

N-2349/3-HHS

HIE REFERENCE SERIES

Volume 3: USER'S GUIDE TO HIE DATA

C. d'Arc Taylor, S. M. Polich, C. E. Peterson, E. M. Sloss

August 1987

HEALTH INSURANCE EXPERIMENT

THE **RAND**
CORPORATION

The research reported herein was performed pursuant to Grant No. 016B-8001 from the U.S. Department of Health and Human Services, Washington, D.C.

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution. Publications of The RAND Corporation do not necessarily reflect the opinions or policies of the sponsors of RAND research.

Published by The RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138

N-2349/3-HHS

HIE REFERENCE SERIES

Volume 3: USER'S GUIDE TO HIE DATA

C. d'Arc Taylor, S. M. Polich, C. E. Peterson, E. M. Sloss

August 1987

HEALTH INSURANCE EXPERIMENT

THE **RAND**
CORPORATION

PREFACE

This Note provides guidance for understanding and using 67 files of research data released to the public by the Health Insurance Experiment (HIE). Conducted by The RAND Corporation from 1974 to 1982 under a grant from the U.S. Department of Health and Human Services, the HIE was a large-scale, controlled trial of cost sharing in health insurance. It assessed how variations in patients' share of health care costs affected their use of health services, their satisfaction with care, the quality of their care, and the state of their health.

The 67 *public-use* files contain a large subset of all data collected by the HIE. The file data pertain to characteristics of the sample, health care services and expenditures during the experimental period, medical/dental health status, and attitudes toward health care. This guide summarizes the experiment; describes each file and associated documentation; discusses the analytic possibilities and limitations of the data; and offers suggestions for choosing analytic subsamples and linking data across files, years, sample groups, and sites for particular analytic purposes.

This document is essential for analysts and programmers who are considering the use of HIE data. It provides an overview unavailable elsewhere, and it points out several minor errors discovered in previously published file documentation.

SUMMARY

From 1974 to 1982, RAND conducted the Health Insurance Experiment (HIE), a large-scale, controlled trial in health care financing funded by the U.S. Department of Health and Human Services. Its purpose was to assess the effects of different types of insurance on patient health and health care delivery, including both fee-for-service and health maintenance organization modes.

Over 8,200 persons were enrolled in the experiment in six sites: Dayton, Ohio; Seattle, Washington; Franklin County and Fitchburg, Massachusetts; and Georgetown County and Charleston, South Carolina. Each family was assigned to one of a number of insurance plans that varied by coinsurance rate, delivery mode, and maximum out-of-pocket expenditure. Data were collected on enrollees' use of health care services and state of health throughout their term of enrollment, randomly assigned at three or five years. A large subset of the collected data has been released to the public in 67 fully documented research files. This Note describes these *public-use* files and provides guidance for using them.

DATA ORGANIZATION

Nearly all of the public-use research files are grouped in topical series covering sample characteristics, enrollee claims (health care use and expenditures), and medical/dental health. Documentation for each file includes a data tape and hardcopy codebook published as a RAND Note.

ANALYTIC LIMITATIONS

The HIE public-use database is large, complex, and rich. It will support many analyses, both point-in-time and longitudinal. Because of the experimental design, however, the data are not appropriate for the following applications:

- Analyses of certain parts of the population: the elderly (and others eligible for Medicare), those with very high incomes, those in the military (and veterans with service-connected disabilities), and those in jail, long-term hospital, and other institutions.
- Analyses that depend on linking illnesses or health care expenditures with socioeconomic or demographic characteristics *at the exact time* of the illness.
- Analyses of health care use at the family level.
- Exact replication of previously published HIE analyses. HIE analysts used files that were slightly different from the public-use versions.
- Analyses of serious illness.

DETERMINING WHICH HIE FILES TO USE

The size and complexity of the database underscore the importance of studying the documentation--including user's guide and codebooks--before attempting to use HIE files in analysis. For example, all information about the sample is grouped in a separate series of files, so users will need at least two files to analyze HIE data: a sample file and a file of claims or medical/dental health data. Often more than two files are needed for analysis of a single subsample or topic; the documentation must be consulted to determine the necessary minimum.

USING SAMPLE DATA

The three main parts of the HIE sample are *baseline-only participants*, *insured enrollees*, and *adjunct enrollees*. The most extensive data, especially longitudinal data on the use of health services, are available for the 8,254 insured enrollees. The 15,411 baseline-only participants provided much demographic and socioeconomic data as well as information on health status, experience with health care, and health-related attitudes. Limited data were obtained for 2,483 adjunct enrollees. Different files address different parts of the sample, though most files are restricted to insured enrollees.

USING CLAIMS DATA

Claims data consist of detailed *primary data* on diagnoses, procedures, and charges for each health care service or hospital stay; and *derived data*, primary data aggregated by health care visit, person-year, or episode of treatment. These data are taken largely from claims that enrollees submitted, so they reflect *billed* services and expenditures, which may differ from *actual* services and expenditures.

For many analyses, the derived data alone will suffice because they incorporate many of the manipulations analysts normally make to primary data. "Visit" and "annual expenditure" data were designed for analyses of how frequently people go to the doctor. "Episode" data indicate when treatment starts, how long it lasts, and how people adjust their demand for health care in the expectation of satisfying/not satisfying the annual deductible or out-of-pocket expense limit specified in their insurance policy.

Some analyses will benefit from the use of derived data augmented by primary data. The primary data provide details, omitted in the aggregation process, on such matters as treatment history and status, symptom or reason for a visit, medical/surgical/dental procedures, and procedure-diagnosis linkage.

USING DATA ON MEDICAL AND DENTAL HEALTH

The medical and dental health files document enrollees' state of health when entering and completing the experiment, a span of three or five years, depending on the assigned enrollment term.

Like the claims data, medical/dental health data exist in primary and derived forms. The primary data are taken from responses to an extensive medical history questionnaire (MHQ) that enrollees completed. One part asked for relatively subjective evaluations of functional limitations, acute symptoms, mental and social health, general health, health habits, and satisfaction with medical and dental care. The other part elicited relatively objective information on enrollees' physical limitations and the presence of specific chronic or acute medical disorders. Different versions of the questionnaire were designed for adults, children, and infants.

Derived data include transformations of MHQ responses or test data from a medical screening examination. One series of files presents measures of health status (e.g., physical, mental, social) and attitudes toward health care. Another series documents the presence and severity of medical disorders in adults (17 common and potentially serious disorders) and children (4 disorders). A final derived file assesses the presence and effects of tooth decay and periodontal disease.

Most users will begin with the derived data and add detail as necessary from the primary data. As an example, someone undertaking a comprehensive analysis of dental health would draw on derived data pertaining to tooth decay, derived data measuring the enrollee's satisfaction with dental care, and primary data on dental health habits.

ACKNOWLEDGMENTS

The authors are indebted to several colleagues on the Health Insurance Experiment for assistance as this guide was drafted. Joseph P. Newhouse provided guidance and support throughout the effort. Others who reviewed all or part of the draft include Willard Manning, Ellyn Bloomfield, Emmett Keeler, Patricia Camp, Carol Edwards, and Janet Hanley. The review by Jeannette Roskamp offered the valuable perspective of an analyst new to the experiment. Mary Stout provided expert assistance with copy preparation. Final production of the Note was supervised by Patricia Bedrosian.

CONTENTS

PREFACE	iii
SUMMARY	v
ACKNOWLEDGMENTS	ix
FIGURES AND TABLES	xiii

Section

I. INTRODUCTION	1
Experimental Design	1
Selection of Enrollees	2
Experimental Treatments	3
Services Provided	5
Terms of Enrollment	6
Data Collection	8
File Development	8
Organization of This Guide	12
II. ORGANIZATION OF HIE DATA	13
Master Sample Series	14
Claims Line-Item Series	15
Aggregated Claims Series	16
HIE Reference Series	18
Medical History Questionnaire Series	19
Health Status and Attitude Series	21
Medical Disorders Series	21
Dental Examinations File	22
Insurance Preference Files	22
III. ANALYTIC CONSIDERATIONS	24
Sample Definition	24
Analytic Limitations	26
Precise Time Linkages	26
"Actual" versus "Billed" Health Care Use	27
Use of Health Care System by Families	27
Replication of HIE Analyses	27
Serious Illness	27
Determining Which HIE Data Files To Use	28
Characteristics of the Sample	28
Health Care Use and Expenditures	29
Medical/Dental Health	30
Comprehensive Information	31
Caveats	31
Data Comparability	31
Insurance Preference Files	32
Presentation of Data	34

IV. USING SAMPLE DATA	36
The HIE Sample	36
Main Sample Groups	36
Variant Subgroups	37
Data Availability	38
Analytic Subsamples	39
Insured Enrollees	39
Never Insured Participants	46
V. USING CLAIMS DATA	48
General Cautions	50
Linking Claims Data	51
Linking within a File: Total Cost of an FFS Outpatient Visit	51
Linking across Files	52
Choosing Analytic Samples	60
FFS- or HMO-Only Participants	60
Users/Nonusers	60
FFS-HMO Issues	62
FFS Usage by HMO Participants	62
Sector Comparisons	62
Adjusting for Inflation	64
VI. USING DATA ON MEDICAL AND DENTAL HEALTH	65
Data Collection Procedures Affecting Data Use	65
Medical History Questionnaire	65
Screening Examination	68
Using Derived Data on Medical Health	70
Health Status and Attitudes toward Health Care	70
Medical Disorders	71
Common Sample	72
Reasons for Missing Data	73
Using Derived Data on Dental Health	74
Content	74
Sample	75
Using Primary Medical History Data	76
Appendix	
A. HIE Benefit Coverage	79
B. Participation Incentive Payments	82
C. Unreleased HIE Data	85
D. Population Change within HIE Experimental Treatments	89
E. HIE Inflation-Adjustment Factors	93
GLOSSARY	97
REFERENCES	101

FIGURES

1. FFS-HMO Relationship of HIE Experimental Treatments	3
2. HIE Participant Sample	36

TABLES

1. HIE Enrollment Periods	7
2. HIE Data Collection Instruments	9
3. Sample Groups in HIE Files Restricted to Insured Enrollees	40
4. Sample Groups in HIE Files Not Restricted to Insured Enrollees ..	47
5. Summary of Claims Data Files	49
6. Summary of Medical and Dental Health Files	66
7. Number of Records in MHQ Series Files	76
C.1. Table 2 Annotated To Indicate Unreleased Data	86
D.1. Population Change within Experimental Treatments	90
E.1. Overall Inflation-Adjustment Factors	94
E.2. Medical Inflation-Adjustment Factors	95
E.3. Dental Inflation-Adjustment Factors	96

I. INTRODUCTION

This section presents an overview of the Health Insurance Experiment (HIE) and its data collection and file development efforts. It provides essential background for understanding this guide.

EXPERIMENTAL DESIGN

The RAND Corporation conducted the Health Insurance Experiment from 1974 to 1982 in six sites across the United States: Dayton, Ohio; Seattle, Washington; Franklin County and Fitchburg, Massachusetts; and Georgetown County and Charleston, South Carolina.¹ The main purpose of the experiment was to assess how varying patients' cost of health services affected their use of services, their satisfaction with health care, the quality of their care, and the state of their health. A related purpose was to study how those outcomes were affected by the mode of delivery--fee-for-service (FFS) or health maintenance organization (HMO).²

Over the course of the experiment, information of some kind was obtained for 26,148 persons. A total of 24,340 persons were administered a baseline interview (*baseline participants*³), of which 7,699 were ultimately enrolled. Of the remaining 16,641 persons, the 15,411 who did not enroll are called *baseline-only participants*; the other 1,230 are part of the adjunct enrollee group defined below. An additional 555 persons were enrolled later, all but a few of them newborns or adopted children under one year of age. Those 8,254 *insured*

¹The sites were chosen to represent the four census regions of the country and both urban and rural areas. They also differed in the amount of delay to obtain an appointment, reflecting different degrees of stress on the ambulatory medical care system. Site selection is described in Ref. 1.

²For a discussion of the purposes and design of the HIE, see Ref. 2. HIE is also called HIS, Health Insurance Study. The terms are synonymous.

³This and other distinctive HIE terms are defined in the Glossary at the end of this document.

enrollees were assigned to an experimental insurance treatment, and data on their use of health services were collected throughout their period of participation. Another 2,483 uninsured *adjunct enrollees* were not assigned to an insurance treatment but resided with insured enrollees or were members of a short-lived control group in Dayton.⁴ In HIE terminology "insured" means only "assigned to an experimental treatment." "Uninsured" refers only to a participant not so assigned, not necessarily someone lacking health insurance altogether.

Selection of Enrollees

Persons offered enrollment in the experiment represent a random sample from each site, subject to certain eligibility restrictions.⁵ They were chosen by a two-stage baseline selection process. In each site an areawide probability sample of dwelling units was drawn. Their occupants were interviewed for eligibility, and those found eligible were questioned in depth about their socioeconomic characteristics and experience with health care (baseline interview).

Eligibility criteria excluded those whose health care delivery systems differed from options available to the general population. The following groups were excluded:

- Those who were eligible for Medicare or would become so during the experiment, i.e., those 62 years of age and older at enrollment, or younger than 62 but eligible for Medicare because of disability or end-stage renal disease.
- Those with family incomes over \$25,000 (1973 dollars).
- Those institutionalized (jail, long-term hospital).
- Veterans with service-connected disabilities.
- Those on active-duty military service or military retirees and their dependents.⁶

⁴This is a very general description of the sample. A full explanation is provided in Sec. IV.

⁵Subject also to slight oversampling of low-income families in Dayton, Massachusetts, and South Carolina.

⁶Details of HIE eligibility requirements are in Ref. 3, Sec. II.

Project staff verified the accuracy of the information given by baseline participants with employers and insurance companies.

In the second selection stage, HIE staff drew a representative sample of eligible persons to be offered enrollment and assigned each family in that sample to one of the insurance plans described below. A sophisticated technique assured that, across plans, families closely resembled each other in 24 health and socioeconomic characteristics.⁷

Experimental Treatments

Sixteen experimental treatments distinguished between coinsurance rates, delivery systems, and maximum out-of-pocket expenditures. All but one of the treatments were health insurance plans, listed below as A-O. Plans A-N entailed different degrees of cost sharing under the FFS system; those in plan O were experimentally assigned to a Seattle HMO. The remaining treatment (not a plan) provided a control group for those in plan O. Figure 1 depicts how the experimental treatments were divided between FFS and HMO systems.

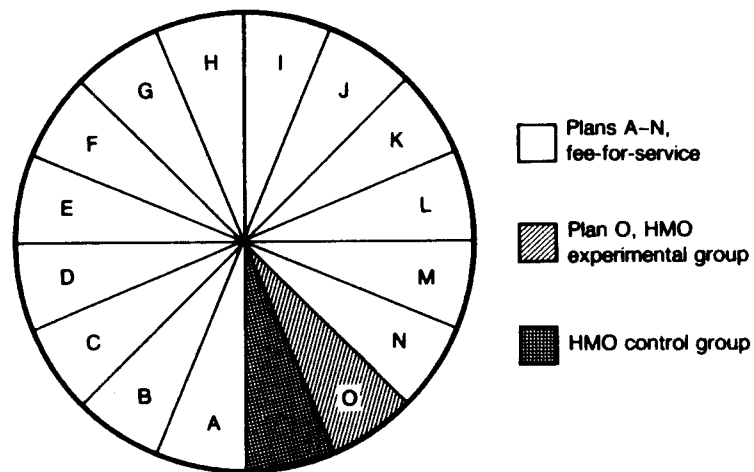


Fig. 1—FFS-HMO relationship of HIE experimental treatments

⁷The logic and technique used to assign individual families to experimental plans are described in Ref. 4.

Insurance Plans. Plans A-N entailed different degrees of cost sharing under the fee-for-service system. Within each cost-sharing group, listed below, plans also differed by the ceiling placed on maximum expenditure. Plan O involved participation in an HMO.

- A. Free care (0% coinsurance) (one plan).
- B-D. Family pays 25% of all health care bills (25% coinsurance) (three plans).
- E-G. 50% coinsurance (three plans).
- H-J. 50% coinsurance for dental and outpatient mental health services and 25% coinsurance for all other services (three plans).
- K-M. 95% coinsurance (three plans).
- N. 95% coinsurance on outpatient services; 0% on inpatient care (one plan).⁸
- O. 0% coinsurance for care received at a Seattle HMO, Group Health Cooperative of Puget Sound; 95% for care preferentially received outside the HMO (one plan).

Plans requiring coinsurance (B-N) placed a ceiling on annual out-of-pocket expenditures, above which care was free.⁹ In all but one plan (N), the ceiling was a specified percentage of the family's income or a dollar limit, whichever was less. The percentage varied with family income and the dollar limit varied with the plan, as indicated below:

⁸During the experiment's first year in Dayton, the provisions of plans A-N differed in two ways: Only plan A covered dental services for adults; and the coinsurance rate on plans K-N was 100 percent instead of 95 percent.

⁹During the experiment's first year in Dayton, expenditures for outpatient mental health care did not apply toward the ceiling.

<i>Plan</i>	<i>Percentage of Family Income</i>	<i>Dollar Limit</i>
B-D	5, 10, or 15	1000/750 ¹⁰
E-G	5, 10, or 15	1000
H-J	5, 10, or 15	1000/750
K-M	5, 10, or 15	1000
N	--	150 per individual; 450 per family

HMO Control Group. Those assigned to plans A-O had gone through the baseline selection process. In contrast, members of the HMO control group were chosen from a random sample of existing members of the Group Health Cooperative, subject to HIE eligibility requirements. The control group was formed to compare HMO use by those who had *chosen* that delivery mode (i.e., members of the control group) with use by those experimentally *transferred* to an HMO from the fee-for-service system (i.e., members of the experimental group). Enrollees in the HMO control group continued with the Group Health Cooperative under their prior arrangements but provided the same data as HMO experimental members.

Services Provided

Plans A-O provided the same comprehensive benefits, including hospital, physician, dental, mental health, visual, and auditory services, drugs (including over-the-counter drugs for certain chronic conditions), and supplies. Services of nonphysician providers, such as audiologists, chiropractors, clinical psychologists, optometrists, physical therapists, and speech therapists, were also covered. The only

¹⁰In plans B-D and H-J the \$1000 limit applied during the first two years of enrollment for Dayton families, who enrolled from November 1974 to February 1975; and during the first year of enrollment for Seattle families, who enrolled from January to September 1976. The \$750 limit applied during subsequent enrollment years for the aforementioned families, and during the entire enrollment period for all other families.

noteworthy exclusions were nonpreventive orthodontic services, cosmetic surgery for preexisting conditions, and outpatient mental health visits exceeding 52 per year.¹¹

Enrollees were able to choose the physicians and other persons who provided their health care. However, if those in the HMO experimental group sought care outside the HMO that was available within, they were responsible for 95 percent of the cost. (For covered services, such as dental or chiropractic, that were unavailable at the HMO, members of the experimental HMO group were fully reimbursed.)

Enrollees in the HMO control group retained whatever benefit package they or their employer had purchased from the HMO. To encourage the reporting of non-HMO care, the HIE reimbursed members of the control group 5 percent for any service otherwise covered by the HIE that they obtained outside the HMO.

Terms of Enrollment

Families who accepted the insurance plan offered from plans A-0 were enrolled in the experiment for either three or five years, the term randomly assigned. All members of the HMO control group were enrolled for five years.

Enrollees assigned any benefits from their existing health insurance policies to the HIE during the time they participated. No family was financially penalized by HIE enrollment. Enrollees were reimbursed for the cost of maintaining their policies, and if their HIE plan could, under any conceivable set of circumstances, provide less coverage than their private policies, they were paid the maximum difference.¹²

Table 1 indicates the timing of enrollment in the experiment and number of enrollees insured immediately after the baseline selection process in each site.

¹¹For details, see Appendix A.

¹²Calculation of the maximum difference is described in Appendix B.

Table 1

HIE ENROLLMENT PERIODS

Site	Number of Enrollees ¹	1974	1975	1976	1977	1978	1979	1980	1981	1982
Dayton	1137	Nov.								Feb.
3-year	533						Feb.			
5-year	604									
Seattle	3111		Jan.							Sept.
3-year	1500						Sept.			
5-year	1611									
Fitchburg	723		July							Oct.
3-year	547						Oct.			
5-year	176									
Franklin Co.	889		July							Oct.
3-year	649						Oct.			
5-year	240									
Charleston	779		Nov.							Feb.
3-year ²	571				Nov.					
5-year	208									
Georgetown Co.	1060		Nov.							Feb.
3-year ³	800				Nov.					
5-year	260									
Total	7699									

NOTE: Timelines mark the month and year in which the first person enrolled in the experiment and the month and year in which the last person left the experiment. Data on use of health services continued to be collected from several groups after the end dates shown here: one year afterward for the Dayton 5-year group and Seattle, Fitchburg, and Franklin County 3-year groups; six months afterward for the Dayton 3-year group.

¹Numbers refer to enrollees insured immediately after the baseline selection process. An additional 555 persons were enrolled and insured later, nearly all of them newborns or adopted children under 1 year of age. Figures for Seattle include the HMO control group. These numbers correct a technical error in Table 1 of all codebooks issued for HIE data files. One enrollee became insured some time (not immediately) after the baseline selection process so does not precisely fit the definition given. Compared with the codebooks, therefore, this table shows one less in the enrollee counts for Seattle, Seattle 5-year, and all-site total.

²Some of these enrollees were also members of a preenrollment group between November 1976 and February 1979. An additional 339 persons participated in the preenrollment phase but did not formally enroll in the experiment.

³Some of these enrollees were also members of a preenrollment group between November 1976 and February 1979. An additional 213 persons participated in the preenrollment phase but did not formally enroll in the experiment.

DATA COLLECTION

Over the course of the experiment, extensive data were collected on participants' demographic and economic characteristics, health status, and use of health services. Background information was obtained on local health care costs, providers, and types of services rendered. The data collection instruments are described in Table 2.

Table 2 shows the types of data gathered from the various participant groups. The most extensive data, especially longitudinal data on the use of health services, are available from the 8,254 insured enrollees. The 15,411 baseline-only participants provided much demographic and socioeconomic data, as well as information on health status, experience with health care, and health-related attitudes. Limited data were obtained for the 2,483 adjunct enrollees.

Several subcontractors to RAND participated in the data collection effort. Until March 1975, Mathematica Policy Research supervised data collection, administered the insurance plans, and processed claim forms. Thereafter, National Opinion Research Center managed data collection and Glen Slaughter and Associates handled insurance administration and claim processing. American Health Profiles, Inc., conducted the medical screening examinations at enrollment (October 1974 through January 1977); CompuHealth administered those examinations at exit (October 1977 through December 1981).

FILE DEVELOPMENT

Subcontractors sent the collected data to RAND, either in hardcopy form or as cleaned data tapes. At RAND the hardcopy data were encoded for machine readability and subjected to computerized checks for logical consistency and adherence to specified response ranges; outliers were checked only for fidelity to the original response and otherwise left unchanged. Limited cross-checking was done to assess logical consistency among a respondent's answers. All identifiers permitting information to be linked to a specific respondent were replaced twice to protect respondents' privacy.¹³ The cleaned records were then arranged

¹³The first conversion was known only to the subcontractor, the second only to RAND. Neither institution could make the full link from the respondent's name to his or her identifier on the analytic files.

Table 2

HIE DATA COLLECTION INSTRUMENTS

Instrument	Topics Covered	Data Collected		
		How	When	From
1. Screening questionnaire [1]	Demographic information to establish basic eligibility	Interview	Beginning of HIE operation in site	Occupants of representative sample of dwelling units on geographic clusters in site
2. Baseline questionnaire, 2 parts	Income, employment Family composition	Interview	4-6 months before enrollment	Baseline participants
	Health status Health care experience and insurance coverage Satisfaction with medical care	Self-administered	4-6 months before enrollment	Baseline participants
3. Enrollment verification form	Changes in family composition, economics, or insurance coverage since baseline questionnaire	Interview	Between administration of baseline questionnaire and enrollment date	Baseline participants determined eligible
4. Medical history questionnaire (MHQ), 3 versions by age group: 0-4 years 5-13 years 14+ years	Form A: health status, attitudes, habits Form B: specific medical disorders	Administered by self or parent [2]	Just before enrollment and exit [3]	Insured enrollees
5. Medical screening examination, 3 versions by age group: 0-2 years 3-13 years 14+ years	Tests of organ system function, including dental	Paramedical personnel	Just before enrollment and exit	Sample of insured enrollees at enrollment; all exiting enrollees
6. Health report	Use of medical or dental services and time spent obtaining them; any restricted activity or bed disability	Administered by self or parent	Biweekly during period of participation	Insured enrollees [4]

1. Administered as a separate questionnaire only in Dayton; part of baseline questionnaire in the other sites.
2. When "parent" appears in this column, a parent was asked to provide data for children 13 and younger.
3. "Exit" refers to normal departure from the experiment after completing the assigned enrollment period, three or five years. Those who "attrited," or voluntarily left the experiment early, received an "attrition" MHQ that was identical to the exit MHQ.
4. In the first year of the experiment in Dayton, the health report was administered weekly to a random half of Dayton enrollees. In the first year of the experiment in Massachusetts and South Carolina, 25 percent of enrollees were exempted to measure the reporting requirement's effect on the use of health services. Also at one point virtually all participants stopped filling out health reports, for budgetary reasons.

Table 2 (cont.)

Instrument	Topics Covered	Data Collected		
		How	When	From
7. Health care questionnaire, 3 versions by age group: 0-4 years 5-13 years 14+ years	Health status, attitudes, habits (subset of MHQ)	Administered by self or parent	Each anniversary of enrollment except at exit	Insured enrollees
8. Annual income report	Amount and sources of family income, taxes paid	Self-administered	Annually (April)	Head of insured family
9. Periodic employment report	Wages, hours worked, family payments for care of children or elderly, government program benefits received	Self-administered	Semiannually	Enrollees (head and family members 16 and older)
10. Assets and debts questionnaire	Family assets and liabilities	Self-administered	Exit	Head of insured family
11. Knowledge of coverage questionnaire	Details of HIE insurance plan	Self-administered	Specified intervals [5]	Insured enrollees
12. Insurance abstraction	Details of selected insurance policies	Abstracted	At time of knowledge of coverage questionnaire	Insurance company brochures
13. Chronic condition questionnaire	Status of condition, correctness of diagnosis, adequacy of treatment	Physician interview	At exit medical screening examination	Sample of insured enrollees found to have certain chronic conditions [6]
14. Evaluation questionnaire	Perceptions and attitudes about HIE and health care system	Self-administered	Exit	Head of insured family
15. Health notice	Use of medical or dental services	Administered by self or parent	Biweekly during preenrollment phase (South Carolina); 6 months-1 year after exit (other sites)	Preenrollees (South Carolina), insured enrollees who have exited (other sites)

5. Intended intervals were enrollment, 18 months, 3 years, and 5 years after enrollment (the last only for the 5-year participants). Actual mailings approximated those intervals in Massachusetts and South Carolina; the first mailing was 2-1/2 years and 1 year after enrollment in Dayton and Seattle, respectively.

6. Hypertension, diabetes, thyroid diseases, chronic heart diseases, chronic lung diseases, joint diseases, ulcers, cerebrovascular disease.

Table 2 (cont.)

Instrument	Topics Covered	Data Collected		
		How	When	From
16. Medical expense report (MER)--fee-for-service claim form, 4 types: Doctors' services and supplies Dental care Hospital and extended care Pharmacy	Each use of medical or dental service, drugs, and equipment; reason or diagnosis; treatment	Administered by self or parent	Time of occurrence	Insured enrollees and providers/suppliers
17. Services rendered report (SERR)--HMO equivalent of MER [7], 2 types: Doctors' services and supplies Hospital and extended care	Each use of medical service provided by HMO; reason or diagnosis; treatment	Abstracted	Annually to cover entire previous year	HMO records for insured enrollees in HMO experimental and control groups
18. Factor price survey	Wages and benefits of selected hospital personnel [8], average daily inpatient population	Phone and mail	Semiannually	Sample of local hospitals
19. Consumer price index	Prices of selected nonmedical products in the six HIE sites	Phone and inspection	Semiannually	Sample of local retailers
20. Physician capacity utilization survey (PCUTS)	Availability of services[9]	Phone	Annually	Sample of local physicians [10]
21. Dentist capacity utilization survey (DCUTS)	Similar to PCUTS	Phone	Annually	Sample of local dentists [11]
22. Insurance preference questionnaire	Willingness to pay higher premium to reduce out-of-pocket expense limit	Self-administered	Exit	Head of insured family

7. Pharmacy data were obtained directly from an HMO-supplied computer tape. Dental care was not available through the HMO; HMO participants reported claims for dental care and other non-HMO services on the MER.

8. Categories of personnel: registered nurses (general-duty), medical technicians, licensed professional nurses, nursing aides, kitchen helpers, general stenographers, and maids or porters.

9. Waiting time for appointments; appointments per hour; patients seen in office, home, and hospital; weekend office hours; office staffing; cost of office visit; whether new patients accepted.

10. Physicians (M.D. or D.O.) specializing in general practice, internal medicine, and pediatrics.

11. Except in Fitchburg, Franklin County, and Georgetown County, where all dentists were surveyed.

in the HIE version of standard computer file format, and the resulting files of *primary variables* made available for HIE analyses.

When an analyst needed information that required manipulation of primary data, *derived variables* were constructed. The analyst and a programmer determined a suitable way of obtaining the information by extracting, aggregating, or transforming primary data, and the programmer wrote the appropriate logic. With the analyst's approval, the new variable was entered on the master file.

A large subset of the data collected during the experiment has been issued to the public in 67 research files.¹⁴ Each file is documented by a data tape and hardcopy codebook published as a RAND Note. The machine-readable tape includes data in both SAS¹⁵ (Statistical Analysis System) and character formats, and an index of character-format variables.¹⁶ The codebook provides an overview of the experiment (nearly identical to the present section), outlines the characteristics of the file and its relation to other HIE files, and describes each variable. Variable descriptions include response codes and response frequencies or summary statistics.

ORGANIZATION OF THIS GUIDE

This document offers guidance for using data in the 67 *public-use* files. Hereinafter, any reference to HIE data must be understood to mean the publicly released subset of HIE data. Section II briefly describes each file. Section III discusses analytic considerations relevant to the data as a whole. Sections IV through VI then address the use of major portions of the data, those pertaining to the sample, enrollee claims, and medical/dental health, respectively.

¹⁴Not all data are being released because of budgetary constraints. Appendix C distinguishes the released and unreleased portions.

¹⁵A registered trademark of the SAS Institute Inc.

¹⁶These are the components of all files issued by RAND. Other institutions (e.g., National Archives) will distribute these files and may alter their contents.

II. ORGANIZATION OF HIE DATA

Nearly all of the 67 research files issued by the HIE are grouped in topical series. The list below provides a bird's-eye view of HIE data organization, showing the general topics covered by the data and the series or files pertaining to them.

Sample characteristics: master sample series

Claims:

- Line-item series (primary data)
- Aggregated series (derived data)
- Reference series

Medical/dental health:

- Medical history questionnaire series (primary data)
- Health status and attitude series (derived data)
- Medical disorders series (derived data)
- Dental examinations file (derived data)

Terms of health insurance: insurance preference files

The rest of this section describes the series and their component files, with three files not fitting in series listed last. Each file is identified by a unique letter-number combination, e.g., MS1 for the eligibility-family changes file. These *file reference numbers* are used throughout the rest of this guide as abbreviations. Each file description indicates the sample addressed, using terms defined in the Glossary and explained in Sec. IV; lists representative variables or topics; and cites the hardcopy codebook.

Codebooks can be ordered from the RAND Publications Department; indicate the desired titles and N numbers. Data files can be ordered from the RAND Data Facility. For each file desired, indicate the file reference number.

MASTER SAMPLE SERIES

MS1: Eligibility-Family Changes File. Data on insurance status, changes in eligibility status, and family relationships. No other file provides this information.

Sample: Insured enrollees (including members of HMO control group) and persons of interest, a subset of adjunct enrollees. Total of 9,142 records, one per person.

Representative Variables (of the total of 54): Insurance status, start/end dates of insurance coverage, reason for losing coverage, length of time insured, family identifier, individual's relation to family/household head.

Codebook: N-2264/1-HHS, *Vol. 1: Codebook for Eligibility-Family Changes File*, by S. M. Polich and C. d'Arc Taylor, May 1986.

MS2: Full Sample Demographic File. Baseline, enrollment, and demographic data. No other file provides this information.

Sample: All HIE participants--insured enrollees, adjunct enrollees, and baseline-only participants. Total of 26,148 records, one per person.

Representative Variables (total, 62): Date of baseline interview, sex, age, race, marital status, education, income, occupation, health insurance, welfare status, hospitalizations, family doctor, medical/dental visits and expenses, assigned experimental insurance plan.

Codebook: N-2264/2-HHS, *Vol. 2: Codebook for Full Sample Demographic File*, by S. M. Polich et al., May 1986.

MS3: Supplemental Data File. Sample data needed for specific analyses. No other file provides this information.

Sample: Overall sample is the same as that in the full sample demographic file; specific sample differs by variable. Total of 26,148 records, one per person.

Representative Variables (total, 19): Changes affecting HMO and FFS-HMO analyses, enrollment refusals, identifier for mothers of newborns, revised death date.

Codebook: N-2264/3-HHS, *Vol. 3: Codebook for Supplemental Data File*, by S. M. Polich and C. d'Arc Taylor, June 1987.

CLAIMS LINE-ITEM SERIES

LI1-LI14: FFS Claims Line-Item Files (14 files). Detailed FFS claims data, including data for HMO enrollees who used medical/dental services in FFS sector. Number of variables in each file shown parenthetically below.

- LI1. Hospital inpatient services (39)
- LI2. Inpatient physician procedures billed by institutions (36)
- LI3. Drugs prescribed by physicians (52)
- LI4. Supplies prescribed by physicians (44)
- LI5. Services rendered by physicians (53)
- LI6. Drugs sold by physicians (63)
- LI7. Supplies sold by physicians (53)
- LI8. Injections administered by physicians (69)
- LI9. Outpatient services billed by institutions (45)
- LI10. Services rendered by dentists (33)
- LI11. Drugs prescribed by dentists (24)
- LI12. Drugs purchased (37)
- LI13. Supplies purchased from pharmacies (17)
- LI14. Supplies purchased from nonpharmacy suppliers (19)

Sample: Insured enrollees who filed claims. Total of 603,998 records, one per line item.

Representative Variables: Diagnoses (multiple), provider ID, inpatient/outpatient procedures/services, drugs prescribed and sold, dosage instructions, symptoms, relation to employment/accident, treatment history, charges, supplies bought.

Codebook: N-2347/1-HHS, *Vol. 1: Codebooks for Fee-for-Service Claims*, by C. E. Peterson et al., June 1986.

LI15-LI25: HMO Claims Line-Item Files (11 files). Detailed services provided or reimbursed by the HMO. Number of variables in each file shown parenthetically below.

- LI15. Hospital inpatient services (34)
- LI16. Inpatient physician services (41)
- LI17. Drugs prescribed by physicians (52)
- LI18. Supplies prescribed by physicians (45)
- LI19. Services rendered by physicians (51)
- LI20. Drugs dispensed by physicians (57)
- LI21. Supplies dispensed by physicians (44)
- LI22. Injections administered by physicians (66)
- LI23. Outpatient services provided by institutions (42)
- LI24. Drugs dispensed (32)
- LI25. Supplies dispensed (13)

Sample: HMO participants, a subset of insured Seattle enrollees. Total of 177,566 records, one per line item.

Representative Variables: Diagnoses, provider ID, procedures performed, drugs/supplies, imputed charges.

Codebook: N-2347/2-HHS, *Vol. 2: Codebooks for Health Maintenance Organization Claims*, by C. E. Peterson et al., August 1986.

LI26-LI29: FFS Claims for FFS-HMO Comparison (4 files). Detailed Seattle FFS claims data with imputed charges for physician services to enable dollar comparisons with HMO data. Number of variables in each file shown parenthetically below.

- LI26. Hospital inpatient services (35)
- LI27. Inpatient physician procedures billed by institutions (32)
- LI28. Outpatient services rendered by physicians (50)
- LI29. Injections administered by physicians (64)

Sample: Insured Seattle FFS enrollees who filed claims. Total of 70,991 records, one per line item.

Representative Variables: Diagnoses, provider ID, imputed charges, procedures performed.

Codebook: N-2347/3-HHS, *Vol. 3: Codebooks for Seattle Fee-for-Service Claims for Comparison with Health Maintenance Organization Claims*, by C. E. Peterson, M. Nelsen, and D. L. Wesley, October 1986.

AGGREGATED CLAIMS SERIES

AC1: FFS Annual Expenditure File. Claims data aggregated by year for insured FFS enrollees. Also covers FFS dental usage by HMO enrollees.

Sample: Insured enrollees. Total of 25,740 records, one per enrollee per year.

Representative Variables (total, 23): Number per year of the following: hospitalizations, physician and nonphysician visits, mental health visits, and dental visits; annual expenditures for inpatient, outpatient, mental health, and dental services.

Codebook: N-2360/1-HHS, *Vol. 1: Codebook for Fee-for-Service Annual Expenditures and Visit Counts*, by C. E. Peterson, M. Nelsen, and E. S. Bloomfield, May 1986.

AC2-AC4: FFS Visit Files (3 files). Claims data aggregated by outpatient, inpatient, and dental visit for insured FFS enrollees. Covers FFS dental visits by HMO enrollees. Number of variables in each file shown parenthetically below.

- AC2. FFS outpatient visits (46)
- AC3. FFS inpatient visits (55)
- AC4. FFS dental visits (16)

Sample: Insured enrollees. Total of 148,123 records, one per enrollee-provider-date of service.

Representative Variables: Type of visit, providers, visit dates, procedures, diagnoses, charges.

Codebook: N-2360/2-HHS, *Vol. 2: Codebooks for Fee-for-Service Visits--Outpatient, Inpatient, and Dental*, by C. E. Peterson et al., June 1986.

AC5-AC6: FFS Episode Files (2 files). Claims data aggregated by treatment episode for FFS enrollees. Description and expenses for each episode (individual file); episode counts and expenses per year (annual file). Number of variables in each file shown parenthetically below.

- AC5. FFS individual episodes (19)
- AC6. FFS annual episodes (42)

Sample: Insured FFS enrollees. Total number of records: 99,001 in individual file, one per episode; 21,094 in annual file, one per enrollee per year.

Representative Variables: Episode description, start/end dates, diagnosis, expense limit at beginning of year, remaining expense limit at start/end of episode, number of episodes per year by type, expenses per episode type per year.

Codebook: N-2360/3-HHS, *Vol. 3: Codebooks for Fee-for-Service Treatment Episodes and Annual Episode Counts*, by C. E. Peterson, C. d'Arc Taylor, and E. S. Bloomfield, June 1986.

AC7: HMO and Seattle FFS Annual Expenditure File. Claims data aggregated by year for insured HMO and Seattle FFS enrollees.

Sample: Insured Seattle enrollees. Total of 11,221 records, one per enrollee per year.

Representative Variables (total, 33): Number per year of the following: hospitalizations, physician and nonphysician visits, mental health visits, imputed expenditures.

Codebook: N-2360/5-HHS, *Vol. 5: Codebook for Health Maintenance Organization and Seattle Fee-for-Service Annual Expenditures and Visit Counts*, by C. E. Peterson et al., December 1986.

AC8-AC9: HMO and Seattle FFS Visit Files (2 files). Claims data aggregated by health care visit for insured HMO and Seattle FFS enrollees. Number of variables in each file shown parenthetically below.

AC8. HMO and Seattle FFS outpatient visits (45)

AC9. HMO and Seattle FFS inpatient visits (53)

Sample: Insured Seattle enrollees. Total of 61,597 records, one per enrollee-provider-date of service.

Representative Variables: Type of visit, providers, visit dates, procedures, diagnoses, imputed charges.

Codebook: N-2360/4-HHS, *Vol. 4: Codebooks for Health Maintenance Organization and Seattle Fee-for-Service Visits--Outpatient and Inpatient*, by C. E. Peterson, M. Nelsen, and D. L. Wesley, December 1986.

HIE REFERENCE SERIES

RF1: Codes (no companion file). N-2349/1-HHS, *Vol. 1: Codes Used in HIE Claims--Diagnoses, Symptoms, Procedures, Drugs, and Supplies*, by M. Nelsen and C. A. Edwards, May 1986. Defines all codes used in HIE claims data, both line-item and aggregated. Includes standard and HIE-created codes: diagnosis (H-ICDA-2), CRVS, supply, reason for visit/symptom, NDC, generic drug, drug therapeutic, American Dental Association procedure.¹

RF2: HIE Provider File: Information about the physicians, hospitals, dentists, and other providers of services to HIE enrollees.

Sample: All providers cited in HIE data. Total of 22,658 records, one per provider identifier.

Representative Variables (total, 26): Provider type, provider specialty, linking identifier.

Codebook: N-2349/2-HHS, *Vol. 2: Providers Cited in HIE Data*, by S. M. Polich, M. Nelsen, and D. L. Wesley, June 1987.

¹H-ICDA-2 refers to the second version of the hospital adaptation of International Classification of Diseases Adapted for Use in the United States; CRVS refers to codes for medical and surgical procedures taken from California Relative Value Studies; NDC refers to National Drug Code.

RF3: User's Guide (no companion file). N-2349/3-HHS, *Vol. 3: User's Guide to HIE Data*, by C. d'Arc Taylor, S. M. Polich, C. E. Peterson, and E. M. Sloss, August 1987. Analytic possibilities and limitations of HIE data; suggestions for choosing analytic subsamples and linking data across series, files, years, sample groups, and sites for particular analytic purposes.

MEDICAL HISTORY QUESTIONNAIRE SERIES

MH1A-MH3A: Adult Form A (3 files). Data from self-administered questionnaire on health status, attitudes, and habits. Number of variables in each file shown parenthetically below.

- MH1A. Dayton adults at enrollment, form A (364)
- MH2A. NonDayton adults at enrollment, form A (373)
- MH3A. Adults at exit, form A (408)

Sample: Adults (14 and older) when enrolling and when completing assigned term three or five years later. Includes insured enrollees, Dayton control group, and PEG-period-only participants. Total of 9,058 records, one per completed questionnaire.

Topics: height and weight, general health, eating habits, sleep and exercise, seat belt use, smoking and drinking, general well being, social activities, life events, symptoms, health perceptions, medical opinions, medical and dental care, effectiveness of health care.

Codebook: N-2485/1-HHS, *Vol. 1: Codebooks for Adults at Enrollment and Exit, Form A*, by C. A. Edwards et al., August 1986.

MH1B-MH3B: Adult Form B (3 files). Data from self-administered questionnaire on verifiable physical limitations and specific medical disorders. Number of variables in each file shown parenthetically below.

- MH1B. Dayton adults at enrollment, form B (282)
- MH2B. NonDayton adults at enrollment, form B (480)
- MH3B. Adults at exit, form B (490)

Sample: Adults (14 and older) when enrolling and when completing assigned term three or five years later. Includes insured enrollees, Dayton control group, and PEG-period-only participants. Total of 9,914 records, one per completed questionnaire.

Topics: Vision, hearing, hay fever, teeth and gums, fluoride treatment, acne, thyroid, joints, heart/lung ailments, hypertension, stroke, stomach, kidney/bladder, cholesterol, anemia, diabetes, cancer, surgical conditions (hemorrhoids, hernia, varicose veins), physical/activity limitations, sleeping pill use, missing limbs, antibiotic allergy, effectiveness of health care, immunization, gall bladder/tonsil surgery,

female organs, medical care, medical appliances, future health expenses, transportation for health care.

Codebook: N-2485/2-HHS, *Vol. 2: Codebooks for Adults at Enrollment and Exit, Form B*, by C. A. Edwards et al., October 1986.

MH4A-MH6B: Child Forms A and B (6 files). Data from two parent-completed questionnaires. Form A pertained to health status, attitudes, and habits. Form B pertained to verifiable physical limitations and specific medical disorders. Number of variables in each file shown parenthetically below.

- MH4A. Dayton children at enrollment, form A (85)
- MH4B. Dayton children at enrollment, form B (63)
- MH5A. NonDayton children at enrollment, form A (151)
- MH5B. NonDayton children at enrollment, form B (224)
- MH6A. Children at exit, form A (147)
- MH6B. Children at exit, form B (235)

Sample: Children 5-13 years old, at family's enrollment and exit three or five years later. Includes insured enrollees, Dayton control group, and PEG-period-only participants. Total of 6,958 records, one per completed questionnaire.

Topics: Form A: height, weight, general health, fluorides, diet, immunizations, safety practices, learning, getting along, general well-being, symptoms, behavior problems. Form B: teeth, fluoride treatment, eyesight, hearing, ear infections, asthma, hay fever, eczema, anemia, lead poisoning, kidney/bladder infection, bedwetting, cancer, convulsions, tonsils, antibiotic allergy, missing limbs, medical appliances, future health expenses.

Codebook: N-2485/3-HHS, *Vol. 3: Codebooks for Children at Enrollment and Exit*, by C. A. Edwards et al., November 1986.

MH7A-MH9B: Infant Forms A and B (6 files). Data from two parent-completed questionnaires. Form A pertained to health status and development; form B pertained to verifiable physical limitations and specific medical disorders. Number of variables in each file shown parenthetically below.

- MH7A. Dayton infants at enrollment, form A (76)
- MH7B. Dayton infants at enrollment, form B (28)
- MH8A. NonDayton infants at enrollment, form A (98)
- MH8B. NonDayton infants at enrollment, form B (122)
- MH9A. Infants at exit, form A (94)
- MH9B. Infants at exit, form B (134)

Sample: Infants (0-4 years old) at family's enrollment and exit three or five years later. Includes insured enrollees, Dayton control group, and PEG-period-only participants. Total of 3,334 records, one per completed questionnaire.

Topics: Form A: height, weight, development, general health, fluorides, diet, immunizations, safety practices, symptoms. Form B: colds, ear infections, eczema, anemia, lead poisoning, cancer, convulsions, tonsils, antibiotic allergy, missing limbs, medical appliances, future health expenses, fluoride treatment.

Codebook: N-2485/4-HHS, *Vol. 4: Codebooks for Infants at Enrollment and Exit*, by C. A. Edwards et al., December 1986.

HEALTH STATUS AND ATTITUDE SERIES

HS1-HS2: Adult and Child (2 files). Data derived from medical history questionnaire on enrollees' state of health and attitudes toward health care at enrollment and exit. Number of variables in each file shown parenthetically below.

- HS1. Adults at enrollment and exit (136)
- HS2. Children at enrollment and exit (28)

Sample: Insured enrollees: adults (14 and older) and children (aged 0-13) at family's enrollment and exit three or five years later. Total of 5,871 records in adult file, 2,840 records in child file--one per person per file.

Representative Variables: Scales of physical health, mental health, social health, perceptions of general health; satisfaction with medical and dental care, cigarette smoking, alcohol consumption, weight, and exercise (adults only).

Codebook: N-2447/1-HHS, *Vol. 1: Codebooks for Adults and Children at Enrollment and Exit*, by E. M. Sloss et al., November 1986.

MEDICAL DISORDERS SERIES

MD1: Adults. Data derived from medical history questionnaire and medical screening examination on 17 disorders: acne, anemia, angina pectoris, chronic obstructive airway disease, congestive heart failure, diabetes mellitus, hay fever, hearing loss, hypercholesterolemia, hypertension, joint disorders, kidney disease and urinary tract infection, peptic ulcer disease, sleeping pill and tranquilizer use, surgical conditions, thyroid disease, and vision disorders.

Sample: Insured enrollees: adults (14 and older) at enrollment and exit three or five years later. Total of 5,871 records, one per person.

Topics (total number of variables, 286): Status and severity of disorder, impact of disorder, results of medical tests.

Codebook: N-2446/1-HHS, *Vol. 1: Codebook for Adults at Enrollment and Exit*, by B. H. Operskalski et al., February 1987.

MD2: Children. Data derived from medical history questionnaire and medical screening examination on four disorders: allergic conditions, anemia, middle ear disease and hearing impairment, and vision impairment.

Sample: Insured enrollees: children (aged 0-13) at family's enrollment and exit three or five years later. Total of 2,840 records, one per person.

Topics (total number of variables, 73): Status and severity of disorder, impact of disorder, results of medical tests.

Codebook: N-2446/2-HHS, *Vol. 2: Codebook for Children at Enrollment and Exit*, by E. M. Sloss et al., March 1987.

DENTAL EXAMINATIONS FILE

DE1: Dental Examinations File. Data from a dental screening examination on tooth decay and its consequences, and periodontal disease and its severity.

Sample: Insured enrollees: persons aged three and older in randomly selected subsample (50-75 percent) of families at enrollment; all persons aged three and older at exit three or five years later. Total of 7,317 records, one per person.

Representative Variables (total, 50): Number of decayed primary and permanent teeth, number of missing or extracted primary and permanent teeth, number of filled primary and permanent teeth, oral hygiene index score, and (for those 12 and older) periodontal disease index score.

Codebook: N-2506-HHS, *Dental Examinations: Codebook for Adults and Children at Enrollment and Exit*, by E. S. Bloomfield, L. Y. Weissler, and A. M. Bell, February 1987.

INSURANCE PREFERENCE FILES

IP1-IP2: Insurance Preference Files (2 files). Data from self-administered questionnaire on willingness to pay a higher health insurance premium in return for a lower annual out-of-pocket expense limit--three hypothetical premium-expense limit combinations. Number of variables in each file shown parenthetically below.

- IP1. Maximum-dollar-expenditure plans (21)
- IP2. Fixed-dollar-limit plan (25)

Sample: Heads of insured enrollee families (except HMO enrollees and enrollees assigned to receive free care) when completing assigned enrollment term. Total of 2,020 records, one per questionnaire recipient.

Topics: Family's expense limit during its last year in HIE; premium-expense limit combination for each of three hypothetical offers; degree of willingness to accept each offer.

Codebook: N-2508-HHS, *Codebooks for Insurance Preference Files: Relation between Expense Limit and Premium*, by E. S. Bloomfield, L. Y. Weissler, and A. B. Holland, October 1986.

III. ANALYTIC CONSIDERATIONS

Many analyses can be done with HIE data, some in greater depth than others. Given the size and scope of the data, it is tempting to think that nearly any analysis of health care use or health status is possible. In reality, as with any database, certain parts of the population are unrepresented in the sample, certain types of analyses are inappropriate with the data, and certain caveats must be understood when using the data. This section is intended to assist analysts by pointing out these restrictions.

SAMPLE DEFINITION

The HIE sample departs from a random national probability sample because the experiment was addressed to the *general* population having a *range* of options for obtaining health care. As a result, HIE eligibility rules excluded the elderly (and others eligible for Medicare), those with very high incomes, those in the military (and veterans with service-connected disabilities), and the institutionalized, as noted on p. 2 above.

The purpose of the experiment was to study the demand for and cost of health care, and we desired to limit out-of-pocket expense to a specified fraction of family income. Therefore, the unit of enrollment was the family--defined as a single self-supporting individual or a group of related persons who shared income. Specifically, after the baseline selection process, enrollment was offered to

- Eligible self-supporting individuals 18 years of age or older.
- Groups composed of any of the following who occupied the same residence:
 - Eligible family head (not dependent on others in the household for more than half of his or her support) and spouse or person whom head claimed to be spouse. Therefore, two heads could be designated. One had to be 18 years of age or older.
 - Other persons financially dependent on the head and closely related to the head by blood, marriage, or adoption.

A family was eligible as long as it contained one eligible person who was at least 18 years old and could be designated head. The ineligibility of other individual members did not affect the eligibility of the family as a whole *except* that the entire family was ineligible if its income exceeded \$25,000 (1973 dollars) at the time of the baseline interview or the head received free or highly subsidized medical care, such as through the armed services or Medicare.

Within a family or household, three categories of members were not eligible for the HIE:

1. Members otherwise ineligible--for example, those over the age limit or veterans with service-connected disabilities.
2. Members financially dependent on but unrelated to the head.
3. Members who entered the family, or moved back home, after the family enrolled in the HIE.

There were exceptions, however. Eligibility was extended to newborns, adopted children under one year of age, and members of families in the Seattle HMO control group who joined the family after it enrolled.

During the course of the experiment, HIE guidelines specified terminating the participation of enrollees whose circumstances changed so that they became ineligible. The guidelines were not rigidly applied, however, but adjusted to fit individual cases so as to maximize the longitudinal data obtained. For example, persons who were jailed were suspended. If they were still jailed on their next anniversary of enrollment, their participation was terminated; if they had been released, they regained their eligibility. The number of persons whose enrollment was terminated for ineligibility--405--amounts to 4.9 percent of the total ever insured. The specific reasons for termination, and their frequency distribution, are itemized in variable RTERM on file MS1.

For more detail on the sample, see Sec. IV.

ANALYTIC LIMITATIONS

As a result of HIE eligibility rules or administrative practices, certain analyses are not possible or recommended with HIE data. Here we describe general limitations; more specific analytic guidance is in Secs. V-VI.

Precise Time Linkages

Data from insurance claims filed during the experiment document health care use and expenditures when they occurred. Data affecting participants' eligibility were also collected continuously. All other publicly released participant data were collected at the following specified times, not necessarily when events occurred:

- During the baseline and enrollment period (demographic and socioeconomic data).
- At enrollment and exit three or five years later (medical history, medical and dental health status/attitudes, and medical disorders).
- At enrollment and anniversaries of enrollment, including exit (family composition data).

As a result, analyses that depend on linking, say, illnesses or health care expenditures with socioeconomic or demographic characteristics *at the exact time* of the illness are not possible. For example, publicly released HIE data do not show family income at the time a new member is born because income data were collected only for the two years preceding enrollment. As another example, there is no direct way of identifying the family of a newborn if the baby died before the family's next enrollment anniversary. The publicly released record of the birth does not show a family identifier, and HIE records at the enrollment anniversary following the baby's birth show that the family composition changed but do not pinpoint how.¹

¹Section IV suggests an indirect means of identifying the family in this case.

"Actual" versus "Billed" Health Care Use

Information on enrollees' health care use comes from the claims they submitted (except for HMO participants, whose usage data were abstracted from records of the Group Health Cooperative). As a result, usage data from the claims files reflect only illnesses, doctor or hospital visits, and services for which charges were claimed, and not necessarily all illnesses experienced, visits made, or services received. This is an important limitation for analysts interested in services that are traditionally billed as a lump sum, such as prenatal and pre-/postsurgical.

Use of Health Care System by Families

Because of ineligibility or refusal to participate, not all members of an enrolled family necessarily provided data to the HIE on their health care use. Some of those nonparticipants were carried on HIE rolls as adjunct enrollees, but no data were collected for them. Analyses of the health care use of such families are inappropriate because of the lack of usage data for their uninsured members.

Replication of HIE Analyses

Even before the site-based part of the experiment ended in 1982, HIE staff were analyzing the data and publishing results in medical journals and RAND publications. HIE analyses illuminate the data and their implications and are often cited in the codebooks as useful resources. However, the data files used by HIE analysts are not the same versions as the public-use files addressed here, and users should not expect to replicate HIE analyses with the data now available. Some errors, including those affecting sample counts, have been corrected in the public-use files.

Serious Illness

The HIE population is generally healthy, and serious illness is rare. For example, there are probably too few cases of cancer to allow a separate analysis of cancer treatment, especially if the focus is on a specific cancer site.

DETERMINING WHICH HIE DATA FILES TO USE

Because of the number of files, as well as the way HIE data are organized, it is not obvious how to choose the necessary minimum number of files for a particular analysis. Here we suggest ways of making that choice efficiently. The general procedure recommended is as follows:

1. Read the user's guide, paying particular attention to the data categories of interest.
2. Consult the hardcopy codebooks for the files of interest to determine the files essential to your analysis. The codebook describes the specific sample addressed by the file, indicates specific data limitations, and gives a preliminary look at response frequencies. It may even point out errors in the file.
3. Order the files you intend to use. See Sec. II for ordering information.

Characteristics of the Sample

The large HIE database required many files and different modes of data organization. To avoid repetition and provide analytic flexibility, we grouped all sample information in a separate series of files--the master sample series--and arranged the rest of the information in other topical series of files.

As a result, you cannot effectively analyze HIE data without reference to the demographic, eligibility, or family information in the master sample series. To attempt it would be to risk including an undesired subsample or excluding a desired subsample in the analysis.

The analyst's first task, therefore, is to define a sample, using the master sample series, then consult the other series for variables of interest for the sample. Section IV provides an introduction to the master sample files.

Health Care Use and Expenditures

If you are interested in health care use and expenditures, you will consult the two series of claims files, first scanning the aggregated series, then turning to the line-item series for greater detail as needed (see Sec. V).

Within the claims data you can of course ignore HMO files if you are interested only in the FFS sector. The reverse is generally true except that some FFS files show FFS sector use by HMO participants (Sec. V clarifies this point).

If the focus is dental usage, you will need only a small subset of the claims files, but the choice is not necessarily straightforward. Oral surgery illustrates the point. The usual tooth extraction done in a dentist's office presents the simplest case: All primary data are in file LI10 (services rendered by dentists), and all aggregated data are in AC4 (dental visits). Complexity arises when the oral surgery is done elsewhere. If in a hospital, the hospital charges appear in LI1 (hospital inpatient services) and the surgeon's charge appears in LI5 (services rendered by physicians) or LI2 (inpatient physician procedures billed by institutions) if the surgeon was considered a hospital staff member. Finally, if the surgery was done as an outpatient procedure in a hospital clinic, the dentist's charge is in LI9 (outpatient services billed by institutions). With such diverse primary data sources, oral surgery data can appear in all three aggregated FFS visit files--inpatient, outpatient, and dental (AC2-AC4). The introductions to the relevant codebooks explain such complexities where they occur.

Users of claims data may find the HIE reference series helpful. The "codes" volume (RF1) defines all the numeric codes--standard and HIE-created--used to denote diagnoses, symptoms, procedures, drugs, and supplies in the claims data. Note that this hardcopy document has no corresponding data tape. The provider file (RF2) gives information about the providers of services to HIE enrollees--type (e.g., physician, hospital, pharmacy) and specialty (e.g., cardiology, endodontia). Only the outpatient and inpatient FFS and HMO visit files (AC2, AC3, AC8, AC9) give the same information; the other claims files show an identifier that must be linked with the provider file to learn the provider's type and specialty.

Medical/Dental Health

Because the topics are more distinct, it is easier to choose the necessary minimum of files from the medical history questionnaire (MHQ) series of primary data or the health status/attitudes and medical disorder data derived from it. After defining a sample, you will probably first consider one of the derived series, according to the topic of interest, then refer to the primary data in the MHQ series if more detail is needed.

The MHQ-derived data (HS1-HS2, MD1-MD2) are in separate files for adults (aged 14+ years) and children (aged 0-13 years). For its age group, each file provides health status/attitude or medical disorder data at both enrollment and exit, covering all sites, and for both forms A and B of the MHQ instrument.

In contrast, there are 18 files of primary data (MH1A-MH9B). Their number reflects the multiple versions of the instrument that were administered. A separate file exists for each combination of differences listed below:

- Two parts--forms A and B--each addressing a different set of topics.
- Three versions by age group: adult (for persons 14+ years old) child (for those 5-13 years old), and infant (for those 0-4 years old).
- Different versions for Dayton and all other ("nonDayton") sites at enrollment.
- Different exit version (all sites).

To illustrate, file MH1A contains form A data for Dayton adults at enrollment; file MH9B contains form B data for infants at exit.

It is obvious, therefore, that you will need several files of primary data to add detail to one file of derived data. And in the derived files, children who reached the age of 14 during the experiment have enrollment and exit data in different files. Section VI provides further detail.

Comprehensive Information

Those interested in comprehensive analyses should use data from several series (besides the master sample series) to link health care use/expenditure information with health history or status information. Note that for comprehensive dental analyses the claims data should be linked with dental health status information in the health status/attitude and dental examinations files.

CAVEATS

The following paragraphs cite distinctive aspects of HIE data of which users should be aware.

Data Comparability

As we have seen (Table 1), HIE enrollment took place at different times in the different sites, with at minimum a one-year difference between enrollment in the first site (Dayton) and the second site (Seattle). As a result, the data differ by site. At the least, data collection for a given instrument was far from concurrent for all participants. At the most, there are substantive differences in enrollment demographic, medical history, health status/attitude, and medical disorder data because of the changes made to the medical history questionnaire after its first administration in Dayton. For some derived variables pertaining to enrollment, Dayton data are either missing or the original questions on which variables are based are so different as to give the impression of noncomparability. We have attempted to compensate for intersite differences in the construction of derived variables and at least point out the differences where they occur. Further detail is in Sec. VI and the relevant codebooks.

On a smaller scale, the time of enrollment data differs within the three-year enrollee group in South Carolina. For reasons explained in Sec. IV, some of the three-year group preenrolled in 1976, then formally enrolled in 1978. They completed two enrollment MHQs, one in each of the years mentioned. A subset also received a medical screening examination in 1976. However, those who did not go through the preenrollment period lack screening exam data because the exams were not conducted in 1978.

As a result, the timing of enrollment data differs by two years within the South Carolina three-year enrollee group. If you are comparing enrollment and exit data for that group, the elapsed time will be three years for some members and five years for others. Further detail is in Sec. VI.

Insurance Preference Files

The insurance preference files (IP1 and IP2) represent a specialized use of the HIE enrollee sample. Heads of families completing their assigned three- or five-year enrollment terms were asked how much they would be willing to pay in higher health insurance premiums to reduce the maximum out-of-pocket expense they were required to pay each year. The questionnaire presented three hypothetical offers of expense limit-premium² pairs:

1. Expense limit two-thirds of the family's HIE limit, in return for a certain randomly generated premium.
2. Expense limit one-third of the family's HIE limit, in return for a higher premium than in offer 1.
3. Expense limit of zero (free care), in return for a higher premium than in offer 2.

The responses are in one of the two insurance preference files: IP1 for families assigned to insurance plans B-M, or IP2 for families assigned to plan N. HMO participants and enrollees assigned to plan A (free care) were not given the questionnaire.

Insurance preference data should not be linked with other HIE data to analyze health care use/expenditures or medical/dental health. HIE sample, claims, or medical/dental health data should be linked with insurance preference data only to explain respondents' answers to the expense limit-premium offers. In fact, insurance preference data should not be analyzed without considering the family's state of health and

²The description of the offers is simplified here. For more detail, see Ref. 5, p. 14.

income. Both would affect the head's judgment of the tradeoff between premium and expense limit.

The insurance preference codebook warns against linking the identifier variable FAMILY with variables in other HIE files. That is because FAMILY identifies the family at exit so it corresponds *only* with either variable IFAMILY4 or variable IFAMILY6 (family identifier at exit from three-year and five-year terms, respectively) on the eligibility-family changes file (MS1).

The following steps describe how you can determine the family members of insurance preference respondents and add information about their health status/attitudes and family income.

1. **Family Members.** In IP1 and IP2, determine each respondent's assigned enrollment term, matching on variables PERSON and ENRTERM.
2. Link IP1 and IP2 with MS1, matching on variable PERSON.
3. For persons with an ENRTERM value of 3, find all other persons with the same IFAMILY4 value in MS1. For persons with an ENRTERM value of 5, find all other persons with the same IFAMILY6 value in MS1. Those with the same IFAMILYi values are members of the same economic family at the time the insurance preference questionnaire was administered.³
4. **Health Status/Attitude Information.** Merge data from enrollment or exit variables of interest on HS1 and HS2, matching on variable PERSON.
5. **Family Income.** Merge data from income variables of interest on MS2, matching on variable PERSON. Remember that income data predate insurance preference data by at least four or six years, depending on the respondent's enrollment term.

³We assume here that analysts are interested in the economic, rather than insurance, families of respondents. The distinction is discussed in Sec. IV.

Presentation of Data

The last three general caveats stem from HIE practices in presenting the data.

Omitted Variables. If a question on a data collection instrument received no responses, the variable representing the question was omitted from the relevant file. To illustrate, HIE line-item files allow two modifiers for the CRVS code representing the medical/surgical procedure recorded on an enrollee's claim form. In assembling the variables for a given line-item file, if variable DEI5608 denoting the second CRVS modifier showed all missing values, the variable was omitted from the file. This explains why some variables may be missing in one of a group of files with otherwise identical variables.

Reading Missing Data. In the SAS version of a file, missing data are represented by the SAS dot "." or a letter denoting a special case. In the character version of a file the representation of missing values depends on whether the variable is alphanumeric, integer, or fixed-decimal. Fixed-decimal variables containing eight characters can have one, two, four, or six digits to the right of the decimal point--denoted 8.1, 8.2, 8.4, and 8.6, respectively. In the character versions of HIE files, "regular" and "special" missing values are written as follows ("b" means blank):

<i>Variable type</i>	<i>Type of Missing Value</i>	
	<i>Regular</i>	<i>Special "A"</i>
Alphanumeric	bbbbbbbbb	Not applicable
Integer	bbbbbbbbb.	bbbbbbbbbA
Fixed-decimal:		
8.1	bbbbbbb.b	bbbbbbbAb
8.2	bbbbbb.bb	bbbbbbAbb
8.4	bbb.bbbb	bbbAbbbb
8.6	b.bbbbbbb	bAbbbbbbb

To obtain the appropriate positive and missing values, read all values as alphanumeric, then convert integer data to integers and fixed-decimal data to the specified floating-point format.

False Look-Alikes. Most variables bearing the same name have exactly the same meaning across files. A few do not, however, so it is important to read the codebook variable definitions carefully.

Following are two examples:

- Variable PLAN always denotes the experimental insurance plan assigned to the enrollee's family.⁴ In the full sample demographic file (MS2), PLAN reflects any changes between the time the family enrolled and the time the individual joined the family. The individual episode file (AC5) does not reflect such changes, and PLAN refers to the family's *original* plan. To illustrate, family X was assigned to plan 30 at enrollment. A year later, it was reassigned to plan 11. Enrollee XY was born shortly thereafter. Enrollee XY's PLAN value is 11 on MS2 but 30 on AC5.⁵
- Variable MDEOFF on the full sample demographic file (MS2) states the maximum dollar expenditure *assigned at enrollment*, whereas MDEOFF on the supplemental data file (MS3) shows the figure *offered* a prospective enrollee *before* enrollment.

⁴From this point on, we refer to insurance plans by their numeric codes in variable PLAN rather than the letters used in Sec. I. The following tabulation shows the correspondences:

<i>Variable PLAN</i>	<i>Sec. I</i>
11	A
5-7	B-D
8-10	E-G
17-19	H-J
2-4,14-16	K-M
1,13	N
30	O

⁵Reassignment of plans was rare, occurring systematically in only two cases: (1) Families in the HMO experimental group (plan 30) who moved outside the GHC service area were reassigned to plan 11, and (2) after the first year of the experiment in Dayton, plans 1-4 requiring 100 percent coinsurance were eliminated and the families reassigned to plans 13-16, respectively.

IV. USING SAMPLE DATA

This section explains how to choose a sample, using files in the master sample series. It describes the HIE participant sample, identifies where to find information about the main sample groups in the master sample series, and provides guidance for choosing analytic subsamples. Sections V and VI cite guidelines for choosing variables of interest once the sample is defined.

THE HIE SAMPLE

Main Sample Groups

The three main parts of the HIE sample are *baseline-only participants*, *insured enrollees*, and *adjunct enrollees*. Figure 2 shows how they are related. The 24,340 persons represented within the dashed border went through the baseline selection process described in Sec. I. The 15,411 who never enrolled are the baseline-only participants. They

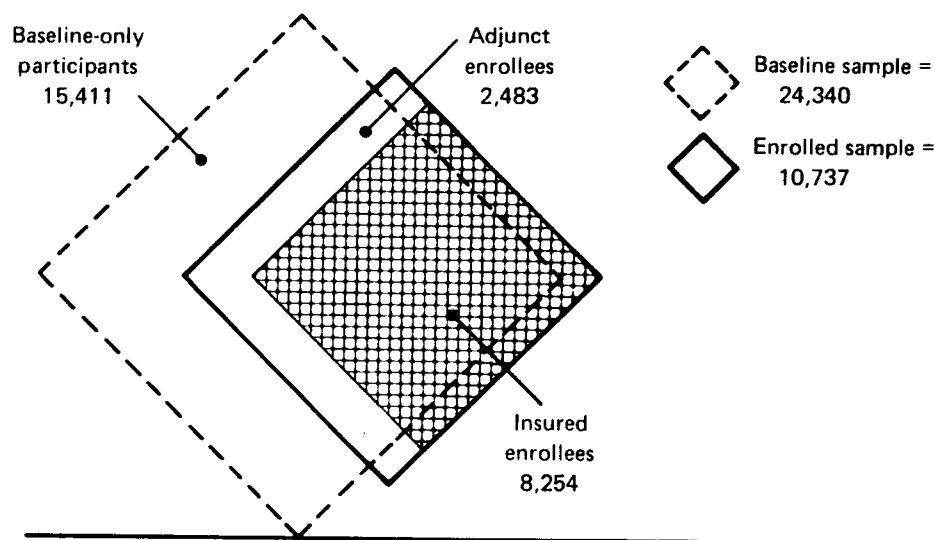


Fig. 2—HIE participant sample

were (1) determined ineligible, (2) found eligible but not offered enrollment, or (3) found eligible but refused enrollment. The rest became enrollees, either insured or adjunct. The enrolled sample is indicated by the solid border; the portion outside the dashed border includes members of the HMO control group and insured or adjunct enrollees who joined the experiment after the baseline period.

Insured enrollees (crosshatched area) signed an enrollment contract and were assigned to an experimental treatment and an enrollment term of three or five years. Their use of health care services was followed throughout the time they were enrolled. Adjunct enrollees were not assigned to an experimental treatment. Most were carried on HIE rolls only because they resided with insured enrollees¹; about 25 percent were members of a short-lived control group in Dayton.²

The full participant sample, without double-counting those in both baseline and enrolled samples, totals 26,148.

Variant Subgroups

Two subgroups depart from the rest of the sample in notable ways. They are the South Carolina PEG-period-only participants and the Seattle HMO control group, part of the baseline-only and insured-enrollee sample groups, respectively.

PEG-Period-Only Participants. In every site except South Carolina, those assigned to three- and five-year enrollment terms enrolled at the same time and exited two years apart. In South Carolina, both groups enrolled two years apart and exited at the same time.

¹If financially dependent on the head of the insured family (or if a member of such a dependent person's family), the adjunct enrollee was called a *person of interest*. If self-supporting (or a member of such a person's family), the adjunct enrollee was called a *family of interest*.

²The Dayton control group was chosen to indicate how the community's use of health services compared with use by HIE-insured enrollees. Control families retained their own health insurance but completed the same questionnaires as insured families. The Dayton control group participated from November 1974 to February 1976; it was discontinued because complete claims data appeared unobtainable from its members. Unlike insured enrollees, Dayton control families had to file claims with both the HIE and their own insurance companies.

Although both groups were chosen simultaneously, the three-year group was not formally enrolled until the beginning of the third year of the experiment in South Carolina. During years 1 and 2, its members served as a preenrollment group (PEG). They were not assigned to an insurance plan but data were collected for them. The two preenrollment years are known as the PEG period. If a PEG family dropped out of the experiment during the PEG period, we consider it part of the baseline-only sample because the family did not formally enroll. Not all South Carolina three-year enrollees have PEG period data. Some who had refused pre-enrollment accepted formal enrollment two years later; others were invited to enroll for the first time at the two-year point.

HMO Control Group. In Sec. I we noted that an unbiased selection model was used to assign families to insurance plans so that, across plans, families resembled each other in demographic characteristics. Members of the HMO control group, however, were chosen randomly from the existing HMO membership, not according to the unbiased selection model. Therefore, the HMO control group does not resemble the other insured enrollees in demographic characteristics.

Special terms signal the differences between the HMO control group and all other insured enrollees. *HMO-insured* refers to the HMO control group, whose members retained their preexisting HMO benefit package when they joined the experiment. *HIE-insured* refers to the other insured enrollees because the HIE assigned them to a new insurance plan when they joined the experiment. Note that the HIE-insured category includes the HMO experimental group, whose members were transferred from the fee-for-service system to the HMO when they joined the experiment.

DATA AVAILABILITY

The three files in the master sample series are the only sources of demographic and socioeconomic information about the HIE sample, but each file provides different information for a different group of participants.

The full sample demographic file (MS2) provides demographic and socioeconomic data for the *entire participant sample*, both baseline and enrolled groups. The supplemental data file (MS3) is also addressed to the entire participant sample but contains highly specialized

information--enrollment refusals, factors affecting sample comparability in Seattle, and identifiers for mothers of newborns--so that any one variable excludes most participants. The eligibility-family changes file (MS1) shows the eligibility status and family composition of *insured enrollees* from the time they enrolled until their participation ended. Since the period of enrollment could be as long as five years, an enrollee's family structure and ability to participate in the experiment could change during that time. MS1 presents data collected at various intervals to document the changes. It also provides limited data for a subset of *adjunct enrollees*, excluding the Dayton control group and self-supporting adjunct enrollees.

We assume that all analysts will use MS2 for demographic information. If your sample is insured enrollees, you will also use MS1. If your sample is baseline participants, you will probably not use MS1. Whatever the sample, you may or may not use the specialized MS3 variables; their relevance will need to be individually determined.

ANALYTIC SUBSAMPLES

For analytic purposes, the HIE participant sample subdivides into two groups: insured enrollees and never insured participants. Each HIE data file contains the variable INSTAT, which distinguishes the groups as follows:

<i>Status</i>	<i>INSTAT Value</i>
Ever HIE-insured (fee-for-service, HMO experimental group)	1
Ever HMO-insured (HMO control group)	2
Never insured (baseline-only, adjunct enrollees)	3

The rest of this section discusses the ever-insured and never-insured groups in turn.

Insured Enrollees

Because of the types and amount of data available, most users will probably restrict their analyses to the 8,254 insured enrollees (INSTAT values of 1 and 2). As a group, insured enrollees are the only ones with longitudinal data on health care use and expenses and

medical/dental health. Table 3 shows the HIE data files containing only insured enrollees and notes the exclusion of certain subsets in the claims and insurance preference files.

If the HIE file you are using contains persons with an INSTAT value of 3, and you want to restrict your analysis to insured persons, you can save time and money by first dropping all persons with INSTAT = 3.

Table 3

SAMPLE GROUPS IN HIE FILES RESTRICTED TO INSURED ENROLLEES

Sample Group	Claims Line-Item Series (LI1-LI29)		
	FFS Line Items	HMO Line Items	FFS Comparison
Fee-for-service	X	--	X
HMO experimental group	X	X	X
HMO control group	X	X	X

Sample Group	Aggregated Claims Series (AC1-AC9)				
	FFS Ann. Exp.	FFS Visits	FFS Episodes	HMO Ann. Exp.	HMO Visits
Fee-for-service	X	X	X	X	X
HMO experimental group	X	X	--	X	X
HMO control group	--	--	--	X	X

Sample Group	Medical Disorders Series (MD1-MD2)		Health Status and Attitudes (HS1-HS2)
	Adult	Child	
Fee-for-service	X	X	X
HMO experimental group	X	X	X
HMO control group	X	X	X

Sample Group	Insurance Preference (IP1-IP2)	Dental Examinations (DE1)
Fee-for-service	X	X
HMO experimental group	--	X
HMO control group	--	X

Note: X means that the sample group is represented in the files indicated. Because of deliberate exclusions or nonresponse, not all members of the group may be represented; details are provided in the relevant codebooks. A dash (--) means that the group is not represented.

Sample Differences by Experimental Treatment. As an indication of sample sizes for analysts who want to compare FFS and HMO delivery systems or study the effects of different coinsurance rates within the FFS system, Appendix D tabulates the number of insured enrollees at the beginning and end of each year by experimental treatment.

Extent of Participation. Though as a group insured enrollees provided the most longitudinal data, several factors altered the extent of individuals' participation, and hence the scope of their data. Users should consider whether those factors, listed below, affect the samples intended for their analyses.

- Term of enrollment. Variable ENRTERM on MS2 indicates whether the enrollee was assigned to a three-year or five-year term.
- Late arrival/premature departure. Newborns and a few others in the HMO control group joined the experiment after their families enrolled. Variable ADDYR on MS1 indicates the administrative year of their arrival, and variable STARTD on MS1 gives the exact date. Similarly, some enrollees left the experiment before completing their assigned term, through attrition (voluntary withdrawal), termination by the HIE, or death. On MS1, variable DROPYR indicates the administrative year of such departures, ENDEXP gives the exact date, and RENDEXP provides the reason. For those who died, RDEATHD in MS3 provides the date of death based on more complete information than DEATHD in MS1. See Appendix D for the number of additions and departures per year.
- Periods of ineligibility. The insured category (INSTAT = 1 or 2) signifies "ever" insured so includes persons who were temporarily uninsured during their enrollment. The most common reason for uninsured periods was suspension, or denial of insurance benefits. The HIE suspended enrollees who became ineligible for the experiment for a period expected to be temporary--by residing outside the United States, for example. Suspension could lead to either reinstatement or termination. A less common reason for temporary ineligibility was beginning

the experiment as a person of interest and later becoming insured (possible only for members of the HMO control group). The easiest way of determining whether an ever-insured enrollee was continuously insured is to refer to variable PERINT (ever a person of interest?) and SUSPD1 (first suspension start date) on MS1. Other variables described below indicate the proportion of time an enrollee was insured every year.

To define a sample of persons who completed their full enrollment terms and were continuously insured, follow the steps listed below:

1. Using MS1, choose persons who meet all of the following criteria:
 - INSTAT = 1 or 2
 - PERINT = 0
 - ADDYR = 0 and DROPYR = "."
 - SUSPD1 = "."
2. Using MS2, check variable ENRTERM to distinguish persons with three- and five-year enrollment terms.

Enrollment-Exit Sample Differences. Medical and dental health data were collected at enrollment and at exit, three or five years later. Enrollees who joined the experiment late or departed prematurely lack enrollment or exit data, respectively. Further detail about enrollment-exit sample differences is in Sec. VI and the relevant codebooks.

Enrollment (hence exit) dates vary by site because the HIE enrollment process took place over a period of two years (see Table 1 above). As a result, the date that enrollment or exit data were collected is not the same for all enrollees.

Variable ENRDATE on both MS1 and MS2 identifies the formal date of enrollment for each person, which is the date the enrollment contract took effect. All members of an insured family have the same ENRDATE value. If an enrollee joined the experiment after the family's enrollment date, he or she will have a different value from other members of the family on variable STARTD (start of insurance coverage or individual's enrollment) in MS1.

Exit dates differ not only because of the intersite enrollment differences but also because of the different enrollment terms.

Families. The primary grouping of enrollees is by family, and the HIE distinguishes an *economic* and an *insurance* family. The economic family consists of related persons residing together who shared income. The insurance family consists of those who were covered under the same experimental treatment (HIE insurance plan or HMO control group). The insurance family, which excludes persons of interest, is thus either identical to or a subset of the economic family.

Variable BFAMILY on MS2 identifies members of the economic family at baseline, and variables IFAMILY1-IFAMILY6 in MS1 identify members of the economic³ family at the beginning of each contract year. (For those who completed the experiment, IFAMILY4 and IFAMILY6 indicate family members at the time of exit from three- and five-year enrollment terms, respectively.)

The contract year was the administrative unit of time in the HIE. The first contract year began on the family's enrollment date (variable ENRDATE on MS1 and MS2) and ended one year later. The second contract year began on the first anniversary of enrollment and extended to the next anniversary date, and so on. Contract years do not necessarily coincide with calendar years. At the beginning of the second and each subsequent contract year of a family's assigned enrollment term, HIE staff formally recognized any changes of eligibility and family composition that had occurred since the last anniversary date.⁴ Examples of eligibility changes are suspension, reinstatement, attrition, and termination. Examples of family composition changes are birth, adoption, moving in or out, divorce, and death. Family composition changes triggered the assignment of a new IFAMILY identifier for all members at the beginning of the *next* contract year.⁵

³Variable labels in the MS1 codebook erroneously define IFAMILY-IFAMILY6 as pertaining to the *insurance* family.

⁴To obtain the actual date on which an enrollee's second and subsequent contract years began, add 10,000 and its multiples to the enrollee's ENRDATE value. That is, add 10,000 to obtain the beginning of the second contract year, 20,000 for the third contract year, and so on.

⁵MS1 can be used to trace changes of family composition from year to year; see Ref. 6, pp. 22-24.

Therefore, to identify members of an economic family at the beginning of contract year i , find all persons with the same value on variable IFAMILY i . Members of the insurance family for that economic family will have a value of 1 or 2 on variable INSTAT and a value of 0 or greater on variable TIME i in MS1. Variables TIME1-TIME5 show the proportion of time an enrollee was insured in contract years 1 through 5.⁶ Remember that newborns or others who joined the family or became insured during year i will not have a value on an IFAMILY variable until year $i + 1$. And if by chance the newcomer moved in and out (or was born and died) *within* year i , the master sample series does not record the person's family. Midyear changes were not formalized because of the administrative and analytic burden entailed.⁷

We illustrate the process described above by showing how to determine insurance families and their members at the beginning of contract year 2, using MS1:

1. Drop all persons with IFAMILY2 = blank. They were not present at the beginning of contract year 2.
2. Drop all persons with values of TIME2 = ".". They were uninsured during year 2.
3. Sort by IFAMILY2 to group the members of insurance families.

Knowing an enrollee's assigned enrollment term (variable ENRTERM on MS2) can be important for interpreting eligibility and family information in MS1, especially for contract years 4 and 5. For example, a blank on variable IFAMILY5, signifying "not present," could mean that

⁶If comparing FFS and HMO enrollees in Seattle, substitute variables TIMET1-TIMET5 in MS3 for the TIME1-TIME5 variables in MS1. The reasons are explained in Ref. 7, pp. 19-21.

⁷We can suggest an indirect way of identifying the family of a baby who was born and died between consecutive enrollment anniversaries. Using MS2, find the head of the newborn's enrollment family (variable XPERSON) and determine the members of XPERSON's family at the beginning of the contract year in which the baby was born. It is possible that the family split between the time of enrollment and the baby's birth, so that XPERSON was no longer in the family at the birth. However, XPERSON is usually the female head, so that possibility is rare.

the enrollee attrited from a five-year term or exited from a three-year term. Similarly, a value of "." on TIME5 covers both those who were uninsured and those who were not present during contract year 5. Knowing that the enrollee's term was three years would remove the ambiguity from those IFAMILY5 and TIME5 values.

A value of 0 on the TIME variables could mean that the enrollee was suspended or insured less than one day. If a suspension was involved, a date falling within the contract year covered by the relevant TIME variable will appear in one of the "suspension date" variables on MS1 (SUSPD1, SUSEND1, SUSPD2, SUSEND2). If the suspension was followed by termination, the enrollee's date on variable SUSEND1 or SUSEND2 (first or second suspension end date) will match the date on variable ENDEXP (date of leaving the experiment). Different dates on those variables indicate that the enrollee was reinstated after the suspension.

Households. Analogous to the identifiers for the economic family are a set of identifiers for the household. Variable BHH on MS2 identifies members of the baseline household, and variables HH1-HH6 on MS1 identify household members at enrollment and the beginning of contract years 2 through 5 (and at exit for those who completed three- and five-year terms). Most *families* and *households* are identical; the household identifier only helps pinpoint multifamily households. If those with the same BHH or HH_i value have a different BFAMILY or IFAMILY_i value, they are members of a household with more than one insured family.⁸ An example is a household consisting of a couple and their self-supporting son.

Relationships. MS1 and MS2 contain information about relationships. At the same times denoted by IFAMILY1-IFAMILY6 and HH1-HH6, variables RELFAM1-RELFAM6 (and RELHH1-RELHH6) in MS1 indicate the enrollee's relationship to the family (household) head--female if present, otherwise male. Note that the relationship codes changed over

⁸Uninsured families (families of interest) are excluded from the sample in MS1. To check whether an uninsured family *might have* resided with an insured family, you could compare BHH and BFAMILY identifiers. Persons having the same BHH values but different BFAMILY values could be a multifamily baseline household that continued into the enrollment period with one of the families uninsured; there is no way of knowing for sure, however.

time, so two codes may have the same meaning (for example, both codes 5 and 21 include "child" and "stepchild").

To illustrate the use of the RELFAM variables, let us say you want to determine the insurance families at the beginning of contract year 2 who had more than one household head:

1. Follow steps 1-3 listed above, p. 44.
2. Check the RELFAM2 values within each family group. If more than one person has a value of RELFAM2 = 1, the group is a multihead family.

Viewing family relationships from another perspective, variable XPERSON on MS2 identifies the female (or if absent, male) head of the participant's family at enrollment.

Never Insured Participants

If you are interested in baseline demographic and socioeconomic data for HIE participants, you might want to add the baseline-only and adjunct-enrollee groups to your analytic sample. Table 4 shows sample representation in the HIE data files that are not restricted to insured enrollees. Tables 3 and 4 together show sample representation in all HIE files.

Like enrollment and exit dates, the dates of baseline information differ according to site. In the MS2 codebook (Ref. 8), Appendix C indicates the dates of administration for the screening, baseline, and enrollment instruments used as data sources for MS2 variables.

Table 4

SAMPLE GROUPS IN HIE FILES NOT RESTRICTED TO INSURED ENROLLEES

Sample Group	Master Sample Series			Medical History
	MS1	MS2	MS3	Questionnaire Series (MH1A-MH9B) ¹
<i>Baseline-only participants</i>				
PEG-period-only	--	X	X	X
All others	--	X	X	--
<i>Insured enrollees</i>				
Fee-for-service	X	X	X	X
HMO experimental group	X	X	X	X
HMO control group	X	X	X	X
<i>Adjunct enrollees</i>				
Dayton control group	--	X	X	X
Persons of interest	X	X	X	--
Families of interest	--	X	X	--

NOTE: X means that the sample group is represented in the files; a dash (--) means that the group is not represented.

¹Entries apply to all files in the series. Members of sample groups represented by a dash may appear in these files; such "accidents" were retained because of the importance of preserving primary data.

V. USING CLAIMS DATA

This section provides guidance for using HIE claims data. Table 5 summarizes the claims files and indicates the shorthand names used in this section. Primary data were taken directly from claims submitted during the experiment or (for Seattle HMO participants) were abstracted from HMO records. Derived data are primary data that have been aggregated in various ways for analysis.

The line-item files consist of primary data on diagnoses, procedures, and charges for each health care service or hospital stay. Those data are appropriate for detailed studies of the treatment for a particular diagnosis or frequency of a given procedure.

For many analyses users will only need to refer to derived data. Variables in the visit, annual expenditure, and episode files aggregate charges by visit, person-year, and episode of treatment. Thus they incorporate many of the manipulations analysts normally make to primary data. Visit and annual expenditure data were designed for analyses of how frequently people go to the doctor. Episode data indicate when treatment starts, how long it lasts, and how people adjust their demand for health care in the expectation of satisfying/not satisfying the annual deductible or out-of-pocket expense limit specified in their insurance policy.

Some analyses will benefit from the use of derived data augmented by primary data. The line-item files preserve details that were omitted in aggregating data for the derived files. Examples of such details are listed below:

- Treatment history/status. For each of four possible diagnoses, the line items indicate the nature of the problem for which the service was rendered--chronic, acute, health maintenance (well care), or pregnancy--and (if a chronic or acute condition) whether this was an initial or repeat visit. Treatment history/status information is carried over to the episode files but not the visit files.

Table 5
SUMMARY OF CLAIMS DATA FILES

Description	Type of Data	Short Name	Files Included ¹
<i>Line-Item Files</i>			
FFS services, drugs, supplies	Primary	FFS line items	LI1-LI14
HMO services, drugs, supplies	Primary	HMO line items	LI15-LI25
FFS data in HMO format for comparison of FFS-HMO costs	Primary	FFS comparison	LI26-LI29
<i>Visit Files</i>			
FFS visits--details of outpatient, inpatient, and dental visits	Derived	FFS visits	AC2-AC4
HMO (and Seattle FFS) visits-- details of outpatient and inpatient visits	Derived	HMO visits	AC8-AC9
<i>Annual Expenditure Files</i>			
FFS annual expenditures	Derived	FFS annual expenditures	AC1
HMO (and Seattle FFS) annual expenditures	Derived	HMO annual expenditures	AC7
<i>Episode Files</i>			
FFS treatment episodes-- details	Derived	Individual episodes	AC5
FFS treatment episodes-- annual counts and expenses	Derived	Annual episodes	AC6

¹Files are denoted by reference numbers from Sec. II.

- Symptom or reason for the visit. The line items indicate up to three symptoms/reasons for each visit.
- Medical/surgical/dental procedures. The line items contain CRVS codes representing all procedures performed during a visit, but only four CRVS codes are carried over to the outpatient visit files and eight to the inpatient visit files.¹ Similarly, the dental line items show both the specific dental procedures performed (e.g., filling, extraction) and the number of each, but the dental visit file shows only the types, indicated by a maximum of four ADA (American Dental Association) procedure group codes.
- Procedure-diagnosis linkage. The line items link the procedures performed with their associated diagnoses, but the visit files do not, so if more than one diagnosis is indicated the procedures associated with each may be unclear.

Specific examples and methods of recapturing such details are given below, in the section on linking data.

GENERAL CAUTIONS

When using claims data, be alert for coding and billing errors. A clerk might easily have mistranscribed a physician's notes regarding diagnoses or procedures performed. The patient or provider could have indicated the wrong date of health care services, complicating attempts to link all services associated with, say, an office visit to a given physician.

For the sake of fidelity, we made very few corrections to the data. In the line-item files only obvious errors in location of service were corrected. For example, "inpatient" was changed to "outpatient" if the patient had not been admitted to the hospital. A few additional errors were cleaned up in the derived data files. For example, the individual episode file (AC5) reflects the inconsistency in the line items that

¹In very few visits did the number of services exceed the CRVS code limit--no more than 4 percent in any visit file. Nearly all of the omitted procedures pertain to radiology or pathology.

maternity services were sometimes entered on the mother's bill, sometimes on the newborn's. The discrepancy is resolved in the annual episode file (AC6) and in the other derived files, where all maternity data are assigned to the mother. Because of such corrections, identified in the derived variable codebooks, it is not a straightforward task to replicate the derived data from the primary (line-item) data.

The HIE claims data were collected for billing purposes. Though valuable for usage analyses, they may not provide a complete picture of health care usage--only usage for which claims were filed.² This means, for one thing, that the service had to be covered by the HIE or (for those in the HMO control group) by the insurance package the enrollee's employer had purchased from the HMO. HIE insurance benefits covered a comprehensive range of services, detailed in Appendix A. In addition, the expected payment for the service had to be thought worth the inconvenience of filling out the claim form. For example, the enrollee or provider might not bother to fill out a claim for a free followup visit because there would be no reimbursement.³

LINKING CLAIMS DATA

Derived data in the HIE claims files were designed to serve general-purpose as well as specific RAND analyses. The following paragraphs cite examples of how users with special interests might tailor the claims data to their analyses by linking records within and across files.

Linking within a File: Total Cost of an FFS Outpatient Visit

Depending on the file or the analysis, you may wish to perform additional aggregation on derived data. For example, in the FFS outpatient visit file (AC2) the variable COVAMT (covered amount) represents the highest level of cost aggregation per visit but may not

²This caution applies more to FFS than to HMO claims data because HMO "claims" were abstracted from GHC records, not filed by the enrollee or provider.

³Reference 9 estimates the percentage of physician claims that were not filed in South Carolina and Massachusetts.

indicate the total cost associated with the visit. That is because COVAMT includes the charges for only a single combination of patient, provider, and date. Costs for x-rays, lab tests, or any other procedures resulting from the visit are excluded if they involved another provider or date.⁴

Outpatient visits most likely to generate a related procedure are the primary-care visits indicated in the variable FTF ("face-to-face" general health visit). FTF visits involved direct evaluation or treatment (excluding mental health). If the FTF physician ordered another procedure, it may appear in the variable RAP (radiology-anesthesiology-pathology-only visit). You can approximate the full costs associated with an FTF visit by the following method:

1. For each FTF visit, devise rules for linking related RAP visits, using enrollee identifiers, later date of service (within a reasonable interval), and referring physician if available (variable REFERBY), which should be the same as the FTF physician (variable PROVID).
2. For each RAP visit identified, assess whether it is likely to be a consequence of the FTF visit.
3. If so, add the COVAMT value for the RAP visit to the COVAMT value for the FTF visit.⁵

Linking across Files

The cross-file link you will probably make most often is between derived and primary data, in particular searching the line-item files for information to add to derived data. We offer several general suggestions to facilitate that process. First, you can save time and money by confining your search to the precise line-item files containing the information sought. If you are interested only in medical information, for example, you would ignore at least the dental line items in files LI10 and LI11.

⁴In the inpatient and dental visit files (AC3 and AC4), COVAMT does not present the same problem. For hospital and dental patients, such related procedures were most likely performed by the same providers and hence are included in COVAMT.

⁵For further discussion and examples of this method, see Ref. 10.

Next, three types of variables are necessary in linking derived and primary files:

- Enrollee or patient identifier, always variable PERSON.
- Provider of the service--for example, physician, emergency room, hospital, pharmacy. The variable name may differ according to the file. Remember that providers are identified by type and specialty only in AC2, AC3, AC8, and AC9. Provider variables in the other claims files show an identifier that must be linked with the provider file (RF2) to obtain more complete information.
- Relevant date--for example, date of service, date a prescription was written, or (regarding inpatients) the period extending from seven days before the admission date to seven days after the discharge date. The variable name may differ according to the file.

Finally, you must observe certain variable naming conventions in file-linking specifications. The names of primary and derived variables have different forms. Most primary variables are numbers prefixed by DEI ("data element indicator"); derived variable names are descriptors. For example, the physician, dentist, or other provider of the service is identified by variable DEI5502 on the line-item files, by variable PROVID on the visit files. If using the SAS version of the data, you must translate the DEI variables into their descriptive counterparts to complete the matching processes indicated below. Variable descriptions in the derived variable codebooks provide the names of counterpart primary variables. If you are using the character version of the data, your input statement must show variable names in the descriptor form.

Visit with Line-Item Data. In the FFS outpatient and dental visit files (AC2 and AC4), variable COVAMT excludes the costs of drugs/supplies bought as a result of the visit. To estimate those costs you must link visit data with line-item data on drugs/supplies. Two methods are given below, one for determining the cost of drugs/supplies obtained from the physician, and the other for drugs/supplies bought elsewhere.⁶

⁶Linking drugs/supplies with visits is also addressed in Ref. 10.

Cost of Drugs/Supplies Obtained from the Physician

1. Match person, provider, and date of service variables for the relevant visit to the corresponding variables in LI6 and LI7.
2. If drugs/supplies are indicated, obtain their HIE-covered costs. Those costs are DEI5558 minus any noncovered costs (DEI5559) if the reason for noncoverage (DEI5560) is listed in Appendix C of the FFS visits (AC2-AC4) codebook.
3. Collapse the costs of the relevant drugs/supplies into one record and add it to the appropriate record on the visit file.

Cost of Purchased Drugs/Supplies⁷

1. Match person, provider, and date of service variables for the relevant visit to PERSON, prescriber (DEI5604), and date prescription issued (DEI5650) or date supply dispensed (DEI5603) in LI12-LI14. If close to the visit date, the date in DEI5650 or DEI5603 provides a reasonable (but not foolproof) indication of relationship to the visit.
2. If related drugs/supplies are indicated, obtain their HIE-covered costs. Those costs are DEI5558 minus any noncovered costs (DEI5559) if the reason for noncoverage (DEI5560) is listed in Appendix C of the FFS visits (AC2-AC4) codebook.
3. Collapse the costs of the relevant drugs/supplies into one record and add it to the appropriate record on the visit file.

To determine which drugs or supplies were ordered in a particular visit, use the following procedure:

⁷This method is appropriate for making the link between a visit for an acute condition and the drugs/supplies bought as a result. It is less effective in making the visit-purchase link for a chronic, health maintenance, or pregnancy condition. Drugs/supplies for those conditions could be purchased at any time, without relation to a specific visit.

Drugs/Supplies Ordered

1. Match person, provider, and date of service variables for the relevant visit to the corresponding variables in LI3 and LI4.
2. Collapse the indicated drug/supply codes into one record showing multiple drug/supply codes and add it to the appropriate record on the visit file.

Another reason for linking visit with line-item data is to add detail about services or diagnoses to visit records of interest. For example, recall that a limited number of CRVS codes representing medical/surgical procedures were transferred from the line items to the inpatient and outpatient visit files. If you want to retrieve CRVS codes for the few visits in which some codes were omitted, the procedure is as follows:

All Medical/Surgical Procedures

1. In the outpatient visit file (AC2), examine CRVS variables 1 through 4 (in AC3, the inpatient visit file, CRVS variables 1 through 8). The visits with a value in every variable are those for which more CRVS codes might be found in the line items.
2. For those visits, match person, provider, and date of service variables to the corresponding variables in LI2, LI5, and LI8. For outpatient visits, the corresponding variables are in LI5 and LI8; for inpatient visits, LI2 and the inpatient records in files LI5 and LI8. For inpatient visits, match to any date between seven days before the admission date and seven days after the discharge date. Check to be sure that any procedure occurring before or after the hospitalization in question was not part of a close but separate hospitalization.
3. Find all CRVS codes (DEI5606) associated with that combination of variables.

4. Collapse the CRVS codes into one record with multiple CRVS codes and add it to the appropriate record on the visit file. For inpatient visits, be sure to include the admission and discharge dates.

Adding procedural detail is more important for dental visits than for inpatient/outpatient visits because only the *types*, not the specific services or their number, were carried over to the derived files. To obtain those details, follow a variant of the procedure listed above:

Specific Dental Services

1. Match person, provider, and date of service variables for the relevant dental visit from AC4 to the corresponding variables in LI10.
2. Find all ADA codes (DEI5625) and associated services (DEI5602) for that combination of variables. Count the number of times the service appears or the number of tooth surfaces if warranted by the ADA code.
3. Collapse the ADA codes and associated numbers into one record and add it to the appropriate record on the dental visit file.

Episode with Line-Item Data. The abundant detail in the line items was radically pruned in aggregating data to the episode level. The episode files contain only summary information about treatment history/status and diagnosis. It is tempting to think of pulling additional data from the line-item files. Two cautions apply: (1) The search of line-item data cannot be limited to a few files, as it can be in linking visit with line-item data, so the procedure is time-consuming and costly. (2) Because of the complex rules for defining an episode,⁸ there is no certainty that a procedure or diagnosis identified in the line items is *part of* an episode, only that it occurred *at the same time*.

⁸Explained in Ref. 11.

Bearing those caveats in mind, you can link episode with line-item data as follows:

Procedural or Diagnostic Detail from Line Items

1. From AC5, select an episode of interest in terms of the enrollee (PERSON), episode start and end dates (variables BEG_DATE and END_DATE), and episode description (variable EPIS_NO).
2. For PERSON, search all line-item files for data between BEG_DATE and END_DATE.
3. For the data identified in step 2, compare treatment history codes (DEI5574, DEI5577, DEI5580, and DEI5583 in all line-item files but LI1 and LI2) with the "type" code reflected in the EPIS_NO value.
4. If treatment history and type codes match, the line-item data could be part of the episode of interest. The probability increases if the enrollee did not have multiple episodes during the period in question.

Unlike the visit data, the expense variables in the two episode files include the cost of drugs and supplies. Therefore it is unnecessary to link episode data with drug/supply line-item data to capture those costs. But if you want to estimate which drugs/supplies were prescribed during an acute or chronic flareup episode, make the following link:

Drugs/Supplies Prescribed

1. From AC5, select an episode of interest in terms of the enrollee (PERSON), episode start and end dates (variables BEG_DATE and END_DATE), and episode description (variable EPIS_NO).

2. For PERSON, find all drugs/supplies purchased (DEI5650 or DEI5603) between BEG_DATE and END_DATE in LI12-LI14.
3. Assess whether they are related to the episode of interest. The likelihood is greater if the enrollee did not have multiple episodes during the period in question. If multiple episodes complicate the assessment, you could look for a match by checking the line-item variables that indicate whether the drug or supply relates to a diagnosis. In LI3, LI4, LI6, and LI7, those variables are DEI5596-DEI5599. If the supply is from LI13 or LI14, it is related to the single diagnosis identified in variable DEI5605.

Episode with Visit Data. Visit files have more procedural and diagnostic detail than do episode files. To add detail to episode records without the expense of searching all line-item files, you could link episode data to visit data as follows:

Procedural or Diagnostic Detail from Visit Files

1. From AC5, select an episode of interest in terms of the enrollee (PERSON), episode start and end dates (variables BEG_DATE and END_DATE), and episode description (variable EPIS_NO).
2. For PERSON, find all visits from AC2-AC4 (with desired procedural or diagnostic data) occurring between BEG_DATE and END_DATE.
3. Assess whether those data belong to the episode of interest. The likelihood is greater if the enrollee did not have multiple episodes during the period in question. If multiple episodes complicate the assessment, you could look for a match by checking the line-item variables that indicate whether the service, procedure, drug, or supply identified in the visit data relates to a diagnosis. In LI3-LI8, those variables are DEI5596-DEI5599. If the supply is from LI13 or LI14, it is related to the single diagnosis identified in variable DEI5605.

Episode files contain more "price" information than do visit files. That is, the individual-episode file (AC5) shows how each treatment episode affected satisfaction of the out-of-pocket expense limit specified in a family's FFS insurance plan. After the family reached that limit--called the maximum dollar expenditure (MDE)--the HIE reimbursed all further charges for covered services until the next contract year.

If you accept the assumption that a patient anticipates the cost of a treatment episode at its start, you could use MDE status information from the individual-episode file, along with visit data, to determine the effects of progressive MDE fulfillment on health care use. The following steps show how to obtain that information.

MDE Fulfillment Data from Individual-Episode File

1. For each enrollee of interest (PERSON), sort visits (from AC2-AC4) by date of service (variable HDATE in the visit files) and sort episodes (from AC5) by BEG_DATE in the individual-episode file. The value for variable MDE on any episode record tells the family's MDE at the beginning of the relevant contract year. The value for variable MDE_END on an episode record is the amount of remaining MDE expected by the patient from that episode's BEG_DATE until the next episode's BEG_DATE.
2. Find an HDATE occurring between two BEG_DATES. The MDE_END value on the episode record represented by the first BEG_DATE is how much MDE the patient thought remained at the time of the visit represented by the HDATE. (This assumes that the patient correctly anticipated future expenditures for episodes then in progress.) If the MDE_END value is a negative number, the MDE has been exceeded.

CHOOSING ANALYTIC SAMPLES

To select analytic subsamples using particular demographic and eligibility criteria, it is necessary to refer to the master sample series of files (MS1-MS3). See Sec. IV for guidance on choosing a subsample of enrollees who were insured for the entire period in question, not suspended or otherwise ineligible to file claims for part of the time.

FFS- or HMO-Only Participants

For some claims-related analyses you will want to confine your sample to FFS or HMO participants exclusively. To do so, use MS2:

- To define an FFS-only subsample, select all persons from the MS2 sample with INSTAT = 1 and PLAN ≠ 30.
- To define an HMO-only sample, select all persons from the MS2 sample with INSTAT = 2 or PLAN = 30. To distinguish between the two groups of HMO participants, again using MS2, the HMO experimental group includes those with PLAN = 30, and the HMO control group includes those with INSTAT = 2.

Users/Nonusers

Analyses of health care use must account for nonusers to accurately determine usage in the total population. For that purpose the annual expenditure (AC1, AC7) and annual episode (AC6) files include a dummy record for each enrollee who had no visits or episodes in a given contract year.⁹ The line-item and visit files have no dummy records. You can add dummy records to those files as follows:

1. From MS1-MS3, use the desired demographic or eligibility criteria to select a population of interest.
2. In MS2, determine which of those enrollees were assigned to three-year enrollment terms and which to five-year (values of 3 and 5 on variable ENRTERM).

⁹The dummy record identifies the enrollee but shows zero or "missing" values on all substantive variables.

- 3a. Drop from your sample the PERSON numbers of enrollees with values other than 1.00 on any relevant TIME variable (time insured during a specified contract year) in MS1. For three-year enrollees the relevant TIME variables are TIME1-TIME3; for five-year enrollees, TIME1-TIME5.
- 3b. Among the people step 3a excludes are the relatively few enrollees, mostly newborns, who joined the experiment after their families enrolled. Unless they happened to be born on the first day of a contract year, they were insured for a fractional part of their first year and have a value other than 1.00 on the TIME variable for the year they joined. If you want to include them, restore the PERSON numbers of those with (1) a value greater than 0 on variable ADDYR (contract year participation began), (2) a value greater than 0 on the TIME variable for their first year, and (3) a value of 1.00 on the TIME variables for all subsequent years of their enrollment term.¹⁰
4. Create a record for each contract year in which an enrollee was insured the full time. The record should contain PERSON and CONTYR (contract year) values.
5. Match those PERSON and CONTYR values with the same variables on the file of interest--line-item (prefix LI files) or FFS/HMO visit (AC2-AC4, AC8-AC9) file. Nonmatches show the PERSON and CONTYR values for which dummy records must be made.
6. Create the required dummy records and add them to the file of interest. Each dummy record should identify the PERSON and CONTYR and show a zero in all "count" variables and "missing" values in all expenditure variables.

¹⁰If comparing FFS and HMO enrollees in Seattle, substitute variables TIMET1-TIMET5 (time insured in years 1-5 before status change) in MS3 for TIME1-TIME5 in MS1.

FFS-HMO ISSUES

FFS Usage by HMO Participants

The FFS line-item, FFS dental visit, and HMO line-item files include services obtained by HMO participants in the fee-for-service sector, so they are useful for studying the frequency with which HMO participants sought care outside GHC, and the charges for that care. Each file treats a different set of FFS services, however. The FFS line items (LI1-LI14) and FFS visits (AC2-AC4) show services (1) unavailable at GHC (e.g., dental or chiropractic) that HMO participants *had* to seek in the FFS sector, and (2) available at GHC but preferentially sought in the FFS sector. The HMO line items (LI15-LI25) show services that may or may not have been available at GHC but that GHC treated as part of its health care and reimbursed completely. An example of the latter is emergency surgery outside Seattle.

In the HMO visit and HMO annual expenditure files, visit counts and charges designated "out of plan" also refer to FFS services, but they are defined differently in each file. In the HMO visit files (AC8-AC9), "out of plan" pertains only to the FFS services for which the HIE reimbursed HMO participants 5 percent. (Those services were different for HMO experimental and control groups; see Appendix A.) In the HMO annual expenditure file (AC7), "out of plan" pertains to any services obtained in the FFS sector.

Sector Comparisons

Comparison of FFS versus HMO expenditures cannot be made directly by comparing the FFS and HMO line item files for Seattle enrollees. A set of comparison files (LI26-LI29) was created to "translate" relevant FFS line-item data into terms comparable with the HMO line-item data.

It can be misleading to compare FFS and HMO *services* by comparing the respective line-item files because of the different bases of the files.

In the FFS line items the data are reported by the enrollee or provider; in the HMO line items the data were abstracted from HMO records.¹¹

It is impossible to compare Seattle FFS and HMO enrollees' *expenditures* by comparing the respective line-item files because the *actual* charges shown in the FFS line items are not comparable to the *imputed* charges shown in the HMO line items. The FFS comparison files provide a basis for comparison because they show charges for Seattle FFS services imputed *by the same method* used to impute charges for HMO services shown in the HMO line items, that is, on the basis of CRVS units.¹²

Charges are imputed for only a portion of the FFS line-item files: inpatient services and physician procedures billed by an institution (LI1 and LI2), physician services (LI5), and injections administered by a physician (LI8). Dental services are omitted because they were not provided by GHC. Outpatient facility services (e.g., charges by an emergency room) and drugs/supplies are also omitted. Costs for outpatient facility services and drugs/supplies cannot be imputed given the lack of standard cost units analogous to CRVS units for medical/surgical procedures.¹³ Therefore, the only expenditures that can be compared, even with the comparison files, are those for inpatient services and hospital physician services, independent physician services, and injections by physicians.

¹¹An extreme example is the distortion you would find by comparing prenatal visits in the FFS and HMO line-item files. At the HMO every prenatal visit was recorded and duly transcribed to the line items. In the fee-for-service sector, however, much prenatal care was billed as a lump sum; individual prenatal visits often went unreported because no reimbursement was expected. As a result, the HMO line items show six times more prenatal visits than do the FFS line items, with only two times more adult women in the HMO than in the Seattle FFS sector.

¹²CRVS units are unit values assigned to CRVS procedures and services based on their time requirements and relative complexity. CRVS unit values are defined in Ref. 12.

¹³Because of the imputation process and omitted services, the record for a given Seattle FFS enrollee on the HMO annual expenditure file (AC7) is not comparable to that enrollee's record on the FFS annual expenditure file (AC1).

Besides permitting the comparison of expenditures by FFS and HMO enrollees, the FFS comparison and HMO line-item files permit comparison of the value of HMO and non-HMO services. For example, you can determine what percentage of all HMO participant expenses was for non-HMO care.

The FFS comparison and HMO line-item files are also appropriate for comparing FFS and HMO services. The only comparison those files do not permit is of drugs/supplies ordered in the FFS versus HMO sectors, for which you must use FFS line items and HMO line items. No file permits comparison of the *cost* of FFS versus HMO drugs/supplies.

ADJUSTING FOR INFLATION

Given the sustained high rate of medical inflation, it is useful to express health care expenditures in constant dollars. Most RAND analyses of HIE data used three kinds of inflation-adjustment factors: *overall* for wage and income data, *medical* for medical expense data, and *dental* for dental expense data. Derived from the consumer price index of the U.S. Bureau of Labor Statistics, the factors vary by month from September 1974 through June 1982. Appendix E lists the factors and explains how to use them.

Data in most HIE files are unadjusted, for consistency with MDE figures. Because the MDE concept is central to the experimental design, it was thought important to express MDE amounts in actual dollars. However, expenditures in the HMO (and Seattle FFS) visit and annual expenditure files (AC7-AC9) were adjusted by the medical inflation-adjustment factors. Those adjustments were necessary because by design there were more HMO than FFS participants with five-year terms. Unadjusted data might have misled users into thinking that HMO use was much higher than FFS use in contract years 4 and 5, whereas inflation and the difference in enrollment term were in fact responsible.

VI. USING DATA ON MEDICAL AND DENTAL HEALTH

This section provides guidance for using a class of data pertaining to enrollees' state of health when entering and completing the experiment. These data have a common origin in the MHQ or medical screening examination (items 4 and 5 on the list of data collection instruments in Table 2 above). Eighteen files contain primary data in the form of the "raw" MHQ responses, and five files contain data derived from the MHQ or screening examination. Table 6 describes the relevant files.

We first review data collection procedures that affect the use of these data. Then, because most users are likely to begin with the derived data and add detail as necessary from the primary data, guidance is offered for derived data on health status/attitudes, medical disorders, and dental health, and then for primary data in the MHQ series.

DATA COLLECTION PROCEDURES AFFECTING DATA USE

Medical History Questionnaire

The MHQ consisted of two parts. Form A (MHQA) asked for relatively subjective evaluations of functional limitations, acute symptoms, mental and social health, general health, health habits, and satisfaction with medical and dental care. Form B (MHQB) elicited relatively objective information on enrollees' physical limitations and the presence of specific chronic or acute medical disorders. The MHQ also had different versions by age group. Persons 14 and older completed the adult version. Parents completed the pediatric version for children 5-13 years of age and the infant version for children 0-4 years of age.

To encourage completion of the instruments, the HIE paid \$5 for each completed adult MHQ and \$2 for each completed child MHQ, up to a family maximum of \$20. If the family was also scheduled for a screening examination, the head received \$20 for completing his/her own MHQ and an additional \$5 for each dependent's completed MHQ, up to a family maximum of \$50.

Table 6

SUMMARY OF MEDICAL AND DENTAL HEALTH FILES

Description	Type of Data	Files Included ¹
<i>Medical History</i>		
Adult health status, attitudes, and habits at enrollment and exit (responses to MHQA)	Primary	MH1A-MH3A
Adult medical disorders at enrollment and exit (responses to MHQB)	Primary	MH1B-MH3B
Child health status, attitudes, habits, and medical disorders at enrollment and exit (responses to MHQA and MHQB)	Primary	MH4A-MH6B
Infant health status, development, and medical disorders at enrollment and exit (responses to MHQA and MHQB)	Primary	MH7A-MH9B
<i>Health Status and Attitudes</i>		
Adult health status and attitudes at enrollment and exit	Derived	HS1
Child health status and attitudes at enrollment and exit	Derived	HS2
<i>Medical/Dental Disorders</i>		
Adult medical disorders at enrollment and exit	Derived	MD1
Child medical disorders at enrollment and exit	Derived	MD2
Prevalence of tooth decay and periodontal disease at enrollment and exit	Derived	DE1

¹Files are denoted by reference numbers from Sec. II.

The experimental design called for every enrollee to complete an MHQ twice--at enrollment and at exit three or five years later.¹ In Dayton, all enrolling participants were asked to complete form A, but only the enrollees scheduled for a screening exam (about two-thirds) were asked to complete form B. In the other sites, all enrollees were asked to complete both forms A and B.

Enrollees scheduled for a screening exam were given form A to complete and bring to the screening exam center, where they were given form B to complete. Enrollees not scheduled for a screening exam were given both forms A and B to complete and mail in prepaid envelopes. At exit, HIE staff mailed MHQ forms to enrollees who lived farther than 100 miles from the screening exam center, for return by prepaid mail.

Departing from the usual pattern, most enrollees in the South Carolina three-year group completed an MHQ at three times: when preenrolling in 1976, when formally enrolling in 1978, and when exiting in 1981. In 1978, MHQ forms were mailed to enrollees, with prepaid return envelopes, because no screening examination was given then. The relevant primary files (MH2A-B, MH5A-B, and MH8A-B) show both 1976 and 1978 data for this group. The derived files are more selective:

- In file HS1, adult health status and attitudes, enrollment variables for the South Carolina three-year group are based on 1978 data collected during formal enrollment.
- File HS2, child health status and attitudes, shows missing values for all children in the South Carolina three-year group on all enrollment variables.
- File MD1, adult medical disorders, uses 1976 MHQ data for those who preenrolled and 1978 data for those who did not preenroll.
- File MD2, child medical disorders, uses 1976 MHQ data for those who preenrolled and assigns missing values to enrollment variables for those who did not preenroll.

¹Variable DATE on each MHQ file indicates when the relevant instrument was received. This date may differ somewhat from the enrollee's formal enrollment or exit date.

After its administration in Dayton at enrollment, the MHQ was revised to clarify the questions and obtain more precise responses. The revised questionnaire was then administered in Dayton at exit and in all other sites at enrollment and exit. Most of the MHQ was untouched, but some questions were deleted and portions of the revised questionnaire are quite different from the original. If the wording of a question was changed substantively, or if the response codes were reworded or renumbered, the variable name (prefix DEI) was nearly always changed. In a few cases, however, the variable name in the Dayton enrollment and the revised MHQs remained the same, despite marked content differences. Therefore, users would be wise to compare the wording of an MHQ question in the original (Dayton enrollment) and revised (Dayton exit/all nonDayton) versions before using a variable derived from it. The codebooks for the MHQ-derived files (Refs. 13-15) append relevant parts of the two MHQ versions to facilitate that comparison.

Screening Examination

The medical screening examination was used to collect data on enrollees' body organ systems at enrollment and exit. At enrollment, a stratified random sample of families was asked to take the screening exam so that the exam's effect on the subsequent use of health care services could be measured. At exit, all enrollees were asked to take the exam.

When the appointments were scheduled, enrollees were given verbal and written instructions about how to prepare for the exam. Adults (14 and older) were asked to fast two hours before the appointment, and enrollees of all ages were asked to bring corrective lenses, hearing aids, and medications with them.

The screening examination included numerous objective tests of organ system function.² They were selected on the basis of their validity, sensitivity, specificity, standardization, reliability, logistics of performance, acceptability to participants, acceptability to the medical community, and cost. The exam was conducted by specially

²For details, see Ref. 16.

trained paramedical personnel, except for the dental portion, which was usually conducted by a dentist. The screening exam could be completed in two hours, including completion of MHQ form B.

The enrollment and exit screening examinations differed in several respects:

- The test method or subgroup screened differed for the periodontal index, blood sugar test, tympanometry, hearing aid evaluation, hand/wrist x-ray, urinalysis, and evaluation of immunization status.
- At enrollment, laboratory analyses were performed by a separate laboratory in each site; at exit, the tests in all sites were analyzed by the same laboratory.
- Only 60 percent of the enrollees were assigned to take the exam at enrollment, whereas all enrollees took it at exit.
- Enrollment screening examinations were performed manually by American Health Profiles of Nashville, Tennessee; exit screening examinations were conducted by Health Testing Institute of Boca Raton, Florida, using an automated testing system.

At exit, persons who lived more than 100 miles from the screening exam center were asked to have a local physician perform a screening examination that included only a small portion of the tests in the regular HIE screening exam. Enrollees who received those out-of-area exams are identified by variable OUTAREAX in MD1 and MD2.³ A dental examination was not part of the abbreviated out-of-area screening exam, so out-of-area enrollees lack exit data in the dental exam file.

The South Carolina three-year enrollee group took the screening examination in 1976 when they preenrolled and in 1981 when they exited. As mentioned above, no screening examinations were conducted in 1978 when the group formally enrolled. Therefore, members of the South Carolina three-year group who did *not* preenroll in 1976 have no enrollment screening exam data.

³No out-of-area exams were given at enrollment because the enrollment sample included only area residents.

USING DERIVED DATA ON MEDICAL HEALTH

Health Status and Attitudes toward Health Care

To facilitate HIE research on how different levels of insurance affect state of health and satisfaction with health care, raw MHQ responses were transformed into measures, primarily in the form of Likert or Guttman scales.⁴ HS1 and HS2 present the measures for adults (14 and older) and children (0-13 years old), respectively. Seven topics are covered:

- Physical health--ability to perform normal activities.
- Mental health--symptoms of mood and anxiety disorders; control over feelings, thoughts, behavior (persons five and older).
- Social health--personal interaction and social participation (persons five and older).
- General health--perceptions of personal health based on objective information plus subjective evaluation by self or parent.
- Health habits--smoking, alcohol consumption, weight, exercise (adults only).
- Satisfaction with medical care--efficacy, cost, physical environment, technical quality, art of care, accessibility/convenience, availability, continuity, and overall satisfaction (adults only).
- Satisfaction with dental care--efficacy, cost, technical and interpersonal skills, pain management, accessibility/convenience, availability, continuity, and overall satisfaction (adults only, exit only).

Many of the scales in HS1 and HS2 are interrelated, and some are subscales of others. Users should therefore check the input variables and construction carefully in the codebook (Ref. 13) before assuming that the scale in a given variable is independent of the scale in another variable.

⁴References 17-29 discuss the method of transforming the raw data and examine the reliability, validity, and power of the resulting measures.

If data were missing in any of the raw variables used to derive a scale, we imputed a value for the scale rather than assign a missing value. The imputation methods are described in Ref. 13, Sec. III.

Beyond HIE research purposes, the data in HS1 and HS2 are useful for testing new hypotheses related to health status and attitudes, conducting comparative analyses of the HIE and other population samples, and studying the relationship between these measures and use or cost measures from HIE claims files.

To adjust health status/attitude data for specific medical conditions, you can link data in HS1 or HS2 with data in MD1 or MD2, matching on variable PERSON. For example, to compare the mental health status of adults with and without an active ulcer, you would merge variable ULCEMHQ or ULCEMHQX (ulcer status at enrollment and exit, respectively) from MD1 with variable MHI or MHIX (mental health index at enrollment and exit, respectively) from HS1, matching on PERSON.

Similarly, you can link primary data from the MHQ series (MH1A-MH9B) to cross-check HS1 or HS2 data, or to create other derived variables related to health status and attitudes.

Medical Disorders

To facilitate longitudinal HIE analyses of enrollees' health, measures were constructed for 17 adult medical disorders and 4 child medical disorders. Derived from MHQA and MHQB responses and from screening exam tests, the measures are presented in the two medical disorder files. Those are the only HIE files containing data from the medical portion of the screening exam.

The particular disorders were chosen because they are easily traceable over time and responsive to changes in the level and quality of medical care. File MD1 presents measures of the following conditions among adults (persons 14 and older):

Acne	Hypercholesterolemia
Anemia	Hypertension
Angina pectoris and ECG abnormalities	Joint disorders
Chronic obstructive airway disease	Kidney disease and urinary tract infection
Congestive heart failure	Peptic ulcer disease and dyspepsia
Diabetes mellitus	Sleeping pill and tranquilizer use
Hay fever	Surgical conditions ⁵
Hearing loss	Thyroid disease
	Vision impairments

File MD2 presents measures for the following conditions among children (aged 0-13):

Allergic conditions
Anemia
Middle ear disease and hearing impairment⁶
Vision impairments

For most disorders a status variable enables analysts to determine whether an enrollee has the disorder, and other variables indicate the disorder's effect on the enrollee in terms of pain, worry, restricted activity, or days in bed.

Common Sample

The enrollee sample is identical in the health status/attitude and medical disorder files. It includes enrollees who

- Participated in the experiment as an insured enrollee (HIE-insured or HMO-insured).
- Completed at least one of the following: enrollment MHQ, exit MHQ, enrollment screening examination, exit screening examination.

⁵Includes hernia, varicose veins, hemorrhoids, and tonsil/adenoid disease.

⁶Also includes data on colds.

The adult files (HS1, MD1) contain 5,871 records, each representing an enrollee who was 14 or older at enrollment or at exit. The child files (HS2, MD2) contain 2,840 records, each representing an enrollee who was younger than 14 at enrollment or at exit. Thus, enrollees who turned 14 between enrollment and exit have records in both adult and child files. There are 566 such records, i.e., records with the same PERSON numbers in HS1 and HS2 and the same PERSON numbers in MD1 and MD2.

Reasons for Missing Data

In the child files (HS2, MD2), each of the 566 records mentioned above shows substantive values in variables based on enrollment data but shows systematically missing values in variables based on exit data. Conversely, in the adult files (HS1, MD1), each of the 566 records has substantive values in variables based on exit data and systematically missing values in variables based on enrollment data.

Other reasons for missing values are outlined in the codebooks for HS1, HS2, MD1, and MD2 (Refs. 13-15). Briefly, they include joining the experiment after the initial enrollment process, not completing the assigned enrollment term, not answering individual questions, not returning the questionnaire (either because of experimental design or nonresponse), not completing any part of the screening examination (either because of experimental design or nonresponse), and not completing individual screening exam tests.

We leave it to users to determine which reasons apply for particular missing data. Merging information from the following variables could be helpful in explaining missing data:

<i>Variable</i>	<i>Meaning</i>	<i>File</i>
AGE	Age at enrollment	MS2
STARTD	Start of insurance coverage or individual's enrollment	MS1
ENRDATE	Family enrollment date	MS1, MS2
RENDEXP	Reason for leaving experiment	MS1
SITE	Site of residence	All files

MHQA	Completed MHQA at enrollment	MD1, MD2
MHQAX	Completed MHQA at exit	MD1, MD2
MHQB	Completed MHQB at enrollment	MD1, MD2
MHQBx	Completed MHQB at exit	MD1, MD2
SCREENED	Completed screening exam at enrollment	MD1, MD2
SCREENDX	Completed screening exam at exit	MD1, MD2
OUTAREAX	Completed out-of-area screening exam at exit	MD1, MD2
PEGFLAG	Member of South Carolina three-year group who did not preenroll	MD1

USING DERIVED DATA ON DENTAL HEALTH

Despite their common origin with other health status data, HIE dental health data are sufficiently different in organization and content to warrant separate discussion.

File DE1, the main repository of dental health data, provides objective information about the prevalence of tooth decay and periodontal disease at enrollment and exit for enrollees three years and older. Data on satisfaction with dental care, available only for adults at exit, are in file HS1, derived from MHQ responses documented in MH3A.

Content

DE1 data are derived from the dental portion of the enrollment and exit screening exams. The file presents several summary indexes:

<i>Index</i>	<i>Significance</i>
DMFT (decayed-missing-filled teeth)	Effects of decay on permanent teeth
DMFS (decayed-missing-filled surfaces)	Effects of decay on permanent teeth
def (decayed-extracted-filled teeth)	Effects of decay on primary teeth
defs (decayed-extracted-filled surfaces)	Effects of decay on primary teeth
OHI-S (simplified oral hygiene)	Degree tooth surfaces are covered by plaque and calculus
PI (periodontal index) ⁷	Severity of periodontal disease

⁷Average of periodontal ratings for individual teeth. Periodontal data were collected only for persons 14 and older at enrollment and only for persons 12 and older at exit.

Other variables present simple counts of teeth (or surfaces) in the decayed, extracted, missing, and filled categories. The total number of teeth (or surfaces) is also given so that users can calculate what proportion of an enrollee's teeth fit each category.

Sample

DE1 contains 7,317 records, each representing an enrollee who

- Participated in the experiment as an insured enrollee (HIE-insured or HMO-insured).
- Was three years or older at exit.
- Completed a dental screening exam at enrollment or exit or both.

There is substantial overlap between the sample in DE1 and that in HS1/HS2 and MD1/MD2. Every PERSON in DE1 should match a PERSON in the HS and MD files. Persons in the HS and MD files who do not have a record in DE1 were either

- Younger than three years at exit.
- Not selected to have a screening exam at enrollment *and* did not complete assigned enrollment term or lived farther than 100 miles from the screening exam center at exit.

As with the other files described above, not all enrollees in the DE1 sample have complete data. Reasons for missing data are described in the DE1 codebook, Ref. 30.

Analysts might find it useful to link dental data with other health files. For example, if you want to study the relation between an enrollee's dental health status and his/her satisfaction with dental care, you can link DE1 and HS1 files, matching on PERSON. Remember that dental satisfaction variables apply only to adults (14 and older) at exit. You might also want to link primary data on dental health habits, matching on PERSON. The MHQB for adults and children had a battery of

questions on teeth and gums and fluoride treatment, and both forms A and B for children and infants (except the Dayton enrollment MHQA) asked about fluoride use.

USING PRIMARY MEDICAL HISTORY DATA

Table 7 shows the number of records per file in the MHQ series. Each record represents a participant--*insured or not*--who completed the relevant MHQ instrument during the experiment.⁸ This differs from the exclusively insured sample in the derived files HS1/HS2, MD1/MD2, and DE1. As noted in Sec. III, the files are organized by instrument, so most members of the MHQ series sample--who completed both enrollment and exit MHQs--have data in four files, one for each MHQ form at enrollment

Table 7
NUMBER OF RECORDS IN MHQ SERIES FILES

Instrument (Files) ¹	Infants	Children	Adults
Dayton enrollment MHQA (MH7A, MH4A, MH1A)	187	351	1213
NonDayton enrollment MHQA (MH8A, MH5A, MH2A)	825	1797	5496
Exit MHQA (MH9A, MH6A, MH3A)	688	1393	5247
Dayton enrollment MHQB (MH7B, MH4B, MH1B)	120	236	840
NonDayton enrollment MHQB (MH8B, MH5B, MH2B)	826	1791	5512
Exit MHQB (MH9B, MH6B, MH3B)	688	1390	5246

¹ File reference numbers refer, respectively, to each of the three files represented by the entries across the row.

⁸Codebooks for the MHQ series of files erroneously imply that the series sample is exclusively of insured persons, although the values on variable INSTAT in each file correctly distinguish insured from uninsured members of the sample.

and one for each MHQ form at exit. For example, participant XYZ, a 15-year-old at enrollment in Dayton who completed both enrollment and exit MHQs, has data in MH1A, MH3A, MH1B, and MH3B.

Systematic exceptions to the generalizations above are as follows:

- Dayton participants (about one-third) who were not scheduled for a screening exam at enrollment therefore did not complete the MHQB and lack data in the enrollment MHQB files MH7B, MH4B, and MH1B.
- South Carolina three-year enrollees who completed an enrollment MHQ in both 1976 and 1978 have records and data for both years (i.e., two records) in MH8A-B, MH5A-B, or MH2A-B. Analysts should choose one of the two records to accurately represent the South Carolina enrollment sample. Variable SOURCIND in the files noted distinguishes the 1976 (SOURCIND = 5), from 1978 (SOURCIND = 3), records of those persons.
- Those who have data only in enrollment files are persons who refused the offer of HIE enrollment or were found to be ineligible, members of the South Carolina PEG who withdrew before formal enrollment, the Dayton control group, and enrollees who did not complete their assigned enrollment terms (died, attrited, or were terminated).
- Those who have data only in exit files are persons who enrolled after the initial enrollment process, most of them newborns.

Some MHQ responses may be improbable because of respondent inaccuracies. Respondents occasionally answered questions they were instructed to skip or failed to answer applicable questions, and parents may have mistakenly responded for themselves in filling out their child's or infant's questionnaire. If a response was outside the normal range for a given question, or answered a question that should have been skipped, the actual questionnaire was inspected to verify that the data file reflected the original answer. If the response duly appeared on the original questionnaire, the response value was left unchanged on the file.

MHQ data are useful with MHQ-derived data and dental health status data for adding detail about health status, attitudes, and disorders. To link the relevant primary and derived files, match on variable PERSON. It is not a good idea to try to replicate the derived data from the primary data; the imputation methods and other transformation procedures make it a complex process.

The MHQB asked about more than the 17 adult and 4 child disorders addressed in the medical disorders series. Therefore, analysts studying use-expenditure issues with the HIE claims data may want to access the primary data to learn more about enrollees' health status than the derived data show. Again, you can link the relevant MHQ series files with claims data by PERSON number.

Appendix A

HIE BENEFIT COVERAGE

The lists below distinguish the services covered and excluded by plans A-0 of the HIE experimental treatments.¹ Those plans represent 15 of the 16 experimental treatments described in Sec. I. Benefit coverage for the HMO control group, the remaining treatment, cannot be summarized because HMO control families retained whatever benefit package their employer had purchased from the HMO.

The same services were covered and excluded for all assigned to plans A-0. The covered service list is subdivided because the location of service differed for members of the HMO experimental group (plan 0), as explained below. For enrollees in the fee-for-service plans (A-N), the charges for covered services (both groups 1 and 2) applied toward meeting a family's annual out-of-pocket expense limit or, when that limit was reached, were completely reimbursed by the HIE. For those in the HMO experimental group (plan 0), covered services in group 1 were available at GHC; covered services in group 2 were available outside GHC and completely reimbursed by the HIE.

To encourage the reporting of non-HMO care, the HIE reimbursed all HMO participants 5 percent of the cost of certain services obtained in the fee-for-service sector. For the HMO experimental group, they included any service in group 1 that the enrollee chose to obtain outside GHC. For the HMO control group, they included any service in either group 1 or 2 that was obtained in the fee-for-service sector.

Covered Services: Group 1

Medical diagnosis and treatment
Surgery and anesthesia
X-ray and laboratory services
Prescribed medicines
Semiprivate hospital accommodations

¹The lists paraphrase the formal certificate of benefits presented to each insured family at enrollment. These rules of coverage were subject to individual appeal.

Outpatient hospital care
Maternity care (prenatal and delivery)
Pediatric care
Vision care:
 Refractive examinations by an optometrist
 Eyeglasses (limit one pair every two years)
 Contact lenses (limit one pair per year)
Prosthetic devices medically prescribed to replace body organ
Speech therapy (medically necessary)
Hearing diagnostic examinations by an audiologist
Hearing aids
Mental health care (limit 52 visits per year)
Physical therapy
Occupational therapy
Private-duty nursing care
Home health care:
 Part-time or intermittent nursing care
 Physical, occupational therapy
 Medical social services
 Part-time or intermittent care by health aide
Medically necessary equipment, appliances, and supplies
Ambulance transportation

Covered Services: Group 2

Dental care
Other health care (e.g., chiropractic, acupuncture, services by Christian Science practitioners)
Semiprivate accommodations in skilled nursing facility
Care for drug addiction and alcoholism
Emergency treatment outside GHC service area (and outside United States)

Excluded Services

Services covered by other insurance (e.g., Worker's Compensation) or provided free
Medically unnecessary "custodial or personal care"
Repairs and adjustments of eyeglasses or hearing aids
Personal convenience items while an inpatient (e.g., television, hairdressing)
Excluded dental services:
 Crowns and jackets made of gold or platinum
 Nonpreventive orthodontia
 Fixed bridge with more than seven units (unless patient is eligible for dentures but prefers bridge)
 Replacement of satisfactory dentures and bridges
 Medically unnecessary cosmetic dental surgery
Sterilization for persons younger than 21 or persons declared incompetent by a judge

Cosmetic surgery for preexisting condition

Nonprescribed drugs

More than 52 outpatient mental health visits per year

Appendix B

PARTICIPATION INCENTIVE PAYMENTS

HIE-insured families were paid a participation incentive (PI) if their HIE plans could conceivably impose a greater financial burden than their existing health insurance policies.¹ Calculated yearly, the PI consisted of (1) an amount calculated to be the *maximum* difference between what the family would have to pay for health care under its HIE insurance plan and what it would have paid under its existing insurance plan, unless (2) the premium a family paid to maintain its existing insurance exceeded the maximum difference. In that case, the family was paid an amount equal to the premium payment.

The calculation of item 1 ignored the family's actual medical expenses. To illustrate, consider family X whose HIE plan specified 95 percent coinsurance up to a maximum out-of-pocket expenditure of \$450, above which care was free.² Family X's existing insurance specified a \$100 deductible, above which the family had to pay 20 percent coinsurance. Under its HIE policy, the family had to spend \$473.68 for medical services (with the 5 percent reimbursement) to reach the \$450 out-of-pocket maximum. For the same charge under its existing insurance, the family would have paid \$100 (the deductible) plus 20 percent of the amount between \$100 and \$473.68. The maximum difference was thus $473.68 - 100 - 0.2 (473.68 - 100) = \298.94 . Family X was entitled to \$298.94 per year for that portion of its participation incentive.

The total PI could not exceed the MDE specified in the family's HIE plan unless the family's share of its insurance premium exceeded the MDE. For example, if family X paid an insurance premium of \$300, its

¹Participation incentive payments were not offered to families receiving free care (plan A, described on p. 4) who had no premium to pay, families who had no health insurance before the experiment, and families whose other policies had equal or less generous terms, under all circumstances, than their HIE plan. Those families did, however, receive \$120 for completing the experiment (see footnote 5 below).

²In HIE terminology, maximum out-of-pocket expenditure is called "maximum dollar expenditure," or MDE.

total PI entitlement was \$450, not \$598.94 (300 + 298.94). If the family paid a premium of \$600, its PI was \$600 because the premium exceeded the MDE of \$450. On the other hand, a family who had a high MDE in its HIE plan and an existing insurance policy with 0 percent coinsurance, no deductible, and an employer-paid premium was entitled to the full MDE amount. The purpose of PI payments was to ensure that a family was no worse off financially by participating in the experiment--whether because of the cost of its insurance premium or the "worse" terms of its HIE insurance plan compared with its existing policy.³

As encouragement for families to complete their assigned enrollment terms, a portion of the family's annual PI was withheld until the last year of the term.⁴ The family received its full annual PI that last year, and the amount previously withheld was paid as part of a completion bonus when the family completed the physical screening examination and medical health questionnaire at exit.⁵

To measure enrollees' responsiveness to PI payments, a subset of families received their full annual PI in the next-to-last, as well as the last, year of their term. That "super PI bonus" was offered to 44.4 percent of the families assigned to insurance plans requiring 95 percent coinsurance, the highest rate (plans K-N, described on pp. 3-4). Super PI

³Calculation of PI is further described in Ref. 3. The formula on p. 20 of that report should read $PI = \max[K \times PG, PR]$.

⁴The percentage of PI withheld depended on the site and assigned enrollment term, as follows:

	<i>3-yr Term</i>	<i>5-yr Term</i>
Dayton	25	15
Seattle	25	15
Fitchburg	33.3	25
Franklin Co.	33.3	25
Charleston	33.3	20
Georgetown Co.	33.3	20

If the discounted PI was not enough to reimburse the cost of the family's insurance premium, however, the family received the full amount of its premium. The difference between the premium and the discounted PI was then subtracted from the withheld amount.

⁵The rest of the completion bonus was the largest annual PI to which the family had been entitled during its enrollment (minus the withheld amount) or \$120, whichever was greater.

recipients represented all sites and both terms of enrollment except Dayton enrollees assigned to three-year terms, who had already begun their next-to-last year when super PI was instituted. Within the 95 percent coinsurance plans, super PI recipients were chosen using the "finite selection model." That model was developed by RAND to assign enrollees to experimental insurance plans so that, across plans, families resembled each other in 24 health and socioeconomic characteristics.⁶

⁶The finite selection model is described in Ref. 4.

Appendix C
UNRELEASED HIE DATA

Because of budgetary constraints, data collected by the instruments marked with an asterisk in Table C.1 are not included in the HIE public-use files, although small portions were used in isolated cases.

Table C.1

TABLE 2 ANNOTATED TO INDICATE UNRELEASED DATA

Instrument	Topics Covered	Data Collected		
		How	When	From
1. Screening questionnaire [1]	Demographic information to establish basic eligibility	Interview	Beginning of HIE operation in site	Occupants of representative sample of dwelling units on geographic clusters in site
2. Baseline questionnaire, 2 parts	Income, employment Family composition Health status Health care experience and insurance coverage Satisfaction with medical care	Interview Self-administered	4-6 months before enrollment 4-6 months before enrollment	Baseline participants Baseline participants
3. Enrollment verification form	Changes in family composition, economics, or insurance coverage since baseline questionnaire	Interview	Between administration of baseline questionnaire and enrollment date	Baseline participants determined eligible
4. Medical history questionnaires (MHQ), 3 versions by age group: 0-4 years 5-13 years 14+ years	Form A: health status, attitudes, habits Form B: specific medical disorders	Administered by self or parent [2]	Just before enrollment and exit [3]	Insured enrollees
5. Medical screening examinations, 3 versions by age group: 0-2 years 3-13 years 14+ years	Tests of organ system function, including dental	Paramedical personnel	Just before enrollment and exit	Sample of insured enrollees at enrollment; all exiting enrollees
*6. Health report	Use of medical or dental services and time spent obtaining them; any restricted activity or bed disability	Administered by self or parent	Biweekly during period of participation	Insured enrollees [4]

1. Administered as a separate questionnaire only in Dayton; part of baseline questionnaire in the other sites.
2. When "parent" appears in this column, a parent was asked to provide data for children 13 and younger.
3. "Exit" refers to normal departure from the experiment after completing the assigned enrollment period, three or five years. Those who "attrited," or voluntarily left the experiment early, received an "attrition" MHQ that was identical to the exit MHQ.
4. In the first year of the experiment in Dayton, the health report was administered weekly to a random half of Dayton enrollees. In the first year of the experiment in Massachusetts and South Carolina, 25 percent of enrollees were exempted to measure the reporting requirement's effect on the use of health services. Also at one point virtually all participants stopped filling out health reports, for budgetary reasons.

Table C.1 (cont.)

Instrument	Topics Covered	Data Collected		
		How	When	From
*7. Health care questionnaire, 3 versions by age group: 0-4 years 5-13 years 14+ years	Health status, attitudes, habits (subset of MHQ)	Administered by self or parent	Each anniversary of enrollment except at exit	Insured enrollees
*8. Annual income report	Amount and sources of family income, taxes paid	Self-administered	Annually (April)	Head of insured family
*9. Periodic employment report	Wages, hours worked, family payments for care of children or elderly, government program benefits received	Self-administered	Semiannually	Enrollees (head and family members 16 and older)
*10. Assets and debts questionnaire	Family assets and liabilities	Self-administered	Exit	Head of insured family
*11. Knowledge of coverage questionnaire	Details of HIE insurance plan	Self-administered	Specified intervals [5]	Insured enrollees
*12. Insurance abstraction	Details of selected insurance policies	Abstracted	At time of knowledge of coverage questionnaire	Insurance company brochures
*13. Chronic condition questionnaire	Status of condition, correctness of diagnosis, adequacy of treatment	Physician interview	At exit medical screening examination	Sample of insured enrollees found to have certain chronic conditions [6]
*14. Evaluation questionnaire	Perceptions and attitudes about HIE and health care system	Self-administered	Exit	Head of insured family
*15. Health notice	Use of medical or dental services	Administered by self or parent	Biweekly during preenrollment phase (South Carolina); 6 months-1 year after exit (other sites)	Preenrollees (South Carolina), insured enrollees who have exited (other sites)

5. Intended intervals were enrollment, 18 months, 3 years, and 5 years after enrollment (the last only for the 5-year participants). Actual mailings approximated those intervals in Massachusetts and South Carolina; the first mailing was 2-1/2 years and 1 year after enrollment in Dayton and Seattle, respectively.

6. Hypertension, diabetes, thyroid diseases, chronic heart diseases, chronic lung diseases, joint diseases, ulcers, cerebrovascular disease.

Table C.1 (cont.)

Instrument	Topics Covered	Data Collected		
		How	When	From
16. Medical expense report (MER)--fee-for-service claim form, 4 types: Doctors' services and supplies Dental care Hospital and extended care Pharmacy	Each use of medical or dental service, drugs, and equipment; reason or diagnosis; treatment	Administered by self or parent	Time of occurrence	Insured enrollees and providers/suppliers
17. Services rendered report (SER)--HMO equivalent of MER [7], 2 types: Doctors' services and supplies Hospital and extended care	Each use of medical service provided by HMO; reason or diagnosis; treatment	Abstracted	Annually to cover entire previous year	HMO records for insured enrollees in HMO experimental and control groups
*18. Factor price survey	Wages and benefits of selected hospital personnel [8], average daily inpatient population	Phone and mail	Semiannually	Sample of local hospitals
*19. Consumer price index	Prices of selected nonmedical products in the six HIE sites	Phone and inspection	Semiannually	Sample of local retailers
*20. Physician capacity utilization survey (PCUTS)	Availability of services [9]	Phone	Annually	Sample of local physicians [10]
*21. Dentist capacity utilization survey (DCUTS)	Similar to PCUTS	Phone	Annually	Sample of local dentists [11]
22. Insurance preference questionnaire	Willingness to pay higher premium to reduce out-of-pocket expense limit	Self-administered	Exit	Head of insured family
7. Pharmacy data were obtained directly from an HMO-supplied computer tape. Dental care was not available through the HMO; HMO participants reported claims for dental care and other non-HMO services on the MER.				
8. Categories of personnel: registered nurses (general-duty), medical technicians, licensed professional nurses, nursing aides, kitchen helpers, general stenographers, and maids or porters.				
9. Waiting time for appointments; appointments per hour; patients seen in office, home, and hospital; weekend office hours; office staffing; cost of office visit; whether new patients accepted.				
10. Physicians (M.D. or D.O.) specializing in general practice, internal medicine, and pediatrics.				
11. Except in Fitchburg, Franklin County, and Georgetown County, where all dentists were surveyed.				

1
8
8
1

Appendix D

POPULATION CHANGE WITHIN HIE EXPERIMENTAL TREATMENTS

Table D.1 shows how the population of insured enrollees in each major type of experimental treatment changed during the five years of the experiment. Here, "year" refers to *contract year*. Experimental treatments are defined in terms of values on variable PLAN in file MS2. Those values reflect the relatively few cases where families were reassigned to different experimental insurance plans after enrollment. Previously published HIE analyses are based on families' *original* plan assignments, with slightly different population figures. See Sec. III, footnote 5 and accompanying text, for explanation of insurance plan changes.

Table D.1
POPULATION CHANGE WITHIN EXPERIMENTAL TREATMENTS

Item	Number of Insured Enrollees				
	Year 1	Year 2	Year 3	Year 4	Year 5
<i>0% Coinsurance (PLAN = 11)</i>					
Start of year	1894	1913	1924	596	602
Departures					
Attrition	3	1	2	0	0
Termination	17	28	25	2	5
Death	3	2	5	3	0
Ineligibility	0	0	1	3	1
Exit	0	0	1337	0	614
Additions	42	42	42	14	18
End of year	1913	1924	596	602	0
<i>HMO Control Group (PLAN = 98)</i>					
Start of year	740	696	642	587	523
Departures					
Attrition	21	31	21	15	3
Termination	44	43	44	56	45
Death	0	1	0	1	0
Ineligibility	0	1	2	2	5
Exit	0	0	0	0	481
Additions	21	22	12	10	11
End of year	696	642	587	523	0
<i>HMO Experimental Group (PLAN = 30)</i>					
Start of year	1148	1133	1123	571	571
Departures					
Attrition	23	13	10	4	2
Termination	6	17	7	5	10
Death	0	0	1	0	2
Ineligibility	0	0	0	0	1
Exit	0	0	554	0	560
Additions	14	20	20	9	4
End of year	1133	1123	571	571	0

SOURCE: HIE administrative files

Table D.1 (cont.)

Item	Number of Insured Enrollees				
	Year 1	Year 2	Year 3	Year 4	Year 5
<i>Coinsurance 95% outpatient, 0% inpatient (PLAN = 1,13)</i>					
Start of year	1276	1234	1210	298	302
Departures					
Attrition	34	27	16	0	1
Termination	25	14	14	2	4
Death	6	7	2	1	1
Ineligibility	1	1	2	0	0
Exit	0	0	904	0	301
Additions	24	25	26	7	5
End of year	1234	1210	298	302	0
<i>25% Coinsurance (PLAN = 5-7, 17-19)</i>					
Start of year	1137	1133	1128	378	374
Departures					
Attrition	6	12	4	4	7
Termination	12	12	22	1	2
Death	1	5	1	2	1
Ineligibility	0	0	0	0	0
Exit	0	0	742	0	365
Additions	15	24	19	3	1
End of year	1133	1128	378	374	0
<i>50% Coinsurance (PLAN = 8-10)</i>					
Start of year	383	378	378	149	143
Departures					
Attrition	9	3	8	3	0
Termination	0	3	4	1	0
Death	0	1	0	2	0
Ineligibility	0	0	0	0	0
Exit	0	0	223	0	146
Additions	4	7	6	0	3
End of year	378	378	149	143	0

Table D.1 (cont.)

Item	Number of Insured Enrollees				
	Year 1	Year 2	Year 3	Year 4	Year 5
<i>95% Coinsurance (PLAN = 2-4, 14-16)</i>					
Start of year	1121	1078	1043	327	319
Departures					
Attrition	49	48	17	9	0
Termination	14	6	13	3	1
Death	0	4	2	1	1
Ineligibility	0	0	0	0	0
Exit	0	0	714	0	324
Additions	20	23	30	5	7
End of year	1078	1043	327	319	0
<i>All Experimental Treatments</i>					
Start of year	7699	7565	7448	2906	2834
Departures					
Attrition	145	135	78	35	13
Termination	118	123	129	70	67
Death	10	20	11	10	5
Ineligibility	1	2	5	5	7
Exit	0	0	4474	0	2791
Additions	140	163	155	48	49
End of year	7565	7448	2906	2834	0

Appendix E

HIE INFLATION-ADJUSTMENT FACTORS

This appendix lists the factors by which HIE data were adjusted for inflation in most RAND analyses. Depending on the type of data, one of three sets was used: *overall* factors for wage and income data (Table E.1), *medical* factors for medical expense data (Table E.2), and *dental* factors for dental expense data (Table E.3). Data in virtually all HIE research files are unadjusted for inflation.¹ File users may want to use these factors in their analyses.

¹Exceptions are the HMO visit and annual expenditure files, in which medical inflation-adjustment factors were applied for reasons explained in Sec. V.

Table E.1

OVERALL INFLATION-ADJUSTMENT FACTORS

Month	1974	1975	1976	1977	1978	1979	1980	1981	1982
January	--	1.561	1.667	1.753	1.872	2.047	2.332	2.605	2.825
February	--	1.572	1.671	1.771	1.884	2.071	2.364	2.632	2.834
March	--	1.578	1.675	1.782	1.898	2.091	2.398	2.651	2.831
April	--	1.586	1.682	1.796	1.915	2.115	2.425	2.668	2.843
May	--	1.593	1.692	1.806	1.933	2.141	2.449	2.690	2.871
June	--	1.606	1.701	1.818	1.953	2.166	2.476	2.713	2.906
July	--	1.623	1.711	1.826	1.967	2.189	2.478	2.744	--
August	--	1.628	1.719	1.833	1.978	2.211	2.494	2.765	--
September	1.517	1.636	1.726	1.840	1.993	2.234	2.517	2.793	--
October	1.530	1.646	1.733	1.845	2.009	2.254	2.539	2.799	--
November	1.543	1.656	1.738	1.854	2.020	2.275	2.562	2.807	--
December	1.554	1.663	1.743	1.861	2.029	2.299	2.584	2.815	--

SOURCE: U.S. Bureau of Labor Statistics, Consumer Price Index. Base year: 1967 (1967 = 1.00).

NOTE: These factors have been used in RAND analyses of HIE data pertaining to income and wages. They have not been applied to the data in HIE research files. To use these factors, divide unadjusted data by the factor for the appropriate date.

Table E.2

MEDICAL INFLATION-ADJUSTMENT FACTORS

Month	1974	1975	1976	1977	1978	1979	1980	1981	1982
January	--	1.610	1.766	1.941	2.112	2.307	2.539	2.795	3.134
February	--	1.630	1.788	1.958	2.133	2.326	2.579	2.826	3.162
March	--	1.646	1.806	1.976	2.145	2.339	2.602	2.847	3.188
April	--	1.658	1.816	1.991	2.157	2.351	2.620	2.870	3.217
May	--	1.668	1.826	2.005	2.169	2.363	2.634	2.890	3.238
June	--	1.681	1.837	2.018	2.179	2.377	2.647	2.915	3.264
July	--	1.698	1.855	2.035	2.194	2.399	2.666	2.956	--
August	--	1.709	1.868	2.049	2.214	2.418	2.684	2.993	--
September	1.552	1.722	1.879	2.063	2.226	2.437	2.706	3.017	--
October	1.563	1.735	1.889	2.072	2.247	2.459	2.728	3.048	--
November	1.575	1.733	1.913	2.081	2.270	2.480	2.745	3.082	--
December	1.590	1.747	1.923	2.093	2.278	2.507	2.758	3.102	--

SOURCE: U.S. Bureau of Labor Statistics, Consumer Price Index. Base year: 1967 (1967 = 1.00).

NOTE: These factors have been used in deriving data pertaining to HMO claims in the HMO visit and annual expenditure files. Their codebooks are Refs. 31 and 32. The factors have also been used in RAND analyses of HIE medical expenses. To use these factors, divide unadjusted medical expense data by the factor for the appropriate date.

Table E.3

DENTAL INFLATION-ADJUSTMENT FACTORS

Month	1974	1975	1976	1977	1978	1979	1980	1981	1982
January	--	1.560	1.672	1.787	1.918	2.079	2.274	2.514	2.758
February	--	1.572	1.682	1.799	1.934	2.094	2.315	2.528	2.768
March	--	1.587	1.694	1.809	1.940	2.103	2.341	2.549	2.786
April	--	1.597	1.698	1.819	1.952	2.114	2.372	2.571	2.811
May	--	1.612	1.706	1.832	1.960	2.124	2.388	2.577	2.821
June	--	1.618	1.716	1.841	1.972	2.143	2.403	2.602	2.839
July	--	1.630	1.720	1.857	1.978	2.152	2.411	2.665	--
August	--	1.636	1.735	1.873	1.988	2.160	2.422	2.699	--
September	1.505	1.641	1.745	1.888	2.006	2.172	2.458	2.708	--
October	1.518	1.651	1.756	1.896	2.029	2.185	2.475	2.716	--
November	1.530	1.658	1.766	1.904	2.039	2.207	2.480	2.723	--
December	1.544	1.665	1.779	1.909	2.047	2.245	2.486	2.739	--

SOURCE: U.S. Bureau of Labor Statistics, Consumer Price Index. Base year: 1967 (1967 = 1.00).

NOTE: These factors have been used in RAND analyses of HIE dental expenses. They have not been applied to the data in HIE research files. To use these factors, divide unadjusted dental expense data by the factor for the appropriate date.

GLOSSARY

Adjunct enrollee	Uninsured member of insured family/household (person/family of interest) or member of Dayton control group.
Attrition	Departure from the experiment by voluntary withdrawal before completing assigned enrollment term. Compare "exit" and "termination."
Baseline participant	Person considered for enrollment at the beginning of the experiment in the site. May or may not have enrolled, either remaining a baseline-only participant or becoming an enrollee.
Contract year	Administrative unit of time for enrollees; year period reckoned from date family signed enrollment contract. First contract year began on enrollment date, second contract year began on first anniversary of enrollment, and so on.
Dayton control group	Group of 669 uninsured enrollees in Dayton who participated in the experiment for 15 months. Formed to compare the community's use of health services with use by insured Dayton enrollees. Members retained their own insurance but were asked to complete the same questionnaires as insured enrollees. Group was discontinued because complete data appeared unobtainable from them.
DEI	Data element indicator. Every variable in HIE primary data files is assigned a unique alphanumeric identifier composed of prefix DEI followed by three or four digits, e.g., DEI1234. Each DEI number represents a single question on a data collection instrument or an item of information on an administrative form.
Derived variable	Aggregation or other transformation of raw data obtained during the experiment. Compare "primary variable."
Economic family	Related persons residing together who shared income. Not all members were necessarily insured HIE enrollees. Compare "insurance family."
Exit	Departure from the experiment after completion of assigned enrollment term, three or five years. Compare "attrition" and "termination."

Experimental insurance treatment	One of 16 groups in which experimental subjects participated. Fifteen were health insurance plans with varying coinsurance rates, out-of-pocket expenditure limits, and both FFS and HMO delivery systems. The sixteenth was the HMO control group.
Family of interest	Uninsured, self-supporting member of insured family or member of uninsured self-supporting family residing in the household of an insured family. Compare "person of interest."
FFS	Fee-for-service mode of health care delivery.
FFS enrollee	Person initially assigned to an FFS experimental insurance treatment (plans A-N listed on pp. 3-4).
File reference number	Letter-number combination (e.g., MS1) that uniquely identifies each HIE public-use file. Defined in Sec. II.
GHC	Group Health Cooperative of Puget Sound, the Seattle HMO that participated in the experiment.
HIE	Health Insurance Experiment.
HIE-insured	Enrollee assigned to an experimental insurance plan paid by the HIE (plans A-O, described on pp. 3-4). Includes members of HMO experimental group. Compare "HMO-insured."
HMO	Health maintenance organization. The HMO that participated in the experiment was Group Health Cooperative of Puget Sound.
HMO control group	Seattle enrollees drawn at random from existing GHC members who met HIE eligibility criteria. The HIE did not pay their insurance premiums.
HMO experimental group	Seattle enrollees experimentally transferred to GHC from the fee-for-service system. The HIE paid their insurance premiums.
HMO-insured	Member of HMO control group. Compare "HIE-insured."
HMO participant	Seattle enrollee who obtained health care from GHC; member of HMO experimental group or HMO control group.
Insurance family	Members of an "economic family" who were covered under the same experimental insurance treatment. All were insured enrollees. Compare "economic family."

Item nonresponse	Missing answer to individual question on an HIE data collection instrument because participant accidentally skipped question. Distinguished from question unanswered because of skip pattern.
Insured	Either HIE-insured or HMO-insured. Compare "uninsured."
Insured enrollee	Person assigned to an experimental treatment; HIE-insured or HMO-insured. Compare "adjunct enrollee."
MDE	Maximum dollar expenditure--maximum out-of-pocket expense to be paid by HIE-insured family per year before health care became free. The amount depended on the family's assigned health insurance plan.
MHQ	Medical history questionnaire, an HIE data collection instrument. MHQA refers to MHQ form A; MHQB refers to MHQ form B.
NonDayton	All HIE sites other than Dayton, Ohio. Refers to Seattle, Washington; Franklin County and Fitchburg, Massachusetts; and Georgetown County and Charleston, South Carolina.
Participant	Anyone with a record in the HIE database; includes baseline-only participants and enrollees.
PEG	South Carolina preenrollment group.
PEG-period-only participants	Subset of baseline-only participants who joined the experiment for only the preenrollment phase in South Carolina.
Person of interest	Uninsured member of insured family/household who was financially dependent on insured family/household. Compare "family of interest."
Preenrollee	Person who participated in South Carolina preenrollment phase beginning in 1976; PEG member. May or may not have formally enrolled in 1978.
Primary variable	Original information in HIE primary data file; unmanipulated raw data. Compare "derived variable."
Provider	Any person, institution, or organization that provided health services, drugs, or supplies to an enrollee.
Public-use files	Sixty-seven data files released to the research community, a large subset of all data collected by the HIE.

SAS	Statistical Analysis System. HIE files contain data in both SAS and character formats.
Suspension	Revocation of HIE-provided insurance benefits because of ineligibility expected to be temporary. Suspended persons remained enrollees.
Termination	Involuntary departure from the experiment. Cancellation of enrollment for permanent ineligibility or failure to fulfill obligations of participation. Compare "attrition" and "exit."
Uninsured	Neither HIE-insured nor HMO-insured. Person/family of interest or member of Dayton control group. Uninsured persons did not necessarily lack health insurance; they were uninsured only with respect to HIE experimental treatments. Compare "insured."

REFERENCES

1. Held, Philip J., *Site Selection Criteria for the Health Insurance Study*, The RAND Corporation, N-2266-HHS, May 1985.
2. Newhouse, Joseph P., "A Design for a Health Insurance Experiment," *Inquiry*, Vol. 11, No. 1, March 1974, pp. 5-27.
3. Clasquin, Lorraine, and Marie E. Brown, *Rules of Operation for the Rand Health Insurance Study*, The RAND Corporation, R-1602-HEW, May 1977.
4. Morris, Carl N., "A Finite Selection Model for Experimental Design of the Health Insurance Study," *Journal of Econometrics*, Vol. 11, 1979, pp. 43-61.
5. Bloomfield, E. S., L. Y. Weissler, and A. B. Holland, *Codebooks for Insurance Preference Files: Relation between Expense Limit and Premium*, The RAND Corporation, N-2508-HHS, October 1986.
6. Polich, S. M., and C. d'Arc Taylor, *Master Sample Series, Vol. 1: Codebook for Eligibility-Family Changes File*, The RAND Corporation, N-2264/1-HHS, May 1986.
7. Polich, S. M., and C. d'Arc Taylor, *Master Sample Series, Vol. 3: Codebook for Supplemental Data File*, The RAND Corporation, N-2264/3-HHS, June 1987.
8. Polich, S. M., et al., *Master Sample Series, Vol. 2: Codebook for Full Sample Demographic File*, The RAND Corporation, N-2264/2-HHS, May 1986.
9. Rogers, William H., and Joseph P. Newhouse, "Measuring Unfiled Claims in the Health Insurance Experiment," in Leigh Burstein, Howard E. Freeman, and Peter H. Rossi (eds.), *Collecting Evaluation Data: Problems and Solutions*, Sage Publications, Beverly Hills, CA, 1985, pp. 121-133.
10. Lillard, Lee A., et al., *Preventive Medical Care: Standards, Usage, and Efficiency*, The RAND Corporation, R-3266-HCFA, August 1986.
11. Keesey, Joan, Emmett Keeler, and William Fowler, *The Episodes-of-Illness Processing System*, The RAND Corporation, N-1745-1-HHS, January 1985.
12. California Medical Association, *California Relative Value Studies*, 5th ed., 1974 revision, San Francisco, CA, 1975.

13. Sloss, E. M., et al., *Health Status and Attitude Series, Vol. 1: Codebooks for Adults and Children at Enrollment and Exit*, The RAND Corporation, N-2447/1-HHS, November 1986.
14. Operskalski, B. H., et al., *Medical Disorders Series, Vol. 1: Codebook for Adults at Enrollment and Exit*, The RAND Corporation, N-2446/1-HHS, February 1987.
15. Sloss, E. M., et al., *Medical Disorders Series, Vol. 2: Codebooks for Children at Enrollment and Exit*, The RAND Corporation, N-2446/2-HHS, March 1987.
16. Smith, Lisa H., et al., *The Health Insurance Study Screening Examination Procedures Manual*, The RAND Corporation, R-2101-HEW, September 1978.
17. Ware, J. E., et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. I, Model of Health and Methodology*, The RAND Corporation, R-1987/1-HEW, May 1980.
18. Stewart, A. L., et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. II, Physical Health in Terms of Functioning*, The RAND Corporation, R-1987/2-HEW, July 1978.
19. Ware, J. E., et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. III, Mental Health*, The RAND Corporation, R-1987/3-HEW, December 1979.
20. Donald, C. A., et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. IV, Social Health*, The RAND Corporation, R-1987/4-HEW, August 1978.
21. Ware, J. E., Jr., A. Davies-Avery, and C. A. Donald, *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. V, General Health Perceptions*, The RAND Corporation, R-1987/5-HEW, September 1978.
22. Ware, J. E., Jr., A. Davies-Avery, and R. H. Brook, *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. VI, Analysis of Relationships among Health Status Measures*, The RAND Corporation, R-1987/6-HEW, November 1980.
23. Rogers, W. H., K. N. Williams, and R. H. Brook, *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. VII, Power Analysis of Health Status Measures*, The RAND Corporation, R-1987/7-HEW, March 1979.
24. Brook, R. H., et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. VIII, Overview*, The RAND Corporation, R-1987/8-HEW, October 1979.

25. Eisen, M. B., et al., *Conceptualization and Measurement of Health for Children in the Health Insurance Study*, The RAND Corporation, R-2313-HEW, May 1980.
26. Stewart, A. L., R. H. Brook, and R. L. Kane, *Conceptualization and Measurement of Health Habits for Adults in the Health Insurance Study: Vol. I, Smoking (Vol. II, Overweight)*, The RAND Corporation, R-2374/1-HEW (R-2374/2-HEW), June 1979 (July 1980).
27. Stewart, A. L., J. E. Ware, Jr., and R. H. Brook, *Construction and Scoring of Aggregate Functional Status Measures: Vol. 1*, The RAND Corporation, R-2551-1-HHS, August 1982.
28. Davies, A. R., and J. E. Ware, Jr., *Measuring Health Perceptions in the Health Insurance Experiment*, The RAND Corporation, R-2711-HHS, October 1981.
29. Donald, C. A., and J. E. Ware, Jr., *The Quantification of Social Contacts and Resources*, The RAND Corporation, R-2937-HHS, October 1982.
30. Bloomfield, E. S., L. Y. Weissler, and A. M. Bell, *Dental Examinations: Codebook for Adults and Children at Enrollment and Exit*, The RAND Corporation, N-2506-HHS, February 1987.
31. Peterson, C. E., M. Nelsen, and D. L. Wesley, *Aggregated Claims Series, Vol. 4: Codebooks for Health Maintenance Organization and Seattle Fee-for-Service Visits--Outpatient and Inpatient*, The RAND Corporation, N-2360/4-HHS, December 1986.
32. Peterson, C. E., et al., *Aggregated Claims Series, Vol. 5: Codebook for Health Maintenance Organization and Seattle Fee-for-Service Annual Expenditures and Visit Counts*, The RAND Corporation, N-2360/5-HHS, December 1986.

