Common Medical Terminology Comes of Age, Part One: Standard Language Improves Healthcare Quality

Jeffrey S. Rose, MD; Bruce J. Fisch, MD; William R. Hogan, MD; Brian Levy, MD; Philip Marshal, MD, MPH; David R. Thomas, MD; Debra Kirkley, PhD, RN

ABSTRACT
It has become abundantly clear that standards of recording clinical terms in human-readable, computer-processable format are indispensable. Controlled medical terminology is the missing link in health information standards (in fact, medical terminology can be viewed as the mother of all standards); its absence interferes with the business of healthcare and impedes the core processes of healing and maintaining health. Medicine has lacked the controlled common medical vocabulary that would enable universal sharing of data at the point of care and ensure reliable information for determining health intervention effectiveness. Simple clinical and code content alone has proven insufficient for healthcare enterprises to successfully manage the terminology problem; the “lexical runtime engine,” formerly called a vocabulary server (VOSER), which manages the vocabulary ontology and serves up the relevant vocabulary to users of applications in the clinical environment, has recently become a reality.

KEYWORDS
• Controlled medical terminology/vocabulary
• Clinical informatics
• Common medical terminology
• Standard clinical terminology
• Standard language
• Healthcare quality
• Sharing of data
• Outcomes
Healthcare delivery today is reliant on information technology. Clinical decision making and the reengineering of business and clinical processes depend on the immediate availability of current and accurate information about patients, diseases, and populations. In fact, the products and processes of information technology (IT), along with strong clinical leadership in the new millennium, are critical to bringing about a badly needed shift in the entire care delivery system from one based on cost to one based on value. Never before have tools existed that could verify, validate, or debunk treatments; narrow practice variation to a cost-effective mean; enable discovery and communication of optimal methods of diagnosis or therapy; and literally prove the value of clinical interventions.

The computer age has clearly shifted the practice of medicine in a wildly different direction than where its austere and archetypal culture would have had it go. But shift it has, and as we appropriately move toward evidence-based practices and shared decision making, it has become abundantly clear that standards of recording clinical terms in human-readable, computer-processable format are indispensable. Controlled medical terminology is the missing link in health information standards (in fact, medical terminology can be viewed as the mother of all standards); its absence interferes with the business of healthcare and impedes the core processes of healing and maintaining health.

The business implications of new “medical intellectual capital” are enormous, because health information is becoming a “commodity.” Improved medical knowledge can enhance care appropriateness across a continuum of environments, narrowing expensive practice variability and creating competitive advantage in the industry for those who possess it. The overwhelming problem healthcare in the United States faces at the turn of the century is the one of transforming our cost-based technology-driven system into one driven by quality and value concerns.

Why “Controlled Medical Vocabulary”?

Excellent, sufficiently detailed information is required to convince practicing doctors that they may alter their routines without sacrificing quality. Even when clinicians and health systems are in command of the information, courage and patience are necessary to change habit and educate patients about the benefits of more appropriate care. To obtain and use this information, and to link clinical actions to the cost of running the business, clinicians and health enterprise managers need some specific tools:

- Computerized patient records, affording collection and presentation of patient-centered medical information to authorized providers at any location
- Standard methods of representation and communication, to make health data universally available to support care, enhance preventive medicine
practices, enable population-based interventions, and populate data repositories for research and outcome determinations

- Medical logic engines, to automatically assess patient data and process alerts, reminders, and communications about patients, ultimately leading to use of artificial intelligence to support clinicians

A new pseudoscience has developed in the information age that we call clinical infonomics. It refers to the state in which care practices, health information systems, and national economics become unified and inseparable. Michael Millenson notes that “We have finally gone from a system that fiercely defended the idea that medical accountability must be defined by the opinions of doctors to a system that accepts the principle that doctors can be held accountable by outsiders using objective data. . . . For a marketplace to work, consumers must have access to reliable and relevant information on everything from health plan rules and procedures to outcomes information from specific hospitals for specific procedures.”

The information technology tools we need so as to transform healthcare work well without controlled medical terminology. Medicine lacks the controlled common medical terminology that allows universal sharing of data at the point of care and ensures reliable information for determining health intervention effectiveness. The complexity of the English language, which is formidable indeed, pales in comparison to the eclectic, ambiguous, redundant idiom of clinicians.

Perhaps the most tragic symptom of the terminology disorder (which is heightened by the advent of the Internet) is the isolation a patient feels when she attempts to understand or educate herself about her condition. The semantic isolation we have created is not only troublesome for our professional lives; it “serves to disenfranchise individual(s) from the healing process.”

The Present Situation

The government and people of the United States spend an astounding amount of money on healthcare, yet we continue to rank poorly among the world’s nations with respect to the infant mortality rate. We also fail to provide adequate health services to some fifty million people. HCFA actuarial estimates note that the average rate of healthcare spending growth was 5 percent since 1993, will be 6.5 percent between 1998 and 2001, and will grow to 7.5 percent between 2002 and 2007. Over the next five years, fueled mainly by private sector spending, healthcare expenses will increase to $2.113 trillion (nearly 17 percent of the gross domestic product). Nearly two-thirds of the potentially controllable rising costs are due to increases in the volume and intensity of medical services (the content of care), which are determined by clinical decisions. Sadly, many of these decisions produce little or no demonstrable benefit. Add to this the observation that medical interventions
often do more harm than good,\textsuperscript{17–18} and one gets a good sense of the urgent need to apply IT infrastructure that can lead to clinical decision making with alacrity. Controlled medical terminology is a vital part of this infrastructure.

The Institute of Medicine study point regarding iatrogenic injury has received a great deal of attention lately: the recognition of 44,000–180,000 potentially avoidable treatment-related deaths and injuries in hospitals every year. The critical point is that most common failures are system errors involving drug knowledge dissemination, drug dosing, patient identity checking, and patient information availability.\textsuperscript{7,19,20–22} In a survey of attendees conducted at the 2000 Healthcare Information Management and Systems Society, 98 percent of respondents believed that controlled medical terminology would be important in mitigating iatrogenic mishaps.\textsuperscript{23}

We mention these expenditures, failings, and errors because the solution of the root dilemma falls firmly in the lap of information technologists, who can give practitioners the tools for revealing and promulgating better (and more cost-effective) treatments, mitigating errors, and elevating the quality of care delivered.\textsuperscript{24} To accomplish this, a three-pronged attack on a number of core issues is needed:

\begin{itemize}
  \item Inadequate patient information at the point of care
  \item Inadequate information about proper courses of medical action
  \item A profusion of bad and misleading data that cloud our thinking and confuse our actions clinically and administratively
\end{itemize}

\textbf{Controlled Medical Vocabulary and Medical Records.} Hospital and out-patient records are disorganized collections of (illegible) personal notes, inconsistently recorded data, and sundry test results. Data are frequently missing or inaccurate; the inconsistent format and fragmentation of information impedes searchability and defies logical grouping of problems. The lack of standard format, content, language, and completeness makes retrieving information from such records difficult. Information that can be harvested is often of dubious value.\textsuperscript{25,26} Illegible writing, arcane acronyms, cryptic personal abbreviations, and misfiled information are the rule. The electronic health record, which captures chart information by optical scanning, makes illegible and incomplete records available everywhere—hardly an improvement. Those EMRs that are entered manually or are encoded by a human solve the legibility problem but still fall short in providing usable information, because they are not uniform, recoverable, or comparable from chart to chart, from provider to provider, from automated system to automated system, and from institution to institution. This makes it nearly impossible to investigate costs, benefits, and outcomes of treatment.\textsuperscript{27}

Dr. Christopher Chute of the Mayo Clinic suggests that “improvement of medical knowledge about best practice depends upon the ability to study practice outcomes and apply them to the patients we see. This implies that we can
generate data about our patients that is comparable, so that it can be used in aggregate analysis, and so clinical decision support resources can be linked to patient data in real time. The single greatest obstacle to comparable data remains medical terminology. Failure to adopt and embrace a common terminology will doom outcomes research and data-driven clinical guideline development.\textsuperscript{28} Common medical terminology is the natural prerequisite for disease and health outcome studies; medical innovation risk-benefit and effectiveness determination; and consistent, timely presentation of evidence-based practice guidelines. All of these have been shown useful in eliminating wasteful activity and in modifying behavior.\textsuperscript{29–33}

One of our most dangerous and pervasive problems is that much of the coded clinical data we currently have is based on ambiguous, biased, incorrect, clinically inaccurate, and often incomprehensible billing codes that are collected from flawed medical records. These codes have been accepted uncritically because they are intended to fulfill financial purposes, not to produce clinical wisdom or truth. In fact, they have recently been mandated for use in the Health Insurance Portability and Accountability Act and the Health Care Finance Administration to simplify healthcare administration.\textsuperscript{34,35}

**The Path Forward: Controlled Medical Terminology.** Over the past two decades, intrepid medical informaticists have grappled with the issues of clinical terminology, and much has been learned.\textsuperscript{36,37} The Computer Based Patient Record Institute suggests that clinical terminology is “standardized terms and their synonyms, which [allow one to] record patient findings, circumstances, events and interventions with sufficient detail to support clinical care, decision support, outcomes research and quality improvement; and can be efficiently mapped to broader classifications for administrative, regulatory, oversight and fiscal requirements.”\textsuperscript{38}

In fact, three separate terminologies must be managed together:\textsuperscript{39}

1. **Application terminology**, also called interface terminology, refers to those terms seen in (and used in) documenting or facilitating care.
2. **Reference terminology**, more academic term classification, is often represented in a complex knowledge base and is rich with rigorously controlled rules and relationships (used predominantly for data analysis).
3. **Administrative terminology**, also called code sets, is a collection of coded expressions used for financial or ancillary system communications; they may or may not have any significant cognitive relationship to the other two terminology types.

Work with the existing terminologies has revealed several other important observations:

- Simple nomenclature (listing) of terms has proven inadequate to serve as a controlled medical terminology schema because it is difficult to manage and nearly impossible to use for knowledge derivation.
• Classification, which is better because it relates terms to each other in various ways, is still less mature than is optimal for efficiently managing, updating, and tracking term and concept changes. Terms in classification are placed in relation to each other because someone decided they should be so related, thus often depending not on knowledge purity but merely on the purpose the terminology was to serve, as with billing codes.

• Knowledge-based ontology, which strictly defines and relates concepts and their terms according to definition. The knowledge-based ontology seems to be best suited to accomplishing the controlled medical terminology task. An ontology uses specific methods to organize abstract symbols (concepts), represented by terms we humans can understand and manipulate. Computers can recognize and manage these abstract concepts by use of assigned numeric codes (meaningless to humans), whereas we humans require readable terms to identify them. An ontology therefore offers the human-readable and machine-processable conceptual framework that diverse computer systems may use for purposes of interoperability. A seamless bridge is built in this manner from our words to the abstract concepts they represent in the computer. Note how closely such an approach approximates the dictum that “the core utility of any language is to represent concepts by terms or words with meanings that are commonly understood.”

Additional critical characteristics of a “good” controlled medical vocabulary set are important as we examine existing vocabularies and code sets:

• **Multiple-use.** The content of the controlled medical terminology should be sufficient to support multiple uses of the structured vocabulary, from application terminology to ancillary code sets, and should be expandable in a methodical fashion so that the core ontologic model is preserved.

• **Unique.** The unit of meaning in the controlled medical terminology should be a unique concept (symbol), represented by a meaningless computer-readable identifier. Each term related to its relevant symbol should have a single conceptual meaning (not vague). Further, this conceptual meaning should refer to one or more unique terms that are not associated with other concepts (not redundant).

• **Unchangeable.** Once defined, a concept should be permanent and immutable. Though it may be made inactive (retired from use or view), the symbol and its terms must remain present in the terminology structure and should retain uniqueness.

• **Hierarchical.** A concept and its terms should be related to each other in the form of a taxonomy—a hierarchy—based on the concept’s essential meaning rather than on an arbitrary or use-specific determinant. Individual terms or concepts may be represented in multiple hierarchies, if appropriate, as long as they remain unique.

• **Logically defined.** A concept should be formally defined by logical relationship to other concepts; such relationships should be through explicit, as
opposed to assertional, definition. (As an example of an explicit definition, pancreatic neoplasm is-a child of gastrointestinal neoplasm, has-morphology neoplasm and has-body-site pancreas. An assertional definition is pancreatic neoplasm has-prognosis poor. The latter may be true, and even characteristic, but poor prognosis does not consistently or explicitly define the concept.)

- **No negative definitions.** The terms “not-elsewhere-classified (NEC):not represented in the terminology in a proper way” and “not-otherwise-specified (NOS):no modifiers of the base concept” should never be used as concepts.

- **Increasingly granular representation.** A concept in the terminology should be represented in multiple levels of granularity, from atomic and logically indivisible (fine-grained) to general or compound (coarse-grained) so that multipurpose use of the terminology is enabled.

- **Multiple-viewed.** The controlled medical terminology, being multipurpose and multigranular, should therefore have multispecific views available to support different uses and applications. (A view to support detailed clinical documentation may reveal different concepts and terms than a view to support billing in the same terminology.)

- **Omitted contextual classification.** The myriad potential contextual uses of a concept should be left to application developers and not modeled within the terminology. (An EMR application must know whether “low back pain” has been used to record a diagnosis in the SOAP note even though it is defined as a clinical finding in the lexicon; or that “asthma” was documented as a chief complaint although it is properly represented in the terminology as a diagnosis.)

- **Evolving.** The terminology should be capable of graceful evolution as new concepts are created, existing concepts are refined, and outmoded concepts are retired. (Controlled terminology refers to a regulated, logical, consistent semantic framework based on accurate and precise meaning, not to one that is fixed, inflexible, or incapable of growth.) The controlled terminology must carefully track changes and notification of commandment violation if the terminology is modified.

Keep these requirements in mind while reviewing the common code and terminology schema in the article “Common Medical Terminology Comes of Age, Part Two” in this volume.

**Tools and Software to Manage CMT.** It is evident that controlled medical terminology (CMT) is optimally modeled and updated in a knowledge-based ontology. Associated with this concept-term representation tool must be a “lexical runtime engine” (what was formerly called a vocabulary server), which serves up the relevant vocabulary to users of applications in the clinical environment. Simple clinical and code content alone has proven insufficient for healthcare enterprises to successfully manage the terminology problem. This section describes some present and emerging software requirements that bring academic exercises into the world of practical application. This section serves as an introduction, but we hope it equips you to understand and select tools for managing your particular needs.
**The Modeling Environment for Vocabulary Services.** A lexicon model implies the existence of a data structure that supports both maintenance and use of controlled vocabulary data. This model is accessed by way of a modeling environment, which may be simple or complex, ranging from a basic list of data with few or no links among elements of the list to a fully realized multidimensional display that facilitates moving within and through the lexicon.

The fundamental architecture of a modeling environment is a relational or object-oriented database core, some application access to this database (such as through an API, or application program interface), and an isolation paradigm. Experience reveals strengths and weaknesses in both database schemas (such as in speed and conceptual clarity). The latest technological approaches enable implementation of object frameworks within relational databases and thus garner the benefits of object-based and relational database systems.

The measure of a modeling environment is its ability to convey the information contained within the model to the user. Here are some of the core facilities offered by a modeling environment:

- **Browsing:** A graphical user interface for looking through the lexicon and moving from concept to concept in a logical fashion.
- **Searching:** Finding and navigating directly to a desired spot in a hierarchy without having known where it was at the start. A viable search interface should support rapid navigation to concept information as well as to neighborhoods of information.
- **Editing:** Making manual or bulk changes to the vocabulary to fulfill local requirements. An important caveat: although editing is critical at the local level, editing systems should have some form of security to keep nationally standardized vocabulary data pure.
- **Configuration management and version control:** As changes are made by administrators and users, careful tracking and auditing of the changes is important. No deletion or change of term should be allowed without keeping comprehensive audit trails. It is through these logs that synchronization and automated updates are facilitated. These tools enable conflict resolution of ambiguous or redundant terminology. They also facilitate frequent, large, or multiple bulk-load changes as are common with a new version of a code set release and with continuous change in pharmacy terminology.
- **List management:** A list creation and management tool set is also necessary to extract the most value from a vocabulary. The vocabulary must be rich in content as a whole, but the user or application may at any time need only a portion of the hundred-thousand-some concepts. Lists accomplish this task. An administrator creates a list to offer choices within an application served by the vocabulary. Another example of list use is an administrator making the starter set of pharmacy items that a group of users may order. A list can be further enhanced by having several instances of common medications, one for each ailment to be treated. Each time a medication is listed, it has
the appropriate prescription details for treatment of a particular disorder (dosage, duration, refills, instructions, warnings, and so on). The listed medication also includes a unique nickname to aid the user in picking the correct medication.

Management tools are needed to create and manage lists, just as they are needed for the lexicon. Advanced management tools are needed when a medication is no longer indicated for an illness, the recommended dosing has been changed, or the medication is removed from the institution formulary altogether. It may be necessary to find all instances of the concept across many lists and modify or delete the entry. Another requirement of the tool set is to create and edit compound terms. Simply having the term “allergic rhinitis” in the diagnosis list would not satisfy an allergist, since severity modifiers are also required: “mild allergic rhinitis,” “moderate allergic rhinitis,” “severe allergic rhinitis.” Instead of filling the vocabulary with all possible permutations of a concept, compound terms can include nicknames and be combined in a list or lists for easy selection.

Conclusion

If we possessed the informatics tools, language, and communication standards outlined here and could act in accordance with evidence-based guidelines and principles; if we could also decrease practice variation and improve knowledge of appropriate and cost-effective techniques for managing health; then we would have the rudimentary tools for controlling the nearly two-thirds of “potentially controllable” rising costs in health expenditure each year.

Examples of significant informatics and evidence-based impact on healthcare cost are abundant. A report of the McKinsey Consultant Group estimated that the U.S. healthcare bill could be cut by $270 billion a year—25 percent of total expenditure—if medical organizations made an annual investment of $50 billion in information systems. A strong case for return on investment for expenditure on informatics tools of a comprehensive nature remains difficult to make because of the myriad human variables involved in cost recovery. However, unless these tools are employed with the courage and conviction to use new information to practice medicine on the basis of evidence, it will be a long while before healthcare woes are successfully addressed. Doing so requires a fundamental shift in how clinicians approach disease and how we treat our patients.

References


**About the Authors**

Jeffrey S. Rose, MD, is a physician executive with the Care Transformation Consulting Practice, Cerner Corp.

Bruce J. Fisch, MD, is chief medical officer of Health Language, Inc., Aurora, Colorado.

William R. Hogan, MD, is medical director of Health Language, Inc.

Brian Levy, MD, is vice president of Health Language, Inc.

Philip Marshall, MD, MPH, is director of products, WellMed, Inc.

David R. Thomas, MD, is physician program manager, Clinical Information Solutions, at IBM.

Debra Kirkley, PhD, RN, is vice president of product management at Health Language, Inc.