

**THE NEW POVERTY MEASURE
ADMINISTRATIVE DATA AS A SOURCE OF MEDICAL EXPENSES¹**

by
Pat Doyle
U.S. Bureau of the Census
and
Meg Johantgen
HSS, Inc.

ABSTRACT In this paper we explore the potential of future health services administrative data systems to improve the measurement of poverty. We first discuss the current and proposed new methods of measuring poverty, focusing on the need and the difficulty in capturing out-of-pocket medical (OOP) costs in the context of a survey focused on economic issues. We then discuss a current data collection effort that may serve as a model for future compilation of administrative data (the Hospital Cost and Utilization Project (HCUP), a federal-state-industry partnership in health care data, directed by the Agency for Health Care Policy and Research (AHCPR)). We conclude with a discussion of how the OOP costs can be integrated into a measure of poverty.

Out-of-pocket medical expenditures cannot be accurately measured with a few items in a survey otherwise focussed on income, poverty, and program participation and thus must be imputed from an external source. Two sources are considered: special purpose surveys and general purpose administrative systems. Special purpose surveys offer the most promise in the short term but the quality of the imputed data suffers due to a highly skewed distribution of OOP costs. General purpose administrative systems, if they can be harnessed, offer the highest quality measure over the long term. The success of HCUP gives us hope that these data can eventually be harnessed thus we recommend a two pronged approach: statistical link to special purpose survey in the short term and development of and direct link to administrative systems over the long term.

KEYWORDS: Administrative data, poverty, out-of-pocket medical expenses

I. INTRODUCTION

Edmonston and Schultze (1995) note: "A major resource, both potential and realized in the development and production of small area estimates is the availability of the vast diversity of administrative records in the United States, both at all levels of government and for all categories of economic and social activities." Furthermore, these records can potentially support analyses of rare events, hard to recall or report information such as medical expenditures, and detailed attributes of events such as complications and comorbidities observed during a hospital visit. With such a rich resource of data in this country, why isn't the social science research community relying heavily on these data for socio-economic research and the construction of key social indicators such as the poverty measure?

In this paper, we explore the potential of administrative record systems for analysis, explore the strides made in the one administrative data project in overcoming the difficulties in collection and use of administrative data, and propose the use of administrative data to improve the measurement of poverty.

A. Background

Generally speaking, administrative data refer to information gathered in the administration of a program or the provision of a service. They are comprehensive in that they reflect the attributes of all persons employed or otherwise affected by the operation, and they are specific in that they contain records of activities that monitor, manage, or facilitate that operation.

The vast array of administrative data collected in the United States share a number of limitations. First, administrative data are highly focussed, capturing only the information needed to administer the program or business to which they pertain. Second, the data systems are locally designed and operated and therefore vary in content and structure across administrative offices. This diversity leads to inconsistencies in units of observation and attributes of those units across local data systems, even across systems with common goals and objectives. Third, as a result of the uneven levels of expertise and available technology in the administrative offices, the local systems are of uneven quality and applicability. Finally, but most important, access to these data is quite often restricted and the nature of the restrictions varies by jurisdiction.

Historically, these limitations have constrained the use of these rich sources of information for the analysis of program policy, health services research, and the construction of key social indicators such as the poverty measure. Notable exceptions to this lack of use include the following:

- The use of IRS tax records and social security earnings records, together with survey data on income and family composition, to study issues of economic well-being. (For examples, see Iams and Sandell, 1996, and Nelson, 1993.)
- The use of quality control data on the Food Stamp Program to analyze characteristics of program participants and to simulate the distributional affects of changes in program benefit formulas. (See, for example, Smolkin, 1994, and Heiser, 1995.)

- The use of Medicare claims data to monitor quality, to detect fraud and abuse in the Medicare program, and as data for health services research.
- The use of program records from subsidized housing programs together with survey data on housing units to analyze characteristics of the housing stock and program participants (Casey, 1992).

Recently a new federal-state-industry partnership produced a data system that further exploits locally administered data systems by integrating the information across systems, placing them in uniform format, and making them available for research. This project, the Healthcare Cost and Utilization Project (HCUP-3), demonstrates the potential of locally administered systems of hospital discharge records as a resource for analysis of a wide variety of health issues including: variations in medical practice, diffusion of medical technology, effectiveness of medical treatments, hospital financial distress, utilization by special populations, and quality of health services.

B. Overview

This paper discusses the potential uses and limitations of a health services data system compiled from locally designed and developed administrative data bases, based on our experience with the HCUP project. We further explore the potential of a system like HCUP (but more comprehensive) in the development of a key social indicator, poverty status. In particular, we focus on the potential of administrative health expense data systems to provide one of the many data elements needed to implement the new definition of poverty in the United States recently recommended by the National Academy of Sciences (Citro and Michael, 1995). This data element is out-of-pocket (OOP) medical expenses, defined to be household payments for "allowable" medical services that are not ultimately reimbursed by an insurance company or a medical provider². While OOP expenses cannot be directly observed in administrative records of medical service utilization, they can be calculated based on the charges for services used and the flow of funds through medical and insurance providers associated with those services.

The following section, Section II, describes a particular need that can be addressed with administrative data—improving the measurement of poverty. This section describes the measure proposed by the National Academy of Sciences in detail. Section III describes the types of health services administrative data available and highlights the advantages and limitations of such data. The section describes the HCUP project in detail as an example of an effort to compile seemingly diverse data into a common data set. Finally, Section IV discusses our recommendations for capturing medical expenses in the measurement of poverty.

II. MEASURING POVERTY: PROPOSED CHANGES AND PERSISTENT OBSTACLES

In today's world, we have sources of data on income and demographic data adequate to measure poverty in the United States as it is currently defined (subject, of course, to measurement errors). But this current definition is inadequate and the underlying survey data does not support its comprehensive improvement. We need to develop tools to implement a more comprehensive poverty measure. Those tools, in all likelihood will include administrative as well as survey data.

This section describes the current poverty measure and the criticism it has received. The section then discusses the potential for a new poverty measure using new health expenditure data. Finally, this section describes unresolved issues concerning the definition and measurement of OOP expenses.

A. The Current Poverty Measure

The poverty measure is a key social/economic indicator used to describe how many persons in the United States have income or other resources that are not sufficient to meet their basic needs. It is also a key determinant of benefit levels and/or eligibility under 27 different federally-sponsored assistance programs (Citro and Michael, 1995). The poverty measure takes into the account the economies of scale achieved by persons who share living quarters and food purchase and preparation. Thus, while the poverty rate is the percent of persons with inadequate resources, it is determined for each person based on the characteristics of the group of cohabitating relatives (the family).

Poverty status is currently determined annually by the Census Bureau based on retrospective income reported in the March Current Population Survey (CPS) for the preceding calendar year.³ Each person's poverty status is determined at the micro level as a function of family income and family composition. A person is considered poor if he or she resides in a family whose cash income is less than a threshold amount designed to represent the cost of providing for the basic needs of the unit. This threshold has its origins based on an assumption that food purchase represents about 30 percent of the cost of living. It was initially defined as 3 times the amount of money needed to provide an adequate diet as defined by the Economy Food Plan and has since been adjusted annually based on changes in the consumer price index (Citro and Michael, 1995).

This current measure is often criticized as being strictly a measure of the adequacy of cash income received by families relative to an out-of-date standard of living.⁴ Numerous presentations and articles (many summarized in Citro and Michael, 1995) have cited major weaknesses in this approach to determine both the level of resources needed to maintain a basic standard of living and the definition of the resources to include in the measurement.

Of particular note for this paper is a concern that the current measure provides inadequate representation of subsidized health insurance and medical costs currently available to individuals at all income levels. The current poverty measure also does not account for the need of severely ill or injured persons to cover expenditures for required medical services. Finally, it does not account for the benefits received through "free" medical services or health insurance subsidized by the government or the employer.⁵

B. A New Poverty Measure Needs New Health Expenditure Data

In response to the criticism of the current poverty measure, the Joint Economic Committee of Congress initiated a review of the poverty measure which led to a request that the National Academy of Sciences establish a panel of experts to review the procedures currently used for measuring poverty in the United States.⁶ The academy panel issued its findings in 1995 in Citro and Michael (1995). Recognizing the inadequacy of the current measure, the Academy reviewed a number of proposals and recommended changing the way in which the federal government determines poverty status. They state in Recommendation 1.2:

"... Family resources should be defined ... as the sum of money income from all sources together with the value of near-money benefits (e.g., food stamps) that are available to buy goods and services in the budget minus expenses that cannot be used to buy these goods and services. Such expenses include income and payroll taxes, child care and other work-related expenses, child support payments to another household, and out-of-pocket medical care costs, including health insurance premiums." (Citro and Michael, 1995.)

Thus, to construct each family's level of countable resources under the Academy's newly recommended approach, the government needs a micro-level (person) data source measuring income, in-kind benefits, taxes, work-related expenses, child support payments, family composition, inter-family transfers of resources and expenses, direct contributions toward health insurance premiums, and nondiscretionary medical expenditures.⁷

The data needs of such a definition of poverty are extensive. Furthermore, development of the needed data base for implementation of the Academy's recommendation will be difficult and will likely involve the use of administrative data, particularly in the area of medical expenses. The HCUP experience discussed later sheds some light on just how difficult compilation of the data can be, but offers some promise that we will eventually overcome the obstacles preventing the direct use of administrative data.

Aside from the practical difficulties associated with using administrative sources to fill in the gap in measurement of medical expenses, there are several unresolved issues surrounding OOP expenses in the academy's recommendations. The following subsections describe two of these issues: the definition of OOP expenses, and the measurement of OOP expenses.

C. Unresolved Issue: Definition of "OOP Medical Expenses"

OOP Medical Expenses refers to expenses paid for by family members for medical services provided and which were not (yet) reimbursed by an insurance company or medical provider. There are two unresolved issues in operationalizing the recommendation that OOP expenses be deducted from income.⁸

The first issue is whether to deduct all or just some direct expenses on medical services from income since medical services include both needed and elective procedures. The Academy acknowledges that not all OOP costs should be deducted from income in determining poverty because some expenditures are discretionary (they go toward medical services that can be foregone without any serious health consequences). However, the determination of need is difficult in many circumstances. For example, having surgery to alter the nose could be a purely frivolous attempt to change someone's appearance or could be medically necessary to correct a defect affecting the flow of air through the nasal passages. Furthermore, some medical conditions (like the common cold) will cure themselves without intervention but persons with colds will contact

medical providers anyway, perhaps to rule out more serious conditions. Because of the lack of clarity, the Academy was not optimistic that a determination of need could be made based on available data and thus did not explicitly recommend that the distinction between discretionary and nondiscretionary services be made.

We assume, however, that a proposal to ignore the issue will not be long-lived and that eventually a distinction needs to be made. Thus, prior to implementation of the new poverty definition, some agency will need to specify how to distinguish discretionary services from other services. This is likely to be a definition based on either 1) a discrete list of services that are discretionary; or 2) a declaration of necessity of services from the physician for each service provided or 3) self-reported necessity of services provided. Once defined, of course, the government will need to capture the pertinent information (service type or declaration for each service) and to measure it at the individual level. This will likely best be handled through administrative sources.

Second, estimates of direct expenditures for medical services and the adequacy of medical benefits vary depending on whether you take a retrospective or a prospective view. Furthermore, as discussed subsequently, our ability to capture OOP costs in a timely manner depends on whether it is defined as a retrospective measure or a prospective measure. An example of retrospective measure is: Did you have any OOP medical expenses last year and, if so, how much? An example prospective measure is: Do you have adequate insurance or resources to cover expenses for catastrophic medical care, should such care be necessary?

The current income-based poverty measure is retrospective—it answers the question of whether you were poor last year or what income you needed during a specific period to meet basic needs. Many of the proposed approaches to measuring poverty, particularly the so-called two-tiered methods, are retrospective except in their treatment of medical benefits.⁹ The Academy's poverty definition is basically retrospective except it is not clear what is intended for the measure of OOP costs. Is it OOP actually incurred (independent of expected future reimbursements); or is it ultimate out of pocket costs incurred for services rendered during the measurement period?¹⁰

D. Unresolved Issue: Measurement of OOP Medical Costs

The data set needed to measure poverty under the proposed definition does not now exist. The Survey of Income and Program Participation (SIPP) comes the closest to providing all the needed input data but even it has deficiencies.¹¹ For example, there is limited information on medical expenses and the information there is targeted to out-of-pocket medical expenses actually incurred last month as recalled by the household after little probing. There is a SIPP module devoted to taxes but unfortunately, the response rate on quantitative variables (like deductions and tax liability) is so low, that the Census Bureau has been forced to issue the data only on a research basis.

Practically speaking, we do not expect a household survey of reasonable cost to provide unbiased estimates of all the components of the recommended poverty measure. It is, however, reasonable to expect a survey like SIPP to provide most of the data elements, with the remaining ones to be supplemented by linkages to information gathered from other sources. In fact, the Academy recommends SIPP as the primary data source for its recommended poverty measure recognizing the need to merge OOP medical expenses from external sources. In the case of medical expenses, the Academy assumed SIPP would be statistically linked to a health expenditure and utilization survey to impute the missing data.

However, before the government can proceed with this link there are several major decisions with the measurement of OOP medical costs to be addressed.

D.1 Measurement Issues Vary by Definition of OOP

The data collection strategy, cost, and complexity vary depending on which definition of OOP costs the government chooses to implement for poverty measurement. Clearly, if the government were to focus on the actual out-of-pocket costs incurred during the measurement period (independent of expected reimbursement or obligations), the information could conceivably be collected directly from the individual in conjunction with other determinants of poverty. This would reduce the complexity of the data base development task.

If, on the other hand, the focus was to be on the ultimate out-of-pocket costs for services rendered during the measurement period, the government could not reliably collect this information directly from the individual. This is more than just a problem of insufficient recall that might be corrected with the use of bracketed response categories. The household often lacks sufficient knowledge of their ultimate obligation due to 1) a lag in resolution of source and amount of payment under fee-for-service-type health insurance plans; and 2) an ill-defined measure of allowable expenditures (necessary versus elective procedures and associated costs, or acceptable charges for necessary procedures). In fact surveys like National Medical Expenditure Survey (NMES) and the Medical Expenditure Panel Surveys (MEPS) base their designs on capturing administrative data for this purpose since it is viewed to be a more reliable source of information on OOP costs.

If, on the third hand, the government were to consider a prospective definition (e.g., determining if an individual has adequate financial protection against a catastrophic illness or injury, should it occur), then the measurement issues resolve around capturing insurance rather than expenses. Insurance, in this case refers to formal private health insurance contracts as well as public insurance, subsidized or free care from providers, and assets that could be converted to (or used as collateral to generate) cash to cover needed costs.¹²

D.2 Measuring Discretionary and Nondiscretionary Costs

The government will eventually need to define discretionary versus nondiscretionary services as discussed above. Unfortunately, households may not recognize the necessity of a recommended medical service, in the context of government policy definition of discretionary spending. If a doctor orders an extra test to confirm a diagnosis or recommends an unnecessary surgical procedure just to be cautious, the patient may not be informed that the procedures are not strictly necessary and thus will have no option to refuse the test or procedure. In fact, there may be some disagreement among medical providers of the need for these services in a particular situation.

Because the medical profession does not typically post its fees for specific services and because the scope of services may not be known in advance, a patient may have little information in advance as to whether the fees for a given medical service are considered reasonable. The exception, of course, is for services covered under prepaid insurance plans where co-payments for services rendered are established in advance. However, the government will need to establish guidelines for "reasonable cost" much in the same way insurance companies have used "reasonable and customary" fees as the upper limit on the level of their reimbursements.

For simplicity, we assume for the remainder of this paper that we want to build a retrospective measure of poverty that excludes from family resources OOP costs for nondiscretionary medical services incurred during a specified period. Under this assumption, the Government faces the following measurement problem:

For a representative sample of persons in the United States, determine money income, in-kind benefits, taxes, work-related expenses, child support payments, family composition, direct premium contributions, and ultimate out-of-pocket costs on nondiscretionary medical services. Furthermore, represent the inter-family transfers of income and benefits as they impact on each of these elements.

III. HEALTH SERVICES ADMINISTRATIVE DATA

Administrative data in health care are plentiful and serve many purposes including: monitoring utilization, determining and charging for the consumption of resources, ascertaining the capacity to supply services, and determining eligibility for benefits or services. Two types of administrative data are particularly useful for health services research: 1) claims data; and 2) encounter data, which may be derived from claims data (AHCPR, 1991).

This section presents an overview of health services administrative databases, particularly claims and encounter level data representing health care provided in hospitals. It discusses the limitations and advantages of these types of data and describes the HCUP project in greater detail—because HCUP represents one attempt to compile data collected at the local and regional level into a uniform data resource.

A. Claims Data

Claims data are gathered and maintained at the level of the patient to report charges and to monitor the use of medical services and resources. Claims data may include the following: demographic information, clinical information, services provided, and payment information.

Unfortunately, these data may be contained in different data systems and, therefore, building a patient-level file or episode of care may be difficult. For example, patient demographic information may be in a file of registration data, hospital or physician payment information may be in a claims-based system, and information about other encounters and ancillary services may be in several stand-alone data systems.

Another concern with claims data is the accuracy and validity of the clinical data. Because claims data are submitted to determine payment, the incentives in the system may be to upcode or target certain diagnoses that increase reimbursement. For example, the coding of complications and comorbidities increased with the introduction of the Medicare prospective payment system, Diagnosis Related Groups (DRGs) (Simborg, 1981). On the other hand, it could be argued that prior to the DRG implementation, there was significant under coding since there was no incentive to code multiple diagnoses.

The most comprehensive claims data are collected by Medicare and health plans. Medicare data have been a rich source of information for the 65+ population and their availability has made it possible to study many health policy issues. Of course, the difficulty with this is the inability to generalize findings to younger populations. Health plans are also becoming a rich data source for information on younger populations, although generalizability is also a problem because plans represent only a subset of the population and selection bias is always a concern.

Medicare, Medicaid, and many health plans are struggling with claims data to try and meet the demand for better measures of utilization and quality. The National Committee on Quality Assurance (NCQA) Report Card Pilot Project examined the implementation of specifically-defined performance measures across a large number of health plans (NCQA, 1994). This project found the availability and quality of the data to be highly variable and made recommendations both to improve the administrative information systems and to develop more sophisticated clinical systems. Problems with plan enrollment data included difficulty identifying true enrollment numbers because members change benefits within plans or employers failed to notify a plan when an employee's status changed. Accurate enrollee identification is also a problem when

one enrollee has two different identification numbers or a family has a single identification number.

Claims data management was also found to vary widely. Clinical coding problems included the use of "home-grown" versions of codes, no requirements for diagnoses or procedure coding before processing, and lack of a quality assurance process to verify the accuracy of codes. Considering that there is little incentive to accurately capture clinical data because they were unrelated to payment determination, missing data are common. Moreover, missing codes may be supplied by abstractors or processing staff with no clinical training. The NCQA study also found that some plans still receive 80 to 100 percent of their claims on paper and there was great potential for error with multiple points of data extraction and entry. Lastly, due to technological constraints, maintaining access to historical data was difficult for some plans.

Despite the problems in integrating diverse data systems, assuring data quality, and the need to represent all people regardless of payer, residence, or services received, claims data are likely to be an increasingly important source of health data for research.

B. Encounter Data

Encounter data are collected to document an interaction with a particular health provider or service and may or may not have a claims component. For this discussion, encounter-level data are those data that are collected for purposes other than reimbursement. These data may be collected by private companies, states, hospital associations, and networks for various purposes. For example, many state data organizations collect data on health services provided in their state. These most often represent inpatient hospitalizations, ambulatory surgery, and emergency care (AHCPR, 1996).

Approximately 42 states collect some data on hospital discharges. Few states include a patient identifier that allows individuals to be tracked across hospitalizations and sites of care. Data collection efforts in most of these systems began in the 1980s when the market-oriented health policies called for collection and dissemination of health information related to cost and quality. Existing data collection efforts expanded from compiling aggregate data to collection of encounter-level data.

Many of these data collection systems developed rapidly without considering data comparability concerns beyond the state, or the potential use of data for other purposes. Yet, there is some consistency in the data representing inpatient hospital care because the data are derived from the Uniform Bill (UB) requirements for submission of Medicare claims. UB-92 requires data elements representing: identifiers (hospital, physician, patient, and insurers); patient demographics (sex, marital status, and birth date); clinical information (diagnoses and procedures, dates of service, and admission/discharge dates); and payment information (payer and charges). This consistency in definition and integration into a single database distinguishes the data from many other administrative data sources.

C. Limitations and Advantages

Although each of these administrative data sources have unique characteristics, common problems with administrative data prevail. Fundamentally, trying to use data for purposes for which it was not intended can lead to problems. For example, you may want detail as to a patient's health coverage, but you may only get an indication of the expected payer in broad categories such as "commercial" or "managed care." Likewise, you may need a measure of an individual's socioeconomic status but resort to expected payer as a proxy measure for lack of better information.

Besides the data not having the information that may be required to answer a particular question, the incentives to maintain accuracy vary considerably. For example, if data are used to determine reimbursement in one state, the incentives to vary coding could be quite different from that in a neighboring state where the data are collected to monitor utilization and trends and little effort is devoted to improving data quality. The increased use of claims and encounter data for measuring utilization and quality has led to improvements in the consistency and quality of the data. Yet, it still varies across hospitals, states, and health plans.

A major advantage of hospital discharge data systems is that all individuals are represented regardless of payer, residence, or type of service provided. The large size of many administrative data systems make them particularly useful for research. Subsamples can be examined and rare occurrences can be studied. This allows researchers to make stronger statements about generalizability. Yet, the comprehensiveness of these systems also make it necessary to implement strict confidentiality and security provisions. Many states have specific restrictions on use of the data in their authorizing legislation while hospital associations may prohibit release all together. Yet, when data are requested for bonafide research and adequate provisions are made to protect confidentiality, many organizations are eager to contribute their data.

D. Example of Data Integration: HCUP

Researchers at the Agency for Health Care Policy and Research (AHCPR) and its predecessor, the National Center for Health Services Research, were charged with exploring patterns of hospital use and cost and analyzing hospital behavior in response to changes in Federal policies and in the structure of the industry. Data sources were developed to support this mission. Hospital discharge data, developed through HCUP, have been the foundation of this effort. A multi-state data base has been created by reconfiguring data for a core set of elements into a uniform format. These data compilations have been possible because: 1) most systems follow some consistent standard (UB); 2) data collection is computerized; and 3) states are eager willing to demonstrate that their data resources are useful.

D.1 HCUP Development

The HCUP data base for years 1970 through 1977, now referred to as HCUP-1, collected all discharge abstract records in a national sample of hospitals. The hospitals had to belong to a discharge abstracting service which restricted the sample. HCUP-2, covering years 1980-1987, was drawn from the universe of short-term non-Federal hospitals and included data from hospitals that process their own data as well as from hospitals participating in discharge abstracting services. However, because the data were collected under special agreements with individual hospitals, HCUP-1 and HCUP-2 have very restricted access. These data bases are described elsewhere (Coffey & Farley, 1988).

As state-based systems grew in the late 1980s, it became feasible to assemble data from individual states. In October, 1992, AHCPR initiated HCUP-3 for the years 1988 through 1994 (SysteMetrics, Inc., 1991). The following are the objectives of HCUP-3:

- To obtain data from statewide information sources, primarily state governments and hospital associations
- To design and develop a multi state health care data base to be used for health services research and

health policy analysis

- To release data to a broad set of public and private users

Unlike HCUP-1 and HCUP-2 which relied heavily on discharge abstracting services, HCUP-3 takes advantage of state and private efforts to collect and edit hospital discharge-level information. It is designed as a federal-state-industry partnership in health care data, under the leadership of AHCPR.

D.2 HCUP-3 Data Source

The data source for the two discharge-level HCUP-3 data bases are existing hospital inpatient discharge data bases maintained by state data agencies, hospital associations, and other private data organizations throughout the U.S. HCUP-3 data are limited to community hospitals, as defined by the AHA. Included among community hospitals are such specialty hospitals as obstetrics-gynecology, short-term rehabilitation, orthopedic, and pediatric hospitals. Not included are long-term hospitals, psychiatric hospitals, and alcoholism and chemical dependency treatment centers.

Written agreements are negotiated with the organizations to assure that data confidentiality and security provisions are upheld. Achieving these agreements may take several months. States may deny release of certain elements all together or may recode them before the data is released. The HCUP implements several measures to comply with state restrictions on release and to enhance confidentiality. Encryption of patient and physician identifiers is completed in the early stages of data processing. Admission and discharge dates are transformed into length of stay and quarter of discharge. Likewise, days of procedure are transformed to days of hospitalization.

HCUP-3 implemented several other measures to assure confidentiality and increase security. File structures were developed such that confidential elements like identifiers and patient zip codes are isolated in separate files and access to these files is restricted. To address state concerns that data might be used for purposes other than research, an HCUP-3 data use agreement was developed.

D.3 HCUP-3 Data Processing

After data are considered complete by the state, data tapes are sent to the HCUP-3 team at AHCPR for processing. Depending on the state's reporting cycle and editing procedures, this can occur 6 to 18 months after the end of the data collection period. Data are reprocessed and translated into a uniform format. This translation is particularly difficult due to the multitudes of ways in which each data system can address the exceptional cases (although not as difficult as it could have been in the absence of the UB-92 standards). For example, one statewide system participating in HCUP created an "other" category for patient gender to classify persons who had sex change operations or to record any ambiguity about the gender of the patient. There is also considerable variation across statewide systems in recording race and ethnicity. For example, three out of the original twelve states in the HCUP project do not record race of the patient and eight of the twelve do not report ethnicity.

Due to varying definitions and coding, it was necessary to use a least common denominator approach in defining these uniform formats. Therefore, the HCUP-3 uniform coding does not represent the optimal coding, but rather what was possible when combining data from multiple sources. For example, the categories of the data element expected payer varied greatly with more than 40 different categories of expected payer defined. Medicare, Medicaid, and commercial insurance are usually distinct, although other

government payers and managed care plans may not be identified at all.

Despite the advantage of using statewide data that has already been compiled and edited, getting the data into a uniform format is expensive and difficult because of the varying definitions, coding, and release restrictions. Moreover, valuable detail is lost because a least common denominator approach must be used. On the other hand, the HCUP-3 has made uniformly-formatted data available for nearly 40% of the hospitals and 50% of the discharges in the United States. The longitudinal nature of the data (1988-1994) also makes them invaluable to study the effects of health policy changes.

As demonstrated by HCUP-3, administrative data are a rich data source and they can be used to examine diverse health care issues for an entire population. Many of the data organizations are continually improving their databases. A consistent definition of what types of hospitals and facilities must report would add to the richness and comparability of the data. Several states are now collecting data from hospital-based ambulatory surgery centers and a few even have emergency room data (AHCPR, 1996). Standard data definitions and coding should also be developed, particularly for outpatient services. As health plans are increasingly pressured to extract comparable information from their administrative data, standardization is likely to improve.

D.4 Availability of HCUP-3 Data

Two discharge-level data bases are available to researchers through HCUP-3:

1. State Inpatient Database (SID). Includes all discharges from all hospitals in participating states.
2. Nationwide Inpatient Sample (NIS). All discharges from a 20 percent sample of U.S. hospitals, drawn from participating states to be representative of the nation.

Both data bases contain the discharge-level clinical and resource-use information included in a typical discharge abstract. These data are often consistent with the UB requirements. A particular advantage of these data bases is the ability to link to hospital-level, county-level, and zip code-level data bases. The potential linkages include the American Hospital Association (AHA) Annual Survey, the Medicare Cost Reports, the Area Resource File (ARF), and census tract or zip-code level data. These linkages expand the types of research questions that can be asked and enhances the value of the data.

Current users of the HCUP data include medical schools, government agencies, consulting firms, managed care organizations, and universities. One of the objectives of the project has certainly been realized—five years of discharge data from a nationally representative sample are available at an amazingly low cost to all users (\$300). To make the data more "user friendly," the data are provided in ASCII format on CD-ROM, with SAS and SPSS tools to help the users.¹³ Data for 1993 will be released soon in a revised format that decreases the cost to the user and improves access to the underlying data.

IV. HOW INDIVIDUAL MEDICAL EXPENSES COULD BE OBTAINED FROM HEALTH SERVICES ADMINISTRATIVE DATA TO IMPROVE THE POVERTY MEASURE

Given the complexities of the health care system, it is unlikely that we can expect to satisfy the data needs of the National Academy of Sciences poverty measure through household data collection alone, regardless of how the exclusion of OOP is eventually defined. Therefore, we anticipate that administrative records will play a role in this process. The questions that need to be addressed are: 1) what kind of role? and 2) what is the potential for satisfying the data needs?

To reiterate from Section II, we concur with the Academy's recommendation that the new poverty measurement be based on SIPP and that SIPP be augmented with medical expenses and direct insurance payments from another source. Two approaches toward supplementing SIPP with medical expenses are a statistical link to a special purpose survey focussed on health expenditures and a direct link to health services administrative data system. In either case, we assume that the cost of collecting medical expenditure data will be borne by the agencies directly concerned with health issues and that the cost to the poverty measurement program will lie predominantly in the linking of the data.

This section first discusses these two approaches further, and then discusses our recommendation. Our preferences in this matter are based on quality of the poverty measure, lag time, and costs to the poverty program.

A. Special Purpose Surveys

The most comprehensive special purpose survey of health expenditures to which SIPP can be statistically linked is MEPS.¹⁴ The objectives of the MEPS are to obtain national annual estimates on health care utilization, expenditures, insurance coverage, and sources of payment for the noninstitutionalized population as well as for policy-relevant subgroups. This is a multifaceted data collection effort with both a household component and a nursing home component. The household component focuses on the civilian noninstitutionalized population consistent with the sample frame of SIPP. The MEPS household component first interviews households (extensively) to collect as much information as is practical on medical utilization and costs as well as to ask permission to contact medical and health insurance providers for more extensive information. Administrative data from a sample of medical providers and the majority of employers and health insurance providers are gathered as part of companion surveys.¹⁵ During post-data-collection data processing, the administrative survey results are directly linked to the MEPS household respondents and analytic measures such as OOP costs are established. To formulate the proposed new poverty measure, SIPP would be statistically matched to public-use data containing these analytic measures.

The statistically linked SIPP/MEPS approach has advantages and disadvantages. One advantage is the improved quality of health related data achieved from the more in-depth focus on health issues than could be allowed in a general purpose survey such as SIPP. An anticipated advantage with the new MEPS is the possibility of continuous expenditure data collection and dissemination.¹⁶ We expect the continuous nature of the new survey to be beneficial in two ways, 1) providing more recent data continuously throughout the decade; and 2) reducing the resources and time required to develop public use data products as the program matures.

The disadvantages of the SIPP/MEPS approach lie in the resources, time lag, and quality of the poverty measure itself. Historically, each time NMES and similar surveys have been fielded, the follow-back surveys were redesigned and tailored to gather information specific to the questions NMES was designed to answer. Furthermore, they were not administered concurrently with the household survey, waiting instead to compile a complete list of providers before fielding the follow-back surveys. This approach was time consuming, resulting in a large delay in availability of analytic files for research and policy analysis. On the bright side, however, this pattern is not likely to reoccur with a new MEPS design¹⁷ and with the continuous fielding of the survey. Finally, the quality of the poverty measure would be affected by the use of imputed data generated from a statistical match and the small sample size of MEPS. Citro and Michael (1995) and Doyle, Beauregard and Lamas (1993) illustrate that the use of imputed (as opposed to actual) data on OOP costs leads to a noticeably different poverty rate because the expenditure distribution is highly skewed. It is difficult to predict the impact of the smaller MEPS sample size on a SIPP-based poverty measure because sample sizes of both surveys are continuously changing in response to budget pressures, priority shifts, and the disproportionate sampling strategies under consideration. However, in the current plans for 1996, SIPP will be fielded to approximately 37,000 households and MEPS will be fielded to approximately 10,500 different households—a reduction of over 70 percent which limits the precision of the estimate of OOP expenses and ultimately the poverty measure itself.

B. General Purpose Administrative Systems

The second approach to capturing OOP costs in poverty measurement is to directly link information on medical costs to the SIPP data base by merging SIPP survey data to an established system of administrative records. Health services administrative systems, if they can be harnessed, can potentially provide a nearly complete audit trail of medical expenses and reimbursements supporting the measurement of ultimate total expenses incurred and distribution by source of payment. Furthermore, the use of actual, rather than imputed, expenses minimizes the impact of the skewed medical expense distribution on poverty.

Operationally, there are a number of difficult issues to be worked out under this proposal, which is its primary disadvantage. First, to date, no comprehensive system of health services administrative data exists. Some data collection efforts exist but they are limited in scope. For example:

- HCUP, discussed previously, focuses on subsets of medical providers
- Medstat's MarketScan data base focusses on subsets of people defined as those affiliated with employers who elect to subscribe to their data base
- Massachusetts Health Data Consortium (1995) has initiated a project to compile comprehensive health care information focussed on a specific geographic area, namely Massachusetts
- Health Care Financing Administration (HCFA)'s administrative data system is a decentralized Medicare claims, validation, and benefits authorization system which focuses on Medicare enrollees and Medicare covered services.

Second, there are issues of confidentiality. While the MEPS/SIPP statistical link can be carried out without

personal information about the respondents or patients, direct linkages to administrative data bases require the use of name, social security number, or other identifying information. However, generally speaking personal information cannot be disclosed across government agencies without prior consent of the subjects (Duncan, Jabine, and deWolf, 1993).

Third, there is a trade off between the time frame for the measurement of expenses and the quality of the OOP measure itself. Health services administrative records will remain open and available as long as the case is active. However, some cases can be active long after the end of the poverty measurement period, particularly if there are complicated insurance arrangements (like dual coverage) or contested bills. Waiting until each case is closed to get the full picture of ultimate OOP costs will delay the production of poverty estimates well beyond the current nine-month time frame.¹⁸ Presumably, an analysis of the time it takes to complete most cases would suggest an alternative approach that maintains the full measure of OOP costs for most, but not all, cases without a considerable lag in the development of the measure itself. Ultimate OOP costs for the outliers can be estimated based on characteristics of the patient, the illness, or the characteristics of any insurance policies.

Despite the admittedly major operational difficulties, there is some precedence for a project such as this. The HCFA plans to use administrative data about Medicare beneficiaries and their Medicare-covered services to enhance and expand survey data collected from Medicare program participants in the Medicare Current Beneficiary Survey and the new Medicare Registry Project.¹⁹ The MEPS survey design contains a provision to use HCFA administrative records about Medicaid and Medicare program participants to validate the information about utilization and costs of services collected directly from the providers of the medical care associated with these participants. MEPS sample cases are given to HCFA, and HCFA extracts all records pertaining to their health utilization and provides the records to AHCPR for integration with the MEPS household survey data.

Furthermore, based on experience with HCUP and on the states' increasing motivation to analyze and monitor all health care costs instead of just hospital costs, we can envision a future where locally administered data from the gamut of health care providers and insurers can be harnessed and fed (virtually if not physically) into a comprehensive system of health care utilization and costs. Such a system would encompass encounter-level data on ambulatory care, pharmaceutical services, and inpatient care fed from providers through the states to a centralized access system sponsored by the federal government. The demand for such a system is clear. For example, the American Health Security Act (the Clinton plan for health care reform), included provisions for uniform administrative health data and a data network for sharing for specific purposes. Indeed, the need to improve the data systems was an area of agreement across most of the health care reform proposals of 1994. Yet even those bills did not provide adequately for statistical uses of the data outside the health care system.

C. Future Possibilities

We recommend a three-pronged approach toward implementation of the proposal to deduct OOP medical expenses from income in calculating poverty. First, resolve the issues surrounding the definition of OOP. This will be facilitated with dialogue established across government agencies. Second, in the short term when there is no comprehensive general purpose health data system to draw on, we recommend the Census Bureau proceed to use MEPS as a source to impute OOP expenses to SIPP. In doing so, we further recommend that the effort include some methodological work to (1) improve consistency between SIPP and MEPS concepts and (2) to determine if the imputation of OOP can be improved to the point where it does

not bias the measure of poverty.

Third, we recommend the Census Bureau and the Department of Health and Human Services jointly pursue the option of developing a general purpose administrative system for health services research and using it to assign medical expenditures to household surveys like SIPP. Such a system needs to be feasible and cost effective while covering all medical and insurance costs incurred by all people, including those in capitated plans. We cannot guarantee its feasibility at this time. However, as we have illustrated in this paper great strides have been made and continue to be made in the collection of administrative data on medical events and their costs.

To advance the compilation of a comprehensive collection of administrative data for purposes such as health services research and the measurement of poverty, several initiatives need to be undertaken. First is the conduct of feasibility study to ascertain the expected state of the art in computing technology and the likelihood the health industry and/or the states will have adequate incentives to refine their administrative systems with a common goal in mind. The Census Bureau cannot, and probably should not, assume sole responsibility for such a feasibility study. Other agencies or foundations that conduct and or sponsor health services research would be better equipped to handle this data collection activity. Furthermore, such agencies would benefit greatly from the availability of an expanded and more comprehensive system of health care administrative data that covers a broad spectrum of the population, encompasses a broad range of health care services, and provides longitudinal person-level information. For example, studies of practice patterns and costs focus on specific diseases and treatments and need the large sample sizes of the administrative data system to capture rare events. Outcomes research could be further advanced with detailed administrative information about treatments during an episode of illness or injury. Finally, since health plans vary in ability to collect and maintain person-level data, the health services research community needs an administrative data set to support analysis of the impact of these controls on plan effectiveness and efficiency.

Second, standards need to be developed in both the definitions and in the coding of data. This is not a trivial matter considering the varied definitions of data elements such as race and ethnicity. Good progress has already been made in standardized information needed to bill for services rendered under the Medicaid and Medicare Programs but more is needed. Standards of clinical data exchange must be developed to ease the burden of creating linkages between disparate data. The American Medical Informatics Association (AMIA) and the American National Standards Institute (ANSI) promote standards to exchange data electronically. ANSI has promoted the use of standards for billing and insurance transmissions that HCFA has adopted.

Third, we need to address confidentiality issues in light of a need to transfer confidential data between agencies. At this time we recommend that SIPP be expanded to obtain signed permission forms from respondents who indicate a willingness to have their medical records extracted. Persons signing the form will have their data extracted from the health services administrative data while persons refusing permission would be treated as nonrespondents, eligible for imputation during post-data-collection processing.

Fourth, the government needs to establish a dialogue between the Census Bureau and the health services research community to accelerate the discussions of issues related to administrative data, its direct use in health services research, and its indirect use in constructing poverty measures and other social indicators. Papers like this one promote an awareness of the potential of administrative data and identify limitations that must be addressed. Continued research and discussion is needed to address these limitations and to continue to explore the mechanisms for compiling these important data.

ACKNOWLEDGMENTS

The authors wish to thank the following people for their thoughtful and constructive comments on draft versions of this paper: Judy Ball, Connie Citro, Rosanna Coffey, Doris Lefkowitz, Dan Walden, and Dan Weinberg. Also, we appreciate all of the clerical and editorial support provided by the staff of Social and Scientific Systems, Inc

ENDNOTES

1. The views expressed in this article are solely those of the authors. No official endorsement by the Bureau of the Census or the U.S. Department of Commerce is implied or should be inferred.

2. As discussed subsequently, “allowable” services have not yet been precisely defined. The notion is that persons would deduct OOP costs from income for services deemed medically necessary but not deduct costs for frivolous services.

3. The CPS is an ongoing nationally representative survey of households and persons in those households.

4. Citro and Michael (1995, table 2-1) illustrate the erosion of the poverty threshold over time relative to the basic assumption that food costs are roughly one-third of the total costs of living.

5. “Free” care, in this case, includes charity care as well as care provided at no cost due to professional courtesy.

6. Funding for the project was provided by the Bureau of Labor Statistics (BLS), the Department of Health and Human Services, and the Food and Nutrition Service of the U.S. Department of Agriculture and technical support was provided by the Bureau of the Census.

7. The issue of discretionary versus nondiscretionary is discussed in Section D.5. Inter-family transfers of resources and expenses is an issue only when a program unit (such as the group of persons covered under Food Stamps) differs from the family unit, particularly when the program unit spans multiple family units.

8. A third issue exists but has been resolved. The term “out-of-pocket” medical expenditures can be ambiguous in the context of poverty measurement regarding the inclusion or exclusion of household direct payments toward health insurance premiums. However, the Academy clarifies this issue by explicitly mentioning that these expenditures be deducted from income along with expenses for medical services provided.

9. Doyle, Beauregard, and Lamas (1993) illustrate the range of options including the two-tiered measure.

10. The analysis of the impact of the poverty measures in Citro and Michael (1995) is based on the ultimate costs but the study references the existing question in SIPP which is based on actual OOP costs last month.

11. SIPP is a recurring longitudinal survey of persons in the civilian noninstitutionalized U.S. population measuring attributes such as monthly household and family composition, monthly income from over 50 different sources, monthly program participation, and monthly labor force and demographic characteristics. The survey encompasses periodic measures of numerous topics including but not limited to asset balances, child care and other work-related expenses, pensions, and taxes.

12. The Academy's recommendations regarding a prospective measure of health insurance are incomplete in that they ignore assets and subsidized care. Their exclusion of assets is based on an assumed annual accounting period for poverty measurement and this makes sense independent of the health insurance issue. However, if the medical aspects of the poverty measure are based on ability to meet future needs for costs of medical services, assets effectively represent part of the health insurance package and should be counted as such.

13. Data from NIS are available through the National Technical Information Service and data from SID are available directly from the states.

14. Starting in 1995, MEPS replaces the series of National Medical Expenditure Surveys (NMES).

15. The medical provider sample includes all hospitals, emergency room and outpatient visits, doctors associated with these visits, and home health care. It also includes a subset of providers of all other services, the subset being defined as providers to a sample of MEPS households and providers to households with at least one Medicaid-eligible person.

16. Traditionally, MEPS' predecessor, NMES, was conducted periodically (every 10 years or so), but MEPS plans call for fielding a smaller sample and introducing a new sample every year.

17. The follow-back surveys will be initiated earlier, beginning with the providers and employers reported in round one of the household component.

18. The information pertaining to poverty status in year x is now collected in March of year $x+1$ and the estimates are made public in the fall of year $x+1$.

19. For more information on the Medicare Current Beneficiary Survey, refer to the Office of the Actuary, Health Care Financing Administration and for more information on the Medicare Registry Project, refer to the Design Contract for the Medicare Beneficiary Health Status Registry, Contract No. 500-95-0038 with Research Triangle Institute, Office of Research and Demonstrations, Health Care Financing Administration.

REFERENCES

- AHCPR. Report to Congress, The Feasibility of Linking Research-related Data Bases to Federal and Non-Federal Medical Administrative Data Bases. (AHCPR Publication No. 91-0003). , Rockville, MD: Public Health Service, 1991.
- AHCPR. Statewide Encounter-level Inpatient and Outpatient Data Collection Activities. Agency for Health Care Policy and Research, Rockville, MD: Public Health Service, 1996.
- Casey, C.H. "Characteristics of HUD-Assisted Renters and Their Units in 1989." HUD-1346-PDR. Washington, D.C.: U.S. Department of Housing and Urban Development, Office of Policy Development and Research, 1992.
- Citro, C.F. and R.T. Michael. eds. *Measuring Poverty: A New Approach*. Washington, D.C.: National Academy Press, 1995.
- Coffey, R. and D. Farley. HCUP-2 Project Overview. (DHHS Pub No. (PHS) 88-3428, Hospital Studies Program Research Note 10.) Rockville, MD: National Center for Health Services Research and Healthcare Technology Assessment, 1988.
- Doyle, P., K. Beauregard, and E. Lamas. "Health Benefits and Poverty: An Analysis based on the National Medical Expenditure Survey." Presented to the annual meeting of the American Public Health Association, 1993.
- Duncan, G.T., T.B. Jabine, and V.A. deWolf, eds. *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*. Washington, D.C.: National Academy Press, 1993.
- Edmonston, B. and C. Schultze (eds.). *Modernizing the U.S. Census*. Washington, DC: National Academy Press, 1995.
- Heiser, N. "Rules of Thumb for producing Food Stamp Program Impact Estimates." Final Report to Food and Nutrition Service, U.S. Department of Agriculture. Washington, DC: Mathematica Policy Research, Inc., 1995.
- Iams, H. M. and S.H. Sandell. "PAST IS PROLOGUE: Simulating Lifetime Social Security Earnings for the Twenty-First Century." Prepared for Presentation at the U.S. Bureau of the Census 1996 Annual Research Conference. March 1996.
- Massachusetts Health Data Consortium. *Health Data News*. September 1995.
- NCQA. NCQA Report Care Pilot Project Technical Report. Washington, D.C.: NCQA, 1994.
- Nelson, C.T. "The Quality of Census Bureau Survey Data among Respondents with High Incomes." Proceedings of the American Statistical Association. Alexandria, VA: American Statistical Association, 1993.
- Simborg, D.W. DRG creep: A new hospital-acquired disease. *New England Journal of Medicine*, 304,

1602-1604, 1981.

Smolkin, S. "Characteristics of Food Stamp Households: Summer, 1993 (Advanced Report)." Alexandria VA: Office of Analysis and Evaluation Food and Nutrition Service, U.S. Department of Agriculture, 1994.

Systemetrics, Inc. AHCPDR Data Base Feasibility Study: Data Sources Evaluation Report. Agency for Health Care Policy and Research, Rockville, MD: Public Health Service., 1991.