. Reports are also prepared on characteristics of the Medicare beneficiary population as well as the providers of services. A few of these reports are available in computer tape format (for example, State and county interim reimbursements).

The Medicare Cost Reports

Content

Under Medicare, institutional providers (hospitals, skilled nursing facilities, and home health agencies) are reimbursed not on the basis of bills submitted, but rather on the basis of "reasonable costs."² To receive reimbursements, providers must submit cost reports to the fiscal intermediaries.³ The cost reports contain detailed information on direct and indirect costs, revenues, aggregate total charges as well as data on the number of full-time equivalent personnel. The cost reports are received by HCFA in printed form (hard copy).

Availability

Data from the full cost reports are not readily available to outsiders. With few exceptions, only hard copy cost reports can be obtained. These can be obtained usually with the hospital's permission from the hospital's Medicare fiscal intermediary. However, under the National Hospital Rate Setting Study, a data file with supporting documentation was prepared on a 25 percent random sample of hospitals from the 48 States plus a 100 percent sample of hospitals from the 15 rate-setting states for a ten-year period (1969-1979). This file contains extensive information from the Medicare Cost Reports and will be supplemented with information from the Area Resource File and the American Hospital Survey. This file will be made available to HCFA contractors.

HCFA has been developing a hospital cost report information system (HCRIS) for maintaining a national longitudinal data base for Medicare hospital cost reports. Implementation of this data base will provide an enormous improvement in the availability of cost report data. Eventually HCRIS data will be available in computer tape format.

Medicaid Data

Content

HCFA does not receive any individual data on Medicaid eligibles, Medicaid recipients, or on payments made for medical services provided. The States are, however, required to provide aggregate statistics on the Medicaid program to the Federal government. For researchers the most important of these are the Medicaid Minimum Data Set (MMDS) and the Medicaid Financial Data Set (Cromwell, Schurman and Adler, 1982).

Medicaid Minimum Data Set (MMDS)

This data set is a subset of the State's Medicaid Management Information System (MMIS) file. It contains the

²Beginning with fiscal year 1984, hospital reimbursements will shift from one based on reasonable costs to prospectively set rates.

³Fiscal intermediaries are those agents that have been selected by each institutional class of provider to act as the link between the provider and HCFA.

statistical data from the adjudicated claims files necessary to comply with Federal statistical reporting requirements. The current requirements include monthly (HCFA-120) and annual (HCFA-2082) reporting forms. Both forms collect data on recipients and on expenditures by type of recipient and by type of medical services. The monthly form also collects numbers of eligibles, claim processing information and premium payments. The annual form collects demographic information and units of service data for long-term care and selected medical providers. The data are submitted by the State in hard copy.

The Medicaid Financial Data Set

This data set consists of information submitted to HCFA for budget and Federal reimbursement purposes. The OA-25 is a budget form submitted and updated quarterly showing projected Medicaid recipients and expenditures for the next three fiscal years. This information is used to compile and update HCFA's Medicaid Budget submission. The HCFA-64 is a quarterly form showing State Medicaid expenditures in various categories of medical services and administration. This information, which is auditable, is used to compute the Federal share of the Medicaid expenditures based on individual State's matching rates and various rates of administrative services performed.

Availability

Only data from the MMDS are routinely made available. A computer tape containing major MMDS data elements over the years 1973 to 1981 is under development and may be available by the end of 1983. Some of the MMDS data are routinely made available in two HCFA publications: the National Annual Medicaid Statistical Report and the Medicare and Medicaid Data Book. The Medicaid financial data set is used primarily for internal management and budget purposes rather than for analytic purposes. Unpublished data from both these sets are available from HCFA. The States have considerable flexibility in establishing their programs. Summary characteristics of each State's Medicaid program are also given in the Medicare and Medicaid Data Book.

The Current Medicare Survey

Content

The Current Medicare Survey (CMS), which was conducted from 1967 through 1977, represented all SMI enrollees in 50 states and the District of Columbia (Moskowitz and Mitchell, 1978). The survey was initiated to supplement the information in the Medicare Statistical System. Its purpose was to collect information on the use of uncovered services and out-of-pocket expenditures for all services. The sample persons were selected for interviews in October of each year and remained in the survey for 15 months. Data were collected by means of monthly personal

interviews. The basic items of information obtained included: name of respondent, date of physician visits, conditions treated, number and kind of prescriptions filled or refilled, the total amount of the bill, and the source of payment. Where no information on charges was available, an imputation method was used that assumed that charges would be the same for similar services rendered in the same area. Information about the characteristics of the sample person, such as age, marital status, and living arrangements, was also collected. Additional information related to the Supplementary Medical Insurance program was obtained on an *ad hoc* basis.

Availability

The CMS data show utilization, charges, and estimated Medicare reimbursement by enrollee demographic characteristics. Geographic tabulations are reliable to the census region level. Published reports of survey findings are available in the Social Security Bulletin for years through 1977. Unanalyzed tabulations are available for years 1975 through 1977. The documentation on survey data for 1977 is available on tape.

National Medical Care Utilization and Expenditure Survey (NMCUES)

Content

NMCUES data reflect the health care experiences of the civilian noninstitutionalized population of the United States during 1980. NMCUES was co-sponsored by HCFA and the National Center for Health Statistics (NCHS) and is the seventh survey of national health care utilization and expenditures that has been conducted since 1953. The most recent survey before NMCUES, was conducted in 1977. NMCUES was designed with special emphasis on the experience of Medicare and Medicaid beneficiaries.

The NMCUES data are obtained from three sources:

- A randomly selected national household sample panel of the civilian noninstitutionalized population.
- A randomly selected four-state Medicaid household sample panel of the civilian noninstitutionalized population (California, Michigan, Texas and New York).
- Medicare and Medicaid administrative records.

Approximately 17,900 people were included in the national household sample panel. The four-state Medicaid household sample panel consisted of approximately 13,700 people. Generally, a single family member supplied the information for all members of the household. A series of five interviews spaced about three months apart was conducted at each household in order to obtain complete utilization, expenditures, morbidity and other data for the calendar year 1980. Additional data were abstracted from HCFA administrative records for people who had Medicare or Medicaid coverage. Survey data on beneficiary utilization expenditures and characteristics eventually will be merged with eligibility and reimbursement data from Medicare and Medicaid administrative records.

A core questionnaire was employed in each interview. This document contained batteries of questions about medical care utilization, expenditures, source of payments, health insurance coverage and employment. In addition, questionnaire supplements were used in the first, third and fifth rounds of interviews. The supplement for the first round contained questions about demographic and social characteristics, limitations in activity and family income. The round three supplement asked questions about access to care. The round five supplement asked detailed questions about employment during 1980, individual income by source, and functional limitations.

Medicare and Medicaid administrative record information collected related primarily to program eligibility and utilization and expenditures. The latter is particularly important because survey respondents seriously underreport Medicaid expenditures made on their behalf.

Availability

Three data files will be produced. The first contains survey data only for the first six months of 1980. This data file, completed in early 1981, was created primarily in order to test file development procedures and to support initial data analysis. No public-use file will be constructed from this data base. A data base containing the full year of survey data was completed by early 1982. This data base, which does not include administrative record information, should be available from NCHS in a public-use format in September 1983. A final "best estimate" data base will merge administrative records data with the survey data to produce a composite picture of coverage, service use and expenditure patterns for Medicare and Medicaid enrollees. These data files will be completed in mid 1983.

Discussion: The Relative Strengths and Weaknesses of the Major HCFA Data Bases From a User's Perspective

The Medicare Statistical System, the annual "2082" Medicaid aggregate statistical, and NMCUES are currently HCFA's major statistical data bases. Of these three data bases, the MSS is the most complete and versatile. Medicare statistics have been collected in a uniform fashion across all States since 1966.

Thus, the MSS will support consistent longitudinal and cross-sectional analyses. Because both the location of the enrollee and provider are known, geographic disaggregation is possible down to the zip code level. Linkages across utilization, enrollee and provider files enable the calculation of utilization rates for numerous cohort groups defined by age, race, sex, geographic residence and type of service used. Utilization data are based on date of service as opposed to date of payment. In these ways, the MSS is an extraordinarily flexible data system.

The MSS hospital, skilled nursing facility, outpatient and home health agency utilization files are relatively complete in terms of utilization and reimbursement. Days of care per 1,000 enrollees or visits per 1,000 enrollees are readily

calculated. Diagnostic and surgical procedure information is less readily available and relatively less accurate. At present, this information is routinely developed and used only on the hospital discharges for a 20 percent sample of Medicare beneficiaries (based on enrollee ID number). As case-mix limits are implemented, diagnostic and surgical procedure information will certainly be more accurate. Currently, only the principal diagnosis, an indication of secondary diagnosis, and principal surgical procedures are recorded. As of 1983, up to five diagnosis and three surgical procedures will be recorded for hospital discharges.

Currently, comprehensive physician data are not collected centrally. Such information is maintained by Medicare carriers but it is very difficult to access. While limited utilization and expenditure information is available for physicians, it is not of the same quality as that for other providers.

The MSS is massive in volume, and even components of the MSS are cumbersome to use. Federal resources are required to access the MSS (compounded by privacy considerations), historically preventing widespread access to MSS primary statistics. Distribution of individual records has been limited primarily to other government agencies and HCFA grantees and contractors. As mentioned above, aggregate statistics have been, and will continue to be available in the form of general tables and other forms of statistical reporting.

Medicaid 2082 data are considerably less powerful than MSS data. They contain only aggregate statistics based on recipients so population-based utilization rates cannot be calculated. Utilization data are date-of-payment-based, thus, Medicaid statistics are widely influenced by financial status of States and billing lags.⁴ Longitudinal and cross-sectional comparisons are limited by the fact that not all States have reported all information since the 2082 data base became reflective of the National Medicaid program in 1973. Linkages across providers and recipients are limited and no diagnostic and surgical procedure information is available.

While the Medicaid 2082 data represent limited analytic capacity, the data base is relatively tractable in terms of its magnitude. A computerized version of major components of this data base spanning the years 1975 to 1982 is available to the public. No confidentiality problem is encountered with this data as it does not contain information on individuals.

While the MSS and 2082 data systems are relatively inexpensive to develop and maintain because they are

⁴At times the data reflect "end of payment" fluctuations that are clearly more a function of budget consideration than health care utilization.

based on the bill paying (administrative) process, this linkage does not impose constraints. Administrative systems are typically limited as they do not provide information on:

- Health status of beneficiaries (apart from diagnosis and procedures).
- Income and employment status of beneficiaries.
- Health insurance coverage (aside from Federal entitlement).
- How beneficiaries' "cost-sharing" obligations were actually met.
- Use of and expenditures for noncovered services.
- Non-beneficiaries

This type of information can only be obtained from personal surveys. Thus, NMCUES augments the types of analysis that can be conducted on Medicare and Medicaid beneficiaries and enables comparisons of beneficiary to nonbeneficiary population. While surveys have decided strengths, they are very expensive and complicated, and hence, are infrequently undertaken. Thus, under current plans, NMCUES will be replicated in 1986 or 1987 at the very earliest. While administrative data bases tend to be large, they can be of relatively simple construction. Survey data bases, however, are complex statistically (weights and variances) and contain large volumes of data. Thus, while NMCUES data will be made available, they will be expensive to use and will require great care in actual use.

In summary, access to MSS primary administrative records data is quite restricted due to the magnitude of the Federal staff required to make the data available to the public and due to complex confidentiality problems. Medicaid's 2082 statistics, while less powerful, may be made generally available sometime during 1983. Similarly, the government has committed itself to make NMCUES data available to the public.

References

- Cromwell, Jerry, R. Schurman, and G. Adler. "Uses, Strengths and Weaknesses of Selected Medicaid Data Bases." Working Paper OR-36. March 1982.
- Dobson, Allen, L. Corder, and J. Scharff. *Analysis of the First Six Months of Medicaid Data from the National Medical Care Utilization and Expenditure Survey*. Forthcoming.
- Goldstein, Irving. *The Medicare Data System*. HCFA Pub. No. 03111. July 1981.
- Moskowitz, Edwin and Robert Mitchell. *Current Medicare Survey Report: SMI Utilization and Charges for the Aged—1974*. HCFA Research and Statistics Note No. 1. June 1978.
- Muse, Donald D. and Darwin Sawyer. *The Medicare and Medicaid Data Book, 1981*. HCFA Pub. No. 03128. April 1982.