## Citation

## Abstract

This publication describes Version 2 of the Clinical Classifications for Health Policy Research (CCHPR), a diagnosis and procedure categorization scheme, and provides descriptive statistics for 1992 hospital inpatient stays illustrating the use of the CCHPR categories. Diagnoses and procedures for hospital stays are coded using the *International Classification of Diseases, 9th Revision, Clinical Modification* (ICD-9-CM), a uniform and standardized coding system. ICD-9-CM consists of over 12,000 diagnosis codes and 3,500 procedure codes. Although it is possible to present descriptive statistics for individual ICD-9-CM codes, it is often helpful to aggregate codes into clinically meaningful categories that comprise similar conditions or procedures.

CCHPR Version 1 was the initial endeavor to construct such clinically meaningful categories. CCHPR Version 2 is based on the Version 1 summary diagnosis and procedure categories. The original categories were modified on the basis of clinical homogeneity, the number of discharges, and ICD-9-CM coding changes.

CCHPR categories can be employed in many types of projects analyzing data on diagnoses and procedures, such as identifying populations for disease- or procedure-specific studies; providing statistical information (such as charges and length of stay) about relatively specific conditions; defining comorbidities; and cross-classifying procedures by diagnoses to provide insight into the variety of procedures performed for particular diagnoses.

Electronic versions of CCHPR classification schemes will be available in a future HCUP-3 Research Note and can be obtained now by contacting the lead author.

## Acknowledgments

## Background

Research Notes are derived from research conducted by the Center for Delivery Systems Research (CDSR), Agency for Health Care Policy and Research (AHCPR). This series provides results of analyses on health policy issues important to the Nation's health care providers and patients.

## Overview

CDSR conducts and supports studies of the structure, behavior, and performance of acute and long-term health care systems. It was created in 1995 by the merger of the Division of Provider Studies and the Division of Long-Term Care Studies. The Division of Provider Studies and its predecessor, the Hospital Studies Program, established in 1978, focused on the use and cost of services provided by various acute care providers, including but not limited to hospitals. The Division for Long-Term Care focused on long-term care providers and services.

The national and statewide databases maintained by CDSR can be used for studies that address a variety of issues, including:

- Variations in medical practice.
- Diffusion of medical technology.
- Effectiveness of medical treatments.
- Hospital financial distress.
- Utilization by special populations.
- Quality of health services.
- Impact of State and Federal health care reform initiatives.

CDSR has the capacity for research at many different levels: hospitalizations, patient care, treatment of diseases, physician practice, hospital group differences, small-area variations, State-to-State comparisons, and changes across time. CDSR's hospital research focuses on cost (hospital, disease, and treatment), quality (volume-outcome, treatment variations, adverse events, and guideline diffusion), and access (uncompensated care, HIV/AIDS, organ transplantation, and vulnerable populations).

These examples do not exhaust the issues that can be addressed with the CDSR databases. CDSR staff consult with industry experts, public officials, and other researchers in selecting topics for study.

## CDSR Databases

The HCUP databases contain the central data supporting the CDSR acute care research program. Each database is built around a core of data on inpatient hospital stays.

*Healthcare Cost and Utilization Project (HCUP-3)*—CDSR is currently in the third phase of the HCUP project, the collection of data for the years 1988-94 from State government and private health data organizations. By integrating data from statewide health data organizations, HCUP-3 is creating a multi-State database in a uniform format that promotes comparative studies of health care services by researchers both inside and outside AHCPR.

The two HCUP-3 hospital inpatient databases are built around core data elements comparable to those in a typical discharge abstract, with safeguards to protect the privacy of individual patients and physicians.

- The *HCUP-3 State Inpatient Database (SID)* contains 100 percent of hospitals and discharges from States with statewide hospital inpatient data systems selected by CDSR. This database contains common inpatient data elements, variables derived from sensitive data elements, and State-specific variables (such as readmission or community characteristics). Each State database in SID is returned to its data source in the uniform HCUP-3 format. Dissemination of SID data is controlled by the data source.

- The *HCUP-3 Nationwide Inpatient Sample (NIS)* includes data from about 900 U.S. hospitals, approximating a 20-percent sample of hospitals. This database includes core inpatient data elements but excludes information that could result, directly or indirectly, in the identification of individual patients or physicians. When data sources consider other variables (such as hospital identifiers) to be confidential, those variables are excluded or encrypted, as required by the data source. The NIS is available to researchers who sign a data use agreement restricting the use of the data to research purposes only. The NIS is being distributed through the National Technical Information Service.

Inpatient databases are supplemented with data on hospitals and local communities from a variety of sources. The American Hospital Association (AHA) has provided data from its Annual Survey of Hospitals and various special surveys since 1970. County-level statistics are obtained from the Area Resource File, compiled by the Bureau of Health Professions of the Health Resources and Services Administration (HRSA). Statistics from the Bureau of the Census at the ZIP-Code level, provided by CACI Marketing Systems, are also used. HCUP-3 also includes statewide encounter data on services other than inpatient hospital care, such as ambulatory surgery.

The CDSR databases will support a variety of studies to be conducted across hospital types, communities, and time. Studies that are national in scope require NIS. State data from SID can be used to study small-area variations, hospital markets, and State health care reforms. These databases permit comprehensive assessment of factors affecting the use and costs of health services.

*Hospital Cost and Utilization Project (HCUP-1 and HCUP-2)*— The HCUP-1 sample spans the period 1970-77; the HCUP-2 sample covers the period 1980-87. These HCUP databases were derived from two national samples of approximately 500 hospitals each and contain nearly 60 million hospital discharges. In HCUP-1, hospitals were selected from clients of discharge abstracting companies; in HCUP-2, hospitals were added to improve representation of the universe of short-term, general, non-Federal hospitals with at least 30 beds.

Because the data for HCUP-1 and HCUP-2 were collected under special agreements with individual hospitals, only CDSR staff have direct access to these data.

## Contributors

AHCPR thanks the following organizations for their contributions of data to the 1988-93 HCUP-3 NIS and SID: Arizona Department of Health Services, California Office of Statewide Health Planning and Development, Colorado Hospital Association, Connecticut Hospital Research and Education Foundation, Inc., Florida Agency for Health Care Administration, Illinois Health Care Cost Containment Council, Iowa Hospital Association, Kansas Hospital Association, Maryland Health Services Cost Review Commission, Massachusetts Rate Setting Commission, New Jersey Department of Health, New York State Department of Health, Oregon Department of Human Resources, Pennsylvania Health Care Cost Containment Council, South Carolina Budget and Control Board, Washington State Department of Health, and Wisconsin Office of the Commissioner of Insurance.

Data files for HCUP-3 were constructed under the technical direction of AHCPR by The MEDSTAT Group (formerly SysteMetrics, Inc.), Santa Barbara, CA, and its subcontractors Abt Associates and the National Association of Health Data Organizations. Social and Scientific Systems, Inc., Bethesda, MD, provides programming support for CDSR researchers.

# Contents

# Clinical Classifications for Health Policy Research, Version 2: Hospital Inpatient Statistics

Anne Elixhauser, Ph.D., Agency for Health Care Policy and Research, and Eileen McCarthy, B.A., National Center for Health Statistics

## Introduction

This publication describes Version 2 of the Clinical Classifications for Health Policy Research (CCHPR), a diagnosis and procedure categorization scheme, and provides descriptive statistics for 1992 hospital inpatient stays illustrating the use of the CCHPR categories. CCHPR represents a way to classify diagnoses and procedures coded using the *International Classification of Diseases, 9th Revision, Clinical Modification* (ICD-9-CM) Fourth Edition (Public Health Service and Health Care Financing Administration, 1991). The ICD-9-CM is a uniform and standardized coding system consisting of over 12,000 diagnosis codes and 3,500 procedure codes. Although it is possible to present descriptive statistics for individual ICD-9-CM codes, it is often helpful to aggregate codes into clinically meaningful categories that group similar conditions or procedures.

Version 1 of CCHPR was the initial endeavor to construct such clinically meaningful categories (Elixhauser, Andrews, and Fox, 1993). The determining factor in creating the categories was the extent to which conditions and procedures could be grouped into relatively homogeneous clusters of interest for public policy research. CCHPR Version 1 consisted of 185 summary diagnosis categories and 172 summary procedure categories. In addition, a five-level hierarchy of increasingly more specific diagnosis categories was constructed.

## Development of the classification systems

The current development effort proceeded in two phases. First, a system of mutually exclusive categories was created for both diagnoses and procedures (CCHPR Version 2). These categories were developed to provide a convenient way to report hospital statistics by diagnosis or procedure. Second, CCHPR Version 2 was expanded into a multi-level, hierarchical system. These expanded categories were developed to support the production of aggregate statistics about larger groupings of CCHPR categories as well as more detailed statistics about subgroups of diagnoses and procedures within CCHPR categories.

## CCHPR Version 2

CCHPR Version 2 is based on the Version 1 summary diagnosis and procedure categories. Version 1 categories were modified to improve clinical homogeneity, to better distribute the number of discharges across categories, and to add ICD-9-CM coding changes. For this first stage in the development process, 1989 data on all-listed diagnoses and all-listed procedures from the California Office of Statewide Health Planning and Development ($N = 3,788,040$) were used.

Trained medical records personnel and a physician with experience in diagnostic classification conducted reviews of the modified CCHPR schemes. The reviewers evaluated the categories and ICD-9-CM code assignments for accuracy and clinical significance.

Version 2 contains more categories than Version 1. Some conglomerate categories (e.g., Other gastrointestinal procedures) and high-frequency categories (e.g., Pregnancy-related conditions) have been divided into more clinically homogeneous groups. The diagnosis classification system aggregates illnesses and conditions into 260 mutually exclusive categories, most of which are clinically homogeneous. Some heterogeneous categories combine several less common individual conditions. External Causes of Injury and Poisoning (E codes) are grouped into a single

1

category because they are not used consistently in inpatient data.

A second classification scheme for procedures contains 231 mutually exclusive categories. Many of the categories represent single procedures; however, some procedures that occur infrequently are grouped according to the body system on which they are performed, whether they are used for diagnostic or therapeutic purposes, and whether they are considered operating room or non-operating room procedures (Puckett, 1994).

All ICD-9-CM coding changes from January 1980 through October 1995 have been incorporated into the coding scheme.

Appendix A provides details and coding for the CCHPR diagnosis classification and Appendix B provides the same information for procedures.[1]

## Expanded diagnosis and procedure systems

After CCHPR Version 2 was completed, expanded hierarchical systems for both diagnoses and procedures were constructed by aggregating CCHPR codes into larger groupings and disaggregating them into smaller groupings of one or more individual ICD-9-CM codes. A four-level system was developed for diagnoses and a three-level system was developed for procedures.

Decisions about aggregating and disaggregating were based on the same criteria employed to develop CCHPR. The first level of both hierarchical systems is most general and conforms, for the most part, to the body systems and conditions that define ICD-9-CM chapters. Examples for diagnoses include 2 Neoplasms, 5 Mental disorders, and 7 Diseases of the circulatory system, and for procedures, 9 Operations on the digestive system and 13 Obstetrical procedures.

The second level consists of more precisely defined categories within the first level, the third level contains even more specific categories, and so on. Eighteen major diagnosis groupings were divided into more than 600 diagnosis categories and 16 major procedure groupings were divided into about 300 procedure categories.

As an example, the first several categories of the diagnosis system are reproduced here. The original CCHPR Version 2 categories are in bold, with the CCHPR number in brackets following the word label. The classification represents a four-level hierarchy of increasingly more specific categories.

1 Infectious and parasitic diseases
   1.1  Bacterial infection
      **1.1.1 Tuberculosis [1.]**
      **1.1.2 Septicemia (except in labor) [2.]**
         1.1.2.1 Streptococcal septicemia
         1.1.2.2 Staphylococcal septicemia
         1.1.2.3 E. Coli septicemia
         1.1.2.4 Other gram negative septicemia
         1.1.2.5 Other specified septicemia
         1.1.2.6 Unspecified septicemia
      **1.1.3 Sexually transmitted infections (not HIV or hepatitis) [9.]**
      **1.1.4 Other bacterial infections [3.]**
   **1.2  Mycoses [4.]**
      1.2.1 Candidiasis of the mouth (thrush)
      1.2.2 Other mycoses

- Level 1 (one digit) roughly follows the ICD-9-CM chapters and always consists of aggregations of CCHPR categories.
- Level 2 (two digits) consists either of aggregations of CCHPR categories (e.g., 1.1 Bacterial infection) or single CCHPR categories (e.g., 1.2 Mycoses [4.]).
- Level 3 (three digits) consists either of single CCHPR categories (e.g., 1.1.1 Tuberculosis [1.]) or ICD-9-CM codes split from a CCHPR category (e.g., 1.2.1 Candidiasis...).
- Level 4 categories (four digits) are always disaggregations of individual CCHPR categories (e.g., 1.1.2.1 Streptococcal septicemia).

Categories within a level are mutually exclusive; hence the number of cases in categories 1.1.1, 1.1.2, 1.1.3, and 1.1.4 will add to the number of cases in 1.1 Bacterial infection. Any minor differences can be attributed to rounding error (because the numbers of discharges are weighted to national estimates).

Appendix C provides details and coding for the expanded diagnosis classification, and Appendix D provides the same information for procedures.

---

[1]Electronic versions of the CCHPR classification schemes will be available in a future HCUP-3 Research Note and can be obtained now by contacting the lead author.

# Data

## Healthcare Cost and Utilization Project (HCUP-3)

The statistics presented in this report are based on data pertaining to hospital inpatient stays from the Healthcare Cost and Utilization Project (HCUP-3). HCUP-3 is a Federal-State-industry partnership in health care data. HCUP's objectives are to obtain data from statewide information sources, primarily State governments and hospital associations; design and develop a multi-State health care database to be used for health services research and health policy analysis; and release data to public and private users.

HCUP-3's Nationwide Inpatient Sample (NIS) Release 1 contains discharge-level clinical and resource use information included in a typical discharge abstract. NIS Release 1 covers 11 States (Arizona, California, Colorado, Florida, Illinois, Iowa, Massachusetts, New Jersey, Pennsylvania, Washington, and Wisconsin), includes nearly 900 hospitals and 6 million discharges per year, and spans the years 1988-92.

NIS is designed to be a 20-percent sample of U.S. "community" hospitals, as defined by the American Hospital Association (AHA). The AHA defines community hospitals as "all nonfederal, short-term, general and other specialty hospitals, excluding hospital units of institutions" (American Hospital Association, 1993). The HCUP-3 sample is a stratified probability sample of hospitals in the frame, with sampling probabilities proportional to the number of U.S. community hospitals in each stratum.

The hospital universe is defined using the AHA Annual Survey of Hospitals. This universe of hospitals is divided into strata using five hospital characteristics: ownership/control, bed size, teaching status, rural/urban location, and geographic region. Hospitals from HCUP-3 participating States (the sampling frame) are selected to represent these strata, and all discharges from sampled hospitals are included in the database. Weights indicate the number of discharges that the sample discharge represents in the universe of discharges from U.S. hospitals for that year in that stratum. The total number of discharges in the universe from that stratum is taken from the AHA Annual Survey of Hospitals.

Because administrative data on inpatient stays were not created for research purposes, there may be problems with the reliability and validity of certain data elements. Green and Wintfield (1993) summarized the literature on coding errors for hospital administrative data and described a decline in error rates during the 1970s and 1980s. Fisher et al. (1992) reported that the accuracy of principal diagnosis and procedure has improved since 1983, when such information became important for determining reimbursement by Medicare and other payers. Green and Wintfield (1993) reported the results of a reabstraction study using records from the California Office of Statewide Health Planning and Development. Information on age and sex was most reliable (error rates less than 1 percent), and principal diagnosis was inaccurate in 9 percent of records. Whittle et al. (1991) reported that estimates of cancer incidence rates based on Medicare claims data were within 6 percent of estimates using the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) data.

Other problems inherent in hospital inpatient data include missing data, underreporting of socially stigmatized conditions such as alcoholism and drug abuse, and underreporting of minor procedures.

Analyses limited to principal diagnoses and procedures will produce an underestimate of diagnoses that tend to appear in secondary positions, such as hypertension, osteoporosis, and Alzheimer's disease (May et al., 1991). Diagnostic and minor therapeutic procedures, which usually appear as secondary procedures, likewise will be under-represented when the focus is on principal procedures. Because the unit of analysis for this study is the inpatient stay rather than the patient, principal diagnoses and procedures are employed to reduce double-counting of stays. Furthermore, the principal diagnosis represents the diagnosis which, after evaluation, was considered the primary reason for admission to the hospital, and thus may be of greater interest than all-listed diagnoses. Similarly, the principal procedure is of interest because it should be the primary therapeutic procedure received by the patient during the stay. Despite these definitions, other diagnoses and procedures are sometimes coded into the principal position. For example, diagnostic procedures are often coded into the principal procedure field.

## Study sample

A 20-percent sample of NIS discharges for 1992 was used for this study ($N = 1,239,148$). These discharges were weighted to obtain estimates that are representative of hospital inpatient discharges in the United States. The estimated total number of discharges represented in these analyses is 34,989,827. This estimate compares favorably with the estimate of 34,640,000 discharges (including newborns) provided by the National Hospital Discharge Survey (Graves, 1994).

The unit of analysis is the discharge, or hospital stay, rather than the patient. Because NIS is limited to inpatient hospital data, conditions treated and procedures performed on an ambulatory basis are not represented here.

## Descriptive statistics

For all analyses, only the principal diagnosis and principal procedure were used. The statistics include:

- Number of discharges.
- Percent of discharges.
- Mean total charges.
- Mean length of stay.
- Percent who died in the hospital.
- Percent male.
- Mean age in years.

This information is shown for diagnoses in Table 1 and for procedures in Table 2.

All charge data are charges for the hospitalization. The charge information shown for procedures (Table 2) is not limited to the charge for the procedure itself. Instead, it refers to the total charge for the hospitalization in which this procedure was listed as the principal procedure. Charges do not necessarily reflect costs, nor are they synonymous with reimbursements. Physician charges are not included in these figures. In the past, missing charge data was often a problem. In this study, charge data were present for 96.5 percent of all discharges, although the percentage of missing data for diagnosis and procedure categories varied. Charge data were available for more than 90 percent of cases in all diagnosis categories except four: 5.8 Preadult disorders [73.] (present for 75 percent of cases), 5.10 Personal history of mental disorder [75.] (present for 87 percent), 17.2.2 Administrative/ social admission [255.] (present for 78 percent), and

17.2.5 Other screening for suspected conditions [258.] (present for 66 percent).

Results are not presented for any diagnosis or procedure category for which the unweighted number of discharges is less than 70. Using a generalized variance technique for proportions, it was determined that a sample of at least 70 discharges is required to assure with 95-percent confidence that the reported proportions had a relative error of less than 30 percent (i.e., if the reported value is $p$, the error is $< 0.3\,p$).

## Using the tables

In general, the diagnosis and procedure categories listed in the tables and appendices follow the order determined by the ICD-9-CM system. For the most part, chapter headings have been maintained and the individual categories are arranged by numeric ICD-9-CM code.

The approach to finding a specific condition or procedure will vary, depending on the information desired. To locate a general condition, body system, or procedure, one approach is to scan the major headings provided as a Table of Contents in front of each table. Then go to that section of the table and browse through the specific categories until the appropriate condition or procedure is found.

To locate a particular ICD-9-CM code, go directly to the appendices. The categories are arranged by the first-listed ICD-9-CM code in the grouping. Scan the column of codes in the appendices to the appropriate location. Identify the corresponding category number and go to that category number in the table.

## Conclusion

This Research Note introduces Version 2 of the Clinical Classifications for Health Policy Research (CCHPR). These classification systems can be employed in many types of projects analyzing data on diagnoses and procedures. The mutually exclusive classification scheme (Appendices A and B) should be useful for those who want to aggregate their own data by a relatively small number of diagnosis and procedure categories for reporting purposes. This scheme can also help to identify populations for disease- or procedure-specific studies.

The expanded categories (Appendices C and D) can be used to provide statistical information

about relatively specific conditions. Because they are hierarchical, they can also be used to examine the contribution of specific conditions to larger diagnosis categories.

In addition, this Research Note provides information on key characteristics of hospitalized patients and outcomes of care by CCHPR category. These statistics should prove useful to health policy analysts, health services researchers, and health care administrators in need of information on hospital discharges by principal diagnosis and procedure.

## References

American Hospital Association. (1993). *American Hospital Association hospital statistics.* 1993-94 Edition. Chicago.

Elixhauser, A., Andrews, R.M., and Fox, S. (1993). *Clinical Classifications for Health Policy Research: Discharge statistics by principal diagnosis and procedure.* (AHCPR Pub. No. 93-0043, Provider Studies Research Note 17.) Rockville, MD: Agency for Health Care Policy and Research.

Fisher, E.S., Whaley, F.S., Krushat, W.M., et al. (1992). The accuracy of Medicare's hospital claims data: Progress has been made, but problems remain. *American Journal of Public Health*, 82, 243-248.

Graves, E.J. (1994). National Hospital Discharge Survey: Annual summary, 1992. Series 13, No. 119. *Vital and Health Statistics.* (DHHS Pub. No. (PHS) 94-1779.) National Center for Health Statistics, Public Health Service. U.S. Government Printing Office.

Green, J., and Wintfield, N. (1993). How accurate are hospital discharge data for evaluating effectiveness of care? *Medical Care*, 31(8), 719-731.

May, D.S., Kelly, J.J., Mendlein, J.M., and Garbe, P.L. (1991). Surveillance of major causes of hospitalization among the elderly, 1988. *Morbidity and Mortality Weekly Report*, 40(SS-1), 7-17.

Public Health Service and Health Care Financing Administration. (1991). *The International Classification of Diseases, 9th Revision, Clinical Modification. Volumes 1, 2, and 3.* Fourth edition. (DHHS Pub. No. (PHS) 91-1260.) Washington: Public Health Service.

Puckett, C.D. (1994). *The educational annotation of ICD-9-CM.* Fourth edition. Reno, NV: Channel Publishing, Ltd.

Whittle, J., Steinberg, E., Anderson, G., and Herbert, R. (1991). Accuracy of Medicare claims data for estimation of cancer incidence and resection rates among elderly Americans. *Medical Care*, 29(12), 1226-1236.