
Diagnosis Clusters Adapted for ICD-9-CM and ICHPPC-2

Ronald Schneeweiss, MB, Daniel C. Cherkin, PhD, L. Gary Hart, Dennis A. Revicki, PhD, Loyd J. Wollstadt, MD, ScM, Michael J. Stephenson, MB, Jack Froom, MD, Earl V. Dunn, MD, Herbert L. Tindall, MD, and Roger A. Rosenblatt, MD, MPH

An ad hoc committee of the North American Primary Care Research Group (NAPCRG) was appointed to adapt the diagnosis clusters instrument for use with the International Classification of Diseases (ICD-9-CM) and the International Classification of Health Problems in Primary Care (ICHPPC-2). This article describes the development and testing of the final roster of 110 diagnosis clusters for family physicians. Almost 90 percent of all diagnoses recorded by family physicians in a variety of settings were included in the clusters. The diagnosis clusters can be used in the analysis of large databases and facilitate comparisons between different providers and practices.

Diagnosis clusters were developed as an instrument to facilitate the analysis of large morbidity (diagnostic) data sets.¹ Briefly, the diagnosis clusters group diagnoses that are clinically related and elicit a similar clinical response from the physician. The clusters incorporate the great majority of the recorded diagnoses used to describe patient visits to ambulatory office-based physicians. Using the clusters, similar diagnoses with low individual frequency but high aggregate frequency are appropriately highlighted, eg, fractures and dislocations. In addition, the clusters were shown to decrease the effect of the idiosyncratic diagnosis labeling and coding patterns of health care providers in different practices.¹

Diagnosis clusters have been used to conduct research into the content of family practice² and ambulatory care.³ They have also been useful as a method to monitor and report on the diagnostic content of family practice residents' ambulatory experiences.⁴

The major contribution of the diagnosis clusters is their utility in reducing the large number of diagnoses used in ambulatory practice to a manageable, yet clinically meaningful, roster that can be utilized to conduct comparative analyses between providers and practices.

A great deal of interest in the diagnosis clusters was expressed by the participants attending the Seventh Annual North American Primary Care Research Group (NAPCRG) meeting held in Columbus, Ohio, in 1982. An ad hoc committee was appointed to standardize the diagnosis clusters for use in family practice. This article describes the work of the committee and presents an expanded diagnosis cluster roster based on the ninth revision of the *International Classification of Disease*,

From the North American Primary Care Research Group (NAPCRG) Ad Hoc Task Committee on Diagnosis Clusters. Requests for reprints should be addressed to Dr. Ronald Schneeweiss, Department of Family Medicine, HQ-30, University of Washington, Seattle, WA 98195.

*Clinical Modification (ICD-9-CM)*⁵ with a cross-tabulation for the second revision of the *International Classification of Health Problems in Primary Care (ICHPPC-2)*.⁶

METHODS

The original roster of diagnosis clusters was tested against the National Ambulatory Medical Care Survey (NAMCS) 1977 and 1978 for all office-based specialties. The diagnoses in this data set were coded according to the eighth revision of the *International Classification of Diseases (ICDA-8)*,⁷ which was in widespread use at that time.

In 1979 the ninth revision of the *International Classification of Diseases* was adapted for use in the United States as ICD-9-CM. The ad hoc committee was given a small grant by the membership of NAPCRG to test the diagnosis clusters against the NAMCS 1980 and 1981 data sets, which were coded in ICD-9-CM, and to derive a comparable roster for ICHPPC-2. The NAMCS data sets were selected as the most recent national database available, and the combined data for both 1980 and 1981 were used to reduce sampling variations.

The resulting list of high-frequency rubrics not included in the original diagnosis clusters were examined by two of the authors (R.S. and D.C.C.). All high-frequency codes or clusters of codes were incorporated into a revised ICD-9-CM-based diagnosis cluster roster according to the original criteria.¹ Briefly, these included all diagnostic categories or clusters of related diagnoses with a recorded frequency greater than 0.1 percent. "Other" residual categories, eg, "Other diseases of the circulatory system," were excluded. Once the revised clusters were determined, a separate computer analysis was run for family physicians to identify any diagnoses that were recorded with a frequency greater than 0.1 percent, but were not included in the roster derived from the data for all office-based specialties. Ten such diagnoses were identified and were added to the roster as a supplementary listing for family physicians.

A cross-tabulation with ICHPPC-2 was prepared and distributed to the committee for comment. The roster was subsequently reviewed by the committee at the ninth annual NAPCRG meeting in Orlando, Florida, in 1984, and a few minor changes were recommended. This final revised roster was sent to the committee for final comment. Committee members were requested to indicate whether any clusters or high-frequency diagnoses should be added to the roster. Three authors (D.R., L.J.W., and M.S.) tested the clusters on ICHPPC-2-coded computer databases in their family practice teaching programs at the Universities of East Carolina, Illinois at Rockford, and McMasters, respectively. The clustered data from these three practices were reviewed to identify any significant omissions. None was found.

RESULTS

The revised roster of diagnosis clusters based on ICD-9-CM consists of 100 clusters plus a supplementary list of ten categories that were added to incorporate additional high-frequency diagnoses recorded by family physicians.* Only two of these supplementary categories exceeded a recorded frequency greater than 0.2 percent, namely, unspecified viral illness and unexplained abnormal biochemical results.

NAMCS 1980 and 1981 included 1,533 separate ICD-9-CM rubrics to describe the conditions seen in ambulatory office-based practice for family physicians. Nine hundred forty-five of these rubrics (62 percent) were included in the diagnosis clusters roster, representing 88 percent of all primary diagnoses recorded by family physicians. The inclusion of the additional supplementary listing of ten clusters for family physicians increased the total to 90 percent of all recorded primary diagnoses.

Eleven cluster categories contained only one

*The complete roster and supplementary list of diagnosis clusters is available upon request by writing to Dr. Ronald Schneeweiss.

TABLE 1. PERCENTAGE OF ICD CHAPTERS INCLUDED IN THE DIAGNOSIS CLUSTERS (DCs): BASED ON NAMCS 1980 AND 1981 FOR FAMILY PHYSICIANS

Chapter Number	Chapter Title	Percentage of NAMCS Primary Diagnoses	Percentage of Chapters Included in the DCs
1	Infectious and Parasitic Disease	3.3	76.0
2	Neoplasms	1.2	100.0
3	Endocrine, Nutritional, and Metabolic Diseases and Immunity Disorders	6.1	90.3
4	Diseases of the Blood and Blood-Forming Organs	0.7	89.0
5	Mental Disorders	2.4	93.1
6	Diseases of the Nervous System and Sense Organs	5.4	86.5
7	Disease of the Circulatory System	13.2	95.5
8	Diseases of the Respiratory System	17.4	95.5
9	Diseases of the Digestive System	5.7	80.4
10	Diseases of the Genitourinary System	5.3	88.9
11	Complications of Pregnancy, Childbirth, and the Puerperium	0.2	93.3
12	Disease of the Skin and Subcutaneous Tissue	4.0	85.9
13	Disease of Musculoskeletal System and Connective Tissue	7.6	80.3
14	Congenital Anomalies	0.08	0.0
15	Conditions Originating in the Perinatal Period	0.03	0.0
16	Symptoms, Signs and Ill-Defined Conditions*	3.8	65.0
17	Injury and Poisoning	9.5	93.8
18	(Supplementary) Classification of Factors Influencing Health Status, Contact with Health Service (V codes)	13.0	95.0
	All recorded diagnoses	100.0	89.8

*Note: The addition of the ten supplementary clusters listing high-frequency fall-throughs for family physicians increased the percentage of diagnoses in Chapter 16 included in the diagnosis clusters from 41.6 percent to 65.0 percent. The overall percentage of recorded diagnoses included in the diagnosis clusters was similarly increased from 87.9 percent to 89.8 percent

ICD-9-CM rubric, but these were high-frequency diagnoses that did not readily fit any of the previously identified clusters, eg, in descending order, obesity (rank 21), viral warts (rank 33), allergic reaction NOS (rank 49), abdominal pain (rank 60).

In 14 out of the 18 ICD chapters, the diagnosis clusters included more than 80 percent of the recorded diagnoses belonging to that chapter (Table 1). The four chapters with less than an 80 percent capture were Chapter 1, "Infectious and Parasitic Diseases," with 76 percent represented in the diagnosis clusters. Not surprisingly, Chapter 14, "Congenital Anomalies," and Chapter 15, "Perinatal Morbidity and Mortality," are almost never recorded in the ambulatory setting.

Only 42 percent of diagnoses in Chapter 16, "Symptoms, Signs and Ill-Defined Conditions," were included in the original diagnosis clusters

roster because of the difficulty in relating these diagnoses in a clinically meaningful way. The addition of the supplementary list for family physicians, which mainly included rubrics from this chapter, raised the inclusions of symptoms and signs from 42 percent to 65 percent. The percentage of all recorded diagnoses included in the diagnosis clusters was similarly increased from 87.9 percent to 89.8 percent overall.

The diagnosis clusters included 193 of the 367 ICHPPC-2 rubrics. These rubrics incorporated 88, 90, and 93 percent, respectively, of all the recorded diagnoses in the three ICHPPC-2 databases that were tested. In the case of five clusters, peripheral neuropathy (rank 48), skin keratoses (rank 53), stricture of urethra (rank 77), strabismus (rank 81), and allergy test (rank 83), no equivalent separate ICHPPC rubric existed, since these diagnoses were hidden in "other" categories in that classification.

DISCUSSION

The diagnostic content of ambulatory office-based practice is of interest, not only in terms of describing the specialty of family practice but also in attempting to understand differences between different providers, practice settings, and geographic areas. Some clinically meaningful way of aggregating the large number of individual diagnoses recorded in ambulatory practice is essential to enable research in this area to proceed. This diagnosis cluster roster represents a tested instrument that appears to achieve this goal.

All disease classifications organize and aggregate the labels physicians attach to patients' health problems. The need for specificity in any morbidity index has made the *International Classification of Disease* (ICD) unsuitable for aggregate analyses that look at the broad spectrum of the diagnostic content of medical practice. The *International Classification of Health Problems in Primary Care* (ICHPPC-2) represents an attempt to reduce the large number of ICD categories to more manageable proportions. However, even the 367 rubrics in ICHPPC-2 are too many to serve as a convenient instrument for comparative analyses. Although often used in reports, the 17 ICD chapters are too broad and blur meaningful distinctions between providers and practices.

The diagnosis clusters presented in this article have been tested on functioning databases and represent the consensus of the NAPCRG ad hoc committee. They include almost 90 percent of all recorded morbidity in different primary care settings and have been carefully constructed to permit a cross-tabulation between ICD-9-CM and ICHPPC-2.

The diagnosis clusters are intended as an instrument designed to facilitate data analysis and

are not presented as a new classification. The supplementary list can be expanded as needed to meet specific regional or research needs and is not meant to be exclusive. The diagnosis cluster roster described in this article represents a standardized version endorsed by NAPCRG that is compatible with both ICD-9-CM and ICHPPC-2.

Acknowledgment

This work was supported in part by grants from the Robert Wood Johnson Foundation and the North American Primary Care Research Group.

References

1. Schneeweiss R, Rosenblatt RA, Cherkin DC, et al: Diagnosis clusters: A new tool for analyzing the content of ambulatory medical care. *Med Care* 1983; 21:105-122
2. Rosenblatt RA, Cherkin DC, Schneeweiss R, et al: The structure and content of family practice: Current status and future trends. *J Fam Pract* 1982; 15:681-721
3. Rosenblatt RA, Cherkin DC, Schneeweiss R, Hart LG: The content of ambulatory medical care in the United States: An interspecialty comparison. *N Engl J Med* 1983; 309:892-897
4. Schneeweiss R: Diagnosis clusters: A new approach for reporting the diagnostic content of family practice residents' ambulatory experiences. *J Fam Pract* 1985; 20:487-492
5. Ninth Revision, International Classification of Diseases, Clinical Modification (ICD-9-CM). Report of the Commission on Professional and Hospital Activities. Ann Arbor, University of Michigan, 1978
6. ICHPPC-2: International Classification of Health Problems in Primary Care. Second Report of the Classification Committee of the World Organization of National Colleges, Academies and Academic Associations of General Practice/Family Physicians. New York, Oxford University Press, 1979
7. International Classification of Diseases, rev 8, adapted for use in the United States. National Center for Health Statistics (Hyattsville, Md). PHS publication No. 1693. Government Printing Office, 1968