Formatting and Character Restrictions in Healthcare Documentation

by Laura Bryan, CMT, AHDI-F

Medical transcriptionists (MTs) are primarily concerned with creating an accurate written record that clearly reflects the provider's thoughts and intentions, and as you probably know, this requires more than simply "typing" what is said. As documentation methods have progressed from typewriters to word processors and now integrated transcription platforms, modern transcription requires a broader understanding of document-creation and document-storage methods. Electronic record systems and software interfaces used throughout the healthcare industry often influence the way documents are created and the way style rules are applied. It is important that those working in the field of healthcare documentation understand the life cycle of transcribed reports so the information can be successfully used to provide optimal patient care.

Healthcare delivery involves many forms of written information—discrete data, documents with paragraphs of narrative "free text," graphs, charts, and images. The healthcare community uses a variety of formats, styles, and notations to clearly communicate health-related information. Greek letters, the degree sign, virgules, ampersands, greater-than and less-than signs, superscripted ordinals, mathematical operators, and other miscellaneous symbols play an important role in the communication and differentiation of technical and scientific information. For example, the hormones associated with the thyroid are typically written with subscript numbers (e.g., T_3 and T_4) while the thoracic vertebrae are indicated with plain arabic numbers (e.g., T3 and T4). Many routinely accepted abbreviations such as *D&C* and *T&A* are so commonly written with the ampersand that one has to pause to recognize the meaning when these terms are written any other way.

Issues of style, such as those published in the AHDI *Book of Style for Medical Transcription*, are primarily concerned with the clear and unambiguous communication of technical and scientific data in the field of healthcare. In addition to technical style rules, character attributes such as bold and underlining, and paragraph attributes such as alignment, indention, and numbering contribute to a document's structure and organization and improve readability. Word processors such as MS Word and WordPerfect make maximum use of formatting and style to increase the readability and even the functionality of a document.

To the extent possible, it is important to apply formatting and rules of style to a transcribed medical report, but within your workplace you may encounter restrictions in the use of certain characters or style guidelines due to technical limitations imposed by the information system that is used. Interfaces that pass information from one computer system to another may impose limitations as well. This article describes the technologies and limitations that influence the application of formatting and style guidelines in medical documentation.

ASCII

To exchange text-based information electronically, every character must be defined and standardized. Computer code must be determined for every character—letters, numbers, symbols, punctuation marks—even the spaces between words. The first standards for exchanging text were developed in the early 1960s for use in Teletype equipment and were later adopted for use in computerized systems including the modern personal computer. The earliest standard for encoding text-based characters is the American Standard Code for Information Interchange (ASCII). The standard is commonly referred to as the ASCII character set or just ASCII (pronounced "askie"). The standard used in the United States is designated US-ASCII. ASCII is a universally accepted standard for encoding characters in written communication and represents the most basic set of characters used in a computing environment.

The ASCII standard includes 128 characters but may be extended to 256 characters. The number of characters in the standard is actually based on the number of characters that can be uniquely encoded using either seven or eight bits per character. A bit is the smallest unit of information that a computer processor can understand. A bit has only two possible values, written as either a 0 or a 1. Code that uses only two possible values is called binary code. Using seven bits per character with only two possible values per bit (i.e., 0 or 1), a total of 128 characters can be encoded ($2^7 = 128$). Eight bits allows for 256 unique codes ($2^8 = 256$). Table 1 shows the characters included in the ASCII character set along with each character's corresponding code number and binary code.

The first 31 characters in the set are actually commands (also referred to as nonprinting characters) for controlling the

To the extent possible, it is important to apply formatting and rules of style to a transcribed medical report, but within your workplace you may encounter restrictions in the use of certain characters or style guidelines due to technical limitations imposed by the information system that is used. Interfaces that pass information from one computer system to another may impose limitations as well.

transmission of text messages as well as very basic layout commands (e.g., new line, carriage return). Most of these control characters were used in the original standard designed for Teletype machines and have since become obsolete. Characters numbered 32 through 127 (as shown in Table 1) include the space, upper and lowercase letters of the Latin alphabet used in U.S. English, arabic numbers 0-9, punctuation marks used in U.S. English, mathematical operators, and basic symbols such as the dollar sign, asterisk, number sign, tilde, underscore character, and the caret. The space (created by the spacebar) is not really considered a character but rather a "nonprinting graphic." The extended set, represented by characters 128 through 256, include Latin letters with diacritics, some Greek letters, several superscripted characters, mathematical operators, and miscellaneous symbols.

An 8-bit system with 256 characters could suffice for most applications in U.S. English, but there are many languages that use other alphabets such as Arabic, Cyrillic, and Hebrew. To accommodate the many characters used throughout the world, a larger and more extensive character set has been defined and is referred to as the Unicode system. Although other character sets have been developed, the first 128 characters of the US-ASCII character set have remained consistent and have been duplicated within other character sets for universal compatibility. As such, these 128 characters represent the most common and widespread set of characters used in a computing environment. Up until 2008, the US-ASCII character set was the most widely used standard for displaying text on the World Wide Web.[1]

Defining character sets is vitally important for creating, displaying, and exchanging text-based information. Software that displays text must be specifically programmed to recognize characters, and only those characters included in the specified character set will be recognized and displayed. In addition to recognizing specific characters, software must also be "aware" of formatting attributes that are added on top of characters. For this reason, a basic understanding of ASCII characters is required if you work in an environment where documents are exported to other information systems or into an archival system with limited character sets available. Systems that display plain text may only take advantage of the first 128 characters in the US-ASCII character set. Characters outside of the ASCII character set may not translate or be recognized correctly. Even though the transcription software or word processor you use to create documents makes a wide variety of symbols and characters available, you must know which characters are allowed and avoid using those that will not be recognized. When instructed to use only ASCII characters, it is implied that you would use only the characters in the first 128 positions of the US-ASCII character set.

The characters contained in the ASCII set can be viewed using the Symbol dialog box in MS Word, as shown in Figure 1. Open the Symbol dialog box (press ALT, I, S) and change the Subset to Basic Latin. In the bottom right-hand corner of the Symbol dialog box, change the From drop-down box to ASCII (decimal). This will display the selected character's ASCII identification code at the bottom of the dialog box. The first 31 characters, originally assigned to control commands when the standard was created in 1963, are not included in the Symbol dialog box. The first character on the grid is character 32, which is the space. Characters with a character code of 126 (the tilde ~) or lower are consistent across all encoding standards and will always be recognized. Character 127 is reserved for the delete command.

Characters listed in the extended set (characters 128-256) may vary depending on the typeface being used. Characters that do not have a corresponding key on the keyboard have a universal "ALT" code, which is the key combination for inserting the character using the ALT key combined with four digits on the keyboard's number pad. This code is also displayed on the Symbol dialog box. Each individual character is assigned the same character code and the same ALT key code regardless of the typeface. Letters may have different styling due to their typeface, but an "M," for example, is an "M" in the same sequential order with the same character code in all typefaces. You can demonstrate this by changing the font in the Symbol dialog box and comparing the order and codes for each character in the Basic Latin subset.

When instructed to limit characters to the "ASCII characters," you can easily reference the Symbol dialog box and use any character on the grid up through the tilde when the From box is set to ASCII (decimal). You can view Unicode character sets in Word (change the From box to Unicode) or use the Windows Character Map. To access the Character Map, go to Start > Programs > Accessories > System Tools > Character Map.

If you study the characters in the Symbol dialog box, you may notice that the extended set includes a superscript 2 and a superscript 3 (character code 178 and 179, respectively). These are defined characters, which are not technically the same as the "regular" characters with superscript formatting applied. In a plain-text editor, the superscript characters will be handled differently than a number 2 with superscript formatting applied. The final outcome will depend on the character set defined in the plain-text editor.

[1]**http://www.w3.org/QA/2008/05/utf8-web-growth.html**, World Wide Web Consortium, Accessed 8/26/2009.

Table 1
ASCII Characters

Decimal	Binary	Value	Description
032	00100000	SP	Space
033	00100001	!	exclamation mark
034	00100010	"	double quote
035	00100011	#	number sign
036	00100100	$	dollar sign
037	00100101	%	percent
038	00100110	&	ampersand
039	00100111	'	single quote
040	00101000	(	Left parenthesis
041	00101001	)	right parenthesis
042	00101010	*	asterisk
043	00101011	+	plus
044	00101100	,	comma
045	00101101	-	minus or dash
046	00101110	.	dot
047	00101111	/	forward slash
048	00110000	0	
049	00110001	1	
050	00110010	2	
051	00110011	3	
052	00110100	4	
053	00110101	5	
054	00110110	6	
055	00110111	7	
056	00111000	8	
057	00111001	9	
058	00111010	:	colon
059	00111011	;	semi-colon
060	00111100	<	less than
061	00111101	=	equal sign
062	00111110	>	greater than
063	00111111	?	question mark
064	01000000	@	AT symbol
065	01000001	A	
066	01000010	B	
067	01000011	C	
068	01000100	D	
069	01000101	E	
070	01000110	F	
071	01000111	G	
072	01001000	H	
073	01001001	I	
074	01001010	J	
075	01001011	K	
076	01001100	L	
077	01001101	M	
078	01001110	N	
079	01001111	O	
080	01010000	P	
081	01010001	Q	
082	01010010	R	
083	01010011	S	
084	01010100	T	
085	01010101	U	
086	01010110	V	
087	01010111	W	
088	01011000	X	
089	01011001	Y	
090	01011010	Z	
091	01011011	[	left bracket
092	01011100	\	back slash
093	01011101	]	Right bracket
094	01011110	^	caret/circumflex
095	01011111	_	underscore
096	01100000	`	
097	01100001	a	
098	01100010	b	
099	01100011	c	
100	01100100	d	
101	01100101	e	
102	01100110	f	
103	01100111	g	
104	01101000	h	
105	01101001	i	
106	01101010	j	
107	01101011	k	
108	01101100	l	
109	01101101	m	
110	01101110	n	
111	01101111	o	
112	01110000	p	
113	01110001	q	
114	01110010	r	
115	01110011	s	
116	01110100	t	
117	01110101	u	
118	01110110	v	
119	01110111	w	
120	01111000	x	
121	01111001	y	
122	01111010	z	
123	01111011	{	left brace
124	01111100	|	vertical bar
125	01111101	}	right brace
126	01111110	~	tilde
127	01111111	DEL	delete

File Formats

Text can be created and saved in a variety of electronic formats using text editors or word processors. The most common formats used in healthcare include plain text (with the extension *.txt), rich text format (with the extension *.rtf), and document format (with the extension *.doc), which is created in Microsoft Word and Word-based processors. Plain-text files are just that—files containing only text. A plain-text editor typically has no provision for storing or displaying character attributes such as bold, italic, or underline. Multiple font faces are typically not supported either. Plain text editors also do not recognize paragraph attributes such as indention or alignment. Line wrapping may or may not be supported.

Rich-text editors are based on the RTF file format originally created by Microsoft in 1987. Microsoft developed this file format for cross-platform use (i.e., to be compatible with

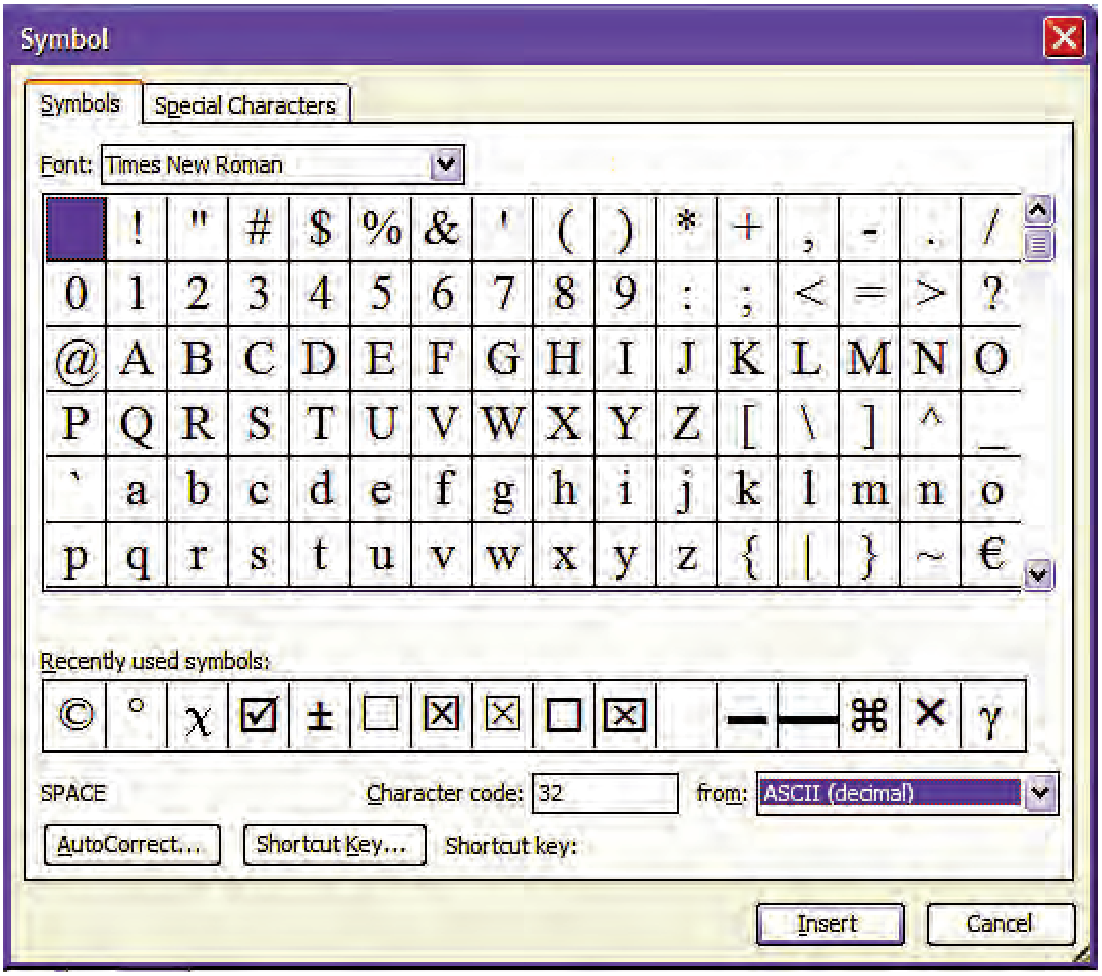


Figure 1. The Symbol dialog box in MS Word showing the ASCII characters.

a Mac, Windows, or Unix system). At a minimum, rich-text editors are capable of supporting bold, italic, and underline formatting. Many typically support paragraph alignment, indents, line wrap, and lists. In contrast to plain-text editors, rich-text editors can also display a variety of font faces and will maintain document margins. Most word processors are capable of opening and displaying files saved in RTF, so this format is favored for its interoperability as well as its ability to retain basic formatting. On the down side, there are several versions of the RTF specification with varying degrees of functionality. This makes it more difficult to predict which characters will be supported and what formatting attributes will be recognized.

Text files, in the form of *.txt and *.rtf, are popular in large enterprises such as hospitals because they have many advantages over proprietary file formats. Text files are cross-platform, generally compatible, and easily exchanged across multiple software applications. Because they contain little-to-no formatting information, text files are much smaller in size. They require less space for long-term storage and can be transferred throughout a network quickly and easily. Text editors do not require the purchase of expensive user licenses, upgrade fees, or ongoing maintenance fees (as is the case for MS Word), making text editors more economical to deploy in a large hospital or transcription service.

Word processors, on the other hand, have a full array of formatting available. MS Word is a robust application for efficiently creating attractively formatted and stylized documents, but it is not always the best format for universal compatibility or for efficient use of storage media. Word's automatic formatting feature, called AutoFormat As You Type, may insert characters that are incompatible with hospital information systems. For example, Word automatically formats quotation marks as "smart quotes." These quotation marks curve inward toward the text so that the opening quotation marks curl to the right and the closing marks curl to the left. Standard quotation marks (included in US-ASCII) are straight, making these curly quotes an entirely different character as far as a computer is concerned. Word also automatically formats single-space fractions for the three most commonly written fractions (¼, ½, and ¾) and automatically inserts an em dash when two hyphens are typed together. The automatic numbered list feature in MS Word uses field codes instead of actual numbers to display the list number, and these fields are not always understood when the text is exported to other information systems. All of these features can be disabled by removing the appropriate check mark on the AutoFormat As You Type dialog box (ALT, T, A > AutoFormat As You Type).

A fully formatted Word document can be saved as a text file (with the extension .txt) but all formatting will be lost. Any images in the file will also be removed. Superscript and subscript characters will be converted to regular characters and unrecognized characters may be substituted or converted to an empty box. Table 2 shows examples of formatted text after converting to text format using different character sets and different conversion rules.

Transcription departments and transcription services typically use software that is designed specifically for creating documents accurately and efficiently. These systems are often based on word processing applications such as MS Word. These applications include important productivity tools for inserting blocks of text, automatically correcting typos, and spell checking. While MS Word is one of the most common processors used for creating transcribed reports, there are also many proprietary software systems on the market. These systems may be built on a rich-text editor or a scaled-back version of MS Word. It is difficult to predict what characters and formatting capabilities are available to the transcriptionist in these proprietary systems.

In many healthcare environments, documents are still printed and placed on the chart. Printed documents benefit greatly from formatting such as bold and indent, and the healthcare facility may prefer a traditionally formatted document because of its improved readability. These same documents may also be transferred into electronic record systems that allow access of the information from multiple points inside or outside the facility. A hospital may be reluctant to install a copy of MS Word on every workstation in the facility because of the high licensing costs, yet users on these stations must still be able to view transcribed reports. Because healthcare facilities use and reuse the text within documents in various ways, characters contained within the document must be compatible with other software applications used throughout the enterprise as well as usable within the long-term storage/archive system.

Connecting Information Systems

Today, large healthcare facilities (hospitals, large multispecialty clinics, ambulatory care centers) use a variety of software applications to manage the entire healthcare enterprise. Each department within a healthcare system may have its own software vendor and a variety of software applications designed specifically for its use. For example, the admissions department uses software to admit patients and gather demographic and billing information. The laboratory uses a Laboratory Information System (LIS) for managing specimens and reporting results, and radiologists use Radiology Information Systems (RIS) for tracking and reporting imaging studies.

Much of the information gathered through these various systems is fed into the Hospital Information System (HIS) that is accessed by the physicians, nurses, and others involved in direct patient care. Each of these separate systems is connected through a series of interfaces, which are software programs that allow disparate systems to exchange information.

Table 2

Formatting commands were applied to the original text in the first column and actual characters (taken from the Symbols dialog box in MS Word) were used to transcribe the text in the second column. Note how characters and formatting are handled differently when converted to text or rich-text format.

Original using Formatting	Original using Characters	US-ASCII character set	MS-DOS character set	US-ASCII with substitutions allowed	RTF
	Temp 98.6°	Temp 98.6?	Temp 98.6ø	Temp 98.6?	Temp 98.6°
	mm² cm³	mm? cm?	mmý cm?	mm2 cm3	mm² cm³
	1¼ 2½ 3¾	1? 2? 3?	1¬ 2« 3?	11/4 21/2 33/4	1¼ 2½ 3¾
	“Hello!”	?Hello!?	?Hello!?	"Hello!"	“Hello!”
mm^2 cm^3		mm2 cm3	mm2 cm3	mm2 cm3	mm^2 cm^3
1¼ 2½ 3¾		1? 2? 3?	1¬ 2« 3?	11/4 21/2 33/4	1¼ 2½ 3¾
“Hello!”		?Hello!?	?Hello!?	"Hello!"	“Hello!”

It is common for people to refer to HL7 as if it is a specific software application or even a brand of software, but this is not actually the case. The HL7 organization creates the standards and protocols that programmers use to write customized code for exchanging information in a given situation.

In these situations, the transcriptionist must be aware of constraints placed on documents by the interfaces that move text between the various systems within the healthcare enterprise.

The most common type of interface used in healthcare facilities is called an HL7 interface. HL7, which stands for Health Level Seven, is a standards developing organization (SDO) that creates standards for exchanging administrative and clinical data in the field of healthcare. According to its website (**www.HL7.org**), "HL7, the leading authority for global healthcare IT standards, provides standards (protocols) for interoperability that improve care delivery, optimize workflow, reduce ambiguity and enhance knowledge transfer among all of our stakeholders, including healthcare providers, government agencies, the vendor community, fellow SDOs and patients."

It is common for people to refer to HL7 as if it is a specific software application or even a brand of software, but this is not actually the case. The HL7 organization creates the standards and protocols that programmers use to write customized code for exchanging information in a given situation. HL7 interfaces are typically unique but are encoded using a standard protocol that defines the structure and content of information to be passed between two computer systems. HL7 interfaces are "event driven" which means a "message" is created and sent to a receiving computer system when a real-world event occurs such as a patient being admitted to the hospital or a lab test being completed.

One of the most common types of HL7 interfaces is the Admission, Discharge, and Transfer (ADT) feed. This interface provides patient demographic information to the various information systems used throughout a healthcare facility so the information does not have to be entered multiple times. Transcriptionists often use ADT feeds as a reference for patient demographic information to properly identify a transcribed report. HL7 interfaces are also used to pass documents from the transcription information system to the hospital information system and/or the document archival system.

The HL7 protocol is a text-based messaging standard, and as such, information flows through an interface as a "stream" of text. A message is created by gathering the necessary elements of a message and lining up the data in a specified order. A message is composed of segments, which are clusters of related information. For example, an ADT message includes the patient identification segment that contains the patient's name, identification numbers, and date of birth. Segments are further divided into composites. The information contained in a segment is identified by its position within the segment. The HL7 standard specifies the name of segments, the type of information included in a segment, and how many composites must be included in each segment. For example, the segment designated PID (patient identification) contains 11 composites and the patient's name is always contained within the fifth composite of the PID segment.

Delimiters are used to mark the beginning and end of composites and the individual components of a composite are also demarcated. As you can see in Figure 2, the vertical bar (called a pipe delimiter) is used to separate composites, and the caret is used to separate components of a composite. This figure shows part of an ADT message for a patient John Doe born on 2/3/1948. The first segment (starting with MSH) is the message header. The second segment identifies the event (EVN) and the third segment contains the patient's identification (PID). In addition to the pipe and the caret, the backslash, ampersand, and the tilde may also be used within messages as delimiters.

HL7 messages are text-based and must use ASCII characters. The only identifiers in an HL7 message are the segment identifiers, so the order of the information within the segments is crucial to identifying the information on the receiving end. The delimiters are also critical for marking the beginning and end of components of the message. A receiving computer is said to "parse" the data, which means the message is deconstructed and the data is stored within the receiving computer according to the data structure of the

```
MSH|^~\&|EPICADT|SMS|SMSADT|19991271409|CHARIS|ADT^A04|1817457|D|2.3|
EVN|A04|199912271408|||CHARRIS
PID||0493675^^^2^ID 1|454721||DOE^JOHN^^^19480203|M||B|254
```

Figure 2. **Part of an HL7 ADT message showing pipe delimiters and carets to separate the components of the message.**

receiving system. If the delimiters are placed incorrectly or removed, the information will be parsed incorrectly and the remainder of the message will be misconstrued.

The delimiters used in HL7 messages (especially the ampersand) may also be used as characters within the actual user data. To prevent characters used within the actual data of the message from being interpreted as a message delimiter, the HL7 standard provides a method for ignoring those characters when they are contained within a message but are not part of the message structure. Programmers refer to this as "escaping" a character, which essentially means the character is ignored when it appears in a given area of the message. In this way, any character in the ASCII standard can be included in the user data of an HL7 message.

When documents are sent through an HL7 interface, the document's text is typically converted to a "string" of text and contained within a segment of the message. Formatting commands may be converted to formatting codes embedded in the string of text and then reinterpreted when the message is processed by the receiving system. Any characters contained within the document's text will be part of the string. If the parser encounters a character used as a delimiter within the body of the document (for example, an ampersand), it will act on that delimiter according to the parsing rules.

Transcriptionists may encounter limitations on the types of characters that may be used within a transcribed report because "the characters will not go through the HL7 interface." Unfortunately, the ampersand is one of the most common characters to be disallowed, yet it is a widely used symbol in transcribed reports. Technically, the HL7 standard does not disallow the use of any ASCII character, but the programmer may fail to instruct the interface to escape characters when located within the body of the transcribed report. Transcriptionists can help to inform information specialists on the importance of style and usage in medical documentation and work cooperatively to improve the quality of the information flowing through a healthcare enterprise.

Conclusions

Transcriptionists play an important role in patient care by accurately conveying the meaning and intent of dictated information. The need to communicate clearly must be balanced with the software's ability to transmit and display the information as well as the storage capacity of the system used to manage the information.

Although technical specifications related to allowable characters and formatting may clash with the AHDI *Book of Style for Medical Transcription*, MTs must understand how to apply technical specifications when creating healthcare documents. Failure to comply may result in incomplete, inaccurate, ambiguous, or even lost reports. Understanding the importance of style to the overall goal of quality patient care is important, and conveying this to software engineers, programmers, and other information technology professionals can lead to better software solutions and better end-user functionality.

1 Technology and Workplace
CEC approved

Laura Bryan, CMT, AHDI-F, is the author of *The Medical Transcriptionist's Guide to MS Word, Technology for the Medical Transcriptionist*, and the soon to be published *Microsoft Word for Healthcare Documentation*, all published by Lippincott Williams and Wilkins. She has worked over 20 years in the healthcare field and currently owns MedEDocs, a transcription technology services company, as well as MTWerks, a technology consulting business focusing on medical transcription.

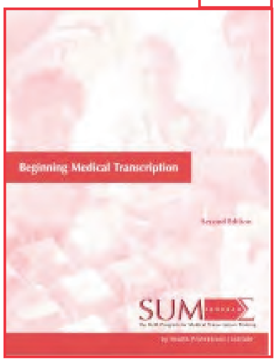
The SUM Program for Medical Transcription Training

Real Dictation • Effective • Comprehensive • Affordable • Teacher Support

Transcription Practice Component

Beginning Dictation

8+12* hours



Dictation from:
- Physician offices
- Clinics
- Hospitals

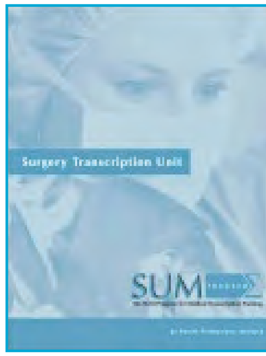
Report types:
- Progress notes
- H&Ps
- Preoperative notes
- Consults
- Discharge summaries
- ED reports

Medical Specialties:
- Cardiology
- Dermatology
- ENT
- Endocrinology
- Gastroenterology
- Gynecology
- Immunology
- Neurology
- Obstetrics
- Oncology
- Ophthalmology
- Orthopedics
- Pediatrics
- Plastic Surgery
- Psychiatry
- Pulmonology
- Reproductive Med
- Trauma
- Urology

*** The 12-hour Classic BMT is included on the BMT2 CD. Total 20 hrs.**

Intermediate Dictation

8 hours



8 hours of original surgical dictation:
- Cardiovascular/Thoracic Surgery
- Gastrointestinal Surgery
- Genitourinary Surgery
- Head & Neck Surgery
- Neurosurgery
- Obstetrics & Gynecology Surgery
- Orthopedic Surgery
- Plastic Surgery

Advanced Dictation

19 hours

New: Transcript Answer Keys have been updated to meet recommendations of AHDI's *The Book of Style for Medical Transcription*, 3rd ed.

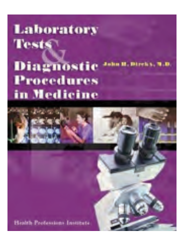
Specialties can be completed in any order

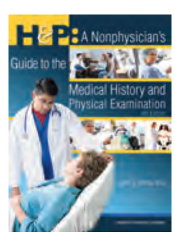
19 hours of authentic, challenging medical and surgical dictation from hospitals and surgery centers.

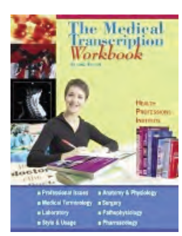
Academic Instruction Component

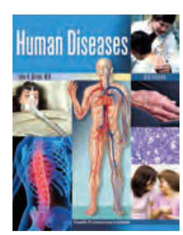
The SUM Program for Medical Transcription Training includes a rigorous academic component with recommended textbooks (some available from HPI) and reading assignments in seven courses.

- **Anatomy and Physiology**
- **Medical Terminology**
- **Human Diseases**
- **Medical Science**
- **Pharmacology**
- **Laboratory Tests**
- **Professional Issues**



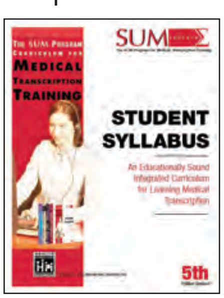






Help for Teachers

Help for Students



The SUM Program Teacher's Manual and the accompanying articles are over 200 pages of information on course offerings and curriculum planning. **Download it for free** at **www.SUMprogram.com**. Numerous articles are grouped into the following sections:
- Curriculum and Program Design
- Teaching Methodologies
- Evaluation and Grading
- Organization Tips
- Educationalese and Marketing
- Professional Growth

SUM PROGRAM
The SUM Program for Medical Transcription Training

www.SUMprogram.com