REVIEW ARTICLE

Medical Coding Dictionary Management & Maintenance

The use of medical coding dictionaries for medical terms data such as adverse events, medical history, medications, and treatments/procedures are valuable from the standpoint of minimizing variability in the way data are reported and analyzed. This chapter discusses the importance of medical coding dictionaries in streamlining and improving the quality of medical terms data obtained during collection and coding. Furthermore, reconciliation of medical terms data between a safety database and a clinical database is improved with the use of medical coding dictionaries during a clinical study. Issues that can affect conversion of reported terms to dictionary terms are considered, including autoencoders, the use of coded terms, and dictionary and software change control and versioning. Due to their widespread use, MedDRA and WHO Drug are discussed in more detail than other dictionaries.

Introduction

Recording and storing data in a controlled, consistent, and reproducible manner for data retrieval and analysis is a necessity for regulatory compliance and clinical study success. To provide control and consistency, a variety of medical coding dictionaries may be used to process, analyze and report collected data. These dictionaries range in size and complexity from simple code lists with a few entries to large and complex dictionary systems containing thousands of entries and related tables. Two examples of commonly used dictionaries are the *Medical Dictionary for Regulatory Activities* (MedDRA) and the *World Health Organization Drug Dictionary* (WHO Drug). Some dictionaries are well established and have been used for years, while others are more recent and may be revised or updated regularly. Establishing and maintaining medical coding dictionaries are important tasks that clinical data management (CDM) personnel or coding specialists must carefully manage.

Transitioning to a new or different coding dictionary presents multiple challenges to CDM. First and foremost is the consistency between the same or related data that are analyzed and reported with different dictionaries (even a different version of the same dictionary). Any lack of familiarity with the content, organization, use, or capabilities of the new dictionary must be addressed prior to its implementation. Processes must be established for managing the release of multiple versions of the same dictionary, handling different dictionaries or versions that have been used, and integrating data coded with different dictionaries or versions.

Scope

This chapter focuses on the management of dictionaries used for coding adverse events, medications, medical history, and other types of clinical data. Although the chapter touches on custom medical dictionaries, the primary focus is on standardized, commercially available medical coding dictionaries and some available options for coding management tools. Any use of the words "dictionary" and "code" within this chapter refer specifically to medical coding dictionaries and medical coding, as opposed to programmatic coding dictionaries and coding.

This chapter does not cover the actual process of coding; for more information on coding please refer to the "Safety Data Management and Reporting" chapter of the GCDMP, as well as the ICH-endorsed guide for MedDRA users, *MedDRA® Term Selection: Points to Consider.*[1]

Minimum Standards

· Select dictionaries that meet project and regulatory requirements.
· Follow established security procedures for dictionary installation and maintenance.
· Ensure user licenses are obtained and kept up to date for any dictionaries and applications used.
· Ensure all sponsor personnel and vendors who will use the dictionaries hold the appropriate licenses. If a vendor has access to the dictionary application, ensure the application license covers vendor access.
· Implement an audit trail for all changes/updates to the dictionaries or synonym listings and support tables associated with the dictionaries.
· Do not modify published commercially available coding dictionaries. If a commercially published dictionary has been modified, then do not refer to it by its commercially available product name.
· Specify the dictionary name and dictionary versions used during coding on all study reports and integrated summaries.
· Store all utilized versions of dictionaries for future reference.

Best Practices
- Select a coding tool to facilitate consistent dictionary use.
- Include the version(s) of utilized dictionaries in metadata.
- Ensure all levels and versions of dictionaries used for coding can be accessed by data management and other dictionary users.
- Establish a process for evaluating term or categorization changes in a dictionary and its effect on previously coded data when moving to a different version.
- Ensure the capability to recode to different versions of a dictionary. For example, this may be needed to allow integrated study analyses to be reported using the same version.
- Ensure individuals who code data have training and professional background appropriate to the dictionary and the version for which they are coding. Training must be completed and documented before coding with the dictionary or version.
- Educate individuals involved in recording, monitoring, reviewing, analyzing and reporting coded data on the functionality and capabilities of the coding dictionaries used.
- Submit requested dictionary changes to the organizations responsible for maintaining the dictionaries using the appropriate approved process of submitting changes.
- Ensure change control processes are in place for all dictionaries, whether commercially available or custom.

Established Standardized Dictionaries in Common Use
MedDRA
Recognizing the increase of global studies and submission of marketing applications to multiple regulatory agencies, the International Conference on Harmonisation (ICH) undertook the development of a global dictionary, which resulted in MedDRA.[2] The US Food and Drug Administration (FDA) is currently using MedDRA in its Adverse Events Reporting Systems (AERS).[3] MedDRA was planned to eventually replace some of the coding dictionaries already in use, such as COSTART and WHO-ART, as well as proprietary variations of those dictionaries that had been developed by sponsors of clinical studies. In many organizations, MedDRA has already replaced other dictionaries that were used in the past.

When MedDRA was initially published, updates were released on a quarterly basis. Following the release of version 4.0 in June 2001, the frequency of updates was reduced to semiannually. The organization responsible for publishing and maintaining MedDRA is MSSO (Maintenance and Support Services Organization). MSSO recognizes the need to stabilize MedDRA terminology to address concerns that an overwhelming amount of resources are needed to maintain frequent version updates and subsequent recoding and reanalysis of adverse events.[4] Since the initial release of MedDRA, revisions have addressed topics in the following areas.

- Updated assignments to system organ class (SOC)
- Consistent use of terminology
- Retirement of terms from current status
- Addition of new terms identified during implementation of the dictionary in clinical studies

A review of the impact of each change and whether an improved coded term is available in a new dictionary version is facilitated by the ability to search within various versions for coded adverse events and dictionary entries at each level. It is possible that an update to a given version will not contain any new terms in a particular grouping or modifications of existing coded terms.

MedDRA is a multiaxial dictionary, meaning that a preferred term (PT) may be associated with multiple SOCs. Each PT, however, is associated with only one primary SOC, regardless of the number of secondary SOCs with which it is associated. Sponsors and contract research organizations (CROs) frequently made changes to dictionaries prior to MedDRA, leading some users to believe the same could be done with MedDRA. MedDRA, however, should not be modified in any way by users. This prohibition of user modifications includes both changes to terms and changes to the assignment of a primary SOC for a term. MSSO has established a detailed process for users to follow, which involves bringing the issue to the attention of MSSO if they find a term to be lacking or in error.

WHO Drug
WHO Drug is one of the more commonly used dictionaries, and was designed by the World Health Organization (WHO) for coding medications in clinical studies.[5] Medications used by study participants prior to or concurrently with a study are commonly coded to facilitate reporting and analysis. A variety of dictionaries or medication references provide information about prescription, generic, and over-the-counter (OTC) medications, as well as herbal supplements. The medication references used for coding medications should be selected based on the scope of medications included, how recently the reference has been updated, the frequency of updates to include the release of new medications, and coding information available in the dictionary. Such coding information may include generic terms, active ingredients, indications, or Anatomical Therapeutic Chemical (ATC) classification. WHO Drug is widely considered the most comprehensive resource for medication coding, and is also associated with a quarterly journal, *WHO Drug Information*, that discusses the most recent news and trends relating to medications and medication development.

In recent years, WHO Drug has been distributed in several formats, known as format B-1, format B-2 and format C. In format B-1, drug names may be repeated within the dictionary if the same name is used for different drugs, which may occur due to each drug being marketed in different languages or countries. Format B-2 is similar to format B-1, except each drug name appears only once within the dictionary. When a drug name appears more than once in the B-1 format, the first entry that was entered into B-1 is typically used as the B-2 entry.

Format C is the newest of the three formats, uses a different file structure than the B formats, and also includes each drug's available dosage formulations (e.g., caplet, liquid, intravenous, etc.) and dosage amounts (e.g., 10 mg, 20 ml, etc.). Format C is much more specific than the other two formats because it can contain many more entries for the same drug, with each entry representing a unique combination of that drug's formulation and strength. Format C was originally intended to replace the B formats, but many companies had difficulties implementing it. As a result, the Uppsala Monitoring Centre (UMC), which is responsible for maintaining and licensing WHO Drug, agreed to continue distributing format B-2 indefinitely. However, the UMC has indicated that starting in 2009 it will no longer distribute the B-1 format, although the files will be available upon request. A WHO Drug license entitles the subscriber to receive both available formats (B-2 and C) of the dictionary.

In 2005, the UMC introduced the WHO Drug Dictionary Enhanced (WHO-DDE). The WHO-DDE combines data from the original WHO Drug Dictionary (WHO-DD) with additional country-specific drug information collected through the UMC's collaboration with IMS Health (an international consulting and data services company). The WHO-DDE is therefore several times larger than the WHO-DD.

New versions of WHO Drug are released quarterly, but companies have the option to receive new versions on a quarterly, biannual or annual basis. The cost for a WHO Drug license is dependent on the frequency that a company chooses to receive updates, with higher costs for more frequent updates. New subscribers only have the option to subscribe to the WHO-DDE, whereas subscribers who are currently receiving the WHO-DD have the option to continue receiving the WHO-DD or upgrade to the WHO-DDE.

Other Dictionaries

Although MedDRA and WHO Drug are the most commonly used dictionaries for clinical studies and postmarket surveillance, the following list briefly describes a few established but not as widely used dictionaries.

- WHO ART—The World Health Organization Adverse Reactions Terminology is a dictionary designed by WHO for coding adverse reactions.
- COSTART—The Coding Symbols for a Thesaurus of Adverse Reaction Terms was developed by the FDA for coding and reporting adverse reactions. It was originally used by the FDA for coding adverse events, although it has since been replaced by MedDRA.
- SNOMED CT—The Systemized Nomenclature of Medicine–Clinical Terms was developed by the College of American Pathologists as a coding system to capture information about medical history, treatments and outcomes.
- CTCAE—Common Terminology Criteria for Adverse Events was developed by the National Cancer Institute as a system for classifying the nature and severity of adverse events. Work is currently underway to integrate CTCAE with MedDRA.

- ICD-9—Published by the WHO in 1977, this dictionary consists of coding for diagnoses and procedures.
- ICD-9-CM—An update to ICD-9, this dictionary became the official system for assigning codes to diagnoses and procedures in hospitals within the United States. Medicare and Medicaid have required the use of ICD-9-CM codes since 1988. These codes are updated yearly.
- ICD-10—Completed by WHO in 1992, and while implemented in most of the world, the dictionary was not adopted in the United States. This dictionary was originally designed to report mortality; however, modified versions have since been created for diagnosis codes (ICD-10-CM) and procedure codes (ICD-10-PCS).

Custom Medical Coding Dictionaries

Custom dictionaries are typically developed to meet company-specific processes. Most custom dictionaries display terminology in a hierarchical pathway ranging from broad terminology to very specific terms. These dictionaries can be used to code adverse event data, medical history data and more commonly, medication data. Some organizations may create a custom dictionary by adapting a commercially available dictionary to better meet the organization's specific needs. If this approach is used, the customized dictionary should not be referred to by the same name as the commercially available dictionary.

There are limitations to using a custom dictionary, such as the lack of a central governing body to maintain the dictionary hierarchy for terminology and classification. Custom dictionaries also may not be consistent with terminology as it evolves over time (e.g., drug formulations may change over time or cease to be marketed). Although versioning is important for all coding dictionaries, some companies may not have a rigorous versioning strategy for custom dictionaries. All relevant regulatory standards should be taken into consideration when developing custom medical coding dictionaries. Additional steps for data reconciliation between different sources might be required when using custom medical coding dictionaries.

Dictionary Application Software Selection

When choosing a coding dictionary, one must also consider the software that will be used to house and search the dictionary. Some dictionaries are already packaged with an accompanying software application, but there are cases where the software must be chosen separately. In all cases the applications need to be validated prior to being put into production. In addition, proper validation of changes and updates needs to be performed prior to any changes or updates being released into production.

Application Service Provider (ASP)

An ASP system can come in many variants depending on the contract with the provider. An ASP may host and manage the implementation and validation of dictionaries, or may provide customized tools for managing and using dictionaries. All types of ASP systems, however, share a

common model, which is that the software is owned by the ASP, usually runs at the ASP's data center using the ASP's servers, and the customer pays a monthly or other contracted fee for service. Most ASPs allow for minimal customization and do not allow for most company-specific items. Support is usually supplied by the ASP, although depending on the contract, some support may be provided by the customer as well.

Commercially Available Applications
Commercially available applications are software packages that are purchased by the user, and may also be referred to as "off-the-shelf" applications. One key difference between commercially available and ASP applications are that with commercially available applications, the customer usually hosts and manages the application on their own servers. Commercially available applications also allow for more configuration options to meet an organization's specific needs. Commercially available software packages are usually more amenable to the use of "add-ons" to allow interaction with other software packages. To make changes to the application software, the request will need to go through the company that owns the software. Application support is typically shared between the software producer and the customer's information technology (IT) support department.

User-Built Applications
Some organizations may choose to build systems that are tailored to the specific needs of the organization (e.g., logistics and workflow). In these situations, the organization hosts the software on their servers and provides all support services. The organization is also responsible for ensuring applications needing validation have followed appropriate software development lifecycle processes to validate the application and the functionality of the application after installation.

The benefit of user-built applications is that they can be customized to meet the organization's specific needs. A risk of user-built applications is that they are dependent upon organization resources. Another risk is that poorly written business requirements may result in an application that does not adequately meet the organization's needs once the application development is completed.

Medical Coding Tools and Methods
In addition to the actual dictionaries and software applications used to house them, CDM personnel and dictionary users should be familiar with the following tools and methods used in dictionary management.

Autoencoders
Autoencoding is a programmatically assisted process for matching a reported term to a dictionary term. A basic autoencoder will take a list of reported terms and look for an exact match with dictionary terms. Various methods exist for autoencoding, such as character string matches with the dictionary, character string matches with synonym lists, and matches found using algorithms.

Within the context of autoencoding, a synonym list is a repository of terms that have previously been coded. Advanced autoencoder designs allow the user to define algorithms to assist with finding suggested "best" matches. These coding algorithms should be evaluated for their ability to handle synonym listings, misspellings, capitalization, word variations, word order, exclusion of irrelevant verbatim text, and other issues that may impede accurate matching. An autoencoder is useful when a large number of entries must be coded, and can expedite consistent coding by eliminating the requirement of manual reevaluation of previously coded terms.

Consistency checks can be performed within some autoencoders. Some autoencoders may also allow for multiple dictionaries and versions of the dictionaries. Added features may also include the ability to access multiple coding jobs on demand; create and maintain synonym lists; configure algorithm lists to support autoencoding; and integrate to safety and clinical databases. These broad-spectrum coding systems decrease regulatory risks and increase efficiency, providing more consistent and high-quality coding output.

Some clinical data management systems include an autoencoder and provide the ability to load electronic versions of coding dictionaries. Other autoencoders may be available as separate applications. Both integrated and standalone autoencoding applications must be fully validated according to current regulatory standards. Other features to be considered when selecting an autoencoder include ease of use, security features and, ideally, the ability to load and store multiple dictionaries or versions. Despite the assistance provided by autoencoders, a manual review of coded data should be performed to ensure consistency and accuracy.

Manual Coding
Manual coding refers to a situation where a person selects an appropriate dictionary entry for each reported term, either in the patient database or in a module of the dictionary application that deals with discrepancies. This method may be used when an autoencoder is unable to code a term or an autoencoder is not being used. Some clinical data management systems have the ability to use manual coding, but standalone manual coding applications are also available. Both integrated and standalone manual coding applications must be fully validated according to current regulatory standards.

Some manual coding applications use the same types of algorithms as autoencoders to provide the user with a list of suggested dictionary terms for a given reported term. Coding applications with this capability should be user-configurable (i.e., allowing for the creation and maintenance of lists of synonyms appropriate to the dictionary) and allow for suitable testing of the configuration to ensure that the suggested terms are accurate and comprehensive.

Ideally, a manual coding application will allow the coder to view all components of a dictionary (e.g., the full hierarchy for MedDRA or the ingredient list and ATC codes for WHO Drug), and also have the ability to see how other reported terms have been coded to ensure consistent

coding of similar terms. Additional features to consider for a manual coding application are the ability to review coded terms for accuracy and consistency, the ability to query a term when it cannot be coded, audit trails that record the user and date/time a term was coded, and extensive, easy-to-use search capabilities.

Hybrid Approaches to Coding

A hybrid approach to coding uses an autoencoder to first automatically code those reported terms that match a dictionary term or that match a term that has previously been coded (i.e., a synonym list). The terms that are not autoencoded are then manually coded. Many clinical data management systems and standalone coding applications support this hybrid approach to coding.

A coding application being used in a hybrid approach should have the same features desired in an autoencoder or a manual coding application. In addition, a hybrid coding application should allow a coder to easily see the terms that did not autoencode, and which will therefore require manual coding. Some hybrid coding applications may provide the ability to distinguish between autoencoded and manually coded terms and a facility to manually override any autoencoded terms, if necessary.

Browsers

A browser is a computerized tool used to aid in accessing terms in a specified dictionary. Browsers are designed to quickly find terms of interest and should be flexible, intuitive, and quick to use.

- Stand-alone browsers—These are applications that allow for the easy search and review of dictionaries. Some also possess a capability for limited linking to external applications (e.g., study databases), where one may not be able to affect a term or coding change from the browser, but would be able to call (or open) the browser from within the dictionary application.
  - WHO Drug—Several WHO Drug browsers with differing feature sets exist, including one produced by the Uppsala Monitoring Centre (which is an entity of WHO that works with international drug monitoring).
  - MedDRA—An application has been provided by the MSSO for searching the MedDRA dictionary, but other vendor-created browsers also exist, with differing feature sets.
- Browsers that are contained within dictionary management systems have enhanced capabilities, although the availability of these enhanced capabilities varies across available systems. Some of these systems can act as a browser, as well as a vehicle for importing and exporting individual reported terms or a batch of reported terms. Various coding approaches outlined above can by performed once the terms are imported into the system.

Dictionary System Validation

Any dictionary application or system for housing dictionaries requires documented validation prior to being placed into production. This validation should include system validation, unit validation (if this level of detail is needed) and user acceptance testing. Full documentation should be maintained for the application, including business requirements, system requirements, design specifications and any other documents or support level agreements that are in place for the system.

The level of validation to be performed by an organization may vary depending on the origin of the system. ASP and commercially available applications may require less validation effort than a user-built application or system. Regardless of whether performed by an ASP, software vendor, or the organization conducting the research, all systems and applications require validation and supporting documentation to meet industry and regulatory standards, as well as to pass audit inquiries.

To prevent any untoward effect on subject data, changes to an application, whether a bug fix or a planned upgrade, may require validation and testing prior to being placed into production. The dictionary application or system that houses the dictionaries also requires documented change control and version control procedures. Change control procedures and version control schemas are usually set by the IT department of the organization to ensure clinical study software needs meet the standards of good clinical practice.

Change Control

The practice of modifying published dictionaries is clearly discouraged by the ICH for the MedDRA dictionary.[1] The organizations that maintain dictionaries have an established process for submitting change requests if, for example, an adverse event term or medication is reported that is not included the dictionary. This process allows for a review of the requested change and dissemination of the request to others using the dictionary. An approved request will appear in a future release of the dictionary. A declined request will not.

Although in-house modifications are highly discouraged, any in-house modifications made to a published dictionary should be clearly stated in reports, so as not to mislead reviewers who are familiar with the published dictionary. Any changes made to dictionary entries should also have documented authorization and be included in an audit trail.

Coding dictionaries may be available in electronic and/or printed format, and multiple versions may be released or published. The dictionary and version used for a given project, time period, or data set should be clearly documented. Where this information is documented may vary between organizations (e.g., in a data management plan or coding guidelines), but the dictionary and version should be referenced in clinical study reports or integrated summaries that report on the coded terms. For multiple ongoing studies, the study team should determine which dictionary and version will be used for coding each study. A systematic process and instructions should be in place to ensure the consistent use of the appropriate dictionary and version.

Processes should be established for evaluation of the extent of changes between versions, the impact of changes

on previously coded terms, and criteria for recoding and implementing the latest version.[6]

Using different dictionaries or versions over a period of time increases the importance of version control, documentation and standardized data reconciliation processes. For example, different versions may be used for coding postmarket safety data versus clinical data, between different studies for the same drug, or even within long-term studies. The impact of version changes can be greater for adverse events, because a term may be deactivated or reassigned to a more appropriate term, rendering the earlier term assignment outdated. Most of the changes to medication dictionaries simply introduce new medications.

Dictionary and version information may be maintained within the clinical database, within the autoencoder as the dictionary files are loaded, or within the metadata of data sets containing coded data. Because version information may be incorporated into electronic files by organizations maintaining published dictionaries, that information may be available without the need for additional in-house action.[7]

Process steps for installing and upgrading to new dictionary versions may vary between organizations and specific dictionaries. However, dictionary installations or upgrades should be subjected to a holistic approach to validation once installed, including processes such as remapping synonym tables and recoding subject data repositories.

Recommended Standard Operating Procedures
· Maintenance of Dictionaries
· Security, Change and Version Control
· Validation and Testing Procedures

Competing Interests
The author has no competing interests to declare.

References
1. **International Conference on Harmonisation Secretariat.** *MedDRA® Term Selection: Points to Consider*, Release 3.2. Geneva, Switzerland: IFPMA; 2003. Available at: http://www.meddramsso.com/mssoweb/docs/PTC_3.2_final.pdf. Accessed December 3, 2008.
2. **MedDRA MSSO Web site.** Available at: http://www.meddramsso.com/. Accessed December 3, 2008.
3. **US Food and Drug Administration.** *Reviewer Guidance: Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review.* Washington, DC: US Government Printing Office; 2005. Available at: http://www.fda.gov/CDER/GUIDANCE/3580fnl.pdf. Accessed December 3, 2008.
4. **MedDRA MSSO.** *Messenger Quarterly Newsletter: May 2002.* Reston, VA: Author; 2002.
5. **Uppsala Monitoring Centre Web site.** Available at: http://www.who-umc.org/. Accessed December 3, 2008.
6. **MedDRA MSSO.** *Recommendations for MedDRA® Versioning for Summary Reporting.* Reston, VA: Author; 2000.
7. **MedDRA MSSO.** *Messenger Quarterly Newsletter: January 2001.* Reston, VA: Author; 2001.

Chapter Revision History

Publication Date	Comments
September 2003	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
May 2009	Revised for content, style, grammar, and clarity.