MEDINFO 2001
V. Patel et al. (Eds)
Amsterdam: IOS Press

Structured Data Entry of Clinical Information for Documentation and Data Collection

N.T. Cheung[a], Vicky Fung[a], Y.Y. Chow [b], Y. Tung[b]

[a]*Hospital Authority Head Office, Hong Kong*
[b]*Tuen Mun Hospital, Hong Kong*

Abstract

Routine databases containing large amounts of clinical data represent a tremendous opportunity for the evaluation of health care practices and outcomes. However, data collected for administrative purposes has limitations in content, accuracy and completeness. Routine entry of clinical information directly into clinical information systems by care providers is one strategy to address this problem. We developed a structured data entry method, the Clinical Data Framework (CDF), which has been used to support the capture of clinical information by clinicians in the normal process of care delivery. A study of the CDF over a two month period showed that it improved the accuracy of completeness of data collection over a coding method which was based on selection of ICD-9-CM codes.

Keywords:

Electronic patient record, data quality, clinical data, structured data entry, coding, clinician/physician data entry

Introduction

One of the promises of the electronic patient record has been that routine collection of detailed clinical information gives researchers a powerful tool for the evaluation of health care practices and outcomes [1,2]. A lot of work has already been done using databases which are based on administrative data, but the accuracy and completeness of such databases is questionable [3] especially with regards to complications and comorbidities [4,5,6]. They also often are lacking in the severity or functional outcomes data necessary for research and evaluation [3].

The key to achieving routine capture of high quality clinical data is for the data to be entered directly by clinicians into the electronic patient record as part of their normal workflow. The Institute of Medicine's report on the computer-based patient record highlights the need for direct entry by health care professionals [7]. There are two main approaches to the direct entry of detailed clinical information. One is to allow free-text input, which gives the clinicians flexibility and expressiveness, and to extract medical concepts from the text using natural language processing. The other is to use dynamic structured entry forms, which are more restrictive in their input, but which ensure that the information is captured in a computable form [8].

The principle of direct clinician entry has been at the core of the development of our Clinical Management System, a clinician workstation architecture. This paper describes our earlier forays into clinician entry, a new method of entry of clinical information (the Clinical Data Framework) and the results of a study into the impact of the Clinical Data Framework on the collection of clinical information.

Background

The Clinical Management System (CMS) project began in 1994. CMS provided an integrated clinical workstation, giving single logon access to all the available clinical information from either the ward or clinic setting. The original functionality included discharge summary generation with ICD-9-CM coding, discharge and outpatient prescription ordering with electronic transfer to pharmacy, clinical notes and appointment booking. Today there are some 2,000 CMS workstations being used by around 19,000 staff to enter or retrieve clinical information.

One of the key features of the CMS has been the direct capture of ICD-9-CM codes by clinicians during the preparation of the discharge summary. This feature has reduced transcription and increased coding accuracy without undue disturbance to the normal workflow of clinicians [9]. Although we chose to use ICD-9-CM as it is the standard in research and analysis applications, we realized its use in a clinical context was going to be difficult. ICD-9-CM was not designed for clinical documentation and has severe limitations in this arena [10,11]. A study on the content coverage of ICD-9-CM and other major coding systems found that ICD-9-CM fails to capture substantial clinical content [12]. Other problems with ICD-9-CM are that clinicians often have difficulty with its precise, rather stilted language, and clinicians are not familiar with the coding rules which accompany ICD-9-CM.

Hospital Authority Master Disease Code Table

To address some of these issues, we extended ICD-9-CM to incorporate locally preferred terminology and to allow greater clinical detail and precision (see table 1). Our extended ICD-9-CM table is called the Hospital Authority Master Disease Code Table (HAMDCT). To complete the discharge summary in the CMS, clinicians simply select the required diagnosis or procedure from the HAMDCT and the respective ICD-9-CM code is automatically generated and the term is included in the documentation. These codes then flow through to our medical record abstracting system, ready for analysis and reporting. The HAMDCT is used in all clinical information systems in the Hong Kong public hospitals, ensuring consistency in reporting clinical data.

Table 1. The HAMDCT extending ICD-9-CM

Code	Extend	Description
188.9	0	Malignant neoplasm of bladder
	1	Carcinoma urinary bladder
	3	Sarcoma of bladder
	4	Transitional cell carcinoma of bladder
567.8	0	Other specified peritonitis
	1	Bile peritonitis
	2	Faecal peritonitis
589.0	0	Unilateral small kidney
	1	Small kidney, left
	2	Small kidney, right
45.13	0	Other endoscopy of small intestine
	1	Oesophagogastroduodenoscopy

Unfortunately, the HAMDCT ran into its own problems. The need for detailed terms is never-ending, and as of July 2000 we had already added over 11,000 new terms to the HAMDCT on top of the original 12,655 terms of ICD-9-CM. Searching for terms became difficult, as a keyword search would often return many similar terms, which would need careful reading before the correct term could be selected. Continued reliance on ICD-9-CM code level analysis meant that the additional clinical information of the new terms was not analyzable.

Clinical Data Framework

The Clinical Data Framework (CDF) is a method of dynamic structured entry which assists clinicians in the systematic documentation and capture of clinical data. The emphasis of the CDF was on clinical documentation rather than the selection of codes. The CDF is integrated with the HAMDCT, gradually taking over some of the functions of the HAMDCT, but for the clinical users the transition is a gradual one.

An axial approach

Under the CDF, data are organized in different "axes". Each disease has a set of axes which allow documentation of different aspects of the disease, including aspects which could not be captured using the HAMDCT (e.g. laterality, staging, severity). Clinicians have the flexibility to document to the appropriate level of detail they wish to report – few axes are mandatory, and many come with default values. The CDF will automatically generate as many ICD-9-CM codes as are necessary to reflect the axis choices. For instance, metastasis codes will be generated in addition to the primary tumour code. Coding rules are incorporated to generate accurate ICD-9-CM codes.

As with codes entered using HAMDCT, a free text comment field was provided to allow entry of extra information.

Figures 1 and 2 show some examples of CDF screens.

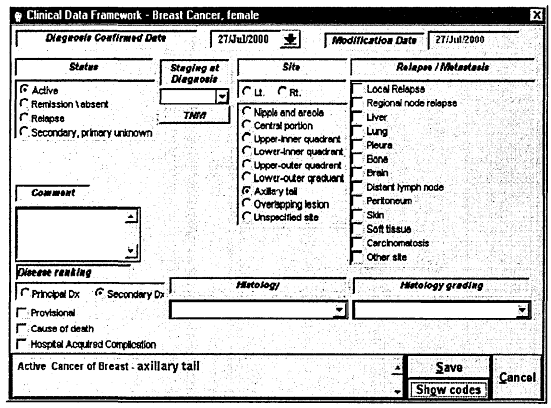

Figure 1. Clinical data framework for breast cancer

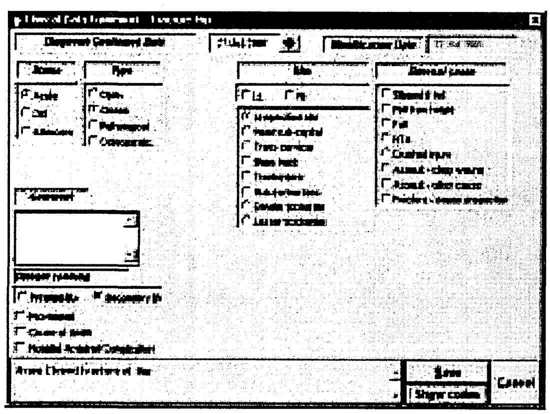

Figure 2. Clinical data framework for hip fracture

There are many benefits to the CDF approach. The CDF follows clinical logic in the collection of disease related information, obviating the need to choose multiple different codes based on the logic of the ICD-9-CM. By allowing capture of information beyond that possible with even extended ICD-9-CM codes, the CDF aims to greatly increase the expressiveness of the electronic record while maintaining a structure amenable to analysis. The CDF ensures a consistent data set for each disease. There is an implicit linkage between the different axes for a CDF, so analysis can use this relationship. The CDF also supports the concept of disease progress. As the disease status changes, clinicians can change the axis values to document the progress of the disease. Unlike codes, which are episode based, CDF is patient-based, spanning episodes of care.

Methods

The CDF was developed for 18 different fractures and cancers (see table 2).

Table 2. The CDF diseases

Cancers	Fractures
Nasopharyngeal cancer	Spine
Female breast	Humerus
Cervix uteri	Forearm
Ovary	Wrist
Lung	Hand
Colon	Hip
Rectum	Femur
	Patella
	Tibia & fibula
	Ankle
	Foot

Tuen Mun Hospital (TMH), a large acute hospital with over 1600 beds, was the site for the two month study. TMH had already implemented the CMS with HAMDCT from 1998, so the clinicians were experienced with discharge summary entry and coding using HAMDCT. All episodes for the CDF diseases were documented on discharge using the CDF during the study period. Apart from an initial briefing to a few senior clinicians and a one page handout, no training was given.

The data from this two month period (Feb/Mar, 2000) was extracted for analysis of entry patterns, and to study the usefulness of the CDF in capturing additional clinical information and in documenting the progress of a disease. The codes generated during this period were compared with those entered using the HAMDCT for an earlier two month period (Sept/Oct, 1999) in the same institution. The free text comment fields and the full text discharge summaries for all these cases were also extracted from the system, and were used as the reference standard to compare accuracy of coding before and after the use of the CDF. Cancers and fractures were examined separately.

Results

Axis analysis

For cancer 398 patients had entries in 493 episodes; 22% of cases had progress updates. Site (60% of patients) and histology (61% of patients) were most commonly reported. 39% of the patient had the cancer status (remission, relapse, etc) captured as a non-default value, and 40% of the patient's cancer staging were reported. Only 32% of patients had histology grading reported. Since laterality is a mandatory item, all cancers that had a laterality axis had this captured in the CDF. Of the 79 breast cancer patients 24 (30%) had additional prognostic factors recorded in the comment field.

Table 3. CDF statistics

	Cancers	Fractures
No. patients	398	293
No. episodes	493	296
Progress entries	88	6

Table 4. Cancer axis non-default entries

Status	156 (39%)
Site	206 (60%)
Staging	159 (40%)
Histology	243 (61%)
Histology grading	126 (32%)
Laterality	197 (100%)

For fractures 293 patients had entries in 296 episodes; 2% had progress updates. The site was reported in 93% of cases. Fracture status and type were mostly reported to the default value: 'acute' (91%) and 'closed' (97%) respectively. Similar to cancer, all fractures with a laterality axis had the laterality captured in the CDF as it was a mandatory item. External cause is a mandatory item for 'acute open' and 'acute closed' fractures. 92% of the patients' data reported to a non-default value for the external cause.

Table 5. Fracture axis non-default entries

Site	272 (93%)
Status	26 (9%)
Type	9 (3%)
External cause – other causes	270 (92%)
External cause – unspecified	25 (9%)
Laterality	282 (100%)

Coding quality - cancer

Due to lack of information in the comment field and the discharge note, the accuracy of the cancer code could not be determined in 61 and 29 cases for the HAMDCT and CDF periods respectively. 90% of the remaining HAMDCT codes were accurate, and 88% of the CDF codes were accurate. The difference was not significant (table 6).

Table 6. Cancer coding accuracy

	HAMDCT	CDF	sig.
Total codes	481	293	
Accuracy determined	420	264	
Accurate codes	378 (90%)	233 (88%)	p=0.47

Using the HAMDCT, 44% of the patients with metastases had a metastasis code entered, vs. 89% for CDF. If we included multiple metastases, then HAMDCT captured 43% of all metastases vs. 92% for CDF. Both were significantly different (table 7).

Table 7. Coding of metastases

	HAMDCT	CDF	sig.
Patients with metastases	127	84	
Codes entered	56 (44%)	75 (89%)	<0.001
Total metastases	146	104	
Codes entered	63 (43%)	92 (92%)	<0.001

Coding quality - fractures

Due to lack of information in the comment field and the discharge note, the accuracy of the cancer code could not be determined in 31 and 53 cases for the HAMDCT and CDF periods respectively. 83% of the remaining HAMDCT codes were accurate, and 90% of the CDF codes were accurate, a significant difference. 2% of HAMDCT fractures were coded to a non-specific site vs. 6% of CDF fractures, which was also significant. As external cause is a mandatory axis, 100% of external causes were captured with CDF, vs. 41% for HAMDCT (table 8).

Analysis

It can be seen that the impact of the CDF depends on the diseases being documented. For cancers, the quality of coding is not affected, but there is a lot of extra clinical information being captured through the CDF. In particular, the capture of metastases is much better with the CDF.

The quality of fracture coding is improved by CDF, but there seemed to be more non-specific codes entered. On closer analysis of the 15 non-specific codes, it was discovered that 9 of these codes were for sites which were codable using HAMDCT, but which had not been included in the CDF. This discovery allows us to improve the CDF by including additional sites. The capture of external causes was much better with CDF, and laterality was captured in addition, but there was less need to capture status and type, probably because the major of fractures at the hospital are acute, closed fractures, which are our default values.

Table 8. Fracture coding accuracy

	HAMDCT	CDF	sig.
Total codes	542	303	
Accuracy determined	511	250	
Accurate codes	422 (83%)	226 (90%)	p<0.01
Non-specific site	10 (2%)	15 (6%)	p<0.01
Missing external cause	212 (41%)	0 (0%)	p<0.001

Discussion

One of the guiding principles of the development of clinical information systems at the Hospital Authority is that where possible care providers should enter data directly into the system as part of their normal workflow, and this data will then flow through the system to be used as needed. There should be no need to reenter any information for the purposes of analysis or research.

We have chosen to use structured data entry to support the entry of good quality data by clinicians. Structured data entry in general is better developed than automated extraction using natural language processing, with other groups adopting this approach [13,14,15,16]. Our approach has the benefit of integrating completely into an ICD-9-CM based system (with the flexibility to move to ICD10 or any other coding system if required), using dynamically created screens (so new diseases can be easily added), and an extremely easy to user interface. There was almost no training and support for the described study, and we are in the process of implementing the CDF in all of our major hospitals – some 2000 workstations – again with minimal training. The free text comment field allows flexibility for further information entry, and also provides valuable feedback for further refinement of the system.

Conclusion

We have found that the structured data entry of the CDF improved data accuracy and completeness. On the basis of these results, we are extending the CDF to all major hospitals. We are also developing the CDF for other important diseases, beginning with diabetes mellitus, myocardial infarction, cerebrovascular accident and peptic ulcer.

Acknowledgements

We would like to thank the members of the Disease Classification Committee and the Hospital Information Records Managers who helped in the development of the Clinical Data Framework, and the Information Technology Department for their work in the development of the CDF and the subsequent data extraction.

References

[1] Selby JV. Linking Automated Databases for Research in Managed Care Settings, *Annals of Internal Medicine*, 1997: 127(8):719-724

[2] Dolin RH. Outcome analysis: considerations for an electronic health record, *MD Computing*, 1997: 14(1):50-6

[3] Iezzoni LI. Assessing quality using administrative data, *Annals of Internal Medicine*, 1997: 127(8 Pt 2):666-74

[4] Newton KM, Wagner EH, Ramsey SD, McCulloch D, Evans R, Sandhu N, Davis C (1999). The use of automated data to identify complications and comorbidities of diabetes: a validation study, *Journal of Clinical Epidemiology*, 52(3):199-207

[5] McCarthy EP, Iezzoni LI, Davis RB, Palmer RH, Cahalane M, Hamel MB, Mukamal K, Phillips RS, Davies DT Jr . Does clinical evidence support ICD-9-CM diagnosis coding of complications? *Medical Care*, 2000: 38(8):868-76

[6] Levy AR, Tamblyn RM, Fitchett D, McLeod PJ, Hanley JA. Coding accuracy of hospital discharge data for elderly survivors of myocardial infarction, *Canadian Journal of Cardiology*, 1999: 15(11):1277-82

[7] Dick RS, Steen EB. The Computer-based Patient Record: An Essential Technology for Health Care, Washington, D.C., National Academy Press, 1991.

[8] Tange HJ, Hasman A, de Vries Robbe PF, Schouten HC. Medical narratives in electronic medical records, *International Journal of Medical Informatics*, 1997: 46(1):7-29

[9] Cheung NT, Wong KC, Chan JTS, Leung CB, Lui SF. Quantifying the Operational Benefits of a Core Clinical Information System, *Proceedings of the Fourth Hong Kong (Asia-Pacific) Medical Informatics Conference*, 1997:147-152.

[10] Shenefelt PD. Limits of ICD-9-CM code usefulness in epidemiological studies of contact and other types of dermatitis, *American Journal of Contact Dermatitis*, 1998: 9(3):176-8

[11] Benesch C, Witter DM Jr, Wilder AL, Duncan PW, Samsa GP, Matchar DB Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease, *Neurology*, 1997: 49(3):660-4

[12] Chute CG, Cohn SP, Campbell KE, Oliver DE, and Campbell JR.. The content coverage of clinical classifications. *Journal of the American Medical Informatics Association,* 1996: 3:224-33.

[13] van Ginneken AM, Stam H, van Mulligen EM, de Wilde M, van Mastrigt R, van Bemmel JH. ORCA: the versatile CPR, *Methods of Information in Medicine*, 1999: 38(4-5):332-8

[14] Henry SB, Morris JA, Holzemer W. Using structured text and templates to capture health status outcomes in the electronic health record. *Joint Commission Journal on Quality Improvement*, 1997: 23(12):667-77

[15] Matsumura Y, Takeda H, Okada T, Kuwata S, Nakazawa H, Hazumi N, Inoue M. Devices for structured data entry in electronic patient record, *Medinfo 9*, 1998: Pt 1:85-8

[16] Kahn CE. Self-documenting structured reports using open information standards, *Medinfo 9*, 1998: Pt 1:403-7

Address for correspondence

Dr N.T. Cheung
Rm 520N, Hospital Authority Building
147B Argyle St
Kowloon
Hong Kong

email: cheungnt@ha.org.hk