

METHODOLOGIC CONSIDERATIONS IN LINKING MULTIPLE DATABASES
FOR THE STUDY OF AIDS EPIDEMIOLOGY

Roselie A. Bright, Marilyn J. Chown, Daniel E. Everitt,
Jerry Gurwitz, and Jerry Avorn,
Harvard Medical School and Beth Israel Hospital
Dr. Bright, 333 Longwood Ave., 3rd floor, Boston, MA 02115

Supported by a cooperative agreement (FD-U-000455-02) with the Food and Drug Administration.

KEY WORDS: Acquired Immunodeficiency Syndrome, zidovudine, Medicaid, Epidemiology, Computerized Databases.

INTRODUCTION

New Jersey Medicaid recipients include many with acquired immunodeficiency syndrome (AIDS) and provide an important opportunity for epidemiologic investigation. In New Jersey, intravenous drug abusers, women, and children form significant segments of the AIDS population which are growing quickly.^{1,2} These groups will probably become the majority of AIDS cases in the near future. The Food and Drug Administration (FDA) is especially interested in the medication use patterns and subsequent survival of AIDS cases. It wants to expand its ability to conduct large-scale observational studies of therapies, since public pressure for rapid approval of medications will not allow the agency to remain solely reliant on traditional clinical trials. The Medicaid system in New Jersey is well-suited to pharmaceuticals research because of an open formulary (that is, any available prescription drug is covered) and complete records of all filled prescriptions.³

Our primary research question is: can all AIDS cases in New Jersey Medicaid be accurately identified from Medicaid claims information? To answer this question, we first needed to identify other available data sources for comparison to Medicaid, short of interviewing the patients themselves. Three were potentially available: original medical records, Uniform Billing hospital discharge data, (a computerized source of data which is available as a commercial service), and the AIDS and ARC (AIDS Related Complex) Registries for New Jersey.⁴

Original Medical Records

The usual gold standard for diagnostic and therapeutic information is original medical records. The hospital discharge summary is written by the primary physician involved with that

hospitalization and may be entered into the record some time after the discharge. It forms the basis for billing, by medical record coders who are employed by the hospital. Because of the expense involved, we were interested in using this source of data as sparingly as possible.

Medicaid

In Medicaid, claims for services are submitted by providers such as physicians, pharmacies, hospitals, and nursing homes (see Figure 1). They are submitted to a fiscal intermediary, which processes them and sends checks back to the providers. All of the claims data are then sent to MMIS, the Medicaid Management Information System, which, in New Jersey, is in the Department of Human Services.^{3,5} Our source of claims data is MMIS in New Jersey.

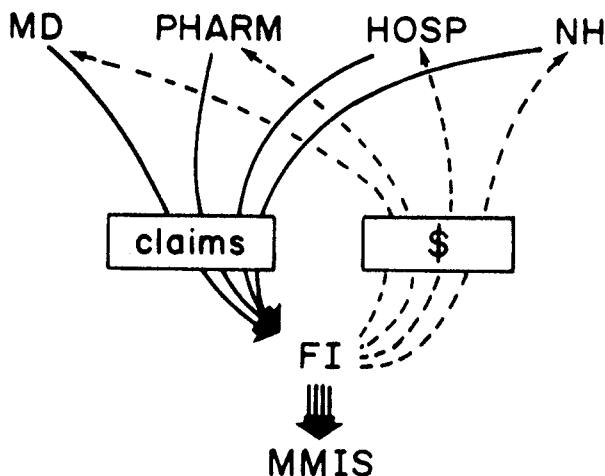


Figure 1. MEDICAID TRANSACTION
PROCESSING

The recipients of Medicaid benefits fall into 3 broad categories: those covered by Aid to Families with Dependent Children, those receiving Old Age Assistance, and the blind and disabled. Women and children with AIDS are most likely to be on Medicaid through the Aid to Families with Dependent Children program, while young men are most likely to qualify through disability.⁶ (See Table 1).

Table 1. Medicaid users and usage on a nationwide basis.^{7,8}

PROGRAM	PEOPLE	SERVICES
Aid to Families with Dependent Children	70%	26%
Old Age Assistance	15%	37%
blind and disabled	13%	35%

Most Medicaid dollars are spent on nursing home care, while the smallest share goes to physician and outpatient service (see Table 2). Claims for prescription drugs are folded into hospital claims, but are separate from nursing home or outpatient care. New Jersey also covers many over-the-counter drugs.

Table 2. Types of Medicaid services and their relative costs.^{7,8}

MEDICAID SERVICE	\$	DRUGS BILLED SEPARATELY
NURSING HOME CARE	44%	YES
ACUTE CARE HOSPITALS	29%	NO
PHYSICIAN AND OUTPATIENT	12%	YES

Information on all claims for service includes: provider type, name and ID number; recipient name and ID number; and the date of service. The data specific to pharmacy claims includes the drug code, days supply, quantity dispensed, date of dispensing, and the prescribing physician. The kind of information that appears on other types of claims includes, as appropriate, the type, place, and date of service,^{3,5} and the diagnosis and procedure codes. Just as important as the claims records are the Medicaid enrollment files, which are crucial to epidemiologic studies of this population. These files include the recipient's ID number, name, date of birth, sex, race, marital status, type of residence (such as nursing home), address, program category, dates eligible for Medicaid, reason for denial of further Medicaid benefits, and coinsurance. For reasons of confidentiality, names and street addresses have not been released to us.

Thus, for a given hospitalization, we have a claim from the hospital, and multiple claims from physicians who provided services. The diagnosis on the hospital claim depends on the discharge summary written by the primary physician. Physicians list diagnoses at their own discretion. It is very important to understand that for any kind of service in Medicaid, whether inpatient or

outpatient, reimbursement depends on procedures performed and medications given, not on diagnoses. Therefore, diagnoses are subject to error.

Uniform Billing Hospital Discharge Data

Uniform Billing Hospital Discharge Data exists for the purpose of formal peer review. Coded abstracts of the discharge reports for all hospital discharges are collected from all of the hospitals. Since this is done commercially, the data tape is expensive.

AIDS and ARC Registries

The last source of comparison data for AIDS cases was the AIDS and ARC Registries. In New Jersey, they are administered by the Department of Health.⁴ For the purpose of disease surveillance, the Centers for Disease Control requires that each state report all AIDS cases on a regular basis, in terms of their numbers and demographics.⁹ New Jersey also has an ARC Registry, run by the same department as the AIDS Registry. To fulfill its mandate, the state remunerates hospitals for each AIDS case reported. The hospital's ascertainment of AIDS cases is based on discharge reports. Each reported case is verified by Department of Health surveillance staff who examine the medical record.¹⁰

Sources of Error

For each data source, there are several sources of error, some specific to AIDS and some more general. The coding of AIDS is difficult largely because it has many manifestations, many of them being opportunistic infections of one type or another. The specificity of ICD (International Classification of Diseases) codes for these formerly rare conditions is very poor. For example, there is no specific code for Kaposi's sarcoma, which, historically, happens to be the disease that first brought AIDS to the attention of the medical community. The ICD codes of choice used by medical abstractors have changed over time, and new codes for AIDS and ARC have been created since the epidemic got underway. Even though there are now codes for AIDS and ARC, they require a complicated history of the number and type of manifestations before they may be officially used. Finally, the physician may wish to protect the patient's privacy by obscuring details of the diagnosis.

Sources of error that are not specific to AIDS cases include those related to the discharge report, such as a lack of precision on the part of the physician, or simple errors made by the coder. As noted, Medicaid does not offer an incentive for diagnostic accuracy on the discharge report. For the AIDS Registry and the Uniform Billing Hospital

Discharge Data, data are generally from in-state treatment. Both Philadelphia and New York City border New Jersey, and many AIDS patients who live nearby get their care in those cities. Although the Pennsylvania and New York registries are required to report New Jersey cases to the New Jersey Registry, we have heard that in practice, this is a low priority task and is not accomplished in a timely fashion.

The relative importance of different errors may vary between data types. The actual error rates may depend on characteristics of the patients, such as age, sex, race, and which AIDS risk factors are present.

METHOD OF INVESTIGATION

Overall Plan

Our overall study plan is to use Medicaid data, the New Jersey AIDS and ARC Registries, and selected medical records. The Uniform Billing Hospital Discharge Data will not be used because it offers no advantages over other data sources. Patient privacy will be strictly protected, as names will never be revealed to the investigators. Only those members of New Jersey's state government who are currently authorized to handle patient-specific AIDS data are permitted to see any of our identified data. In our overall plan, the Department of Health will cross reference the list of all Medicaid enrollees against the AIDS and ARC Registries to create a data table like the one in Table 3.

Random samples will be taken of the groups in each table cell and the hospital medical records will be reviewed by surveillance staff in the New Jersey Department of Health, to see if the patient really has AIDS, ARC, or neither.

Table 3. Data table to be created by our overall study plan (see text).

MEDICAID -----	REGISTRY -----			N
	AIDS	ARC	NEITHER	
AIDS				
ARC				
NEITHER				

Pilot Study

Because of the sensitivity of this project's subject matter, a pilot study is being done. We have provided a limited list of the identification numbers of Medicaid enrollees in 1987 who had an AIDS or ARC diagnosis. We do not have access to names. The list was forwarded through the Department of Health to the Department of Human Services, who added names and returned the list directly to the Department of Health. The Department of Health has linked the list to their registries, which were based on name. When the names haven't matched exactly, they have

examined potential misspellings and used birthdate, sex, race, and county of residence information.

Once the Department of Health is ready to complete this pilot study, they will sample each of the cells in Table 4 marked with an "X" and conduct a medical record review to find out what the AIDS and ARC status of these patients really is. Then they will remove all names from the data file and send it back to us.

We will to analyze the various error rates by each of the following variables: age, sex, race, marital status, zidovudine use, place of residence, and receipt of care out-of-state. If all has gone well by this point, we will be able to study other categories of Medicaid enrollees with the AIDS and ARC Registries.

Table 4. Data table to be created by our pilot study.

MEDICAID -----	REGISTRY -----			N
	AIDS	ARC	NEITHER	
AIDS		X	X	360
ARC			X	85

CONCLUSION

In conclusion, creative solutions can be found when attempting a new type of medical database record linkage system in an environment of extreme political sensitivity. However, one must have more patience than is ordinarily required for working with administrative data.

REFERENCES

1. Berkelman RL, Heyward WL, Stehr-Green JK, et al. Epidemiology of human immunodeficiency virus infection and acquired immunodeficiency syndrome. *Amer J Med* 1989; 86:761-70.
2. Update: acquired immunodeficiency syndrome - United States, 1981-1988. *MMWR* 1989; 38:229-36.
3. Bright RA, Avorn J, Everitt DE. Medicaid data as resource for epidemiologic studies: strengths and limitations. *J Clin Epidemiol* 1989. (in press)
4. Altman R. The epidemiology of AIDS in New Jersey. *NY State J Med* 1988; 88:236-9.
5. Ray W, Griffin M. Use of Medicaid data for pharmacoepidemiology. *Am J Epidemiol.* 1989; 129:837-849.
6. Kopantais G, Powell-Griver E. Characteristics of persons dying from AIDS. *Advance Data.* National Center for Health Statistics. 1989; 173.

7. Health Care Financing Program Statistics: The Medicare and Medicaid Data Book, 1981. Baltimore, MD: Health Care Financing Administration. Office of Research and Demonstrations: 1981.
8. Health Care Financing Program Statistics: Analysis of State Medicaid Program Characteristics, 1984. Baltimore, MD: Health Care Financing Administration. Office of the Actuary; 1985.
9. HIV infection reporting - United States. MMWR 1989; 38:496-99.
10. Department of Health, New Jersey. Personal communication. 1989.