

Managing programs for the elderly: Design of a social information system

by Howard Birnbaum, Robert Burke, and Frederick Pratter

This paper describes a comprehensive approach to assembling a health care information system to monitor programs for the elderly and disabled in a cost effective manner. The Social Information System (SIS) described in the paper was implemented for the evaluation of the New York State Long-Term Home Health Care Program (LTHHCP).

This evaluation required the collection and organiza-

tion of large amounts of client specific data, including claims, clinical and programatic data. Sources for these data included client medical records, Medicare, Medicaid, and the New York State Food Stamps, Public Assistance, Title XX, and Energy Assistance Programs. Recommendations are made regarding client identification, data elements, access, and structure of the data base.

Introduction

Public policy research on the financing of health programs for the elderly, the sick, and the disabled typically focuses on the costs of Medicaid and Medicare. A current major concern of this research is the high cost of institutionalization because, increasingly, a substantial portion of Medicaid expenditures goes for nursing home care. Although there have been numerous studies of the effectiveness of various home care policy options to reduce the costs of institutionalization (Skellie et al., 1981; Applebaum et al., 1980; Weissert et al., 1980), rarely have studies looked at the combined costs of various policy options across different social support programs and funding sources. Yet options that are cost-effective for one program may increase costs for another. For example, if a Medicaid home care program reduces institutionalization, it also may increase Food Stamps, Energy Assistance, and Supplemental Security Income (SSI) payments, as well as the cost of other programs that provide support to people at home. A home care program could reduce the utilization of nursing home services, but savings in nursing home reimbursement might not be greater than both the extra outlays for home care services and the additional spending on health and social services and SSI payments.

To evaluate the combined costs of various social support programs, it is necessary to develop a comprehensive approach to establishing the cost-effectiveness of these programs for the elderly, the sick, and the disabled. Such approaches to public financing of health care rest fundamentally on the accessibility of large amounts of financial and administrative data. The purpose of this paper is to present a framework for developing such a data base for managing programs for the elderly and to describe the implementation of such a data system.

This research was supported by the Health Care Financing Administration. Contract No. 500-79-0052, Long-Term Home Health Care Program Evaluation. The views expressed here are those of the contractor and not of the Department of Health and Human Services or of the Health Care Financing Administration.

Reprint requests: Howard Birnbaum, Health Data Institute, 7 Wells Ave., Newton, Mass. 02159.

In the past two decades, the emergence of large-scale information management technology has led to discussions about the appropriate role of such systems in the management of health care. Clearly, properly designed health care information systems can be of great value to providers and policymakers, who want to ensure the cost-effectiveness as well as the adequacy of care provided to the United States population. At the same time, the magnitude of the information flow has led to a specialization of information systems that makes it difficult if not impossible, to get an overview of what kind of care is being provided and at what cost.

In general, three types of data are available from health care providers:

- Claims data, such as benefits or entitlements provided,
- Clinical data, such as individual patient assessments or medical histories, and
- Program management data, such as monthly admissions and discharges.

Typically, these data are collected in an uncoordinated manner and generally are available only to one or another of the several potential users of the data. Even the best designed management information systems (for example, the Welfare Management System [WMS] in New York State) collect data on only some of the benefits received by individuals. There is currently no single data source that combines data on health programs from Federal, State, and local sources.

As part of its ongoing evaluation of the New York State Long-Term Home Health Care Program (LTHHCP),¹ Abt Associates Inc., in cooperation with New York State, has designed an information system that joins these three different types of data into a data base, called a Social Information System (SIS). In the process of assembling this data base, much has been

¹The LTHHCP is an innovative program designed to provide home care to chronically ill individuals who otherwise would be in nursing homes. The program is also known as the "Nursing Home Without Walls." In conjunction with a budget cap to contain costs, the LTHHCP features comprehensive assessments of patients' conditions, case management, 24-hour service availability, and a broad offering of medical, psycho-social, and environmental services. See Birnbaum et al. (1983) and New York State Senate Health Committee (1981) for further details on the LTHHCP.

learned about the principles underlying such a system, the methods by which it can be implemented, and some implications of an SIS for improving management of health as well as other social support programs. In this paper the status of the LTHHCP SIS data base is reviewed and the lessons learned are explored.

Need for a social information system

Current discussions about the financing of long-term care have raised issues concerning the expansion of public-funding alternatives to institutional care and in-home services. Public policymakers and program planners need conclusive and generalizable evidence that policy initiatives such as the New York State LTHHCP do in fact save public dollars. At present, the research on the fiscal impacts of alternative approaches to home care has not provided this evidence (U.S. General Accounting Office, 1981).

In inaugurating alternative care programs, State and local decisionmakers must realize that while they are experimenting with new concepts, they are also making choices about what their programs will look like. The effectiveness of the LTHHCP and other in-home service delivery models is not well understood (U.S. General Accounting Office, 1981). Nevertheless, numerous new programs are starting across the Nation, as in Oregon and Missouri, in response to the community-based services provisions, Section 2176, of the Omnibus Budget Reconciliation Act of 1981. All are experimenting with different arrangements of the wide variety of available program options.

The process of local experimentation brings with it the need to evaluate the successes and failures of all innovative alternative care programs. Indeed, the Section 2176 legislation requires that management and evaluation information on the impact of community-based programs be reported annually to the Department of Health and Human Services. Fine tuning, as well as complete overhauling, of these social programs may be necessary. Management and policy improvements will require empirical information of the sort contained in the SIS described here.

The information retrieval capacity needed for program evaluation is similar to that needed for good ongoing program monitoring. Both tasks require collection of large amounts of comprehensive data concerning program activities at a client-specific level. A recent administrative directive (New York Department of Social Services, 1982) illustrates these needs:

"In order to assure that the Long-Term Home Health Care Programs are providing a quality, cost-efficient alternative to institutionalization, the New York State Department of Social Services and Department of Health require accurate information on the LTHHCP patient's characteristics, lengths of stay, services provision, monthly expenditures, third-party reimbursement, etc. To date, none of the existing reporting (or MMIS claim) requirements provide the necessary information."

Aware of this need, New York State is developing a program management information system for the

LTHHCP. The LTHHCP's new information system, when implemented, will improve on the existing situation, but it will not be integrated to take full advantage of New York's extensive WMS. The New York WMS collects and organizes data on the full range of New York's social support programs into a readily manipulable data base; the data in the WMS include Income Maintenance (such as Food Stamps and Public Assistance) and Medical Assistance (Medicaid), as well as Title XX and Child Support Enforcement.

The prototype SIS described in this paper was developed to collect data needed for evaluating the LTHHCP. The SIS is also relevant for other community long-term care programs where information on both health and other social support programs is required. Although the data already exist independent of the SIS, they are too disparate to be comparable. As a result, the data have never been used for the purposes described here or to their full potential. The reason is that data bases are invariably set up for one or another specific reason under a variety of severe time and resource constraints. There is no general framework or perspective for integrating a data set across social support programs. The problems in building the SIS described here are specific to New York; however, similar ones would be present elsewhere.

Three types of data useful for an SIS already are gathered by various health care providers, as noted earlier: claims data, clinical data, and program management data. However, there are disparate purposes for collecting each kind of data. A number of problems, illustrated by the needs of the New York LTHHCP evaluation, occur as a result of these disparate purposes.

First, since the LTHHCP relies heavily on pre-existing systems and forms, there is no uniform set of basic information on patients comparable to the long-term health care minimum data set suggested by the National Committee on Vital and Health Statistics (U.S. Department of Health and Human Services, 1980).

A second problem is that the patient information collected by the LTHHCP program is not integrated into the Medicaid claims payment data system. While New York's Medicaid Management Information System contains utilization and expenditure data, such data are not easily linked to other patient records. It is desirable to link these data to determine the utilization and cost of all Medicaid services, including non-LTHHCP-provided care. Non-LTHHCP-provided care may cost more than LTHHCP care or vice versa. The LTHHCP sites maintain large amounts of data to document the service needs of patients, but there is little readily available data on LTHHCP expenditures.

The third problem is that claims data are inappropriately organized for other than fiscal and administrative purposes. Claims data, while containing useful information, are organized so that claims for authorized services for eligible patients can be scrutinized prior to payment. As a result, the data are cumbersome to manipulate and must be aggregated by case for analysis.

The fourth problem is lack of linkages and common patient identifiers between the various social State support programs. Data forms used to monitor and manage

programs often collect summary data, but that data may not be linkable to other data at a patient-specific level. Much of the necessary identification data can be found in the WMS.

The fifth problem is the lack of analytically oriented evaluation data. Program monitoring and management information systems typically collect data about resource allocations. This collection of items is not conducive to analytical investigation of causal patterns.

In addition to these problems, a major activity for the evaluation and for the SIS involved collection of the broad base of social support data necessary to trace the interrelated expenditures by the State government associated with LTHHCP participation. Nonprogrammatic data (such as Medicare, Food Stamps, Public Assistance, SSI, Title XX, and Energy Assistance) are necessary for an evaluation of the LTHHCP, since enrollees may incur governmental expenditures for services not provided by the LTHHCP. For example, if LTHHCP patients are hospitalized more often than if they had been cared for in a nursing home, then participation in the LTHHCP may lead to a net increase in government expenditures. Either Medicare or Medicaid might cover these hospital episodes. To understand the full cost implications of the LTHHCP, therefore, it is necessary to collect data on the Medicaid and Medicare health care expenses of the LTHHCP patients, including services covered as well as not covered by the LTHHCP.

In addition, by remaining at home, LTHHCP patients also may receive social support services other than health, paid for by the State and the Federal Government. Although these costs are not paid for by Medicaid, it is important to account for governmental expenditures on programs such as Public Assistance, Food Stamps, Energy Assistance, and SSI. Governmental budgets involve such program level transfers. It may be the case that the total health care expenditures for home care are less than the expenditures for costs of institutionalized care or vice versa. However, comparisons of "health only" data are unnecessarily limited.

Design and contents of a social information system

The evaluation of New York State's LTHHCP required the development of a comprehensive data base, combining program and case management data with records of claims processed. This data base is termed a Social Information System (SIS) both because its data elements are systematically organized, so as to be easily stored and readily retrievable, and because the data elements encompass many of the characteristics of individuals in their social environment. That is, the data in the SIS measure the individual and the home environment: how the individual functions physically, socially, and emotionally, and how much support is provided to maintain the individual.

The SIS includes all the types of data in the Long-Term Health Care Minimum Data Set (U.S. Department of Health and Human Services, 1980), as well as additional utilization and financing data. Data are available about the evaluation samples' acute and long-term health

care episodes, as well as their involvement in other social support programs. The data base includes a full year of data for the sample of participants enrolled in the LTHHCP (approximately 700 persons) and for a sample of comparison nonparticipants (also approximately 700 persons). The comparison sample was identified by data collectors from among nonparticipants in the program. This sample included those who could have participated by meeting all the LTHHCP eligibility requirements. That is, comparison persons had to be medically eligible for institutionalization; Medicaid eligible; living in an environment suitable for home care; and budgeted for services within a cost cap equal to 75 percent or less of the costs of appropriate nursing home care.

The SIS data include the following:

- Demographic measures;
- Medical information;
- Measures of psycho-social, environmental, and physical functioning;
- Health services utilization and costs (Medicaid and Medicare);
- Food stamp assistance;
- Public assistance;
- Title XX assistance;
- Energy assistance.

The design of the SIS allows for inclusion of SSI data; however, these data have not yet been included. The other omission in the SIS was the lack of comprehensive Social Security data. Since Social Security retirement and disability payments are not affected by institutionalization, these data were not gathered for the evaluation. A comprehensive SIS would include all SSI and Social Security payments.

The difference between the SIS and previously existing sources of information about LTHHCP participants is that most of the data in the SIS, although included in claim or program files, was not organized into a system useful for either management or evaluation purposes. The advantage of the SIS is that it organizes data that was previously stored in agency file drawers and various computers, and arranges it into a system of records that can be readily accessed. Data for the SIS were collected from over 100 Federal, State, county, city, and provider sources; data for a single participant can come from eight to twenty sources. There are two Federal sources, two State sources, and several public sources in each of the ten counties, plus data from numerous providers.

Although the SIS offers great potential for research and program management, an information system of this type also has potential for abuse. In order to ensure data confidentiality, the LTHHCP SIS is managed by a data base administrator whose primary responsibilities include protecting the identity of the persons from whom data is collected (*Federal Register*, May 15, 1981.) This administrator evaluates requests for information and provides tables and other arrangements of SIS data for legitimate research and management goals. In this way, it is possible to ensure the confidentiality of the data. (see "Recommendations" section.)

The SIS described here provides data for the evaluation of the financial impacts and health outcomes of the LTHHCP. It also could support more general research

concerning health care patterns of the elderly and the ill. For example, the data base is sufficiently rich as to allow cross-cutting analyses of the tradeoffs between the array of New York State and Federal programs that support the elderly, the disabled, and chronically ill patients.

As developed and implemented for the evaluation of the LTHHCP, the SIS is designed for operation during a limited time period and without a capacity for routine updating of data. Incorporation of routine updating capacity would allow continual policy management analysis of the LTHHCP. It also would be feasible to adapt the SIS for routine use by New York in monitoring the utilization and expenditure patterns of all recipients of publicly funded social support programs.

Sources of data

The development of a comprehensive data base such as the LTHHCP SIS implies the following: 1) that there are well-defined data sources; 2) that data elements of interest are available for specific time periods; and 3) that data elements may be aggregated into logical analytic units. This section discusses each of these implications and describes the process used to develop the SIS for policy analyses of long-term care issues.

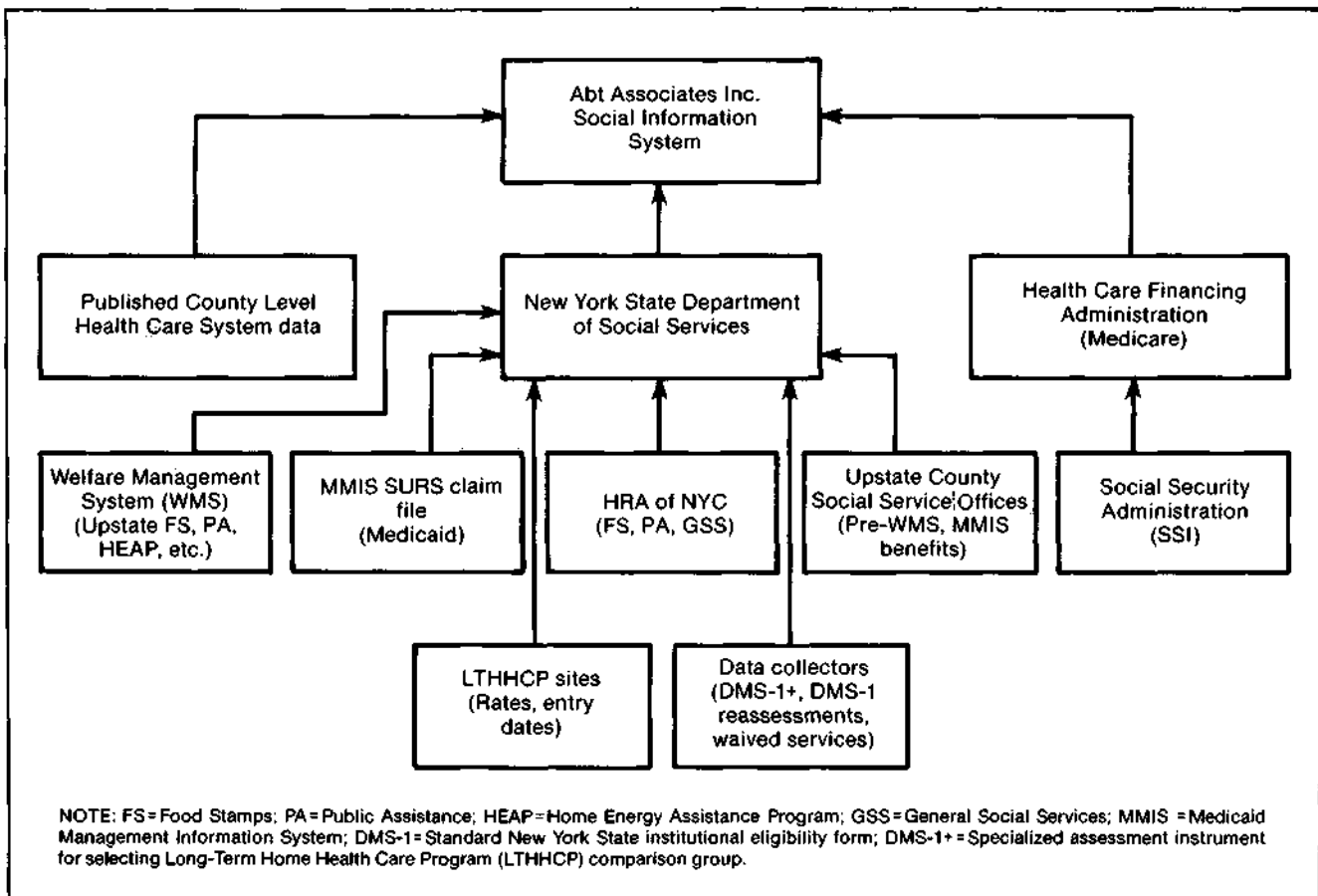
There are approximately 1400 persons in the SIS developed for the LTHHCP evaluation; the sample intake period was between October 1, 1980 and January

15, 1982. Each case was followed for twelve months from date of entry into the study. In the case of program participants, this is the date of initial admission to the LTHHCP; for the nonparticipants, it is the date of screening for study eligibility.

The flow of data from the various primary and secondary sources to the LTHHCP SIS is depicted in Figure 1. The remainder of this section discusses these sources of data.

The maximum potential number of data claims or data forms in the data base as a result of the data-gathering process (summarized in Figure 1) is illustrated in Table 1. The unit of measure is the person/program/month. This corresponds to the multiple measures of one person's entitlement to one social support program for one month. For example, a patient enrolled in Medicaid would have 12 months of Medicaid data. Data on seven support programs will be included: Medicaid, Medicare, SSI, Food Stamps, Energy Assistance, Public Assistance, and Title XX. Other data include program records and patient assessments. In addition, a specialized patient assessment instrument, a long-term placement form entitled the DMS-1+, was developed for selecting comparison patients in the study. The DMS-1+ is based on the standard DMS-1 New York form used to ascertain a patient's medical eligibility for institutional care. The DMS-1+ incorporates patient medical, functional, and mental status data on the New York DMS-1 and also includes information on the Home

Figure 1
Long-Term Home Health Care Program Social Information System data sources



Assessment Abstract used by the LTHHCP to ascertain the appropriateness of the patient's home environment for home care.

Table 1
Number of Social Information System forms, by data source type

Source	Persons eligible ¹	Number of claims or forms
Medicaid	1,372	167,808
Medicare	1,071	25,444
Food Stamp	448	3,601
Public Assistance	49	292
DMS-1 + ²	1,384	1,384
DMS-1 ²	1,384	3,584

¹ Data were requested on 1,384 patients for a 1-year period following study entry date. Eligibility was determined by the existence of a valid case or client number for that data type.

² DMS-1 is a long-term care placement form.

Data for the SIS are of two types: primary and secondary. Primary data include those data sets where data collectors, acting under evaluation and New York State staff direction, have collected data, usually by patient interview. The primary data elements are mainly patient status items. The abstraction of patient data from his/her medical record, in lieu of a direct assessment, is also considered primary. Three major sources of primary data are incorporated into the SIS. The data elements from these sources are listed in Table 2.

Table 2
Social Information System primary data, by source¹

Source	Data elements
LTHHCP data recorded on Patient Intake Form ² or Master Lists ²	Patient's name Site Admission/discharge dates Social security no. Medicare no. Medicaid no. Food stamp Date of assessment Control no. Patient's name Medicare no. Medicaid no. Site no.

See footnotes at end of table.

Table 2—Continued
Social Information System primary data, by source¹

Source	Data elements
Initial DMS-1 and DMS-1 + ²	DMS-1 predictor score Qualifies for LTHHCP Original patient site Date of assessment Date of birth Marital status Race Patient's current residence location Patient's living arrangement: stairs, etc. live alone/with someone Source of income Home/Place where care could be provided Patient and family characteristics Measures of activities of daily living ³ (ADL) Budget Assessment details: Patient transition Sources of information Diagnostic group Date of assessment Date of latest hospital stay Provider name Diagnoses Health and functional status scale scores: Nursing care Skilled nursing needs Functional status Mental status Impairments Skilled therapies required
DMS-1 reassessments Collected two times at 6-month intervals; halfway through the study year and at the end of 1 year from admission.	DMS-1 predictor score Current patient location Patient's name Patient's sex Social Security no. Medicare no. Medicaid no. Date of birth Date latest hospital stay Provider name Diagnoses Health and functional status scale scores: Nursing care Skilled nursing needs Functional status Mental status Impairments Skilled therapies required

¹ Collected by means of patient interview, abstraction, and interpretation of patients' medical records, or the Home Assessment Abstract. Requires clinically-trained data collector.

² These are data collection forms developed especially for this evaluation project.

³ ADL measures include independence in bathing, dressing, eating, toileting, transferring, and so forth.

Secondary data incorporated into the SIS encompass mainly patient utilization and financial data that are collected routinely by various agencies. The secondary data are listed in Table 3 by each social support program available for any sample member.

The SIS was conceptualized during 1980, and planning was continually revised for three years to incorporate New York's changing data systems. Ultimately, there were four general data sources for the SIS information:

- Patient and program medical and other records;
- County systems containing both hard copy and computerized data;
- MMIS; and
- WMS.

The plan for collecting primary data involved extensions of several routine New York State data collection forms used in long term care placement, specifically, their DMS-1 and Home Assessment Abstract (HAA) (Abt Associates Inc., March, 1981).

Table 3
Social Information System
secondary data

Public programs	Data type
Food Stamp (FS)	Eligibility status: amount of dollars authorized to patient's case during study period.
Public Assistance (PA)	Eligibility status: amount of dollars authorized to patient's case during study period.
Title XX (TXX)	Eligibility status: amount of units of service utilized by the patient during study period.
Energy Assistance (HEAP)	Eligibility status: amount of dollars authorized during study period.
Medicaid (MA)	Service utilization and amount of dollars paid to each provider type on behalf of the patient for the study period.
Medicare (MC)	Service utilization and amount of dollars paid to each provider type on behalf of the patient for the study period.
Supplemental Security Income (SSI)	Eligibility and amount of dollars paid directly to the patient during the study period.

The plan for secondary data collection relied heavily on county-supplied Medicaid, Food Stamp, and Public Assistance data (Abt Associates Inc., December, 1981). Until recently, most New York health and social support data were located in local (county) social services offices. These data often were in filing cabinets or on a variety of computers of different types in data files of widely varying levels of accuracy and ease of utility. The gradual adaptation of the MMIS and WMS across the counties in New York has led to standardized statewide data bases.

Tracking patients

The major problem in using the secondary data, aside from accessibility, was that the program and county-level files were not linkable across individuals because there was no coordinated or central source of identification numbers. It was necessary to devise an elaborate crosswalk to link patient data from the various data sources. An important component of the SIS is this extensive crosswalk of patient identification numbers for each of the relevant programs and data sources, which is named ID.CENTRAL. This crosswalk is the key to linking data collected for periods before the development of New York's MMIS and WMS as well as to aggregating data across the various New York social support programs.

During the 1970's, New York began development of a centralized state-wide Medicaid Management Information System (MMIS). Gradually, counties throughout the State entered into agreements with the MMIS. By early 1982, almost every county in the State had joined the central MMIS. Prior to the statewide MMIS period, Medicaid county data were available only in nonstandardized formats, including hard copy and several computer systems.

The WMS includes eligibility and authorization data on the numerous social support programs funded by New York State, including Food Stamp, Public Assistance, Title XX, and Energy Assistance. However, as with MMIS data, there are pre-WMS and post-WMS periods. Post-WMS data are linkable across programs; but pre-WMS data were maintained at a county level, with no uniform or readily known set of patient identifiers. Here, too, in the pre-WMS period, data were available in various modes, including hard copy and computerized data systems of various levels of availability and utility.

Tracking data

A second problem in using the range of SIS secondary data involves its wide dispersal among numerous sources. The majority of secondary data sources is available only from the State (MMIS for Medicaid and WMS for Food Stamp, Title XX, Public Assistance, and Energy Assistance). As increasing numbers of counties became operational on the State data bases during the

evaluation period, it became appropriate to shift emphasis away from county-supplied data to State-supplied data. Thus, reliance on and need for county-level data collection was reduced. Nevertheless, certain data elements remain available only at the county level.

The cooperation of the State and local Departments of Social Services has been crucial in efforts to assemble the data. Elaborate and clearly specified data protocols were implemented for each county and for each data source. The development of the SIS involved extensive negotiations and procurement processes regarding data from numerous local and State social service and other agencies. Any state interested in pursuing the development of an SIS should be prepared for similar major developmental efforts.

A case example of the LTHHCP SIS

To help the reader visualize the complexity and magnitude of the SIS, data for one case, Mrs. Johnson, an LTHHCP program patient in Onondaga County, is presented. Although the data are real, Mrs. Johnson's name is fictitious.

In Table 4 a summary of the data, obtained for Mrs. Johnson from primary and secondary sources is shown. The DMS-1 predictor score and the average dollar amount per month of services appropriate for this case are indicated. The predictor score is used as a summary of primary data. This score is a weighted addition of elements of the DMS-1. Thus, although the DMS-1 collected over a hundred data elements, they can be aggregated into a manageable number of scales; data on only one of these numerous scales are presented in Table 4. The DMS-1 was administered at entrance into the program, and at two additional times, 6 and 12 months. For clarity of exposition, the scores are written next to the month the assessment occurred.

Tradeoffs in funding sources are illustrated in this table. For example, Mrs. Johnson's Food Stamp entitlement amounts remained constant throughout the year. Public Assistance amounts, however, declined at the time her Title XX services began. The shift in benefits from one public assistance program to another could be due to many reasons, including changes in the case composition or in program or eligibility criteria. It is important to study the shift from several perspectives. A researcher, for example, may want to know if this shift resulted in a decrease in the total public dollars utilized by Mrs. Johnson. A program monitor may want to know if the home care program is meeting its objectives. A case manager may want to know if the Title XX program better meets Mrs. Johnson's needs.

Table 5 is a month-by-month breakdown of Mrs. Johnson's Medicaid and Medicare expenditures. Several important problems typically encountered in conducting research with these data are illustrated. The pre-MMIS data were hand-calculated by the county Department of Social Services staff and provided no breakdown by units of utilization for each service. Therefore, estimates of pre-MMIS units of service must be calculated. Fortunately for this SIS construction, only a small portion of the data base was so affected. Furthermore, since Mrs. Johnson has both Medicare and Medicaid coverage, the total costs per services require reconciliation of both Medicare and Medicaid data. In the example, Mrs. Johnson was hospitalized in Month 11 for four days. The first day was covered by Medicaid (\$304); days two, three, and four were covered by Medicare (\$1,225). This hospitalization also resulted in both Medicaid and Medicare paying a proportion of the drug/supply bill and in Medicare paying the entire physician bill. It is only when Medicaid and Medicare data are appropriately included that the correct utilization and dollar amounts can be presented.

Recommendations

Lessons learned about an SIS, based on the LTHHC study, can be translated into recommendations for other data-base developers. These recommendations are grouped into four categories:

- Client Identification;
- Data Elements;
- Access; and
- Structure

Client identification

A unique client identification number should be given to recipients of all State or Federal assistance programs administered by the State. This identification procedure will assure that each individual can be rapidly and correctly identified. Where possible, each separate program should share the common identifier and the identifier should be used for one and only one individual. These restrictions require that the definition of identifier be accepted across all data bases; that is, the identification number must not be comprised of numbers that relate only to the program. For example, Medicaid patient data would only be accessed by the common identification number and not the Medicaid number. The identification would be maintained as a separate file and would function as the major link to other program data files. In addition, this client identification number could be

Table 4
Social Information System summary data for Mrs. Johnson

Month ¹	DMS-1 score predictor	Public Assistance ³					Title XX			
		Medicaid	Medicare	Food Stamp ²	Public Assistance	Special utility	Total P.A.	Energy ⁴	Service type	No. of units
1 January 1981	368	\$ 678		\$ 10	\$ 181.85	\$ 60	\$ 241.85			
2 February 1981		1,548		10	201.55	60	261.55			
2 March 1981		1,847		10	207.95	60	267.95			
4 April 1981		1,662		10	207.00	60	267.00			
5 May 1981		116		10	207.00	60	267.00			
6 June 1981		840		10	207.00	60	267.00			
7 July 1981	404	920		10	230.45	60	290.45			
8 August 1981		918		10	123.15	60	213.15			
9 September 1981		1,568		10	213.15	60	273.15		Counseling	6
10 October 1981		1,666		10	66.09	60	103.09		Counseling	5
11 November 1981		1,046	\$ 1,729	10	66.09	60	111.09	\$155.00	Counseling	1
12 December 1981		81		10	58.09	60	118.09		Counseling	8
13 January 1982	408	—							Counseling	2
Total		\$12,890	\$ 1,729	\$ 120			\$ 2,681.37⁵	\$155.00		

¹The pre-MMIS period corresponds to Jan. 1981–Aug. 1981. The post-MMIS period begins Sept. 1981. The pre-WMS period corresponds to Jan. 1981–Oct. 1981. The post-WMS period begins Nov. 1981.

²Mrs. Johnson's Food Stamp case was reported for herself only.

³Public Assistance: the amounts need to be divided by number of family members in the case (for Mrs. Johnson it is 4).

⁴Home Energy Assistance Program: these dollars distributed once per year.

⁵Total for family = \$2,681.37 ÷ 4 persons = \$670.34 per person per year.

Table 5
Illustrative Medicaid and Medicare data for Mrs Johnson¹

Provider	Months											
	1	2	3	4	5	6	7	8	9	10	11	12
	\$	\$	\$	\$	\$	\$	\$	\$	Units \$	Units \$	Units \$	Units \$
Hospitals												
Medicaid									--	--	1	304
Medicare									--	--	3	1225
Nursing homes									--	--	--	--
Physicians												
Medicaid									1	15	1	8
Medicare									--	--	4	60
Clinics									--	--	1	86
Home health agencies									200	1516	212	1608
Home health agencies/ registered nurses									1	21	1	21
Drugs/supplies												
Medicaid									--	16	--	30
Medicare									--	--	--	140
Transportation									--	--	--	--
Totals	[\$678	\$1568	\$1847	\$1662	\$116	\$840	\$920	\$918]	[\$1568	\$1666	\$2775	\$81]
	Pre-MMIS						Post-MMIS					
	Breakdown of service by month not available											

¹Admitted to LTHHC on January 20, 1981.

embedded into any and all other unique program identifiers by using prefix and suffix routines, although it would clearly be preferable to use a single common identifier for each client. Furthermore, it is desirable that these identification numbers be readily updated so that they are current. That is, additional identifiers should be added when new clients enter a program. When a client is discharged, this fact should be noted in the file and the date entered. Such a procedure would permit corrections to the data set to be carried out quickly.

The structure of the data base can be designed to take maximal advantage of this client linkage. For example, the SIS could contain a synopsis file for each client that would contain his/her name, date of birth, sex, social security number, and an indicator for each data base noting whether the client was program eligible and, if so, the appropriate case number. The synopsis file also might contain the total benefits paid in the previous calendar year for each data type, and the number of months in that year during which the client was program eligible.

Data Elements

The second set of recommendations focuses on the data elements. One major concern is that not all the files in the data base are in the same format. For example, some data bases (such as the Medicaid Claims File) have the individual medical *claim* as the unit. For LTHHCP-specific data the *client* (or *participant*) is the unit; and for Public Assistance and Food Stamps the *case* is the unit. The case contains at least the client and often the members of the household in which the client resides. Thus, merely having the correct data from each data base does not necessarily provide accurate information at the client level. Using Mrs. Johnson as an example, her claims on Medicaid would have to be aggregated to get her monthly Medicaid expenditures. To obtain Mrs. Johnson's average monthly share of public assistance and food stamp dollars, however, the dollars allocated each month to her and her family would have to be divided by the number of members of her household. For analytic purposes, the marginal allocation of Food Stamps for Mrs. Johnson is also relevant, since family allocations are non-linear with respect to household size.

Another problem involves reconciliation of benefits authorized with amounts redeemed. Some programs, Food Stamps for example, record the dollar amounts authorized for the case in one file and the amount of dollars actually redeemed in another file. Even if the authorized file is available, the redemption file may not be. Because of a time lag in receipt of the redemption information, it may prove too difficult to merge it with the authorization amounts in a timely manner. If the correlation between these two figures (authorization versus redemption) is determined and if, over time, it is

demonstrated to approximate a constant, then only one of these files—the authorized amount file—need be used as an integral component of the SIS. It is also necessary to obtain service dates for each record so that the time period covered in Month 1 of Medicaid can be matched with the same Month 1 in each of the other data bases.

Access

There are two access problems: data input and data retrieval. In order to assure the integrity of the data base, safeguards must be created to make sure that only authorized individuals can update or modify existing data. Another set of safeguards is necessary to make sure that only authorized individuals read or retrieve data. Yet another is necessary to prevent accidental alteration or destruction of data items. There are several ways of arriving at these safeguards. First, the data base should have a single administrator who acts as gatekeeper for both entry and retrieval. Second, a security system should be created such that some persons can achieve access only to selected segments of the data base. Third, only an aggregated file should be made available for researchers or others (such as planners and policymakers) who would need to use the data; this aggregated file would not contain client identifiers, ensuring that unauthorized individuals would not have access to confidential client information. This file would contain data aggregated to some specific units such as month and treatment site, but contain no other client identifiers. Fourth, the integrity of the SIS can be maintained when direct approval from the system administrator is needed to write on specific files. With all these safeguards, unauthorized and intentional or unintentional changes in the files are rendered more difficult.

Structure

While numerous analyses of the SIS data are possible, to be feasible they must be possible within reasonable resource constraints. Since it can be very costly to analyze data measuring the micro activities of clients, the data base must be organized parsimoniously. Both computer and staff resources can be minimized if the data base is appropriately set up at the outset. This section reviews the principles that the LTHHCP SIS uses to allow cost-effective data manipulation.

If data for a client covers each of the variables from each of these sources, he or she will be considered to have a "rectangular" data base. The term rectangular is used to mean that for every possible variable or data item for each observation (that is, client), there exists a valid entry or value (or a confirmation that no data apply, for example, that the client is ineligible for Medicare).² If values are applicable but missing, appropriate imputation techniques can be used to provide a valid substitute value. Both visually and in the computer

²In designing such a data base, there is a tradeoff between condensing the data matrix through "sparse matrix" techniques and the consequent storage and retrieval costs. The costs of using data compression techniques are excessive if the data are used on a continuing basis, as they are in the SIS.

the data base looks like Table 6. The resultant rectangular data base would be cumbersome and extremely expensive to manipulate. In the case of the New York SIS, it holds thousands of data items on each of 1,400 individuals.³ This data base encompasses the expenditures incurred for this person over twelve months by social support programs. Altogether, the data base includes over one and a half million data elements.

³Preliminary data currently in house contain, for a single individual, up to 400 Medicaid data elements, up to 52 Title XX elements, 12 Food Stamps elements, 36 Public Assistance elements, 1 Energy Assistance element, as well as Medicare claims, SSI information and patient specific data at three points during the study year.

Table 6
Model rectangular data base

Sample ID	Variable 1	Variable 2	Variable M	Variable N
1	a	d	g	j
2	b	e	h	k
•	•	•	•	•
•	•	•	•	•
•	•	•	•	•
1400	c	f	i	l

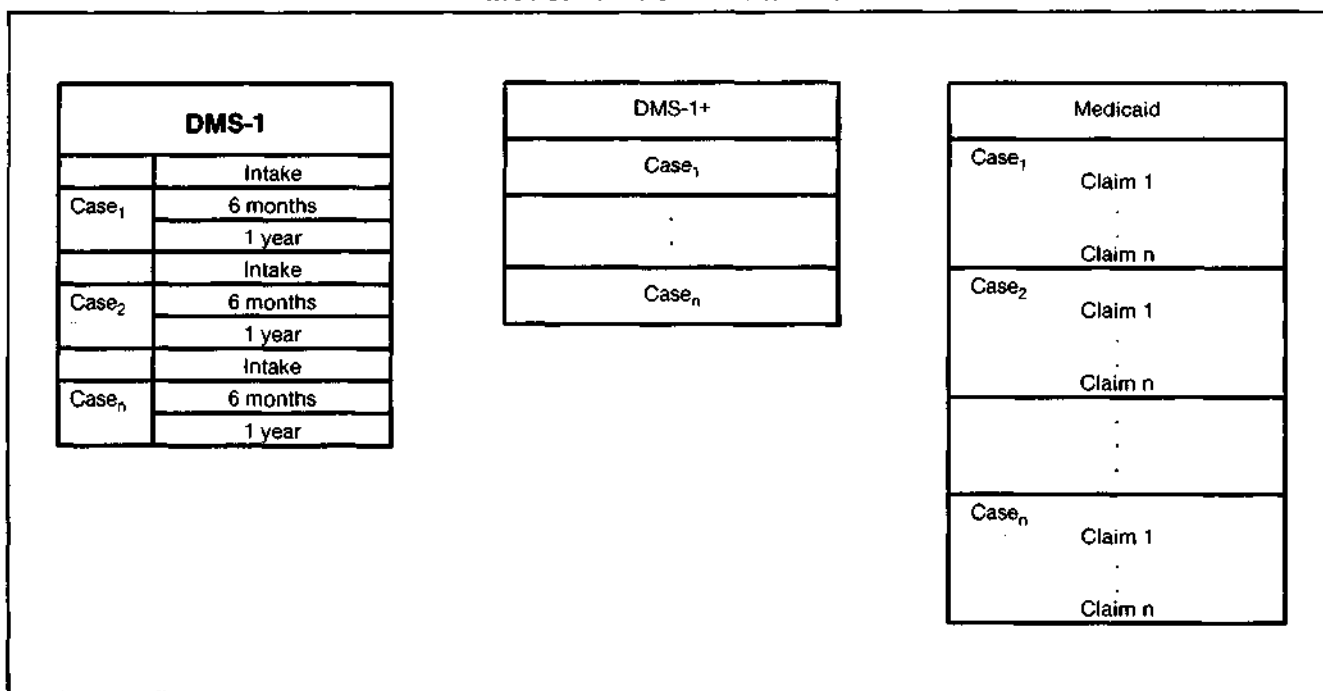
There are two steps to developing a cost-effective SIS. Both were implemented for the SIS built for the LTHHCP Evaluation.

The first step requires the maintenance of separate data files for different data sources, in a "relational" file format.⁴ For example, in this type of structure, program data would be stored at one level of observation (the LTHHCP programs), while other data might be at the patient year, half-year, or month level. Figure 2 illustrates this relational model for three of the sources. For each analysis, the relations could be "joined" to create analytic subfiles that could be rectangular, but the need for storing redundant data at the lowest common level would be eliminated.

The second step is to reduce or aggregate the data to a level that results in a data file that is economically feasible to manipulate. For example, one could ag-

⁴A relational data base allows for maintenance of separate data files for each different data source. This approach provides that each data file retain its own units (e.g., claims, cases, patients).

Figure 2
Model relational data base



gregate Medicaid claims data into monthly or annual amounts for selected carefully defined variables. Or, one could aggregate all hospital expenses and ancillaries into a single "hospital expenditure variable," all physician claims into a "physician variable," and so forth.

The technical requirements of the SIS follow logically from the evaluation design, and are thus specific to that project. However, the general principles of the system design are relevant to any such integrated data system. The data base has been designed from the point of view of analytic utility. At the same time, the realities of scheduling and budget have imposed certain constraints on the system development process.

The key decision, both for cost-effectiveness and utility, was to maintain the data base in a relational format. Data from each of the sources are stored as separate files, linked by a common identifier (study number). Then each of these files is aggregated or adjusted to the level of the individual study participant, while time series information is expressed positionally.

For example, in the Primary Health Status file there is one record per case, continuing the data from the initial health status review (DMS-1+), the 6-month and 12-month follow-ups (DMS-1), and admission and discharge data obtained from the LTHHCP for program participants. The other files in the data base are the Medicaid, Medicare, and auxiliary benefit history files, including Public Assistance, Food Stamp, and so forth.

During the aggregation process several preprocessing steps took place. Data items from each source were combined to form scales, and redundant or contradictory information was edited out. Where this could not be accomplished, the data item was left blank.

The key to combining data from the various sources is the structure of these analytic files. Each file is at the same level of observation (the study participant); each contains the common identifier (study number, referenced through ID.CENTRAL); each is in a fixed record length format and is stored as a sequential file. When file extracts are required for some particular purpose, such as analysis or report generation, these can readily be produced by special purpose extract and merger software developed by Abt Associates Inc., using the COBOL programming language. Reports are generated by using a combination of this software, SPSS (for statistical tabulation and testing) and COBOL report writing software.

None of the currently available data base management system packages⁵ has been used, since these packages are more appropriate to applications where records for an individual need to be accessed on a routine basis. In contrast, the analysis of the SIS files requires a consecutive (rather than a random access) format. The needs of the LTHHCP evaluation require relatively infrequent access to a high percentage of the observations in the file. This type of file access, characterized by low volatility (items are modified infrequently) and high activity (many records are accessed in one run), is best served by a sequential file structure.

In order to access the information stored in the SIS, it was possible to construct index reference files containing

status information and the relative sequence numbers of the observations in the master files. The indices can be small direct access files keyed on the status information. For example, in order to select all the program participants in upstate New York and extract their benefit history information, it is necessary only to join two index files (client status and site code). The common subset of identifiers, each associated with a relative address in the benefit history file, is used to abstract only those records desired, without the necessity of processing the entire data base. This technique is well-suited to processing even very large data bases, as the ease of retrieval of a particular data item is not affected by the total number of records but only by the size of the index files that must be joined. Obviously, it is critical to anticipate the kinds of data request that will be made of the SIS in order to construct the needed indices. This is done at the time the files are created or updated, when the marginal cost of writing the index files is insignificant compared with the processing requirements for accessing every record in the master file.

Data security and confidentiality are inherent in this type of file structure. Since the primary data files (the master records for each data source) contain no identifying information other than the study identifier, obtaining a copy of these files could not compromise the confidentiality of the data. The index files, on the other hand, contain no confidential information. All files, whether primary data or index, are password protected, and can be accessed only with a significant amount of software support.

Updating the files can be done in several ways. Since the primary data files are stored sequentially, periodic updates could be accomplished by passing a transaction file (containing the set of changes, keyed by study identifier) against the master records, creating a new up-to-date master. The older master is retained for back-up. Another way the files could be maintained for routine updates would be to change the addresses in the index files to reference the storage location of the new data. Either approach is well within the limits of current data processing technology, and does not require major software development efforts.

The SIS developed for the evaluation is intended to serve the particular needs of the LTHHCP evaluation. To this end, the SIS has been structured so as to be cost-effective to create and to maintain, to provide data security, and to be useful to the analysts working on the program. The techniques employed, however, are of general utility and would be applicable to a wide range of research and management information reporting tasks.

Summary and conclusions

In summary, the LTHHCP SIS was developed by Abt Associates Inc. under contract to the Health Care Financing Administration, in conjunction with New York State and with the assistance of several of the counties in New York. Its development was made possible by forward thinking New York State planners who established the WMS. It contains data on Medicaid patients living in New York City, and Westchester, Albany, Herkimer, Cattaraugus, Erie, and Onondaga

⁵Such as Total, 1022 or SIR.

Counties. Persons whose data are in the SIS either have been LTHHCP participants or are comparison individuals similar to LTHHCP patients. Data have been collected from both hard copy and computerized sources including both State-wide and county-only systems. Although there are some gaps, generally the data include comprehensive information for previously described data elements for a one-year period between time of entry into the evaluation sample through December 1982.

An SIS capacity is not unique to New York. Similar systems could be developed by other States which possess both automated Medicaid and Welfare Management Systems, such as Michigan. An ongoing, continually updated SIS could allow State officials to track patterns of utilization and expenditure across different social support programs by means of a common set of identifiers. It would allow analyses of the tradeoffs across social support programs for health care, social services, food, energy assistance, and other such programs. Such cross-cutting analyses are necessary to evaluate what happens when legislative mandates change the eligibility for and benefits of these programs. In addition, one support program, at its own officials' discretion, may initiate or continue to offer services formerly provided by another; Title XX and Medicaid have a history of such arrangements.

It will take a major commitment from the State and the Federal Government to devote sufficient resources to continue developing the SIS data system. Moreover, the technical and political problems in implementing unique client identifiers may be extremely difficult or in some cases impossible.

Administrators of health and welfare programs have a different perspective from analysts and evaluators. They may be reluctant to change administrative systems, particularly if they must bear the cost or staff burdens. The task is not just to collect large amounts of data, but also to summarize these data carefully in order to focus effectively on important evaluation and program monitoring issues. This may be one of the most difficult tasks for data system designers. The production of data where units of measurement have the same "format" is extremely complicated, often time-consuming, and requires constant attention to detail. For a variety of reasons, the data systems designers (not the reporting components) may have to accomplish this task. There are limits in creating a data system that extracts component parts from a variety of sources. Because data are gathered for different purposes (workload management, cost accounting, eligibility determination, benefit payments, and so forth), definitions, data elements, reliability, and structure may all vary in the source files. Therefore, files like SIS may not be as "rich" in detail as some of the source files, although they have a much broader scope. For this reason, no one data system can answer all questions or solve all problems.

Despite these drawbacks, it is hoped that government officials will recognize the strength and importance of the data arrangements described here and attempt to implement them. Even if an SIS system is not implemented, the principles discussed here should be useful to architects of claims and eligibility data systems for

social support programs for the elderly, the sick, the disabled, and the needy. At a minimum, administrative program data systems can be made more relevant and useful to public policy research.

Acknowledgments

The cooperation of the following was crucial to assembling the data base described here as part of the Long-Term Home Health Care Program Evaluation: New York State and local Departments of Social Services in New York City and Albany, Cattaraugus, Erie, Herkimer, Onondaga, and Westchester Counties. This paper would not have been possible without the efforts of William Mossey, Richard Nussbaum, Dan DeSisto, and Shelley Goldman (New York State Department of Social Services), Margaret Sager (New York State Department of Health), Kathy Ellingson and Leslie Saber (Health Care Financing Administration, Office of Demonstrations and Evaluations), and Elizabeth Axelrod, Mark Reichin, Saul Franklin, and Sally Stearns (all of Abt Associates Inc.).

References

- Abt Associates Inc.: Long-Term Home Health Care Program Evaluation Data Collection Manual. Mimeographed. Cambridge, Mass. Abt Associates Inc., Mar. 1981.
- Abt Associates Inc. Long-Term Home Health Care Program Evaluation Secondary Data Plan. Mimeographed. Cambridge, Mass. Abt Associates Inc., Dec. 1981.
- Applebaum, R., Seidle, R. W., and Austin, C. D.: The Wisconsin community care organization: Preliminary findings from the Milwaukee experiment. *The Gerontologist* 20 (1980).
- Birnbaum H., Swearingen C., Dunlop, B., Burke, R.: Case Study: New York Long Term Home Health Care Program. Cambridge, Mass. Abt Associates Inc., Mar. 1983.
- Federal Register*: Systems Notice for the Long-Term Home Health Care Program Evaluation. No. 09-70-0025. Washington, D.C., May 15, 1981.
- New York State Department of Social Services: Draft Administrative Directive, Appendix VI, p.1, 1982.
- New York State Senate Health Committee: *Nursing Home Without Walls Program*. Albany, New York, August 1981.
- Skellie, F. A., Mobley, G. M., and Coan, R. E.: Cost-Effectiveness of Community-Based Long-Term Care: Current Findings of Georgia's Alternative Health Services Project. Atlanta, Ga. Georgia Department of Medical Assistance, 1981.
- U.S. Department of Health and Human Services, Public Health Service: *Long-Term Health Care Minimum Data Set*. DHHS Publication No. (PHS) 80-1158. August 1980.
- U.S. General Accounting Office. Improved Knowledge Base Would Be Helpful in Reaching Policy Decisions on Providing Long-Term, In-Home Services for the Elderly. Pub. No. HRD-82-4. October 26, 1981.
- Weissert, W. G., Wan, T. H., and Livieratos, B.: Effects and Costs of Day Care and Homemaker Services for the Chronically Ill. NCHSR Research Report Series. DHHS Publication No. (PHS) 79-3258. February 1980.