e-Health – For Continuity of Care
C. Lovis et al. (Eds.)

doi:10.3233/978-1-61499-432-9-720

Abbreviations in Swedish Clinical Text - use by three professions

Elin LÖVESTAM[a,1] , Sumithra VELUPILLAI[b] and Maria KVIST[b, c]
[a] *Dept. of Food, Nutrition and Dietetics, Uppsala university, Uppsala, Sweden*
[b] *Dept. of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden*
[c] *Dept. of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Sweden*

Abstract. A list of 266 abbreviations from dieticians' notes in patient records was used to extract the same abbreviations from patient records written by three professions: dieticians, nurses and physicians. A context analysis of 40 of the abbreviations showed that ambiguous meanings were common. Abbreviations used by dieticians were found to be used by other professions, but not always with the same meaning. This ambiguity of abbreviations might cause misunderstandings and put patient safety at risk.

Keywords. Abbreviations, health records, dietetic records, ambiguity

Introduction

Abbreviations and acronyms are frequently used in patient records, and become a patient safety risk as many abbreviations are ambiguous [1, 2]. The word abbreviation originates from *brevis*, latin for short, and this is precisely the purpose of abbreviations in several ways. Not only are they quicker to write and read for the professionals using electronic health records (EHRs), they also make the whole text shorter and easier for the reader to overview. Abbreviations in EHRs are often domain specific but can also belong to general language use [3]. There are standard acronyms established in the medical community that can be found in medical terminologies, but often abbreviations are created ad hoc, not following standards, and ambiguous. For English medical abbreviations, a third of the short abbreviations in the UMLS terminology are ambiguous [3]. Furthermore, a word or expression can be shortened in several different ways. These features can depend on subdomains such as specialty or profession.

Interpreting abbreviations and acronyms can be troublesome even for different health care professionals and they are one of the biggest comprehension barriers for patients reading their EHRs [4, 5]. Ambiguous and unexplained abbreviations in patient records can lead to misunderstandings that put patient safety at risk [2]. According to Swedish legislation, the patient record must be understandable for safe health care, and should also enable the patient to follow her/his health care process.

There are not many studies on abbreviations in Swedish EHRs, and for an emerging profession such as dietitians there is almost nothing written about documentation. A categorization of abbreviations is informative when developing

[1] Corresponding Author.

automated abbreviation detection systems for improved information extraction. The aim of this study was to analyse the use of some abbreviations in clinical text by three professions in Swedish health care; dietitians, nurses and physicians.

1. Methods

Abbreviations from dietetic EHRs were collected by the first author in a dietetic corpus. These abbreviations were used to extract matching abbreviations from EHRs written by three different professions. A sample of these extracted abbreviations were then analysed in their context to study the level of ambiguity.

1.1. Data

The original dietetic corpus consisted of 147 systematically collected notes, about 30 000 words. They were all written by dietitians in 2009-2010 in hospitals and primary care centers in three different counties in central Sweden.[2]

Three subsets from the Stockholm EPR Corpus[3] [6] from the years 2009-2011 were used in this study: dietetic notes (DIET), daily notes written by nurses (NURSE) and radiology notes written by physicians (X-RAY). The total number of words was 9 516 625 (DIET), 13 925 899 (NURSE) and 15 370 134 (X-RAY), respectively.

1.2. Extracting and matching abbreviations

A manual content analysis of the original dietetic corpus was performed by a dietitian, identifying and interpreting all abbreviations found in the 147 dietetic notes. The analysis resulted in a list of 266 abbreviations. The three subsets (DIET, NURSE, X-RAY) were tokenized using Stagger [7], a Swedish tokenizer and part-of-speech tagger. Modifications were made in Stagger in order to properly handle tokenization of the domain-specific abbreviations. Each word was then matched with the list of abbreviations in order to compute the frequencies in each subset.

1.3. Context analysis

In a context analysis of the original dietetic corpus, the abbreviations were classified in different categories depending on their meaning. From the list of 266 abbreviations, 40 abbreviations were selected on the basis of being promising for analysis; either because they were short (1-2 letters) and therefore could be expected to be ambiguous, or because they were highly subdomain specific for dieticians and therefore could be expected to mean something else for the other two professions. For each of these 40 abbreviations, 20 sentences were randomly extracted from each of the three subsets (DIET, NURSE, X-RAY), and read by two of the authors (a dietitian and a physician) taking note of the different meanings. The disambiguated abbreviations were then classified in regards to their meanings, using the same categories as for the original dietetic corpus.

[2] Deemed exempt by the Regional Ethical Review Board in Uppsala

[3] Ethical approval by the Regional Ethical Review Board in Stockholm, permission number 2012/2028 31/5

2. Results

2.1. Frequency analysis

Most abbreviations from the original corpus were found in all three subsets (72%, see table 1). Some abbreviations (8%) were used exclusively by dieticians, almost all concerned food and diet. There was also a small number of abbreviations (n=24) that were not found in the other subsets. Most of these concerned variations of non-standardized general abbreviations, medical assessment and intervention, and food/meals.

Table 1. Frequencies of shared abbreviations, not taking into account that they may be ambiguous. Abbreviations from the original dietetic corpus were matched against DIET, NURSE AND X-RAY.

SEPR Corpora with matched abbreviations	Number of instances	%
DIET + X-RAY + NURSE	191	72
DIET + X-RAY	9	3,4
DIET + NURSE	12	4,5
X-RAY + NURSE	7	2,6
DIET only	21	7,9
X-RAY only	0	0
NURSE only	2	0,8
None (original dietetic corpus only)	24	9,0
total	266	100

2.2. Context analysis of the original dietetic corpus and the three subsets

The classification of the 266 abbreviations in the original dietetic corpus is shown in table 2. Some abbreviations had multiple meanings and therefore five abbreviations were placed in more than one category. Table 2 also shows the classification of the three subsets regarding the analysis of the 40 selected abbreviations, where the abbreviations in the DIET subset more often concerned food and medical assessment or treatment while the same abbreviations in the X-RAY subset more often concerned diseases and body parts.

Almost all of the 40 selected abbreviations had ambiguous meanings, both within and between professions (1-13 meanings, mean 4.3). The abbreviations in the DIET corpus were least ambiguous (mean 1.8 meanings/abbreviation), while the NURSE (mean 2.7) and X-RAY (mean 2.7) subsets showed more ambiguity, figure 1.

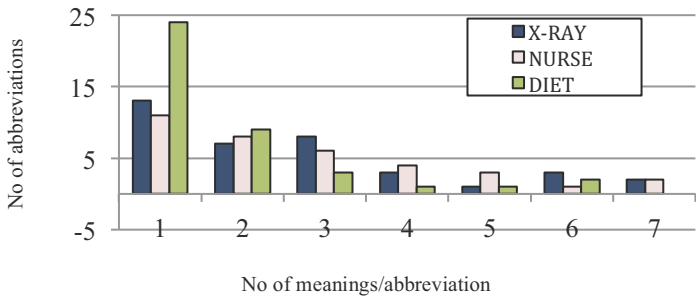

Figure 1. Multiple meanings is manifested for many of the 40 abbreviations for each of the three subsets

Table 2. Categorization of abbreviations found in the original dietetic corpus and the three subsets.

Categories	No of abbreviations				Explanation
	Original corpus (n=266)	DIET (n=40)	NURSE (n=40)	X-RAY (n=40)	
General non-standardized	90	20	27	21	Not classified in any of the specific categories below, non-standardized but common in everyday Swedish
General standardized	12	4	5	6	Found in the Swedish Writing Rules (SWR) [8]
General standardized, punctuation mistakes	25	2	2	1	Found in the SWR, with minor punctual deviations
Food/meals	36	10	8	1	Meals, food and cooking
Medical assessment and intervention	33	13	11	3	Assessment methods and scales, treatments, medicines
Diseases	18	7	6	17	Diseases
Units	17	4	9	8	Length, volume and other units
Institutions/hospitals	16	8	3	3	Names of institutions
Laboratory values	16	3	2	2	Blood lipids, iron status etc
Health care professionals	5	1	1	1	Professions within health care
Body parts	0	0	1	6	Organs and other body parts
Names of persons or products	0	6	8	6	Names of persons or products
Probable spelling mistakes	3	5	13	16	No identified meaning, interpreted as spelling mistake

An example of this ambiguity is the abbreviation “gr” which has in total seven different meanings: gram (*gram*), Celsius degree (*grader*), group (*grupp*), times (*gånger*), angular degree (*grader*), level of disease (*grad*), due to (*p gr a, på grund av*).

A certain expression can have many different abbreviations, and the abbreviations may be dependent on profession, for example six different abbreviations for the disorder Diabetes Mellitus were found in the three subsets (table 3). While dietitians and physicians used different abbreviations for the two main types of the disorder, nurses predominantly used “diab”. There may be other abbreviations specific for nurses and physicians, e.g. DM 1 and DM 2, not found in this analysis.

Table 3. Frequencies of abbreviations for different expressions of the disorder diabetes mellitus. Abbreviations from the original dietetic corpus were matched against DIET, NURSE AND X-RAY. Note that the subsets are of different sizes and that the different frequencies therefore are not directly comparable.

		No of instances found		
Abbreviation	Original expression	DIET	NURSE	X-RAY
diab	diabetes	118	965	158
dm	diabetes mellitus	432	38	290
DM	Diabetes Mellitus	597	69	4219
DMT1	Diabetes Mellitus type 1	71	0	0
DMT2	Diabetes Mellitus type 2	39	0	0
IDDM	Insulin Dependent Diabetes Mellitus	416	0	36

3. Discussion

In this study we have characterized the use of abbreviations in Swedish patient records among three health care professions. Ambiguous abbreviations were common both within and between the different professions. A limitation of the study is that it was based upon an exact matching of the 266 abbreviations found in the original dietetic corpus, meaning that no other abbreviations than those we looked for were found. A more open collection of all abbreviations in the corpora, followed by categorization, might have revealed even more synonymous abbreviations as well as more ambiguity.

There seems to be somewhat different abbreviation cultures among the three professions, such as the use of different abbreviations for Diabetes Mellitus, or the more frequent food abbreviations among dieticians. However, there are also many similarities, and many of the different meanings of the analyzed abbreviations were common for all three professions.

The frequent ambiguity of abbreviations, as well as the use of many different abbreviations for the same word, may have implications for patient safety as it increases the risk of misunderstandings of the EHR. It also hampers the possibilities for patients to understand their own patient records. Our main contribution is the categorization and analysis of abbreviation types used by different professions, which could be informative when developing automated abbreviation detection systems.

Acknowledgement

The authors thank Maria Skeppstedt for technical assistance and fruitful discussions. This work was supported by a grant from the Vårdal Foundation.

References

[1] Friedman C, Kra P, Rzhetsky A. Two biomedical sublanguages: a description based on theories of Zellig Harris. *J of Biomed Informatics* 2002; 35: 222–235.

[2] Paparella, S. Avoiding dangerous abbreviations and dose expressions. *J of Emergency Nursing* 2004; 30(1):54-58.

[3] Liu, H, Lussier, YA, Friedman, C. A study of abbreviations in the UMLS. In *Proceedings of the AMIA Symposium* (p. 393): 2001. American Medical Informatics Association

[4] Laippala V, Danielsson-Ojala R, Aantaa K. et al. Vocabulary in Discharge Summaries–The Patients' and the Nurses' Perspective. The 4th International Louhi Workshop on Health Document Text Mining and Information Analysis: 2013

[5] Keselman A, Slaughter L, Arnott-Smith C et al. Towards consumer-friendly PHRs: patients' experience with reviewing their health records. In *AMIA Annual Symposium Proceedings* (Vol. 2007, p. 399): 2007. American Medical Informatics Association.

[6] Dalianis H, Hassel M, Henriksson A et al. Stockholm EPR Corpus: A Clinical Database Used to Improve Health Care. Proceedings of the Fourth Swedish Language Technology Conference, (SLTC-2012), Lund, Sweden, October 25-26, 2012, pp. 17-18.

[7] Östling R. Stagger: an Open-Source Part of Speech Tagger for Swedish. Northern European Journal of Language Technology, 2013; 3:1–18.

[8] The Swedish Language Council (Ed.). *Svenska skrivregler* [Swedish Writing Rules]. Stockholm:Liber;2000.