This column is the third in a four-part series providing observations and insights from the author’s experiences with ambulatory electronic medical record (EMR) implementation. It is intended to provide a glimpse “behind the veil” of an EMR-based care environment, with a particular focus on issues that have hitherto received little attention in the literature.

“Hey, computer guy!”

These words, which reached my ears on a recent Monday morning, were not spoken in the tone of joyful recognition with which one might, for instance, hail the hot dog guy at one’s local ballpark. They were, in fact, more growled than spoken, in the baleful manner of a disappointed sports fan — or, as fate would have it, an overworked, if goodhearted, doctor. Discretion being the better part of valor, I valiantly started walking very fast in the other direction, but was soon overtaken.

“This darn computer system is driving me nuts,” said my colleague. “I saw a patient with a migraine this morning, and did you know it won’t let you just put down a diagnosis of ‘migraine’? It makes you pick from all those cockamamie choices. The only one that fit the patient was ‘migraine not otherwise specified without mention of intractability.’ What kind of nonsense is that? How is someone supposed to make sense of a patient’s problem list when it’s filled with that kind of gobbledygook?”

“Oh, ah,” I explained.

“And another thing,” he continued. “The choices for diagnoses don’t even make sense a lot of the time. You and I both know how asthma is divided into intermittent versus persistent forms, right? Why does that computer program make you choose

“Many of the advantages of EMRs over paper-based records derive from things the EMR can do with structured patient information.”

between ‘intrinsic asthma’ and ‘extrinsic asthma’? That’s a totally archaic way of thinking about the disease! You should get that software company to fix those lists, or better yet, just let us type in the diagnosis. Why do we have to choose from a list, anyway?”

“Well, the software company isn’t really responsible for the list. It’s the World Health Organization, and the National Center for Health Statistics, and as for choosing from a list…”

“Yeah, whatever. I gotta go. All this great technology has me running unbelievably late.”

Patient information stored in an EMR is often described as being either “structured” or “unstructured.” “Unstructured” information, of which plain narrative text is the most common example, lacks any standardization of format, and its meaning generally cannot be automatically “understood” by a computer system. “Structured” information consists of abbreviated representations of ideas (often numerical “codes,” with accompanying text descriptors or “terms”), whose meaning is standardized, i.e., agreed upon by all users of the representation format (“coding system”). This standardization facilitates automatic processing of the information by a computer.

Most EMRs allow for entry of patient information in a combination of structured and unstructured formats. For instance, the record of a patient visit might include both a free-text narrative (unstructured) and a code indicating the final diagnosis for the visit (structured).

Many of the advantages of EMRs over paper-based records derive from things the EMR can do with structured patient information. These have been detailed extensively in the literature, and include the following:

• Prevention of medical errors by alerting users to patient conditions that require intervention — also known as clinical decision support (CDS)
THE PHYSICIAN PERSPECTIVE

- Powerful tools for viewing information in a patient’s record (e.g., viewing numerical information graphically or viewing records of only those visits where a particular diagnosis was addressed)
- Generating data for billing processes as a by-product of clinical documentation
- Database reports across populations of patients for clinical, research, or administrative purposes

All of these benefits involve the output of information from an EMR. Getting structured information into the EMR, on the other hand, involves certain difficulties. This issue’s column focuses on some of those difficulties, and how our organization has attempted to overcome them.

Taxonomic Issues

The coding systems underlying structured data entry in EMRs determine the choices from which the users must select information to enter. These lists of codes impose their own conceptual framework (“taxonomy”) as to how the things they represent (symptoms, diagnoses, medical procedures, etc.) should be categorized. If this conflicts with the taxonomic framework with which the user is familiar, confusion and frustration can result.

The situation regarding asthma codes mentioned in the introduction illustrates this. The International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) is a commonly used coding system for medical diagnoses. We use it in our EMR to represent diagnoses for specific patient encounters as well as to populate patients’ “Problem Lists.”

As my frustrated colleague pointed out, the categorization scheme for asthma-related diagnoses in ICD-9-CM is based on the underlying cause. In contrast, current medical opinion emphasizes degree of severity as the most salient characteristic for categorizing asthma. Although our EMR allows users to append free-text comments to a diagnosis (e.g., a statement as to the severity of a patient’s asthma), this information cannot be effectively used by any automated processes in the EMR.

Semantic Issues

The information that physicians and other clinical personnel are trained to enter in their patient’s records is full of subtlety and nuance, which even unrestricted prose is sometimes inadequate to capture. It is unsurprising, then, that the use of structured data poses challenges having to do with the meaning of the data itself (“semantics”).

For example, many standardized coding systems do not provide a built-in way to express uncertainty. When one of our clinicians determines that a patient may have diabetes, it is possible to record a diagnosis of “diabetes” as structured information, but any statement as to the tentativeness of the diagnosis must be entered as free text. The EMR’s CDS modules, consequently, will generate the same alerts and reminders as for confirmed diabetics, and the patient will appear on the automated reports of diabetic patients we periodically generate.

Sometimes, a coding system offers a term that is congruent with what a user wishes to express, but is so awkwardly worded that it undermines the clarity of the patient record. The headache-related term about which my colleague was complaining is one example (from ICD-9-CM). This problem is compounded when terms must be abbreviated due to string-length limitations imposed by the database structure. For example, in our system, ICD-9-CM code 491.21, “Chronic bronchitis, obstructive type, with acute exacerbation” is denoted by the term “Obstr Chr Bronchitis W Ac Exacerb” — hardly the most comprehensible of diagnoses.

Lexical Issues

Although most coding systems assign numerical identifiers to structured data terms, it is impractical for users to memorize these numbers. Therefore, entry of structured data must rely on some interaction between the user and the text-based terms in a coding system. When the list of choices is short, this can be achieved by having the user choose terms from a “menu”-type list. When the list of choices is long, however, that approach is impractical, and it is necessary to have the user type free text into and then choose from among the matches retrieved by the EMR.

This process often does not work perfectly, even if a term exists that matches the meaning the user has in mind. If the text typed in by the user differs in the precise words used, or in the word order, from the term the coding system uses (“lexical differences”), the EMR will be unable to match the two. For example, in our EMR we use the Current Procedural Terminology (CPT) coding system for storing information on patients’ past surgical histories. Suppose a user wishes to record that a patient has had an incisional breast biopsy. The CPT term for that procedure is “biopsy of breast, incisional” (code #19101). If the user were to type in “breast biopsy,” “incisional breast biopsy,” or even “biopsy breast,” there would be no match.

There are ways to mitigate the lexical issues in entering structured data. In many EMRs, it is possible to assign additional text strings to each code such that those text strings, when entered by a user of the EMR, will result in a match to that code. We have found this to be an immensely valuable tool and have put substantial effort into populating our coding system databases in this manner.

We have also found it valuable to train users in the “art” of text pattern-matching. This consists in typing in...
just enough text to indicate what type of code one is searching for — with enough specificity to avoid generating an unwieldy number of matches, but without so much that minor lexical variation will prevent a match.

**User Interface Issues**

As mentioned above, when a structured datum is to be acquired and the allowable choices are few, it is possible to display them all as a “menu” and have the user select the appropriate item(s). One example of this common to many EMRs is the use of documentation “templates,” which prompt the user for responses to a number of menus, each of which details some aspect of a patient encounter (e.g., duration of symptoms). These templates compose a visit note from those choices, but in many EMRs also store the choices as structured data.

The design of the user interface has a substantial impact on the usability of menu-driven structured data entry in an EMR. Ideally, the EMR should allow users to switch easily between structured and unstructured data-entry modes, so they can “opt out” of structured data entry when the choices before them do not meet their semantic needs. In addition, there should be minimal need to switch between keyboard and mouse in the data-entry process.

**Conclusion**

Much of the promise EMRs hold for improving health care hinge on structured information. The difficulties in acquiring data in structured format, however, are substantial.

Limitations in existing coding systems, such as the taxonomic issues discussed above, can undercut the value of collecting structured data in the first place. However, improvements in coding systems, particularly the development of more flexible, multiaxial coding systems that can better adapt to changes in the conceptual constructs of medicine, can mitigate this.

The lexical issues in entry of structured data are also largely surmountable. Several companies, and one large government project (the Unified Medical Language System Metathesaurus) have undertaken the task of comprehensive lexical indexing of the major clinical coding systems. It seems likely that the eventual standard approach for addressing lexical issues will involve dedicated, regularly updated products, which can be “plugged in” to an EMR for linking free text entered by users to standardized codes.

The semantic issues described above are probably inherent in the process of giving structure to data. There will undoubtedly always be a role for unstructured text in EMRs to convey the details for which no coding system less complex than language itself is sufficient. It is incumbent upon EMR developers and implementers to recognize which structured data fields should be supplementable with free text, and to ensure that such free text is displayed in the EMR along with the structured data.

Lastly, careful attention should be given to user interface issues in the entry of structured data. Appropriate interface design can substantially speed the data entry process, so that the advantages of structured data entry are realized with minimal burden on busy clinical personnel.

**About the Author**

Eric Rose, MD, is the associate director for clinicalinformatics with the University of Washington Physicians Network. His home page is at http://faculty.washington.edu/momus/infodoc.htm.

---

**THE PHYSICIAN PERSPECTIVE**

“We have also found it valuable to train users in the "art" of text pattern-matching.”

— Eric Rose, MD