

REFERENCE 1

Core Health Data Elements

*Edited Version of the Report of the National Committee
on Vital and Health Statistics
U.S. Department of Health and Human Services
August 1996*

Table of Contents

1. Executive Summary	129
2. Introduction.....	133
3. Background	134
4. Process followed by the Committee	136
4.1. Review State-of-the-Art of Widely-Used Core Data Sets	136
4.1.1. Background	136
4.1.2. Current Activities	136
4.1.3. Compendium of Core Data Elements	137
4.1.4. Other Widely Used Data Sets.....	137
4.2. Obtain Input, Through Hearings and other Means	138
4.3. Progress, Issues and Problems Raised.....	138
4.4. Interact Closely with Recognized Standards-setting Groups	140
4.5. Promote Consensus by Identifying Areas of Agreement on Data Elements	140
4.5.1. The Core Health Data Elements	141
4.5.2. The Core Health Data Elements (Listed by Readiness for Implementation)..	156
4.5.3. Additional Data Items	157
4.5.4. Related Data Set Activities	158
5. Conclusions	158
6. Recommendations to the Department	159
7. Future Activities	160

1. Executive Summary

Introduction

The identification, definition, and implementation of standardized data in the health care and health care information fields are long overdue. The increasing use of electronic data, the evolving managed care field, and the growing requirement for performance monitoring and outcomes research have made it imperative that all health data collection activities, where possible, utilize standardized data elements and definitions.

The National Committee on Vital and Health Statistics (NCVHS) has undertaken a first step in bringing together leaders in the field to seek consensus on a small set of data elements that are often considered the core of many data collection efforts. The Committee's goal has been to develop a set of data elements with agreed-upon standardized definitions that, when needed in a data collection effort, can be used to collect and produce standardized data. The intent is not to specify a data set for mandated external reporting; the list of recommended data elements is by no means exhaustive, and, unlike earlier activities, is not a "data set" to be used in a specific setting.

It is the expectation of the Committee that the health care field will find these recommended data elements to be fundamentally important for any collection of person and health care encounter data and will consider these elements and standardized definitions for inclusion in their data collection efforts wherever possible. Favorable input has been received from a wide range of experts, and these elements should be compellingly useful both to states and to provider organizations.

Background

In August 1994, the Department of Health and Human Services asked the Committee to provide information and advice that will help maximize the utility of core person and encounter data for meeting the Department's responsibilities. Specifically, the Department charged the Committee to:

- Review state-of-the-art of widely-used core data sets in the United States and other countries (including coding and formatting features that allow for flexibility);
- Obtain input, through hearings and other means, from the diverse parties who will report and use standardized data sets;
- Interact closely with recognized standards-setting groups; and
- Promote consensus by identifying areas of agreement on data elements and data sets among different stakeholders and areas that will require further research and development before consensus can be reached.

In developing a strategy for accomplishing these tasks, the Committee described a context in which the project would be undertaken that included the following issues:

- Why such data sets are needed in the current and evolving health care arena;
- What multiple functions they might accomplish for a variety of different users;
- What data elements (including definitions, vocabularies and coding structures) they might contain; and
- What potential problems, such as assuring data quality and preserving confidentiality of identifiable records, can be expected and what approaches might be used to address these problems.

The Core Health Data Elements

The following list of data elements contains those elements selected for the first iteration of this process. Consensus has been reached on definitions for some of these elements; for others, there is much agreement, but definitions must still be finalized; and for a third group, additional study and testing are needed.

These elements apply to persons seen in both ambulatory and inpatient settings, unless otherwise specified. For the first 12 elements, with the exception of unique identifier, information may not need to be collected at each encounter. Standard electronic formats are recommended to the extent that they have been developed.

CORE HEALTH DATA ELEMENTS PROPOSED FOR STANDARDIZATION

- 1. Personal/Unique Identifier 2/**
- 2. Date of Birth**
- 3. Gender**
- 4. Race and Ethnicity**
- 5. Residence**
- 6. Marital Status**
- 7. Living/Residential Arrangement 1/**
- 8. Self-Reported Health Status 2/**
- 9. Functional Status 2/**
- 10. Years of Schooling**
- 11. Patient's Relationship to Subscriber/Person Eligible for Entitlement**
- 12. Current or Most Recent Occupation and Industry 2/**
- 13. Type of Encounter 2/**
- 14. Admission Date (inpatient)**
- 15. Discharge Date (inpatient)**
- 16. Date of Encounter (outpatient and physician services)**
- 17. Facility Identification 1/**
- 18. Type of Facility/Place of Encounter 1/**
- 19. Health Care Practitioner Identification (outpatient) 1/**
- 20. Location or Address of Encounter (outpatient)**
- 21. Attending Physician Identification (inpatient) 1/**
- 22. Operating Clinician Identification (inpatient) 1/**
- 23. Health Care Practitioner Specialty 1/**

24. Principal Diagnosis (inpatient)
25. Primary Diagnosis (inpatient)
26. Other Diagnoses (inpatient)
27. Qualifier for Other Diagnoses (inpatient)
28. Patient's Stated Reason for Visit or Chief Complaint (outpatient) 2/
29. Diagnosis Chiefly Responsible for Services Provided (outpatient)
30. Other Diagnoses (outpatient)
31. External Cause of Injury
32. Birth Weight of Newborn
33. Principal Procedure (inpatient)
34. Other Procedures (inpatient)
35. Dates of Procedures (inpatient)
36. Procedures and Services (outpatient)
37. Medications Prescribed
38. Disposition of Patient (inpatient) 1/
39. Disposition (outpatient)
40. Patient's Expected Sources of Payment 1/
41. Injury Related to Employment
42. Total Billed Charges 1/

Footnotes: 1/ element for which substantial agreement has been reached but for which some amount of additional work is needed; 2/ element which has been recognized as significant but for which considerable work remains to be undertaken. A lack of footnote indicates that the element is ready for implementation.

Additional Data Items

While reviewing the draft list of data elements, respondents indicated a number of additional data elements that they felt were important core elements. Examples include information on health behaviors, such as smoking and alcohol consumption; information on preventive services; language ability; severity of illness indicators; provider certainty of diagnostic information; information to link a mother's and infant's charts; information on readmissions and complications. Future projects may undertake to seek consensus among some of these items.

Conclusions

As a result of the process followed in the conduct of this project and based on careful analysis by its members, the Committee has reached the following conclusions:

- The response to the Committee's activities through both participation in meetings and written comments indicates that the health care information field is solidly in favor of the identification and use of standardized data elements and definitions.
- The number of standards-setting organizations is growing; however, all that addressed the Committee are actively seeking participation by a 'recognized' leader/group who can forge consensus for the health care information field. But time is short; decisions are being made by organizations now.
- Response was significant and positive to the Committee's request to review a set of core data elements that were identified after a series of hearings and other information-gathering

efforts were completed. Most organizations were supportive in wanting to 'get on board' with standardized data elements.

- There is already consensus among data collectors and users for a significant number of data elements, especially elements related to person descriptors and to selected information on inpatient and ambulatory encounters.
- There is less agreement on data definitions, even for data items that have been in the field for years. Definitions must be refined and made available in standardized formats to data collectors.
- There are data items, such as health status and functional status, that are considered crucial elements, but for which substantial additional evaluation and testing must be undertaken to reach consensus on standardized content and definition.
- Because they recognize the significance of this project, respondents also recommended a number of additional items that they would like evaluated and possibly included in a core set of standardized data elements.

Recommendations to the Department

1. The Committee recommends that the HHS Data Council:

- Circulate the report within the Department for review and constructive criticism.
- Investigate the formation of leadership sites within the Department for each of the standards-setting organizations.
- Refer the core health data elements recommendations to the National Uniform Claim Committee for their consideration as they study the issue of uniform data elements for paper and electronic collection in Fall 1996.
- Provide stable resources to the project to establish an interagency work group, with DHSS taking the lead, to work with the key standards-setting organizations in the area of core health data elements.

2. The Committee recommends the following actions specifically related to the core data elements:

- For those elements that the Committee recommends as being ready to standardize, request each of the data collection entities within the Department to review the set of data elements and to match data contents and definitions with similar items that they are currently collecting or plan to collect. Report to the HHS Data Council on the viability of these elements and definitions being adopted in their program. If a reporting entity is using a different element or definition, explain why their current usage is preferable.
- Support implementation and testing activities for those data elements for which agreement on definitions has been reached and those for which minimal additional work is needed on definitional agreement. Public and private participants have indicated a willingness to work together to disseminate information, test data elements, and utilize electronic means to ensure the widest dissemination of these activities.

- For those data elements, which have been recognized as significant core elements, but for which there is not consensus on definition, support the formation of a public-private working group to conduct or coordinate additional study or research and to further refine definitions. This group, or a separate group, could also be the focus for evaluating additions to the list of core data elements and for setting up methods for testing and promulgating the final products.
 - Place the Committee's report, elements and definitions on an appropriate departmental Home Page as guidance to the field and as a means of encouraging use and soliciting further comments and suggestions while the report is under review within the Department.
3. Because agreement on a unique personal identifier has been recognized as a key element to the successful establishment of core data elements, and their use, support the formation of a public-private working group to study and provide recommendations in this area.
4. Support the NCVHS in continuing its work in this area, especially using its expertise to discuss research issues, to assist in consensus building, and to participate with the Data Council in the implementation of the core health data element project recommendations.

Core Health Data Elements Report

2. Introduction

The identification, definition, and implementation of standardized data in the health care and health care information fields are long overdue. Information is collected by a wide range of users and in a myriad of different formats. Work has been undertaken in the past to try to bring some semblance of order to selected areas of health data collection, especially in the areas of hospital inpatients and physician office visits.

The ever-expanding sites of care, combined with the increasing use of electronic data, make it imperative that all health data collection activities, where possible, utilize standardized data elements and definitions. Standardized data elements will be vitally important in the evolving managed care field, where there is a need to follow individuals through a continuum of care and at multiple sites. Performance monitoring and outcomes research are two additional areas that are currently hampered by the inability to link data sets from various sources due to varying data elements and definitions.

The National Committee on Vital and Health Statistics has completed a two-year project requested by the Department of Health and Human Services to review the current state of health-related core data sets; obtain input on their collection and use; interact with data standards-setting groups; and, most importantly, promote consensus by identifying areas of agreement on core health data elements and definitions. The Committee's goal has been to develop a set of data elements with agreed-upon standardized definitions that, when needed in a data collection effort, can be used to collect and produce standardized data. The intent is not to specify a data set for mandated external reporting; the list of recommended data elements is by no means exhaustive, and, unlike earlier activities, is not a "data set" to be used in a specific setting.

It is the expectation of the Committee that the health care field will find these recommended data elements to be fundamentally important for any collection of person and health care encounter data and will consider these elements and standardized definitions for inclusion in their data collection

efforts wherever possible. Favorable input has been received from a wide range of experts, and these elements should be compellingly useful both to states and to provider organizations.

3. Background

The National Committee on Vital and Health Statistics (NCVHS) and the Department of Health and Human Services, which it advises, have initiated and completed the first iteration of a process to identify a set of core health data elements on persons and encounters or events that can serve multiple purposes and would benefit from standardization. In August 1994, the Department recognized the National Committee's unique history in promoting standardization of health information when it asked the Committee to provide information and advice that will help maximize the utility of core person and encounter data for meeting the Department's responsibilities. More recently, the Department has been asked by the Vice President to play a leadership role, working with the Committee, in accelerating evolution of public and private health information systems toward more uniform, shared data standards.

Charge and Context

Specifically, the Department charged the Committee to:

- Review state-of-the-art of widely used core data sets in the United States and other countries (including coding and formatting features that allow for flexibility);
- Obtain input, through hearings and other means, from the diverse parties who will report and use standardized data sets;
- Interact closely with recognized standards-setting groups; and
- Promote consensus by identifying areas of agreement on data elements and data sets among different stakeholders and areas that will require further research and development before consensus can be reached.

In developing a strategy for accomplishing these tasks, the Committee described a context in which the project would be undertaken that included the following issues:

- Why such data sets are needed in the current and evolving health care arena;
- What multiple functions they might accomplish for a variety of different users;
- What data elements (including definitions, vocabularies and coding structures) they might contain; and
- What potential problems, such as assuring data quality and preserving confidentiality of identifiable records, can be expected and what approaches might be used to address these problems.

In accepting these challenges, the Committee seeks to facilitate consensus development and incorporate the concepts of multiple use, continued change, and long-term evolution of core data elements into general thinking and practice. The goal is to see what commonalities already exist and to what extent there can be further movement toward greater commonality of terms and consistency of definition.

Uniform Data Set Development

The National Committee on Vital and Health Statistics has been a sentinel organization in the area of uniform data efforts. Promoting the standardization of health information has been a consistent and defining Committee activity for 25 years. The Committee's efforts, first in the area of inpatient hospital data (the Uniform Hospital Discharge Data Set or UHDDS) and later in the area of ambulatory care (the Uniform Ambulatory Care Data Set or UACDS) have moved the country in the direction of achieving comparability in the health data collected by federal agencies, states, localities and the private sector, as well as in the international community. The Committee recognizes the need for uniform, comparable standards across geographic areas, populations, systems, institutions and sites of care to maximize the effectiveness of health promotion and care and minimize the burden on those responsible for generating the data. To this end, the Committee has advised the Department on such matters as Federal-state relationships, nomenclatures and classification systems, core data sets, and access and confidentiality issues.

The data sets promulgated by the NCVHS have become *de facto* standards in their areas for data collection by Federal and state agencies, as well as public and private data abstracting organizations. They have influenced the claim forms on which Medicare and Medicaid data sets are based. Both the UHDDS and UACDS have been reviewed and updated by the NCVHS and the Department in recent years. In addition, the Committee and Department have been involved in activities related to standardizing the collection of data in the long-term care setting.

The UHDDS currently in use was promulgated by the Department in 1985; the NCVHS recommended and circulated a revision in 1992, with additional recommendations from an Interagency Task Force in 1993. The UACDS has never been officially promulgated by the Department, but a 1989 revision by the NCVHS and an Interagency Task Force has been widely circulated, as has a further refinement by the NCVHS in 1994. No decisions have been made by the Department on any of these recommended revisions of either the UHDDS or the UACDS.

Standardization and Confidentiality

In recent years, the Committee has recognized the importance of electronic standardization efforts, which are taking place in the business community. The Committee has appointed a liaison to participate in selected meetings of the American National Standards Institute (ANSI) Accredited Standards Committee (ASC) X-12, a private sector coalition that is developing transmission standards for health data. The focus of the NCVHS effort has been on the content of the data to be transmitted, rather than the method of transmission.

The Committee has recognized that data confidentiality is a major concern in the collection of health data from an increasing number of sites, and the Committee has long been concerned with personal privacy and data confidentiality issues. In the early 1990's, it formed an Ad Hoc Work Group on Confidentiality to study issues related to confidentiality, unique personal identifiers and data linkage across time and systems. Thus, the NCVHS was the natural locus of the continuing efforts of DHHS to investigate the further standardization of health data.

4. Process Followed by the Committee

4.1. Review State-of-the-Art of Widely-used Core Data Sets

4.1.1. Background

Standardized data sets can serve many purposes in the current and future health care arena. In the evolving managed care field, the need to follow individuals through a continuum of care and at multiple sites will become increasingly necessary. Performance monitoring and outcomes research are two areas that are currently hampered by the inability to link data sets from various sources.

Standardized data sets, starting with the UHDDS developed by the NCVHS, have been in use for more than two decades. There has been substantial agreement on data elements in these sets, but less agreement on data definitions. In addition, these efforts have concentrated on individual sites of care, ie., hospital inpatient, physician office, and nursing home, which, until recently, were the traditional sites of most care. In recent years, the focus of health care has been shifting to hospital outpatient and other outpatient care, including clinic, hospice and home care, sites for which standardized data collection had not been developed.

The transference of data sets from the traditional sources has not fully met the needs of these sites. Additionally the move in the health care payment system to managed care has increased the need to be able to link data sets and individual records across time, facility, and broader geographic locations. Thus to meet the needs for standardized data, movement must be made toward standardized definitions for those data sets that are already in use, and for an increased use of standardized data elements and definitions by those data collection efforts for which no current standardized data sets exist.

The data sets that are currently standardized are prime examples of satisfying multiple purposes with a single data set. With the use of UHDDS-defined data, for example, state and private abstracting systems have been providing comparable state and local data for health planners for many years. These same data bases are being used to provide input to Federal surveys such as the National Hospital Discharge Survey (NCHS) and the Hospital Cost and Utilization Project (AHCPR). However, AHCPR is in the process of publishing findings indicating definitional discrepancies even within the organizations collecting the UHDDS.

4.1.2. Current Activities

The National Committee is well aware of the numerous efforts currently underway in both the public and private sectors to standardize health data, especially the progress made during the past 10-15 years in developing uniform data sets (Uniform Hospital Discharge Data Set and the Uniform Ambulatory Care Data Set) as well as common claim forms (Uniform Bill 82 and its successor UB 92 and HCFA 1500). To document the current status of activities in the field, the Committee awarded a contract to produce a Compendium of Core Data Elements. In addition, information was solicited by the NCVHS through two large-scale mailings, and public meetings were held with agencies and organizations which are currently collecting health data sets.

4.1.3. Compendium of Core Data Elements

To measure the current state of the use of various data sets, the Committee contracted with the Center for Health Policy Studies (CHPS) in Columbia, MD to begin identifying major data sets already in existence, especially in the private field. A range of organizations was contacted including health plans/insurers, trade or professional associations, employers, data standards organizations, and Government. Although 61 requests were made regarding data sets, almost one-third of respondents indicated that they did not have a set of health data items that they collected. Of 18 trade or professional associations contacted, only four submitted data sets. Half of the ten major employers who were asked to participate declined; only four actually sent in a data set. However, in the three remaining areas of health plans/insurers, government, and data standards organizations, the vast majority supplied data sets. A total of 31 responses were received.

Data sets received were assessed for their consistency with other data sets, particularly minimum data sets such as the UHDDS and the UACDS, the HCFA 1500 and the UB 92 data sets, and also with other current and future data sets under development by data standards organizations (ANSI). These comparisons also included consideration of the general availability, reliability, validity, and utility of data elements. A series of matrices were prepared that arrayed individual data elements in use or proposed for use by different organizations with the type of organization.

Producing the compendium was a much more involved effort than was originally envisioned, and probably is representative of problems to be overcome in the future when standardization implementation is planned. Just trying to obtain data from some large organizations was quite difficult; responses were not received in a timely fashion, and when received, the data layouts often were computerized lists rather than lists of data items with their definitions. In some instances, lists of items were received with many basic data items not included. In these cases, it is possible that the data items, such as person characteristics, are part of a more basic file kept by the organization, and the information for that file was not included.

In a number of instances, lists of data items were obtained, but without definitions. Previous experience indicates that at least some, if not many, of these data items have differing definitions. The Committee recognizes the importance of having both data items and identical definitions in order to compare and analyze data elements.

One problem that was encountered was that of requesting what the private organizations consider proprietary information. It was thought that this was one of the reasons why some organizations, especially private employers, declined to participate.

From the respondents, a total of 138 different data elements were obtained. A large number were collected by only a few of the data sets. Also, although different data sets may include the same data element, in most cases it was not possible to verify that the data collection instructions and definitions were the same.

Based on the compendium effort, a working list of 47 data elements frequently collected or proposed for collection regarding eligibility, enrollment, encounters and claims in the United States was prepared.

4.1.4. Other Widely-used Data Sets

The Committee works closely with the National Center for Health Statistics, the Agency for Health Care Policy and Research, and the Health Care Financing Administration (HCFA). Updates of activities in each of the agencies are presented to the Committee on a regular basis. HCFA has also

provided information on its efforts to define a core data set for states and managed health care plans (McData), which is undergoing review at this time. At the October 1995 meeting of the NCVHS, a session was held at which the Department of Veterans Affairs, the Georgia State Department of Health, and others, demonstrated their institutions' integrated financial, clinical, consumer, and public health information systems that are currently in place or being tested.

During the October 1995 and March 1996 NCVHS meetings, Dr. Don Detmer, University of Virginia, updated the Committee on international progress in data standardization and computerized patient records. He had visited a number of western European countries speaking with experts in health information infrastructure, and reported that several countries now have a national policy of support for the computerized patient record. Dr. Detmer identified four overarching issues: privacy and confidentiality, computerized patient records, standards and classification, and knowledge-based management. Also in March, a consultant to the NCVHS updated the World Health Organization on the core data element activity and returned with input to the process.

4.2. Obtain Input, through Hearings and other Means

Armed with the extensive listing of potential data elements culled from the Compendium, in September 1995, the NCVHS contacted approximately 2,000 individuals and organizations in the health care utilization and data fields to seek their input in identifying those basic elements most in need of collection and/or in need of uniform definitions. In addition to requesting a written response from these experts, they were invited to participate in one of two special meetings organized by the Committee to discuss the project and to seek input. In order to have as wide a participation at the meeting as possible, both East and West coast meetings were held in Oakland, CA, in early November, and in Washington, DC, in early December.

Both meetings were successful at bringing together experts in the field and expanding the knowledge base of the Committee. Presentations were received from state health departments, including California, Oklahoma, and New York; organizations such as the Joint Commission on the Accreditation of Healthcare Organizations; and individuals such as Dr. James Cooney, Associate Director, Georgia Center for Health Policy, who had participated in earlier Committee efforts to define uniform data sets.

4.3. Progress, Issues, and Problems Raised

Several major issues were raised that were broader than the discussions of specific data elements. Virtually all saw the need for uniform data items and definitions, and the issue of a unique identifier was a frequent topic. This issue represents more than just what item or set of items the identifier will include; it opens up the whole issue of data linkage, privacy, and data confidentiality with its relevant benefits and risks. Another issue was the role of the National Committee itself as the source of information on common data elements. Most participants eagerly supported an independent committee, such as this, to gather input and advise the public health and health care communities. However, the activities envisioned by many participants go much farther than an advisory committee can handle. These discussions led to the issue of needing DHHS staff dedicated to participating in the meetings of numerous data standards committees, advising the Department, and producing further iterations of data elements as future agreement is reached. Currently, such a staff does not exist.

Some states and organizations are on the cutting edge of multiple use of standardized data. For example, the State of California, in testimony to the NCVHS, described its efforts in improving health and health care delivery by linking data collected through medical facilities, school-based health and educational data bases, as well as need-based data bases such as eligibility listings for the Special Supplemental Nutrition Program for Women, Infants and Children (WIC) or reduced school-lunch programs. This project has brought together efforts from several state agencies, including education (for the school data), agriculture (the source of WIC data in some states), as well as health departments. Consensus building on data elements and definitions was, as always, a complex issue. Data quality is a perennial issue. Although the UHDDS has been in the field for two decades and its data items are widely used by government and private organizations, issues of quality and comparability remain. A presentation by AHCPH reported on a study of 10 state data organizations and two statewide hospital associations participating in the Healthcare Cost and Utilization Project (HCUP-3). (Currently approximately 40 states collect health data on inpatient hospital stays.) AHCPH compared the 12 systems with the UB-92 and monitored deviations at 3 levels - easy, moderately difficult, and difficult to correct problems.

A detailed report of these findings is in the process of publication by AHCPH, but findings have shown that even well-recognized standards are not consistently followed. Any new data items, as well as the old, must be produced with clear instruction on data collection and coding.

Confidentiality of identifiable records is another critical issue. Currently, data are often shared within a facility in an identifiable format. However, identifiers are commonly removed when a data set is provided outside of a facility, such as to a state health data organization. And now, with movement toward HMO's, PPO's, and other types of managed care, there may be a greater need to share identifiable data. States have varying laws to protect the confidentiality of these data, and often the laws do not protect data that have crossed state lines. Sufficient penalties for breach of confidentiality either do not exist or are not enforced. There have been several proposals for Federal legislation in recent years; however, to date, no Federal legislation protecting the confidentiality of health records exists.

Several states, including California, Oklahoma, and New York presented findings on using a combination of key data items to perform probabilistic matches. Using items such as first name of mother; first digits of last name; date of birth; place of birth, etc., matches could be obtained without identifying the individual. It appeared that some types of data linkage could be obtained in states with smaller populations, but might not work nationwide. New York, using the last 4 digits of the Social Security Number, with other characteristics (such as date of birth), indicated a match rate exceeding 99 percent.

Problems could arise from adding and modifying data items and definitions too frequently. James Cooney, Ph.D., former member, NCVHS, described the burden to organizations from the addition of a single data item. Each item that is recommended must be considered carefully. Additionally, too frequent modification of items or definitions will cause confusion, overlapping data definitions in a single data year, and add to the burden of the facility or organization.

In addition to the presentations at the meetings, more than 100 written responses to the solicitation letter were reviewed and considered. Of these, approximately 70 percent provided information about their data elements. Approximately 30 percent of respondents were from state and local governments, followed by professional associations and the Federal Government with 18 Percent and 17 percent respectively. Providers, Insurers, and universities represented about 7 percent each.

The Committee reviewed all of the input received from the hearings, meetings, letters and other communications. In addition, the historical knowledge of the NCVHS and its earlier decisions in the area of data standardization played a role in the preparation of a listing of core data elements and, where possible, recommended definitions. The draft listing was again disseminated in early April

1996 to the original mailing list and especially to those who had provided earlier assistance. To assure the widest possible distribution, the document was also placed on the DHHS and NCHS Home Pages in an electronic format. More than 150 responses to this second request were received, including responses from the leaders in the health care and health care information fields.

4.4. Interact Closely with Recognized Standards-setting Groups

The importance of participating in meetings of the various standard-setting groups has been recognized by the Committee. Members of the Committee and DHHS staff participate when possible, however, the increasing numbers of groups and meetings is problematic from a staff and budget point of view. To identify the large number of organizations involved in various aspects of health data standards, staff at NCHS produced a report describing the various groups by type of organization. The report provides important background information on coordinators and promoters of standards development; lead standards-development organizations; organizations developing performance measures indicators; departmental organizations; international organizations; and others.

To obtain the latest plans, at its October 1995 meeting, the NCVHS held a session focused on Standards Development Organizations and related organizations. Participating organizations included:

- ANSI (American National Standards Institute)
- ANSI HISPP (Health Informatics Standards Planning Panel),
- ANSI ASC X-12 (Accredited Standards Committee),
- HL-7 (Health Level 7)
- WEDI (WorkGroup on Electronic Data Interchange)
- ASTM (American Society for Testing and Materials),
- NUBC (National Uniform Billing Committee),
- NUCC (National Uniform Claim Committee), and
- NCQA (National Committee for Quality Assurance).

Although Committee members were aware in a general way of ongoing standards developments activities, this session focused on the need for action being required now and in the near future if the health care community is to obtain and maintain a presence as data standards are developed and finalized. At the very minimum, there need to be "place holders" provided to standards organizations to inform them that certain data elements are critical elements, even when the specific format of the items is currently undecided. It became obvious that staff dedicated to participating in and monitoring the activities of these organizations is crucial if all relevant voices (including public health and epidemiology) are to be heard. At the March 1996 NCVHS meeting, many of the same standards-setting groups were present and indicated their support of the Committee's efforts.

4.5. Promote Consensus by Identifying Areas of Agreement on Data Elements

The major output of this project to date has been the recommendation of core data elements, definitions, vocabulary and classifications. This effort, described below, is the culmination of input from the historical knowledge and work of the Committee, including the uniform basic data sets already developed; and information provided in meetings, hearings, and through correspondence with Federal, state and local health agencies, private organizations, universities, etc.

The goal has been to develop a set of data elements with agreed-upon standardized definitions that, when needed in a data collection effort, can be used to collect and produce standardized data. The Committee's intent is not to specify a data set for mandated external reporting; not every element may be needed in a specific collection effort, and these data elements do not represent all of the important data items that are collected in the field or needed for specific applications. They do represent those items that are routinely collected in many efforts, such as basic person information, as well as items specific to inpatient or ambulatory care settings, such as provider information, diagnoses, and services.

It is hoped that, as data collection evolves, certain data items, such as personal data, (i.e., date of birth, race, occupation) will only need to be collected at time of entry into a health care plan or to be updated on an annual basis, to reduce the burden of data collection. Other data items are related to a specific episode of care and will be provided at each encounter.

4.5.1. The Core Health Data Elements

The following list of data elements contains those elements selected for the first iteration of this process. Consensus has been reached on definitions for the majority of these elements; for others, there is much agreement, but definitions must still be finalized; and for a third group, additional study and testing are needed. These elements apply to persons seen in both ambulatory and inpatient settings, unless otherwise specified. For the first 12 elements, with the exception of unique identifier, information may not need to be collected at each encounter. Standard electronic formats are recommended to the extent that they have been developed.

The Committee recognizes that this is an iterative process and has included in these recommendations several elements that have been proposed for standardization, even though no consensus currently exists concerning appropriate or feasible definitions. The description of the element indicates this present lack of agreement. The Committee has chosen to include these elements because it believes that the need for the type of information they contain will continue to increase. The Committee encourages the Department and its partners to give high priority to conducting evaluation and testing on such elements and also seeks to alert organizations developing standards or data sets to leave place holders for their inclusion. In addition, a number of elements for which consensus is close, must be field tested to confirm their definitions and collectibility. A listing of the Core Health Data Elements grouped by level of readiness for implementation is provided after the section with the definitions of each data element.

The NCVHS has undertaken parallel efforts to identify elements specific to mental health, substance abuse, disability and long-term care settings. Some recommendations in the area of mental health and substance abuse are included here. Other recommendations will be circulated for comment at a future time.

TABLE OF CORE HEALTH DATA ELEMENTS

1. Personal/Unique Identifier 2/
2. Date of Birth
3. Gender
4. Race and Ethnicity
5. Residence
6. Marital Status
7. Living/Residential Arrangement 1/
8. Self-Reported Health Status 2/
9. Functional Status 2/
10. Years of Schooling
11. Patient's Relationship to Subscriber/Person Eligible for Entitlement
12. Current or Most Recent Occupation and Industry 2/
13. Type of Encounter 2/
14. Admission Date (inpatient)
15. Discharge Date (inpatient)
16. Date of Encounter (outpatient and physician services)
17. Facility Identification 1/
18. Type of Facility/Place of Encounter 1/
19. Health Care Practitioner Identification (outpatient) 1/
20. Provider Location or Address of Encounter (outpatient)
21. Attending Physician Identification (inpatient) 1/
22. Operating Clinician identification 1/
23. Health Care Practitioner Specialty 1/
24. Principal Diagnosis (inpatient)
25. Primary Diagnosis (inpatient)
26. Other Diagnoses (inpatient)
27. Qualifier for Other Diagnoses (inpatient)
28. Patient's Stated Reason for Visit or Chief Complaint (outpatient) 2/
29. Diagnosis Chiefly Responsible for Services Provided (outpatient)
30. Other Diagnoses (outpatient)
31. External Cause of Injury
32. Birth Weight of Newborn
33. Principal Procedure (inpatient)
34. Other Procedures (inpatient)
35. Dates of Procedures (inpatient)
36. Procedures and Services (outpatient)
37. Medications Prescribed
38. Disposition of Patient (inpatient) 1/
39. Disposition (outpatient)
40. Patient's Expected Sources of Payment 1/
41. Injury Related to Employment
42. Total Billed Charges 1/

Footnotes: 1/ element for which substantial agreement has been reached but for which some amount of additional work is needed; 2/ element which has been recognized as significant but for which considerable work remains to be undertaken. A lack of footnote indicates that these elements are ready for implementation.

Person/Enrollment Data

The elements described in this section refer to information collected on enrollment or at an initial visit to a health care provider or institution. It is anticipated that these elements will be collected on a one-time basis or updated on an annual basis. With the exception of the personal/unique identifier, they do not need to be collected at each encounter.

1. Personal/Unique Identifier - the unique name or numeric identifier that will set apart information for an individual person for research and administrative purposes.

A. Name - Last name, first name, middle initial, suffix (e.g., Jr., III, etc.)

B. Numerical identifier

The personal/unique identifier is the element that is the most critical element to be collected uniformly. The NCVHS recommends the use of Social Security Number with a check item such as date of birth, while at the same time undertaking the study and evaluation needed to confirm this use or the recommendation of another identifier. More emphasis on the confidential use of SSN is essential. Standards groups should be consulted regarding setting criteria for recording of names.

Rationale and discussion

Without a universal unique identifier or a set of data items that can form a unique identifier, it will be impossible to link data across the myriad of healthcare locations and arrangements. In the 1992 revision of the Uniform Hospital Discharge Data Set (UHDDS), the NCVHS recommended "using the Social Security Number (SSN), with a modifier as necessary, as the best option currently available for this unique and universal patient identifier." However, recent testimony has led the Committee to investigate this issue further, in light of perceived inadequacies of the SSN (e.g., lack of a check digit, multiple SSN's, etc.), particularly when used alone, and impediments (legal and otherwise) to its use. Other potential problems include lack of numbers for newborns, legal and illegal non-citizens and persons who wish to hide their identity, as well as a recommendation that a system would need to be established to assign and track dummy numbers.

New York State presented testimony that indicated that the last four digits of the SSN combined with the birth date were capable of linking data to a very high degree of probability. The State of California has tested the use of a series of data items that are readily known by individuals and which can be combined to link data. By January 1998, all California State Department of Health data bases will contain five data items to facilitate linkage. These data items include birth name, date of birth, place of birth, gender, and mother's first name. Seven confirmatory data items (including SSN) should also be collected when possible.

Those present at the November and December 1995 NCVHS regional meetings agreed that the establishment of a unique identifier is the most important core data item. A unique identifier such as the SSN in conjunction with at least one other data item or, alternatively, an identifier drawn from another distinct set of data items routinely collected presently would seem the most viable. Whichever number is chosen, attention must be paid to which data linkages will be permitted and for what purposes.

Development of a unique identifier does not necessarily mean that the individual is identifiable to users. The NCVHS recognizes the vital importance of maintaining confidentiality and emphasizes that

any public use of a unique identifier should be in an encrypted form. The unique identifier must be developed and protected in such a way that the American public is assured that their privacy will be protected.

2. Date of Birth - Year, month and day - As recommended by the UHDDS and the Uniform Ambulatory Care Data Set (UACDS). It is recommended that the year of birth be recorded in four digits to make the data element more reliable for the increasing number of persons of 100 years and older. It will also serve as a quality check as the date of birth approaches the new century mark.

3. Gender - As recommended by the UHDDS and the UACDS.

1. Male
2. Female
3. Unknown/not stated

4. Race and Ethnicity - The collection of race and ethnicity have been recommended by the UHDDS and the UACDS, and these elements have a required definition for Federal data collection in Office of Management and Budget (OMB) Directive 15. The definition has been expanded slightly from the OMB requirement:

4A. Race

1. American Indian/Eskimo/Aleut
2. Asian or Pacific Islander (specify)
3. Black
4. White
5. Other (specify)
6. Unknown/not stated

4B. Ethnicity

1. Hispanic Origin (specify)
2. Other (specify)
3. Unknown/not stated

It is recommended that this item be self-reported, not based on visual judgment or surnames. Whenever possible, the Committee and participants recommended collecting more detailed information on Asian and Pacific Islanders, as well as persons of Hispanic Origin.

Rationale and discussion

The collection of this element allows for the investigation of issues surrounding health and health care by a person's race and ethnic background. Although it is best understood in conjunction with a socioeconomic indicator, researchers may gain a better understanding of the trends and impact of care on racial/ethnic minorities in the U.S. It remains unclear whether the modest health gains seen in low-income and racial/ethnic minority populations in the last thirty years will continue, considering the changes in the U.S. health care system. These data assist in the examination of disparities in stage of illness, care, and outcome, some of which have been documented in the past among racial and ethnic groups.

OMB is currently investigating the possibility of changes to this classification, and the Committee will await the OMB recommendations. The Committee is concerned about the possible inclusion of a

"multiracial" category, without an additional element requesting specific racial detail and/or primary racial identification, because of its anticipated impact on trend data and loss of specificity.

The National Association of Health Data Organizations has also opposed such an inclusion. A recent Bureau of Labor Statistics study found that only 1.5 percent of respondents will choose the multiracial category. The study also found that with the multiracial option there was a considerable decline in percentage terms (approximately 29 percent) of respondents choosing American Indian, Eskimo or Aleut. However, there is some evidence that the number of interracial marriages is accelerating.

5. Residence - Full address and ZIP code (nine digit ZIP code, if available) of the individual's usual residence.

Rationale and discussion

This recommendation is in accord with the 1992 UHDDS and the UACDS, as well as recommendations by the NCVHS Subcommittee on State and Community Health Statistics. The Subcommittee determined that residential street address has the advantage of enabling researchers to aggregate the data to any level of geographic detail (block, census tract, ZIP code, county, etc.) and is the best alternative to insure the availability of small area data. In addition, home address will allow the application of GIS (Geographic Information Systems) technology to the analysis of health issues. Some thought needs to be given to completing this item for persons with no known residence or persons whose residence is outside of the United States. Because the full residential address could serve as a proxy personal identifier, confidentiality of the complete information must be safeguarded in public use of the data.

6. Marital Status - The following definitions, as recommended by the NCVHS, should be used.

1. Married - A person currently married. Classify common law marriage as married.

A) living together

B) not living together

2. Never married - A person who has never been married or whose only marriages have been annulled.

3. Widowed - A person widowed and not remarried.

4. Divorced - A person divorced and not remarried.

5. Separated - A person legally separated.

6. Unknown/not stated

Rationale and discussion

The Committee recognizes that a person's social support system can be an important determinant of his or her health status, access to health care services, and use of services. Marital status is one element that is sometimes used as a surrogate for the social support system available to an individual and can be important for program design, targeting of services, utilization and outcome studies, or other research and development purposes. It also may be required to verify benefits.

7. Living/Residential Arrangement - The following definitions are recommended by the NCVHS:

7A. Living Arrangement

1. Alone
2. With spouse
3. With significant other/life partner
4. With children
5. With parent or guardian
6. With relatives other than spouse, children, or parents
7. With nonrelatives
8. Unknown/not stated

Multiple responses to this item are possible. This element refers to living arrangements only. Marital status is discussed in element 6.

7B. Residential Arrangement

1. Own home or apartment
2. Residence where health, disability, or aging related services or supervision are available
3. Other residential setting where no services are provided
4. Nursing home or other health facility
5. Other institutional setting (e.g. prison)
6. Homeless or homeless shelter
7. Unknown/not stated

Rationale and discussion

The usual living/residential arrangement of an individual is important for understanding the health status of the person as well as the person's follow-up needs when seen in a health care setting. Together with marital status, this element provides a picture of potential formal/informal resources available to the person. The element also provides information on patient origin for health resource planning, and for use as an indirect measure of socioeconomic status.

A key distinction to be ascertained in "residential arrangement" is whether organized care-giving services are being provided where the patient lives. The Committee encourages the use of the above definition, while continuing to study and evaluate other residential categories, such as those used by the Bureau of the Census.

8. Self-Reported Health Status - There was much interest in documenting health status, one element that can precipitate the demand for health care and help determine the prognosis, although there was no consensus on how its definition should be standardized. A commonly used measure is the person's rating of his or her own general health, as in the five-category classification, "excellent, very good, good, fair, or poor."

Used in the National Health Interview Survey and many other studies, this item has been shown to be predictive of morbidity, mortality, and future health care use, when collected in a general interview type of setting. This item would be collected at first clinical visit and periodically updated, at least annually. Additional evaluation and testing are needed on standardizing the health status element. At the present time, standards-setting organizations should assign place holder(s) for this element.

9. Functional Status - The functional status of a person is an increasingly important health measure that has been shown to be strongly related to medical care utilization rates. A number of scales have been developed that include both a) self-report measures, such as the listings of limitations of Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) and the National Health Interview Survey age-specific summary evaluation of activity limitations, and b) clinical assessments, such as the International Classification of Impairments, Disabilities and Handicaps (ICIDH) and the Resident Assessment Instrument (RAI) (widely used in nursing homes). In addition, there are some disabilities, such as severe mental illness or blindness, where ADLs and IADLs are not sufficient measures. Self-report and clinician measurements are each valuable, and having both available is especially informative. Whichever method is used should be designated. Particular scales are more appropriate for measuring different functions or disabilities and should be selected on the basis of the needs of the patient population (such as, use of social functioning scales for those with mental disorders and substance abuse). Functional assessment scales must also be age-appropriate. At present, there is no widely recognized instrument for measuring the functional status of children. Periodicity of assessment also is an issue.

Consideration of these various issues and additional study and evaluation are needed before recommendations can be made for standardizing functional status measurement. Work on this topic is currently ongoing in the NCVHS Disability and Long-Term Care Statistics Subcommittee. It is possible that the description of functional status may entail more than a single measure, thus needing space for more than one measure and/or an additional element to document the scale used. At the current time, however, it is crucial that standards-settings organizations set aside place holder(s) for this element.

10. Years of Schooling - Highest grade of schooling completed by the enrollee/patient. For children under the age of 18, the mother's highest grade of schooling completed should be obtained.

Rationale and discussion

Collection of years of schooling has been recommended by the NCVHS and others as a proxy for socioeconomic status (SES). Years of schooling has been found to be highly predictive of health status and health care use.

Ideally, one would also collect income to more fully define socioeconomic status. However, income questions are often considered intrusive, whereas years of schooling are more acceptable to respondents. The NCVHS Subcommittee on Ambulatory and Hospital Care Statistics commented in the 1994 UACDS revision that years of schooling completed is the most feasible socioeconomic element to collect in the UACDS.

11. Patient's Relationship to Subscriber/person eligible for entitlement -

- A. Self
- B. Spouse
- C. Child
- D. Other (specify)

Rationale and discussion

This relationship (i.e., self, spouse or child of subscriber) is often obtained and can be of importance for payment and research purposes.

12. Current or Most Recent Occupation and Industry - This data item is very useful to track occupational diseases as well as to better define socioeconomic status. Standardized coding schemes, such as the Census Bureau's Alphabetical Listing of Occupation and Industry and the Standardized Occupation and Industry Coding (SOIC) software developed by the National Institute for Occupational Safety and Health, should be reviewed. In some situations, it is possible that a free-form narrative will be collected in place of the codes, to be coded at a later point. The Committee feels that, over time, there will be increasing attention focused on this item and reaffirms its recommendations in the 1994 revisions to the UACDS that additional study and evaluation be conducted on the feasibility and utility of collecting and periodically updating information on a person's occupation and industry. In addition, the usefulness of both current/most recent occupation and industry as well as the addition of usual or longest held occupation and industry must be evaluated. All have significant value and could result in the collection of four separate data elements.

Encounter Data

The elements described in this section refer to information related to a specific health care encounter and are collected at the time of each encounter.

13. Type of Encounter - This element is critical to the placement of an encounter of care within its correct location, i.e., hospital inpatient, outpatient, emergency department, observation, etc. However, there was no clear-cut listing of mutually exclusive encounter locations or definitions to draw upon. This term is one that needs study and evaluation before it can be implemented. However, a place holder for this element is recommended to the standards-setting organizations.

14. Admission Date (inpatient)- Year, month, and day of admission as currently recommended in the UHDDS and by ANSI ASC X12. An inpatient admission begins with the formal acceptance by a hospital of a patient who is to receive health care practitioner or other services while receiving room, board, and continuous nursing services. It is recommended that the year of admission contain 4 digits to accommodate problems surrounding the turn of the century.

15. Discharge Date (inpatient) - Year, month, and day of discharge as currently recommended in the UHDDS and by ANSI ASC X12. An inpatient discharge occurs with the termination of the room, board, and continuous nursing services, and the formal release of an inpatient by the hospital. Four digits are recommended for the discharge year.

16. Date of Encounter (outpatient and physician services) - Year, month, and day of encounter, visit, or other health care encounter, as recommended by the UACDS and ANSI ASC X12. Each encounter generates a date of service that can be used to link encounters for the same patient over time. Grouping of similar services provided on different dates,

as is often the case under batch billing, can be problematic if specificity of data elements is lost; the objective is to encourage identifying a unique date of record for each encounter. However, for services billed on a batch basis, two dates would be required to encompass the range of dates from the beginning of all treatments included under the batch (global) code to the end, with a check box to indicate that this is a batch-based encounter.

Health Care Facility and Practitioner Identifiers

Each provider should have a universal unique number across data systems. The National Provider Identifier and National Provider File (NPI/NPF), currently under development by the Health Care Financing Administration (HCFA) and intended for implementation in 1997, could and should meet this need, if all providers are included. The NPI/NPF will provide a common means of uniquely identifying health care providers, including institutions, individuals, and group practices, both Medicare providers and those in other programs. Participation in the system will be voluntary for non-HCFA providers at first. Currently some states are using state facility identifiers, but the Committee recommends that these identifiers be superseded by the NPI/NPF.

The immediate goal of the NPI/NPF project is to support HCFA's Medicare Transaction System initiative by providing a single, universal method for enumerating the providers who serve Medicare beneficiaries. It will do so by assigning a unique identifier to each provider. In the future, the system will integrate non-HCFA subscribers. It is planned that enumeration of Medicare providers will begin in calendar year 1996. The draft systems requirement definition was issued in January, 1995. It is recommended that the NPF be the source of all unique provider identifiers, for institutions and individuals. Systems may also choose to collect other identifiers (e.g., tax number), which they can link to the NPI. Items shown below with an asterisk (*) indicate that this type of information can be obtained from linking the NPI with the National Provider File and may not need separate collection. The Committee recognizes that all practitioners may not be included initially in this system, but ultimately all should be included.

17. Facility Identification - The unique HCFA identifier as described above. This identifier includes hospitals, ambulatory surgery centers, nursing homes, hospices, etc. If the HCFA system does not have separate identification numbers for parts of a hospital (i.e., Emergency Department, Outpatient Department), an additional element (such as element 13) will need to be collected along with the facility ID to differentiate these settings. The Committee recommends that the HCFA identifier be adopted when completed.

18. Type of Facility/Place of Encounter - As part of the NPI/NPF system, described above, HCFA is defining a taxonomy for type of facility. This taxonomy builds on previous NCVHS and departmental work and should be reviewed by the NCVHS and standards organizations. The Committee encourages the development of one taxonomy and will monitor progress.

19. Health Care Practitioner Identification (outpatient) - The unique national identification number assigned to the health care practitioner of record for each encounter. There may be more than one health care provider identified:

- A. The health care practitioner professionally responsible for the services, including ambulatory procedures, delivered to the patient (health care practitioner of record)
- B. The health care practitioner for each clinical service received by the patient, including ambulatory procedures

Initial enumeration by HCFA will focus on individual providers covered by Medicare and Medicaid; however, the system will enable enumeration of other health care practitioners, as identified by system users. The Committee recommends that the HCFA identifier be adopted when completed.

20. Location or Address of Encounter (outpatient) - The full address and Zip Code (nine digits preferred) for the location at which care was received from the health care practitioner of record (see 19A.). As recommended by the UACDS, address should be in sufficient detail (street name and number, city or town, county, State, and Zip Code) to allow for the computation of county and metropolitan statistical area.

21. Attending Physician Identification (inpatient) - The unique national identification number assigned to the clinician of record at discharge who is responsible for the discharge summary, as recommended by the 1992 UHDDS.

22. Operating Clinician Identification - The unique national identification number assigned to the clinician who performed the principal procedure, as recommended by the UHDDS.

23. Health Care Practitioner Specialty - As part of the NPI/NPF system, HCFA has identified a very detailed list of specialties for health care practitioners. This listing should be reviewed by the NCVHS and standards organizations and, if found acceptable, recommended for use.

24. Principal Diagnosis (inpatient) - As recommended by the UHDDS, the condition established after study to be chiefly responsible for occasioning the admission of the patient to the hospital or nursing home for care. The currently recommended coding instrument is the ICD-9-CM.

Rationale and discussion

Principal diagnosis is required by most systems for inpatient reporting. The Committee acknowledges that there are differences in coding guidelines for reporting diagnosis in inpatient and outpatient settings, and this may result in a lack of comparability in data between the two settings. It is recommended that convergence of these guidelines be investigated.

25. Primary Diagnosis (inpatient) - The diagnosis that is responsible for the majority of the care given to the patient or resources used in the care of the patient. The currently recommended coding instrument is the ICD-9-CM.

Rationale and discussion

The primary diagnosis is not part of the UHDDS, and in most diagnostic situations, the principal and primary diagnoses will be identical. Respondents have indicated a mixed use of this item for inpatients. There is also concern that medical personnel may be confusing the definitions/uses of

principal versus primary diagnosis. Some respondents incorrectly interpreted this item as a means of classifying primary site for cancer, utilizing ICD-O (oncology). The NCVHS notes that the Department of Veterans Affairs routinely collects this element, and thus approves the continued inclusion in this core list, pending a review of uses and users of this element.

26. Other Diagnoses (inpatient) - As recommended by the UHDDS, all conditions that coexist at the time of admission, or develop subsequently, which affect the treatment received and/or the length of stay. Diagnoses that refer to an earlier episode that have no bearing on the current hospital or nursing home stay are to be excluded. Conditions should be coded that affect patient care in terms of requiring clinical evaluation; therapeutic treatment; diagnostic procedures; extended length of hospital or nursing home stay; or increased nursing care and/or monitoring. The currently recommended coding instrument is the ICD-9-CM.

27. Qualifier for Other Diagnoses (inpatient) - The following qualifier should be applied to each diagnosis coded under "other diagnoses," as was recommended in the 1992 revision of the UHDDS:

Onset prior to admission

- 1.Yes
- 2.No

Rationale and discussion

This element is currently being collected by California and New York hospital discharge data systems; there is an indication that use of this qualifier can contribute significantly to quality assurance monitoring, risk-adjusted outcome studies, and reimbursement strategies.

Ambulatory Conditions

The elements for ambulatory conditions contain information on the Patient's Stated Reason for Visit and the Problems, Diagnosis, or Assessment, both of which were recommended by the UACDS. The latter element, which describes all conditions requiring evaluation and/or treatment or management at the time of the encounter as designated by the health care practitioner, has been divided into two elements: 1) the diagnosis chiefly responsible for services provided, and 2) other diagnoses.

28. Patient's Stated Reason for Visit or Chief Complaint (outpatient) - Includes the patient's stated reason at the time of the encounter for seeking attention or care. This item attempts to define what actually motivated the patient to seek care and has utility for analyzing the demand for health care services, evaluating quality of care and performing risk adjustment. The NCVHS recommended this as an optional item in the UACDS but that high priority should be given to conducting additional study as to the feasibility, ease and practical utility of collecting the patient's reason for encounter, in as close to the patient's words as possible. There is not one agreed-upon coding system for this item; the International Classification of Primary Care, and the Reason For Visit Classification used by the National Ambulatory Medical Care Survey are two such systems. Additional evaluation and testing are warranted for this important information.

29. Diagnosis Chiefly Responsible for Services Provided (outpatient) - The diagnosis, condition, problem, or the reason for encounter/visit chiefly responsible for the services provided. Condition should be recorded to the highest documented level of specificity, such as symptoms, signs, abnormal test results, or other reason for visit, if a definitive diagnosis has not been established at the end of the visit/encounter. The currently recommended coding instrument is the ICD-9-CM.

Rationale and discussion

Information on all patient problems and diagnoses requiring attention at the encounter are needed to assess the quality of care delivered, to determine what types of health problems are being seen and treated in the different types of ambulatory care facilities, and for assessing the appropriateness of the setting used to perform the services. During the NCVHS review of core health data elements, discussion arose regarding the specificity of diagnoses reported. The official national outpatient/physician coding and reporting guidelines provide instruction that a suspected or rule out condition not be reported as though it is a confirmed diagnosis.

The instruction clarifies that only what is known to the highest level of specificity should be reported. In some instances this may be a symptom or an abnormal finding. Medicare and many other payers adhere to these guidelines. Some third party payers, however, have ignored the guidelines and required facilities and health care practitioners to report a diagnosis that justifies the performance of services being provided. This has resulted in inconsistent data found in many outpatient databases and has skewed patient outcome studies. It is anticipated that the introduction of ICD-10 will alleviate this problem. The NCVHS recommends continued monitoring of provider practices with regard to coding and revision of these recommendations if current guidelines continue to be ignored.

30. Other Diagnoses (outpatient) - The additional code(s) that describes any coexisting conditions (chronic conditions or all documented conditions that coexist at the time of the encounter/visit, and require or affect patient management). Condition(s) should be recorded to the highest documented level of specificity. The ICD-9-CM is the recommended coding convention.

Rationale and discussion

Information on multiple diagnoses is important for developing severity indexes and assessing resource requirements and use.

31. External Cause of Injury - This item should be completed whenever there is a diagnosis of an injury, poisoning, or adverse effect. The currently recommended coding instrument is the ICD- 9-CM. The priorities for recording an External Cause-of-Injury code (E-code) are:

1. Principal diagnosis of an injury or poisoning
2. Other diagnosis of an injury, poisoning, or adverse effect directly related to the principal diagnosis.
3. Other diagnosis with an external cause.

Rationale and discussion

The collection of this element has been recommended by the UHDDS and the UACDS, and a separate element for its collection is included on the UB 92. The information that this element provides on the causes of patients' injuries or adverse effects is considered essential for the development of intervention, prevention and control strategies. Compelling evidence presented by the Indian Health Service, states and nonprofit organizations demonstrates that effective intervention strategies can be implemented in response to available data on external causes of injury.

Procedures

All significant procedures, and dates performed, are to be reported. A significant procedure is one that is:

1. Surgical in nature, or
2. Carries a procedural risk, or
3. Carries an anesthetic risk, or
4. Requires specialized training.

Surgery includes incision, excision, amputation, introduction, endoscopy, repair, destruction, suture, and manipulation. A qualifier element is recommended to indicate the type of coding structure used, i.e., ICD, CPT, etc.

32. Birth Weight of Newborn (inpatient) - The specific birth weight of the newborn, recorded in grams.

Rationale and discussion

Birth weight of newborn is readily available in the medical record and has singular importance for risk-adjustment outcome studies and health policy development related to maternal and infant health.

33. Principal Procedure (inpatient)- As recommended by the UHDDS, the principal procedure is one that was performed for definitive treatment, rather than one performed for diagnostic or exploratory purposes, or was necessary to take care of a complication. If there appear to be two procedures that are principal, then the one most related to the principal diagnosis should be selected as the principal procedure. ICD-9-CM Vol. 3 is required; however NCVHS strongly advocates a single procedure classification for inpatient and ambulatory care.

34. Other Procedures (inpatient) - All other procedures that meet the criteria described in element 33.

35. Dates of Procedures (inpatient) - Year, month, and day, as recommended in the UHDDS and by ANSI ASC X12, of each significant procedure.

36. Procedures and Services (outpatient) - As recommended by the UACDS, describe all diagnostic procedures and services of any type including history, physical

examination, laboratory, x-ray or radiograph, and others that are performed pertinent to the patient's reasons for the encounter; all therapeutic services performed at the time of the encounter; and all preventive services and procedures performed at the time of the encounter. Also, describe, to the extent possible, the provision of drugs and biologicals, supplies, appliances and equipment. The HCFA Common Procedure Coding System (HCPCS), based on CPT-4, is required for physician (ambulatory and inpatient), hospital outpatient department, and free-standing ambulatory surgical facility bills; however, NCVHS strongly advocates a single procedure classification for inpatient and ambulatory care. The Committee recognizes the importance and desirability of linking services with diagnoses, wherever feasible.

37. Medications Prescribed - Describe all medications prescribed or provided by the health care practitioner at the encounter (for outpatients) or given on discharge to the patient (for inpatients), including, where possible, National Drug Code, dosage, strength, and total amount prescribed.

Rationale and discussion

The collection of information on medications is crucial to understanding the health care encounter and the services provided to a patient. The Committee recognizes that not all providers are obtaining this detail, but it is anticipated that these data will be more frequently collected in the near future with the growth of computerized prescription information.

38. Disposition of Patient (inpatient) - As recommended by the UB 92 and as an expansion of the 1992-93 UHDDS data element:

1. Discharged Alive
 - A. Discharged to home or self care (routine discharge)
 - B. Discharged/transferred to another short term general hospital for inpatient care
 - C. Discharged/transferred to skilled nursing facility (SNF)
 - D. Discharged/transferred to an intermediate care facility (ICF)
 - E. Discharged/transferred to another type of institution for inpatient care or referred for outpatient services to another institution
 - F. Discharged/transferred to home under care of organized home health service organization
 - G. Discharged/transferred to home under care of a Home IV provider
 - H. Left against medical advice or discontinued care
2. Expired
3. Status not stated

Rationale and discussion

In addition to documenting whether the patient was discharged alive or died during the hospitalization, the patient disposition is an indicator of the patient's health status at the time of discharge and need for additional services.

39. Disposition (outpatient) - The health care practitioner's statement of the next step(s) in the care of the patient. Multiple responses are possible. At a minimum, the following classification is suggested:

1. No follow-up planned (return if needed, PRN)
2. Follow-up planned or scheduled

3. Referred elsewhere (including to hospital)
4. Expired

Rationale and discussion

The critical distinction here is whether follow-up is planned or scheduled, as an indicator of continuing health problems and continuity of care. Expired has been added because the outpatient setting includes a wide range of sites, including Emergency Departments and ambulatory surgery centers.

40. Patient's Expected Sources of Payment - The following categories are recommended for primary and secondary sources of payment:

40A. Primary Source - The primary source that is expected to be responsible for the largest percentage of the patient's current bill.

40B. Secondary Source - The secondary source, if any, that will be responsible for the next largest percentage of the patient's current bill.

- Self-pay
- Worker's Compensation
- Medicare
- Medicaid
- Maternal and Child Health
- Other government payments
- Blue Cross
- Insurance companies
- No charge (free, charity, special research, or teaching)
- Other
- Unknown/not stated

Rationale and discussion

The categories in this element were recommended by the UHDDS for primary and secondary sources of payment. The Committee recognizes the ongoing discussion of discrepancies between 'expected' and 'actual' sources of payment. Source of payment categories, as recommended in the past, are no longer sufficient. The continuing expansion of types of payments and the combination of payments within groups is ever changing. However, the information is still considered useful to collect for trend purposes and for some indication of patients' coverage by third-party payers.

HCFA is developing a new system, called the HCFA PAYERID project, which will assign a unique identifier to every payer of health care claims in the United States. Participation is voluntary, and HCFA, which is funding its development, has been working to get consensus about the kind of system that would be useful. The database will contain payer names, billing addresses and business information. The information, which is already in the public domain, will be accessible by names and ID numbers, and available in several formats. Who will have access to the database for research purposes, and to what data, has yet to be determined. "Payers" are defined as public and private entities that have contract responsibility for health care payment.

Medicare decided a PAYERID was needed because of the difficulty its contractors were having in transferring claims to other insurance companies, due to incomplete information or multiple names for payers. It is hoped that the system will improve the coordination of benefits, as well as providing access to information about health insurance and making it easier to track third party liability

situations. HCFA, however, has estimated that there are approximately 30,000 individual payers in the U.S. They currently are not developing a system of categories to accompany the IDs. Such a system would be helpful to the extent that it is feasible in the current highly dynamic market.

Because the PAYERID system is still being developed, and because HCFA currently has no plans to categorize payers, the Committee recommends the current UHDDS categories while encouraging continued study and evaluation of categories used by other data collectors.

41. Injury Related to Employment

Rationale and discussion

Whether an injury is work related or not can be of significant importance both in the area of injury prevention and in medical care payment. During the discussion on including External Cause of Injury in the 1994 revision to the UACDS, CDC and labor and business groups urged collection of whether or not an injury occurred at work or was work-related. This element is currently collected on the HCFA 1500 form.

42. Total Billed Charges - All charges for procedures and services rendered to the patient during a hospitalization or encounter.

Rationale and discussion

The UHDDS and UACDS have recommended the collection of all charges for procedures and services rendered to the patient during a hospitalization or encounter. This item already is collected by most state health data organizations collecting hospital discharge information and offers the only readily available information on the fiscal dimensions of care and the relative costs of different types of care. Although there is agreement that "payments" or "costs" are needed, most participants agreed that it is virtually impossible to collect these items consistently across time and locations. Moreover, in the electronic format, in most instances, payments would not be available at the time that patient and medical data are entered. It might not be feasible to expect the record to be updated to include payment data when it becomes available. Therefore, billed charges should be collected, at a minimum.

4.5.2. Core Data Elements Listed by Readiness for Implementation

Elements Ready for Implementation:

- 2. Date of Birth**
- 3. Gender**
- 4. Race and Ethnicity**
- 5. Residence**
- 6. Marital Status**
- 10. Years of Schooling**
- 11. Patient's Relationship to Subscriber/Person Eligible for Entitlement**
- 14. Admission Date (inpatient)**
- 15. Discharge Date (inpatient)**
- 16. Date of Encounter (outpatient and physician services)**
- 20. Location or Address of Encounter (outpatient)**

- 24. Principal Diagnosis (inpatient)
- 25. Primary Diagnosis (inpatient)
- 26. Other Diagnoses (inpatient)
- 27. Qualifier for Other Diagnoses (inpatient)
- 29. Diagnosis Chiefly Responsible for Services Provided (outpatient)
- 30. Other Diagnoses (outpatient)
- 31. External Cause of Injury
- 32. Birth Weight of Newborn
- 33. Principal Procedure (inpatient)
- 34. Other Procedures (inpatient)
- 35. Dates of Procedures (inpatient)
- 36. Procedures and Services (outpatient)
- 37. Medications Prescribed
- 39. Disposition (outpatient)
- 41. Injury Related to Employment

Elements Substantially Ready for Implementation, but need some added work:

- 7. Living/Residential Arrangement
- 17. Facility Identification
- 18. Type of Facility/Place of Encounter
- 19. Health Care Practitioner Identification (outpatient)
- 21. Attending Physician Identification (inpatient)
- 22. Operating Clinician Identification (inpatient)
- 23. Health Care Practitioner Specialty
- 38. Disposition of Patient (inpatient)
- 40. Patient's Expected Sources of Payment
- 42. Total Billed Charges

Elements which require a substantial amount of study and evaluation:

- 1. Personal/Unique Identifier
- 8. Self-Reported Health Status
- 9. Functional Status
- 12. Current or Most Recent Occupation and Industry
- 13. Type of Encounter
- 28. Patient's Stated Reason for Visit or Chief Complaint (outpatient)

4.5.3. Additional data items

While reviewing the draft list of data elements, respondents indicated a number of additional data elements that they felt were important core elements. Some of these included information on health behaviors, such as smoking and alcohol consumption; information on preventive services; language ability; severity of illness indicators; provider certainty of diagnostic information; information to link a mother's and infant's charts; information on readmissions and complications, to mention a few. Future projects may undertake to seek consensus among some of these items.

4.5.4. Related Data Set Activities

Concurrent with these activities being undertaken by the full Committee, there are two related projects undertaken by the Subcommittee on Mental Health Statistics and the Subcommittee on Disability and Long Term Care Statistics. With the assistance of the Center for Mental Health Services, SAMHSA, and a contractor, Webman Associates, a study was undertaken to identify and survey a representative sample of mental health, managed care, substance abuse, disabilities and long term care experts who would be willing to offer recommendations about the content of an ideal minimal data set for a health care record that is inclusive of the relevant information. Over three dozen data sets were studied, among them two nationally approved data sets, the Mental Health Statistics Improvement Program Data Set (MHSIP) and The Adoption and Foster Care Analysis and Reporting System (AFCARS) data set. After review of the data elements collected, the subcommittees decided to study in-depth six data clusters:

1. Disability
2. Mental Health and Substance Use History of Consumer and of Consumer's Family Members
3. Guardianship/Caregiver
4. Living Situation
5. Categorization and Coding of Wrap Around Services (including community-based services, housing assistance, job training, etc.)
6. Functional Assessment Criteria

The preliminary results of this project have been prepared. The process for these specialized areas is ongoing and final recommendations for specific elements have not yet been submitted to the full Committee. Respondents to this project welcomed the notion of a core data set and standardized forms in this area. It is important to note for this report, however, that the two subcommittees are in agreement with the core data elements that are described herein. Their continuing study is involved with more detailed data elements that relate specifically to the areas of mental health, substance abuse, and long term care.

A second study is currently underway, one which will investigate core data elements in common use in data sets on persons with disability and/or persons receiving long term care. The major objectives of this project include the production of a report assessing existing data for care provided to persons with disabilities in institutional and community long term care settings, as well as in rehabilitation. Common data elements and areas for standardization will be considered as well as criteria for selection of data elements. Recommendations and linkage with the current project will be discussed.

5. Conclusions

As a result of the process followed in the conduct of this project and based on careful analysis by its members, the Committee has reached the following conclusions:

- The response to the Committee's activities both through participation in meetings and written comments indicates that the health care information field is solidly in favor of the identification and use of standardized data elements and definitions.
- The number of standards-setting organizations is growing; however, all who addressed the Committee are actively seeking participation by a 'recognized' leader/group who can forge consensus for the health care information field. But time is short; decisions are being made by organizations now.

- Response was significant and positive to the Committee's request to review a set of core data elements that were identified after a series of hearings and other information-gathering efforts were completed. Most organizations were supportive in wanting to 'get on board' with standardized data elements.
- There is already consensus among data collectors and users for a significant number of data elements, especially elements related to person descriptors and to selected information on inpatient and ambulatory encounters.
- There is less agreement on data definitions, even for data items that have been in the field for years. Definitions must be refined and made available in standardized formats to data collectors.
- There are data items, such as health status and functional status, that are considered crucial elements, but for which substantial additional study and evaluation must be undertaken to reach consensus on standardized content and definition.
- Because they recognize the significance of this project, respondents also recommended a number of additional items that they would like evaluated and possibly included in a core set of standardized data elements.

6. Recommendations to the Department

1. The Committee supports the HHS Data Council in its formation of the Health Data Standards Committee to focus attention on the needs for standardized data both within the Department and in the health care community at large and to foster collaboration and consensus with the major standards-setting organizations. To this end, the Committee recommends that the Data Council:

- Circulate the report within the Department for review and constructive criticism.
- Investigate the formation of leadership sites within the Department for each of the standards-setting organizations.
- Refer the core health data elements recommendations to the National Uniform Claim Committee for their consideration as they study the issue of uniform data elements for paper and electronic collection in Fall 1996.
- Provide stable resources to the project to establish an interdepartmental work group, with DHHS taking the lead, to work with the key standards-setting organizations in the area of core health data elements.

2. The Committee recommends the following actions specifically related to the core data elements:

- For those elements that the Committee recommends as being ready to standardize, request each of the data collection entities within the Department to review the set of data elements and to match data contents and definitions with similar items that they are currently collecting or plan to collect. Report to the HHS Data Council on the viability of these elements and definitions being adopted in their program. If a reporting entity is using a different element or definition, explain why their current usage is preferable.
- Support implementation and testing activities for those data elements for which agreement on definitions has been reached and those for which minimal additional work is needed on

definitional agreement. Public and private participants have indicated a willingness to work together to disseminate information, test data elements, and utilize electronic means to ensure the widest dissemination of these activities.

- For those data elements which have been recognized as significant core elements, but for which there is not consensus on definition, support the formation of a public-private working group to conduct or coordinate additional study or research and to further refine definitions. This group, or a separate group, could also be the focus for evaluating additions to the list of core data elements and for setting up methods for testing and promulgating the final products.
- Place the Committee's report, elements and definitions on an appropriate departmental Home Page as guidance to the field and as a means of encouraging use and soliciting further comments and suggestions while the report is under review within the Department.

3. Because agreement on a unique personal identifier is recognized as a key element to the successful establishment of core data elements, and their use, support the formation of a public-private working group to study and provide recommendations in this area.

4. Support the NCVHS continuing its work in this area, especially using its expertise to discuss research issues, to assist in consensus building, and to participate with the Data Council in the implementation of the core data element project recommendations.

7. Future Activities

As highlighted earlier, the Committee has identified a number of areas that should be considered for implementation by the HHS Data Council. These include the review and implementation of core data elements and definitions within departmental data collection activities; formation of public-private work groups to assist in promulgating data elements for which consensus has been reached or for undertaking additional study on critical elements for which there are no standardized definitions. Additionally, a consensus must be reached on the unique personal identifier.

Participants in the various meetings had discussed ways to disseminate new data items, seek input, and inform data collectors of recommended elements and definitions. It was felt that the Committee should consider designing a WEB page on the Internet that could be used for these activities. The Committee could recommend such an activity, but it would require departmental staff to actually design, input data, and monitor and update the site.

Several organizations have volunteered to facilitate dissemination and feedback of the core data elements project. These activities could take several forms. One would be through the use of a state-level or regional-level organization that already has a line of communication with other organizations. An example of this could be NAHDO which could undertake to work with its members. Another form would be through an organization that already has a WEB page; several organizations indicated that they would be willing to test the sharing of this information through their Internet sites.

It is of vital importance to participate in and/or be members of the numerous data standards groups. Currently there is little or no input from the public health field for several reasons. One major reason is the staff and dollar resources required to travel to and participate in several meetings per group per year. Another problem is that, although the HHS Data Council has recently established a Health Data Standards Committee, until the past few months, there has been no central location within the Department for monitoring the activities of the data standards groups. Throughout the meetings it

became apparent that many standards-setting groups are moving ahead without broader input, for example, from those in the public health and epidemiology fields. Place holders will be set, and, in some cases data items and definitions decided on, before national and local public health agencies and organizations will be able to act.

It became obvious early in the meetings that the identification of core data elements, their definitions, and the consensus-building needed to encourage use of these items would be an ongoing and full-time activity for several years. Although the Committee serves a very useful purpose in bringing together the experts to discuss and consider these elements, it takes dedicated departmental staff to keep the process underway on a day-to-day basis. The Committee recommends that the Department fund these activities on an ongoing basis.