

## Research Article

# Comprehension of Consumer Health Information: Classification of Linguistic Barriers. A Scoping Review

Dirk van Nimwegen<sup>1,2\*</sup>, Pascal Coorevits<sup>1</sup>, Robert Vander Stichele<sup>1</sup> and Els Lefever<sup>2</sup>

<sup>1</sup>Department of Public Health and Primary Care, University Ghent, Belgium

<sup>2</sup>Department of Translation, Interpreting and Communication, University Ghent, Belgium

## Abstract

**Background:** Patients often encounter linguistic barriers in their understanding of medical texts, particularly in discharge letters exchanged among healthcare professionals.

**Objectives:** The primary aim of this study is to establish a comprehensive classification of these linguistic barriers.

**Method:** A scoping review was conducted using the bibliographic databases Pubmed and EmBase. Articles focusing on the classification of barriers hindering layman's comprehension were included, with the exclusion of research solely reliant on readability tests.

**Data extraction:** A total of 137 key terms associated with barriers in the literature underwent rigorous analysis, involving exclusion, inclusion, concatenation, and iterative rounds of discussions with experts. Comparative analysis guided the reorganization of these key terms into a new and refined classification of barriers.

**Results:** The outcome of this process is a novel classification comprising thirty-three newly labelled barriers categorized into four classes: Lexis, Semantics, Syntax, and Coherence.

**Conclusion:** This newly developed classification serves as a foundational framework for future research endeavours focused on the automated support of patients in comprehending complex professional medical communication.

**Keywords:** Patient empowerment; Linguistics; Comprehension; Comprehension barrier; Classification; Consumer health information; Health communication; Consent forms

## Introduction

Participatory medicine calls for patients to take an active role in their healthcare. This requires that patients are provided with health information they can easily comprehend. This puts pressure on medical documents, such as referral letters, which must not only be available [1], but also really understandable or made understandable for the patient. It should be clear that lay texts have different characteristics from specialised medical reports. Several recommendations relate to vocabulary, intelligibility and structure within the text [2,3].

There is no conclusive methodology for making medical texts, more specifically referral letters and/or discharge letters, understandable to patients. Nevertheless, many attempts at simplifying medical texts, one way or another, have been described in the literature. Where some studies attempt simplification through lexical simplification [4,5], others focus on complex linguistic constructions and carried out their simplification at syntactic, lexical and semantic levels [6-9]. There also are studies that include the structure of the document in the desired simplification [10].

The need for informed consent is widely accepted in the context of clinical trials and procedures. In addition, more and more medical information is available for the patient, starting with the electronic health record, personal health record, referral and discharge letters, etc. However, the patient's understanding of the specialised materials is often hampered e.g. by medical jargon [11-13]. Shared use of medical documents by experts and patients seems promising for patient empowerment, patient-centeredness, and improved patient health outcomes, but the actual impact has not been established yet [12,14]. For this, a common approach and standardised evaluation of online health information is needed, with standardised definitions, harmonised categories and through sound measuring levels of comprehension [15,16].

A first orienting step is to classify possible barriers for patients to really understand the content of texts that concern them, but are directed at other health professionals. The question is whether or not the provision of shared (electronic) medical documents really serves the patient, when the information given by medical experts is not adequately understood. Patients who do not understand medical information provided to them, often turn for assistance to the digital media. Not only is the quality of these media heterogeneous, but there is also a problem with the attitude of information seekers. Users are not sufficiently sensitive to the quality of information; information supporting their own opinion is preferably processed; users who feel threatened focus on positive information; and vigilant users focus on negative information, which might result in cyberchondria [17]. Information processing skills, and also health-related information literacy play an important role in the comprehension of online health information by consumers or patients [18].

**Citation:** Nimwegen D, Coorevits P, Stichele RV, Lefever E. Comprehension of Consumer Health Information: Classification of Linguistic Barriers. A Scoping Review. *J Med Public Health*. 2024;5(2):1103.

**Copyright:** © 2024 Dirk van Nimwegen

**Publisher Name:** Medtext Publications LLC

**Manuscript compiled:** Apr 16<sup>th</sup>, 2024

**\*Corresponding author:** Dirk van Nimwegen, Faculty of Medicine and Health Sciences, Department of Public Health and Primary Care University Ghent, 9000 Ghent, Belgium, Tel: +32-476248083

Norman & Skinner [19] presented an overview of 'essential skills for Consumer Health' with emphasis on eHealth literacy. Their 'Lily-model' contained six 'petals': computer literacy, media literacy, science literacy, information literacy, traditional literacy, and health literacy. These six literacy types "combine to form the foundational skills required to fully optimize consumers' experiences with eHealth". A prerequisite is that all these skills are present in one individual, with a high level of competence in each of the skills, which is seldom the case [20,21].

For a good comprehension of complex medical information, efforts are needed from health care providers who produce the information and from patients who must process the information. However, much of the medical information in Electronic Medical Records (EMRs) stems from communication between experts (e.g. discharge letters from specialist to general practitioner). This communication is not intended by their authors to be read by patients. Paradoxically, such documents are increasingly accessible by patients on eHealth Platforms. Readability as represented by readability scores (such as Flesh-Kincaid, SMOG, and others) is not the same as 'understanding' or 'comprehension' [22,23]. Patients either need the skills to 'translate' the expert text, lexically, syntactically, and semantically, either needs to be supported in this process. To develop the latter approach, adequate computer linguistic support in the translation of specialised medical communication needs to be developed, and that necessitates full comprehension of the linguistic barriers to comprehension is needed.

## Objectives

The aim of this scoping review is to explore and identify the linguistic barriers to patient understanding of medical texts and to provide a classification of these barriers. This classification will be used as the basis for an extensive lexical analysis of selected words and phrases from referral letters from cardiologists to general practitioners in a use case involving patients with heart failure. Based on the conclusions of the above, the following is a first exploration of applications that can help to automate the translation process as much as possible.

## Materials and Methods

Scoping reviews have to identify and map available evidence [24]. In this scoping review evidence on the existence of barriers for patient's understanding of medical texts will be collected. The search was focused on articles containing enumerations, classifications, and taxonomies of such barriers. The search strategy was developed according to the Prisma Guideline Extension for scoping reviews [25]. The search string for both databases was composed by using a triad of key concepts in order to keep a balance between specificity and sensitivity in the search: i) Setting: e.g. electronic record and synonyms, ii) Concept: e.g. comprehension, understanding and synonyms, iii) Usability: e.g. lay-friendliness and connected concepts. The full search string for PubMed is given in Box 1 and the string for Embase in Box 2.

The following inclusion criteria were used to guide the collection of information:

- Publication language: English
- Containing an abstract
- Research articles focused on informed consent and health literacy

- Quantitative/qualitative research that focused on the understanding of patients in reading medical documents such as Electronic Patient Records, referral letters, and Electronic Health Records.
- Articles that delivered an enumeration, classification or taxonomy of barriers (as exhaustive as possible)

The exclusion criteria applied were the following:

- Studies only using readability scores of medical documents or practical barriers for laypeople such as access to electronic records, knowledge of IT, or technical problems
- Barriers concerning multilingualism of patients
- Barriers proper to ethnical conditions
- Unclear or general propositions concerning readability.

For each of the retrieved relevant studies a similarity search in PubMed and a citation analysis in the Web of Science and Google Citations was performed to complete the collection of relevant studies.

**Data extraction:** all articles were analysed on the occurrence of key terms used for barriers, and their categories (if any). Every term for each mentioned barrier was noted, together with its class term if present in the original text, and annotated with the context wherein this term was used. All this information was presented in a work file with the original context and source. Each collected term for a barrier (and its original categorisation) from this work file was entered by the first author in a new draft classification of barriers (with a new label and definition if needed), completed with a categorisation into new classes.

**Data analysis:** This draft classification was presented to two controlling experts (EL, RVS) with expertise in linguistics and medical lay language communication. The method used was based on a grounded theory approach [26]. After gathering of the data, they were separated, sorted and synthesized through qualitative coding. Coding means that labels were attached to segments of data that depict what each segment is about. This coding distills data, sorts them, and gives a handle for making comparisons with other segments of data [27]. Labels for barriers and classes were iteratively discussed with the experts and eventually renamed, in order to obtain a consensus about a new classification of barriers, most suitable for automated or semi-automated technological solutions. This process was repeated in iterative rounds using comparative analysis until a final consensus was achieved.

Though this review is based on the data provided by the selected articles, the experts were asked to add possible barriers that - in their opinion - were aspects of expert language (Figure 1).

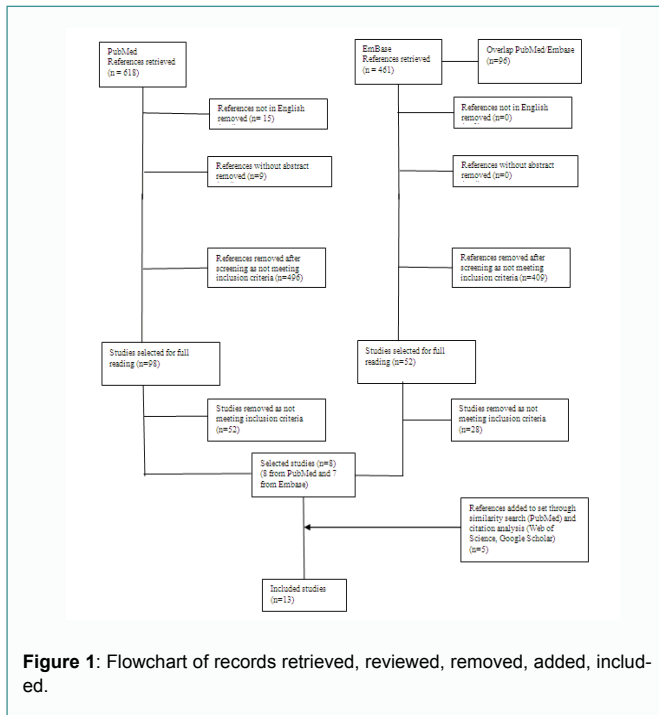
## Results

### Search results

From PubMed, 376 articles were retrieved, while 362 articles were retrieved from Embase. From both databases 7 common articles were selected and 1 additional from PubMed. Through similarity search and citation analysis, 5 more studies were added. Hence, 13 articles were included in the scoping review (Figure 1) [12,28-38].

### Data extraction

From the selected articles, 137 barriers in different contexts were identified and entered in a work file. Twenty-two barriers were



excluded as 'non applicable'. Three selected articles provided a specific classification for barriers [12,39,40], which served as the basis for a new draft classification. In Annex 1, the work file with the original 137 barriers is provided. In addition, the restructuring into a new classification is given, with the link to the original key term, its context and source.

### Data-analysis

Some barriers were identified repetitively in a number of articles, or were very similar and thus concatenated to one new barrier. After five iterative rounds with the experts, the 137 identified barriers were reduced to 33 barriers (either because disregarded or concatenated) and categorized in four classes (Lexis, Semantics, Syntax and Coherence).

### Non applicable

The 'not applicable' category contains features that in fact are barriers for patients but which cannot be considered as linguistic in nature. We disregarded 19 different key terms as shown in Annex 1. Typographic errors were excluded as well as impediments due to explicit incorrect content, wrong medicine, wrong age of the patient, test results or medication dosage reported in wrong units, etc.

### Description of the four classes

Lexis includes barriers pertaining to lexical units of vocabulary, such as words, sub-words, sub-units, compound words or multiple word units, and phrases (Table 1). The use of abbreviations and acronyms (correct and incorrect) and terminology (expert terms and foreign words) prevail in literature. Some examples: "alz" for Alzheimer disease, "postop" for post-operative, "excoriation" (for Skin Picking Disorder), and "appendicitis" for "blindedarmontsteking" (Dutch). In general, it has to do with ambiguity in meaning or lack of clarity for the layman-reader. Syntax refers to barriers pertaining to the grammatical structure of a text, i.e. the set of rules, principles, and processes that govern the structure of a sentence, including word order (Table 2).

Unexplained syntactical constructions regularly occur in the medical narrative and hamper the comprehension for the non-expert reader. In the prior category we defined ambiguity as ambiguity of meaning. Polysemic nouns or adjectives are considered as lexical. The ambiguity encountered here is considered as referential and occurs when a word or phrase can be interpreted to refer to more than one item. The experts advised to add the 'discontinuous main clause' as a syntactical barrier (in Dutch 'tang-constructie'), as it is a common aspect of expert language. It is highly linked to formal syntax (ex. "You should read all the instructions which you receive in the course of the day with the greatest attention"). Semantics is used in terms of meaning of concepts (by term or description), but where it was considered as a terminology problem, the keyword was included in the lexis category (Table 3).

Confusion is often caused by insufficient knowledge of medical concepts, misinterpretation of test results (is a positive test good or bad news?), inaccurate or ambiguous instructions. The key term 'meaning' covers mostly unexplained medical terms, insufficient knowledge of what is 'normal' for measurements and of medical shorthand. Coherence is considered as the micro- and macro-structural organization of the text, which supports the possibility of a patient to navigate safely and cohesively through the medical document (Table 4).

Coherence of a text on the macro-level can be improved by inserting a heading above every subject change in the text or a brief explanation of the relation between the topic and the previous text content. At the micro level, a narrative can gain cohesion, among other things, by linking sentences through the repetition of key words or overlap of arguments.

Finally, Table 5 lists the new classification with its four classes and the retained thirty-three barriers, relabeled if needed to form a coherent terminology, and illustrated by definitions or examples. In the final categorization of the thirty-three barriers in 4 classes, Lexis grouped nineteen barriers, linked to seventy-five mentions in the original articles. Syntax contained eleven barriers (linked to sixteen mentions), while Semantics clustered two barriers (linked to fourteen mentions) and Coherence one barrier (with ten mentions) (Table 5).

## Discussion

### Main findings

This scoping review on barriers for lay comprehension of medical texts included thirteen articles, ranging from 2002 until 2018, yielding 137 terms for barriers. These items were reclassified in an iterative process with experts to a new categorisation in 4 classes (lexis, syntax, semantics, coherence) with 33 barriers. Three of the thirteen articles were considered crucial as a basis for classifying the defined barriers. Two studies presented a taxonomy as a starting point for the task of categorizing and reformulating barriers 39 and as a tool for the documentation of the linguistic aspects of the e-record [12]. In a third article Smith et al. classified barriers according to the degree of coherence of texts [31]. All three emphasize the provisional or partial state of their classifications. It must be clear that the goal of comprehension support is not to position patients as professionals, but to enable them to work with professionals [39]. As a consequence, it was necessary to collect all observed linguistic barriers in medical texts that affect a reader's understanding.

As more articles (not subject of this review, due to exclusion arguments) emphasize, linking notes to vocabularies and other

**Table 1:** Lexis: final barriers and definitions.

Reduced key term	Definition
Abbreviations	a shortened form of a written word or phrase used in place of the whole word or phrase
Ambiguity (of meaning)	a word or expression that can be understood in two or more possible ways: an ambiguous word or
Colloquialism	Used in or characteristic of familiar and informal conversation
Multiple word units	a word consisting of components that are words
Confusing terminology	confusing technical or special terms used
Context	the parts of a discourse that surround a word or passage and can throw light on its meaning
Defining phrase	adding descriptive well-known terms to difficult new concepts
Describing phrase	unclear, or informal words in combination with formal descriptive expression
Exemplar	a typical or standard specimen
Expert terminology	technical or special terms used by experts
High formality	use of words and phrases performed in accordance with tradition or a set of rules
Informality	characteristic of or appropriate to ordinary, casual, or familiar use
Misspelling	an incorrect spelling
Misspelling (linguistic error)	a noticeable deviation from the adult grammar of a native speaker
Neologism	a new word, usage, or expression
Nominalisation	the process or result of forming a noun or noun phrase from a clause or a verb
Spelling	the forming of words from letters according to accepted usage (orthography)
Synonymy	the names that have been used to designate the same taxonomic group
Unclear qualifier	a word (such as an adjective) or word group limits or modifies the meaning of another word (such as a noun) or word group that is not clear or unidirectional
Unexplained numbers	use of numbers without any context of meaning

**Table 2:** Syntax: final barriers and definitions.

Reduced key term	Definition
Explicit causality	a clearly defined relation between a cause and its effect or between regularly correlated events or phenomena
Explicit actor	a clearly defined subject of the verb in a sentence
Sentence	a word, clause, or phrase or a group of clauses or phrases forming a syntactic unit which expresses an assertion a question a command a wish an exclamation or the performance of an action
Sentence length	the length of a sentence influences the readability
Main verb to front	putting the main verb in the beginning of sentences, so that the reader would not have to keep too much information in working memory before encountering the verb to which it should be "related."
Negation	the action or logical operation of negating or making negative
Passive voice	asserting that the grammatical subject of a verb is subjected to or affected by the action represented by that verb
Expert terminology	the technical or special terms used in a business, art, science, or special subject
No/unclear pronouns	An unclear pronoun reference occurs when it is not clear which noun or verb a pronoun refers to
Sentence connectives	connectives are used as connecting words within a sentence. They are used to indicate different purposes, such as addition sequence, consequence and or contrast reason and time
Discontinuous main clause	discontinuity is a grammatical phenomenon in which a constituent of a sentence is split into two parts due to the insertion of an element which is not part of it or the movement of a constituent to a position which is not its canonical one.

**Table 3:** Semantics: final barriers and definitions.

Reduced key term	Definition
Confusion	the act of disturbing in mind or purpose
Meaning	the thing that is conveyed especially by language

**Table 4:** Coherence: final barriers and definitions.

Reduced key term	Definition
Macrostructure	A text is coherent at the macro level when the order of topics is logical as opposed to random and if the smaller and larger sections of text are clearly related to each other and to the overall topic.
Microstructure	At the micro level, a text is coherent if each sentence is explicitly related to the next.

information sources could also improve the understandability of medical terminology. Furthermore, providing a standard information structure could help patients find the information they need [10]. The observed variability in the approach to categorization of barriers in the selected literature justifies the creation of a new classification, focusing on barriers from the viewpoint of the patient, and suitable for automated, technological solutions.

### Appropriateness of the classification

In the study of Collins-Thompson [40], all barriers were classified following the specific - slightly simplified - categories of readability feature types: lexis, syntax, discourse (renamed to coherence) and

semantics. Keselman & Smith [39] also indicate that blurring can occur at categorical borders. By focusing on four simplified categories, the risk on blurring was mitigated to some extent as suggested by Hansen and Zethsen [12], but some categories may differ in "depth", mostly in the semantics category.

In most of the selected studies the ambiguity of abbreviations was the most notable lexical comprehension barrier. Grigonyté et al. [29] call compound words (multiple word groups) containing both abbreviations and misspellings the most urgent area for future improvement. Lövestam et al. [28] presented a categorization of abbreviation types as useful for automated abbreviation systems and identified different abbreviation types. Most of the selected studies pointed to ambiguity as leading to confusion and incomprehension. "Ambiguity" appeared twice as a category: as lexical and as a syntactical barrier. We call a barrier lexically ambiguous if it can have more than one meaning, without clues from the context about which meaning to assign (ambiguity of meaning). When a sentence's ambiguity derives from its structure, we have a syntactically ambiguous claim (referential ambiguity). This arises more often as the sentence grows more complex.

### Limitations and strengths

This study is a scoping review, with the aim to clarify and

**Table 5:** Classification of barriers for patient comprehension.

Class	Barrier	Appearance
Lexis	Abbreviation	Abbreviation from general language (ex. “gr.” for “gram” or “group” or “grade”), abbreviation from medical terminology (ex. “Rp. kolpo om 4 mdr.”), abbreviations and unexplained numbers, acronyms (ex. “ICD implant”), body parts, diseases, food/meals, general non-standardized (shared abbreviations, not taking into account that they may be ambiguous), general standardized (abbreviations found in writing rules), General standardized - punctuation mistakes, (abbreviations found in writing rules with minor punctual deviations), health care professionals, Institutions/hospitals, laboratory values (ex. “HDL cholesterol”), medical assessment and intervention, names of persons or products, probable spelling mistakes ex. “HVI” for “HIV”), shortened form (ex. “Pt.” for “Patient”), units (ex. “tsp” for “teaspoon” or “5 ml” or “5 cc”)
	Ambiguity (of meaning)	ex. “WBC” for “White Blood cell” or “White Blood Cells” or “White Blood Cells Count”. ex. incorrect characters in brand name. ex. incorrect characters in abbreviation
	colloquialism	ex. “pee” for “urination”
	compound words	ex. “Cardiotoxicity”
	confusing terminology	ex. “ICV implant” for “ICD implant”
	context	ex. “cut out” (1. restriction - 2. excision)
	defining phrase	ex. “breast removal” for “masectomy”
	describing phrase	ex. “ticking bomb” for “occurs suddenly”
	exemplar	ex. “aspirin” for “analgesic”
	high formality	level of formality in terms of lexis, ex. “if you drink alcohol” for “in connection with alcohol consumption”
	informality	metaphor (ex. “we give everything we have in the drawer against encefalopati”), euphemism/analogy (ex. “faeces” / “the stomach is working”)
	Misspelling (linguistic errors)	clinical term (ex. “aterosclerosis” for “atherosclerosis”), brand name misspelling, generic name misspelling (ex. “gaulbladder” for “gallbladder”)
	neologism	ex. “ADD-lets” (ADD patients)
	nominalisation	ex. “insulinisation”
	officialese	ex. “he resumes food intake”
	synonymy	Ex. “BS/blods. [blood sugar, two different abbreviations]”
Unclear qualifiers	ex. “slightly more”	
Semantics	terminology (expert)	common expressions with an expert meaning (ex. “Therapy”), officialese (ex. “he resumes food intake”), presupposition (ex. “bone-friendly lifestyle”), doctor-ese (ex. “Postop”), incorrect explanation (ex. “pacemaker” for “ICD implant”), transformation (ex. “break” for “fracture”), unidiomatic expressions, foreign words, inconsistent terminology (“Women of childbearing age” and “fertile woman” in same record)
	unexplained numbers	ex. “HbA1c is 57”
Semantics	confusion	brand name confusion (ex. “Flomax” for “Flovent”), explanation of medical terms, misunderstandings (ex. “a disorder of glucose production” vs. “glucose metabolism”), non-sensical (ex. “trouble finding a green expectorant”), inaccuracy (ex. “return to the office in a few months”)
	meaning	information/meaning/significance on tests and results, lack of specificity of concept
Syntax	ambiguous reference	ex. “Can take freely per os and supplement with glucose 20% 35 ml/hour” (“tangconstructie” dutch) – added by experts: discontinuity is a grammatical phenomenon in which a constituent of a sentence is split into two parts due to the insertion of an element which is not part of it or the movement of a constituent to a position which is not its canonical one. (ex. “You should read all the instructions which you receive in the course of the day with the greatest attention”) Highly linked to formal syntax.
	discontinuous main clause	ex. “High-risk pt. Not ready for parent ward. Is recommended to take Protein drink”
	ellipsis	ex. “alveoli give the oxygen to the blood” for “the oxygen is given”
	explicit actor	ex. “If you come in contact with certain substance” for “People with an allergy react differently to certain substance”
	explicit causality	ex. “If you feel tired, do not drive a motor vehicle” for “carefulness is recommended in connection with car driving”
	formal syntax	moving the main verb to the front of the sentence helps readers in the sense that they quickly encounter the meaning that connects two concepts
	main verb to front	ex. “No stasis”
	negation	ex. “Må tage frit per os og supplere med Glucose 20% 35 ml/time.” [Can take freely per os and supplement with Glucose 20% 35 ml/hour]
	no pronouns	Passive to active sentence structure
	passive voice	sentence length, incomplete sentences, sentence connectives
Coherence	Sentence	Use of headings by subject change, explaining link between
	Macrostructure	Argument overlap, given-new order, redundancy, headers and topic sentences, micro and macro structure

restructure concepts of barriers for laymen in reading medical texts. Because of the empirical nature of this research building on existing literature, it is possible that specific linguistic aspects were not taken in consideration.

Each of the studies focuses on a specific aspect (e.g. terminology, abbreviation, aligning lay and expert texts, etc.), but a complete method facing all of the observed barriers does not exist yet, and a clear vision on comprehensive computational solutions is missing. The strength of this review is that it resulted in a classification that takes in account and structures all the observed barriers. Based on this classification, it should be possible to implement a thorough

analysis of discharge letters, in order to develop a generic method to ‘decode’ such texts as completely as possible for laymen. This method has to tackle the diversity in problems of syntactical, lexical, semantic and structural nature. A use case with discharge letters from newly diagnosed heart failure patients will be the next step in bridging the existing gap in the communication between the medical professionals, on the one hand, and the lay patient, on the other.

### Short discussion of possible applications and methods to overcome barriers

It was not the aim of the study to give an exhaustive view on already published work concerning applications for solving lexical,

Table 6: Selected articles.

Id	Article	Year	Cita-tions	Method	Focus	Description
1	Hansen & Zehtsen [12]	2018	51	Qualitative contextual and linguistic analysis of electronic Record	Characterization of language in e-records, with a view to potential readability	Hansen & Zehtsen [12] made an extensive qualitative analysis of contextual and lexical elements in Electronic Records. It is interesting to notice that they use readability and potential lay-friendliness interchangeably and that they considered the used e-records from the patient's perspective. Besides general comments as layout, spelling, cohesion, and coherence, the authors evaluated syntax (sentence length, density of information, voice, nominalisation and premodification), lexis (expert terminology, Latin, abbreviations, euphemisms, personal pronouns and officialese) and stylistic markers (irony, humour, metaphors, metonymy, analogy). These parameters were chosen on the basis of their significant role with regard to the ability of patients to read and understand their records.
2	Lövestam et al. [28]	2014	4	Context analysis of abbreviations in dietetic corpus	Categorization of abbreviations	Lövestam et al. [28] focused their research on abbreviations and acronyms, which are frequently used in patient records. The authors consider them as a patient safety risk as many abbreviations are ambiguous, and thus as one of the biggest comprehension barriers for patients reading their EHRs. Besides, interpreting abbreviations and acronyms can be troublesome even for different health care professionals. Their categorization and analysis of abbreviation types used (though only based on a dietetic corpus) and the frequent ambiguity of abbreviations, as well as the use of many different abbreviations for the same word, indicate the implications for patient understanding.
10	Grigonytė et al. [29]	2014		Analysis of problems with the lexical simplification of Electronic Health Records	Enumeration of elements in lexical simplification	Grigonytė et al. [29] distinguish 4 important barriers for lexical simplification: abbreviations, compound words, terminology and spelling.
3	Kvist and Velupillai [30]	2013		Quantitative corpus statistics analysis. Content analysis	Classification of problems in standard phrases in radiology reports	Kvist and Velupillai [30] characterize the clinical language used in radiology reports. They conclude that there was a need for lexical exchange for terminology, abbreviation expansion, compound splitting, and syntactic simplification
4	Keselman and Smith [39]	2012	15	Classification of errors based on content analysis of medical documents	Lay error classification scheme based on retellings	Keselman & Smith [39] introduced a taxonomy of errors in healthcare, as a tool for understanding the causes of errors in lay comprehension, and consequently developing support. This requires an in-depth understanding of the nature and causes of errors that laypeople make when comprehending clinical documents. This taxonomy is based on a content analysis of medical texts, by laypersons retelling the content of such documents. Keselman & Smith distinguish as main categories of errors: understanding of clinical concepts, clinical research, medications, devices, procedures, terminology, medical findings, and one rest category. Most common misspellings appear with the use of medication name, medication brand name, Medication Partial Memory (errors in medication brand names, which render the source name recognizable if one is familiar) and abbreviations.
5	Smith et al. [31]	2011	14	Analysis of recall of propositions based on clinical trial and visit note	Enumeration of methods improving coherence in medical texts	Smith et al. [31] confirm the principles of text coherence in their research and show that improving the coherence of typical clinical documents has a small effect on consumer comprehension. The lack of positive effect of a vocabulary-enhanced text is counterintuitive.
6	Bhavnani et al. [32]	2011	34	Questionnaire on patients' use, understanding, and impact of electronic patient records	Enumeration of difficulties with understanding reported by patients	Bhavnani et al. [32] asked patients what they thought about the information in their health records. The most commonly reported difficulties were: "understanding abbreviations, medical words and terminology and the meaning and significance of test results"
7	Kools et al. [33]	2004	17	A qualitative study of coherence in Health Education Brochures	Classification of changes made according to writing principles	Classification of writing principles to maximize text coherence. By consistently checking and adjusting a text in accordance with an objective set of coherence principles, an important part of intuitive and subjective writing of understandable texts is overcome.
8	Soergel et al. [34]	2004	52	Analysis of consumer health expressions. Analysis of expressions not mappable with UMLS	Categorization of health expressions not in agreement with professional terminology	Soergel et al. [34] distinguished categories for characterizing "non-regular" consumer health expressions, and conceptual errors, by analysing mismatches between consumer and professional concepts and forms, based on postings on web-based health discussion forums

9	Askehave and Zethsen [35]	2002	7	Analysis of translations of package inserts	Enumeration of factors that impair the quality of the translation of inserts	Askehave & Zethsen [35] considered the readability and user-friendliness of translated patient package inserts. They conclude that the author's expert status should not determine the linguistic expression of texts for the population at large and that focus must be on the receivers and their needs. They consider six global barriers for better understanding
11	Pyper et al. [36]	2004	79	In-depth patient interviews	Enumeration of major patient difficulties in understanding	Pyper et al. [36] interviewed patients of a Health Centre and noted medical terms, abbreviations, acronyms, information and test results as important problems.
12	Keselman et al. [37]	2007	98	Survey	Enumeration of experienced Comprehension barriers	Keselman et al. [37] conducted a survey regarding the comprehension and information needs related to medical records. Sections that were perceived as most difficult were physicians' and nurses' notes, radiology reports and lab test results. The qualitative analysis resulted in the conclusion that the most notable specific comprehension barriers included professional terms, abbreviations, difficult concepts, and data ordering.
13	Lee et al. [38]	2017	4	retrospective, cross-sectional review of 400 patient History and Physical examination notes, with a quantification of the presence of medical shorthand, jargon, subjective descriptors and typographic errors.	Subdivision of five major categories of interest: acronyms and abbreviations, medical jargon, subjective descriptors, mental and personal health and typographical errors	The most prevalent note characteristics across categories were General Medical Acronyms, Medical Jargon, Typographical Errors, and Lab Tests & Infectious Diseases.

syntactical and semantical barriers, but several promising initiatives are already present. Kandula et al. [41] developed already in 2010, a lexical text simplification tool, and Wang et al. [42] illustrated possible future applications. More recently, other initiatives try to deliver lexical resources, sometimes based on medical classification systems [43-45], text mining [18,46,47] or sentence alignment from comparable

corpora [48]. Zelina et al. [49] extract, label, cluster from non-English clinical notes, and semantically map segments, which has effect on the basic structure of the document, using recurrent neural networking. Frei et al. [50] use pretrained non-English datasets in order to retrieve sufficiently large datasets for training smaller and more efficient models for use-case-specific tasks. They also note that for German medical NLP, the field has made notable advanced in terms of available datasets. Ondov et al. [51] distinct two kind of tools, the ones chiefly procedural, the others chiefly neural. The limits of artificial intelligence are far from being reached today [52,53], but to offer a complete syntactical, lexical, semantic and structural reprocessing of a discharge letter (in compliance with the classification of barriers), will continue to pose a major challenge to developers (Table 6).

## Conclusion

This scoping review provides the basis for technical and lexical remediation to make referral letters from medical specialists to general practitioners more accessible and comprehensive for lay people. With this scoping review a new classification was created, not to summarize earlier stated barriers, but to support future research in applications to remediate barriers.

## References

- Åhlfeldt H, Grabar N, Hallett C, Hardcastle D, Kakkinakis D, Mancini C, et al. Literature Review on Patient-Friendly Documentation Systems. Milton Keynes, UK: Centre for Research in Computing, The Open University; 2006.
- Singh N. Writing lay summaries what medical writers need to know. *Med Writ*. 2018;27(2):49-54.
- Clinical Trials Expert Group. Good Lay Summary Practice. *Eur Forum Good Clin Pract*. 2021:1-85.
- Silpa KS, Irshad MA. Survey of lexical simplification. In: Vanchipura R, Jiji KS, editors. *Emerging Trends in Engineering, Science and Technology for Society, Energy and Environment*. United States: CRC press; 2018.
- Leroy G, Endicott JE, Kauchak D, Mouradi O, Just M. User evaluation of the effects of a text simplification algorithm using term familiarity on perception, understanding, learning, and information retention. *J Med Internet Res*. 2013;15(7):e144.
- Chmielik J, Grabar N. Comparative study between expert and non-expert biomedical writings: Their morphology and semantics. *Stud Health Technol Inform*. 2009;150:359-63.
- Cohen T, Widdows D. Embedding of semantic predications. *J Biomed Inform*. 2017;68:150-66.
- Miller T, Leroy G, Chatterjee S, Jie F, Thoms B. A classifier to evaluate language specificity of medical documents. *Proc Annu Hawaii Int Conf Syst Sci*. 2007:6.
- Pylieva H, Chernodub A, Grabar N, Hamon T. Generalizability of readability models for medical terms. *Stud Health Technol Inform*. 2019;264:1327-31.
- Kujala S, Hörhammer I, Väyrynen A, Holmroos M, Nättiäho-Rönnholm M, Häggglund M, et al. Patients' Experiences of Web-Based Access to Electronic Health Records in Finland: Cross-sectional Survey. *J Med Internet Res*. 2022;24(6):e37438.
- Zeng QT, Tse T, Crowell J, Divita G, Roth L, Browne AC. Identifying Consumer-Friendly Display (CFD) Names for Health Concepts. *AMIA Annu Symp Proc*. 2005;2005:859-63.
- Hansen MM, Zethsen KK. Is Electronic Access to Medical Records an Empowering and Patient-centered Initiative? - A Qualitative Contextual and Linguistic Analysis of Danish Electronic Record. *Hermes (Denmark)*. 2018;58:157-73.
- Messari R, Simonet M, Mousseau M. Bridging the Gap between Lay Language and Medical Terminology. Conference: Proceedings of the 2008 International Conference on Information & Knowledge Engineering: USA; 2008.
- Archer N, Fevrier-Thomasm U, Lokker C, McKibbin KA, Straus SE. Personal health records: A scoping review. *J Am Med Informatics Assoc*. 2011;18(4):515-22.
- Cardillo E, Hernandez G, Bodenreider O. Integrating consumer-oriented vocabularies with selected professional ones from the UMLS using semantic web technologies. *Lect Notes Inst Comput Sci Soc Telecommun Eng*. 2011;69:54-61.
- Diviani N, Van Den Putte B, Giani S, Van Weert JCM. Low health literacy and evaluation of online health information: A systematic review of the literature. *J Med Internet Res*. 2015;17(5):e112.

17. Chen J, Druhl E, Polepalli Ramesh B, Houston TK, Brandt CA, Zulman DM, et al. A natural language processing system that links medical terms in electronic health record notes to lay definitions: System development using physician reviews. *J Med Internet Res*. 2018;20(1):e26.
18. Chen J, Jagannatha AN, Fodehm SJ, Yu H. Ranking medical terms to support expansion of lay language resources for patient comprehension of electronic health record notes: Adapted distant supervision approach. *JMIR Med Inform*. 2017;5(4):e42.
19. Norman CD, Skinner HA. eHealth literacy: Essential skills for consumer health in a networked world. *J Med Internet Res*. 2006;8(2):e9.
20. Hemsley B, Rollo M, Georgiou A, Balandin S, Hill S. The health literacy demands of electronic personal health records (e-PHRs): An integrative review to inform future inclusive research. *Patient Educ Couns*. 2018;101(1):2-15.
21. Moser DK, Robinson S, Biddle MJ, Pelter MM, Nesbitt TS, Southard J, et al. Health Literacy Predicts Morbidity and Mortality in Rural Patients with Heart Failure. *J Card Fail*. 2015;21(8):612-8.
22. Bossert S, Strech D. An integrated conceptual framework for evaluating and improving 'understanding' in informed consent. *Trials*. 2017;18(1):482.
23. Zheng J, Yu H. Readability Formulas and User Perceptions of Electronic Health Records Difficulty: A Corpus Study. *J Med Internet Res*. 2017;19(3):e59.
24. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E, et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol*. 2018;18(1):143.
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Ann Intern Med*. 2018;169(7):467-73.
26. Chun Tie Y, Birks M, Francis K. Grounded theory research: A design framework for novice researchers. *SAGE Open Med*. 2019;7:2050312118822927.
27. Charmaz K. *Constructing Grounded Theory A Practical Guide Through Qualitative Analysis*. Sage Publications. 2006.
28. Lövestam E, Velupillai S, Kvist M. Abbreviations in Swedish Clinical Text--use by three professions. *Stud Health Technol Inform*. 2014;205:720-4.
29. Grigonytem G, Kvist M, Velupillai S, Wirén M. Improving Readability of Swedish Electronic Health Records through Lexical Simplification: First Results. In: *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*. Gothenburg, Sweden: Association for Computational Linguistics; 2014.p.74-83.
30. Kvist M, Velupillai S. Professional language in swedish radiology reports--characterization for patient-adapted text simplification. *Scand Conf Heal Inform*. 2013.
31. Smith CA, Hetzel S, Dalrymple P, Keselman A. Beyond Readability: Investigating Coherence of Clinical Text for Consumers. *J Med Internet Res*. 2011;13(4):e104.
32. Bhavnani V, Fisher B, Winfield M, Seed P. How patients use access to their electronic GP record--a quantitative study. *Fam Pract*. 2011;28(2):188-94.
33. Kools M, Ruiters RAC, Van De Wiel MWJ, Kok G. Increasing readers' comprehension of health education brochures: A qualitative study into how professional writers make texts coherent. *Health Educ Behav*. 2004;31(6):720-40.
34. Soergel D, Tse T, Slaughter L. Helping Healthcare Consumers Understand: An 'Interpretative Layer' for Finding and Making Sense of Medical Information. *Stud Health Technol Inform*. 2004;107(Pt 2):931-5.
35. Ashehave I, Zethsen KK. Translating for laymen. *Perspect Stud Transl*. 2002;10(1):15-29.
36. Pyper C, Amery J, Watson M, Crook C. Patients' experiences when accessing their on-line electronic patient records in primary care. *Br J Gen Pract*. 2004;54(498):38-43.
37. Keselman A, Slaughter L, Arnott-Smith C, Kim H, Divita G, Browne A, et al. Towards consumer-friendly PHRs: patients' experience with reviewing their health records. *AMIA Annu Symp Proc*. 2007;2007:399-403.
38. Lee EH, Patel JP, Fortin AH 6th. Patient-centric medical notes: Identifying areas for improvement in the age of open medical records. *Patient Educ Couns*. 2017;100(8):1608-11.
39. Keselman A, Smith CA. Classification of errors in lay comprehension of medical documents. *J Biomed Inform*. 2012;45(6):1151-63.
40. Collins-Thompson K. Computational assessment of text readability: A survey of current and future research. *ITL - Int J Appl Linguist*. 2014;165(2):97-135.
41. Kandula S, Curtis D, Zeng-Treitler Q. Semantic and syntactic text simplification tool for health content. *AMIA Annu Symp Proc*. 2010;2010:366-70.
42. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. *J Biomed Inform*. 2018;77:34-49.
43. Fellbaum C, Hahn U, Smith B. Towards new information resources for public health--from WordNet to Medical Word Net. *J Biomed Inform*. 2006;39(3):321-32.
44. van Mens HJT, van Eysden MM, Nienhuis R, van Delden JJM, de Keizer NF, Cornet R. Evaluation of lexical clarification by patients reading their clinical notes: a quasi-experimental interview study. *BMC Med Inform Decis Mak*. 2020;20:278.
45. van Mens HJT, Martens SSM, Paiman EHM, Mertens AC, Nienhuis R, de Keizer NF, et al. Diagnosis clarification by generalization to patient-friendly terms and definitions: Validation study. *J Biomed Inform*. 2022;129:104071.
46. Elhadad N, Sutaria K. Mining a lexicon of technical terms and lay equivalents. 2007:49-56.
47. Grabar N, Hamon T. Automatic Extraction of Layman Names for Technical Medical Terms. *IEEE Int Conf Healthc Informatics*. 2014:310-19.
48. Cardon R, Grabar N. Parallel sentence retrieval from comparable corpora for biomedical text simplification. *Int Conf Recent Adv Nat Lang Process*. RANLP.2019:168-77.
49. Zelina P, Halamkova J, Novacek V. Extraction, labelling, clustering, and semantic mapping of segments from clinical notes. *IEEE Trans Nanobioscience*. 2023;22(4):781-88.
50. Frei J, Kramer F. Annotated dataset creation through large language models for non-english medical NLP. *J Biomed Inform*. 2023;145:104478.
51. Ondov B, Attal K, Demner-Fushman D. A survey of automated methods for biomedical text simplification. *J Am Med Inform Assoc*. 2022;29(11):1976-88.
52. Zernikow J, Grassow L, Gröschel J, Henrion P, Wetzel PJ, Wetzel PJ. [Clinical application of large language models: Does ChatGPT replace medical report formulation? An experience reports]. *Inn Med (Heidelb)*. 2023;64(11):1058-64.
53. Kugic A, Kreuzthaler M, Schulz S. Clinical Acronym Disambiguation via ChatGPT and BING. *Stud Health Technol Inform*. 2023;309:78-82.



**Box 1: PubMed search string**

(health records, personal [MeSH Terms] OR "medical record"[All Fields] OR "medical records" [MeSH Major Topic] OR health records, electronic [MeSH Terms] OR "record access" [All Fields] OR "clinical text" [All Fields] OR "EHR" [All Fields] OR "PHR" [All Fields] OR information services [MeSH Major Topic])

**AND**

(comprehension [MeSH Terms] OR understanding [All Fields] OR understanding [Mesh Terms] OR terminology as topic [MeSH Terms] OR vocabulary [TIAB] OR health knowledge, attitudes, practice [MeSH Terms])

**AND**

(lay-friendliness [Title/Abstract] OR layman [All Fields] OR laymen [All Fields] OR lay [All Fields] OR "consumer-friendly" [All Fields] OR "patients/statistics and numerical data" [MeSH Major Topic] OR "patient-centered" [All Fields] OR ambiguity [Title/Abstract] OR readability [Title/Abstract] OR "professional language" [All Fields] OR "expert language"[All Fields] OR jargon [All Fields] OR "patient empowerment" [All Fields] OR "patients' experiences" [All Fields])

**Box 2: EmBase search string**

('health records' OR 'medical record'/exp OR 'medical record' OR 'health record' OR ('health record' AND 'electronic') OR 'record access' OR 'clinical text' OR 'ehr' OR 'phr' OR 'information services'/exp OR 'information services')

**AND**

('comprehension'/exp OR 'comprehension' OR 'understanding'/exp OR 'understanding' OR 'terminology'/exp OR 'terminology' OR 'nomenclature'/exp OR 'nomenclature' OR 'health knowledge'/exp OR 'health knowledge')

**AND**

('lay-friendliness' OR 'layman'/exp OR 'layman' OR 'laymen'/exp OR 'laymen' OR 'lay' OR 'consumer-friendly' OR 'patient centered communication'/exp OR 'patient centered communication' OR 'patient centered outcomes research'/exp OR 'patient centered outcomes research' OR 'patient centeredness'/exp OR 'patient centeredness' OR 'ambiguity'/exp OR 'ambiguity' OR 'readability'/exp OR 'readability' OR 'professional language' OR 'expert language' OR 'jargon' OR 'patient empowerment'/exp OR 'patient empowerment' OR 'patients experiences')

**AND**

[english]/lim AND [abstracts]/lim AND ([embase]/lim OR [medline]/lim OR [pubmed-not-medline]/lim)

## Annex 1

Id	Key term in article	Detail in article	Context in article	Class-detail proposition	Class_head proposition	Article_id	Class Def.	Barrier
1	General comments	length of record	The sheer length of the records is likely to contribute negatively to readability. The longest records consisted of 268 pages, which is a rather large number of pages to navigate, especially if the patient in question is not a strong reader (Dickinson et al. 2001: 154, Askehave/Zethsen 2000a: 38). Much information is simply not relevant for the patient and the level of redundancy is high.	redundancy	Coherence	1	Syntax	sentence Length
2	General comments	high redundancy	Much information is simply not relevant for the patient and the level of redundancy is high.	redundancy	Coherence	1	Coherence	microstructure
3	General comments	clear and logical structure of e-record	Generally, the e-records have a clear and logical structure, which is likely to have a positive influence on lay-friendliness (Keselman et al. 2007: 402, Jensen 2013: 99-102, Dickinson et al. 2001: 156).	structure	Coherence	1	Coherence	macrostructure
4	General comments	headlines are dominated by expert language	However, the many headlines are dominated by expert language and the headlines do not always cover the actual content of the section	structure	Coherence	1	Lexis	expert terminology
5	General comments	linguistic errors	The records are furthermore characterised by numerous linguistic errors, typically spelling mistakes which could, in some cases, result in comprehension problems (Jensen 2013: 117-118, Göpferich 2009: 42-43)	spelling	Lexis	1	Lexis	spelling
6	General comments	inconsistent font or pitch	Use of a particular font or pitch is not consistent; neither is the use of capital letters and punctuation. This could potentially confuse the patient as it is, for example, difficult to assess which passages should be emphasised (Jensen 2013: 112-113, Göpferich 2009: 40-42).	fonts	Typo	1	not applicable	not applicable
7	Syntax	sentence length	Sentences are mostly either very long or very short, consisting of one word only. Long sentences are not necessarily problematic from a readability point of view, but when coupled with other expert features, the typically long and complex sentences must be presumed to hamper understanding. When sentences become very long (there are examples of seven-line sentences in the records), they are likely to be problematic solely on the grounds of the amount of information included. As regards the very short sentences, they are in fact typical for expert language as well: "håndkøbsmedicin" [over-the-counter medicine] The reader needs to be able to interpret this compound noun which forms a one-word sentence (Jensen 2013: 66-88, Askehave/Zethsen 2000b: 71-72). In this case, the context shows that the patient can be treated for pain with over-the-counter medication instead of prescription painkillers.	sentence length	Syntax	1	Syntax	sentence
8	Syntax	voice	The passive voice is the norm in the e-records, thus making it harder for the reader to find out who the agent is (Jensen 2013: 48-52, Göpferich 2009: 46-48, Askehave/Zethsen 2000b: 71-73): "Der skal holdes gang i maven" [the stomach should be kept in working order]	passive voice	Syntax	1	Syntax	passive voice
9	Syntax	nominalisation	The records contain many nominalisations which, like the passive voice, hide the agent of the sentence. Nominalisations are not necessarily problematic if they are commonly used in ordinary language, but when they are infrequent, and even sound rather artificial to the non-expert as is the case in the records, they are typically harder to process: "Selvkategorisation" [Selfcatheterisation] "Insulinerings" [Insulinisation] "Elevation" [Elevation] "Depression" [Depression] The two latter examples are in the context of elevation or depression of the shoulder. These nominalisations are not common in Danish and especially "depression" may be confused with the common noun denoting a mental illness.	nominalisation	Syntax	1	Lexis	nominalisation

10	Syntax	premodification	The e-records contain a very large number of premodifications. These can be challenging to most lay readers as it is often difficult to be sure what modifies what when expert knowledge is not present. The Danish language exacerbates the problem because of an orthographic tendency to write words together (Jensen 2013: 63-64, Askehave/ Zethsen 2000b: 72-73): "transplantations during blod prøver" [transplantation crossmatch blood test] "den efter hånden ret komplekst sammensatte antiencephalopati- behandling". [The eventually rather complexly constructed antiencephalopathy-treatment]. Adding to the complexity is the fact that the premodifications themselves often contain other expert language characteristics such as expert terms, nominalisations and officialese.	officialese	Syntax	1	Syntax	expert terminology
11	Syntax	ellipsis	Especially in the moves about future plans, ellipsis is very common and sentences are short and condensed: "Højrisiko pt. Ikke klar til stam. afd. Tilrådes at indtage Protein drik". [High-risk pt. Not ready for parent ward. Is recommended to take Protein drink] It is thus up to the patients to guess the verbs and pronouns left out.	ellipsis	Syntax	1	Syntax	ellipsis
12	Lexis	expert terminology	Practically all sentences in the e-records are dominated by expert terminology mostly of Latin/ Greek origin. In contrast to English, medical Danish often has a Latin-based expert term and a Danish lay term for the same phenomenon (appendicitis is, for example, called appendicitis in Danish expert language, but blindtarmsbetændelse when mentioned in layman terms. The expert term is simply never used by non-experts. This means that Danes in general do not have the same Latin-based medical vocabulary as English-speaking people (Zethsen 2004: 125-142, Askehave/ Zethsen 2000b: 69-70). These are just a few of the many examples of expert terminology": "ekskoriation" [excoriation] "Terapeutisk acitespunktur" [Therapeutic ascites puncture] "Orienteret x 3. ABC-stabil." [orientated x 3. ABC- stable] "biomekanisk bevægelsesterapi" [biomechanical movement therapy] "Smertedækket med Epi. Breiviks blanding. 12ml". [pain-relieved with Epi. Breivik's mixture. 12 ml.] "pater" [The Latin "pater" referring to a father]	expert terminology	Lexis	1	Lexis	expert terminology
13	Lexis	common expressions with an expert meaning	The word 'therapy' has clear psychological connotations in Danish and is not usually applied to any form of physical treatment. "Dettes gøres under dække af 10 ml lidokain." [this is it simply informs the reader of the anaesthetic used, but to the non-expert it sounds rather dubious. Generally speaking, very few examples or analogies are used to explain the expert content of the records. The text is much more abstract than concrete, which is especially problematic in the moves which set out plans for the future – plans which may actively involve the patient (Jensen 2013: 65-66 and 85-87). done under cover of 10 ml lidokain]	expert terminology	Lexis	1	Lexis	expert terminology

14	Lexis	abbreviations and unexplained numbers	The many abbreviations in the e-records may cause readability problems as the reader is unable to decipher them (Short 1986: 1317, Bhavnani et al. 2010: 3, Göpferich 2009: 46-47)”: Rp. kolpo om 4 mdt.” [the three abbreviations here mean “recipie colposcopy in 4 months”] “CS” [The abbreviation “CS” stands for the English term” cervical screening” which is an added complication. The English expression is very likely not understandable to most Danes even if they were able to write out the abbreviation]” Pt. [“patient” is abbreviated to “pt” which in ordinary Danish is the abbreviation for the Latin expression “pro tempore” meaning “for the time being”. Only once in 689 pages is “patient” spelled out. Numbers are rarely explained and may therefore be nonsensical to patients. An exception is the following example (Keselman et al. 2007: 402, Jensen 2013: 85-88)”: Pt.s HbA1c er 57, og det er også ganske fint for en diabetespt.” [Pt’s HbA1c is 57, and this is actually quite good for a diabetes patient]	abbreviations	Lexis	1	Lexis	abbreviations
15	Lexis	abbreviations an unexplained number	The many abbreviations in the e-records may cause readability problems as the reader is unable to decipher them (Short 1986: 1317, Bhavnani et al. 2010: 3, Göpferich 2009: 46-47)”: Rp. kolpo om 4 mdt.” [the three abbreviations here mean “recipie colposcopy in 4 months”] “CS” [The abbreviation “CS” stands for the English term” cervical screening” which is an added complication. The English expression is very likely not understandable to most Danes even if they were able to write out the abbreviation]” Pt. [“patient” is abbreviated to “pt” which in ordinary Danish is the abbreviation for the Latin expression “pro tempore” meaning “for the time being”. Only once in 689 pages is “patient” spelled out. Numbers are rarely explained and may therefore be nonsensical to patients. An exception is the following example (Keselman et al. 2007: 402, Jensen 2013: 85-88)”: Pt.s HbA1c er 57, og det er også ganske fint for en diabetespt.” [Pt’s HbA1c is 57, and this is actually quite good for a diabetes patient]	unexplained numbers	Lexis	1	Lexis	abbreviations
16	Lexis	abbreviation from general language	Abbreviations were found to be of two kinds; abbreviations from general language and abbreviated medical terminology. The expansion of clinical abbreviations is not a trivial task and will require domain-adapted Natural Language Processing (NLP) tools, preferable context aware for disambiguation.	abbreviation	Lexis	1	Lexis	abbreviations
17	Lexis	abbreviation from medical terminology	Abbreviations were found to be of two kinds; abbreviations from general language and abbreviated medical terminology. The expansion of clinical abbreviations is not a trivial task and will require domain-adapted Natural Language Processing (NLP) tools, preferable context aware for disambiguation.	abbreviation	Lexis	1	Lexis	abbreviations
18	Lexis	officialese	“Patienten informeret om telefonisk svar fra sygeplejerske om 2 uger”. [patient informed of telephonic answer from nurse in 2 weeks]” Han genoptager fødeindtagelsen” [he resumes food intake]	expert terminology	Lexis	1	Lexis	expert terminology

19	Lexis	synonymy	Synonymy, especially in the form of using both the Latin and the Danish terminology interchangeably and in the form of inconsistent abbreviations, is common in the records": Hernie/brok" [hernia from Latin and layman Danish, respectively] BS/blods. [blood sugar, two different abbreviations]. Synonymy is generally not advised in texts directed at lay people as the phenomenon is likely to create confusion. The layperson may simply not know that two expressions stand for the same, as in the case with the two expressions for "hernia", and they may assume that a second condition is involved (Jensen 2013: 56-62, Askehave/Zethsen 2010: 106).	synonymy	Coherence	1	Lexis	synonymy
20	Lexis	presupposition	Expert terminology, abbreviations, unexplained numbers in themselves involve the presupposition that the reader is familiar with the terms and expressions and will understand the content. Otherwise, the communication would be meaningless. However, there is also another kind of pre-supposition in the e-records involving expressions which are not difficult to understand as such, but where the lay reader may not know what the expressions entail (Jensen 2013: 64-66)": Vi tager nye blodprøver i dag, for at se om han er på vej den forkerte vej". [we will take new blood samples today to see if he is taking the wrong direction] In this example, the reader may not know what "the wrong direction" actually indicates": Forslag til evt. yderligere udredning og behandling: - Knoglevenlig livsstil" [Proposal for possible further investigation and treatment: - bone-friendly lifestyle] In this example, it is presupposed that the reader knows what a "bone-friendly lifestyle" entail". Henvist [til gynækologisk ambulatorium] fra egen læge pga ASCUS. Kendt med psoriasis." [Referred [to gynecological outpatient clinic] from own doctor due to ASCUS. Known with psoriasis.] In this rather complex example, it is presupposed that the reader understands the abbreviation ASCUS (a form of abnormal cells), but also the medical connection between the presence of abnormal cells in the cervix and the skin disease psoriasis.	expert terminology	Lexis	1	Lexis	expert terminology
21	Lexis	pronouns	These are frequently left out with the same effect as the passive voice and nominalisation. The agent is hidden and the sentence becomes more impersonal (Göpferich 2009: 43-44, Askehave/Zethsen 2000b: 71-72): "Må tage frit per os og supplere med Glucose 20% 35 ml/time." [Can take freely per os and supplement with Glucose 20% 35 ml/hour]	no pronouns	Syntax	1	Syntax	no pronouns
22	Stylistic markers	metaphor/euphemism/ analogies	In connection with the e-records, stylistic markers are notable through their absence. There are a few metaphors like: "Vi giver alt hvad vi har i skuffen imod encefalopati." [we give everything we have in the drawer against encefalopati] but generally the text is very formal and does not contain metaphors, wordplay, and so on. This absence is of course characteristic for expert language, though metaphors and analogies could be used to make difficult content more accessible to the layman reader. The text is not characterised by euphemisms with the exception of the mention of bodily functions. In the cases where the euphemism is created by means of Latin": faeces" [faeces]. It may not be understood by the average patient, whereas the euphemisms which make use of everyday expressions are much easier for the patient to understand (Jensen 2013: 61-62)": der er gang i maven" [the stomach is working]	metaphor	Style	1	Lexis	informality

22	Stylistic markers	metaphor/euphemism/ analogies	In connection with the e-records, stylistic markers are notable through their absence. There are a few metaphors like: "Vi giver alt hvad vi har i skuffen imod encefalopati." [we give everything we have in the drawer against encefalopati] but generally the text is very formal and does not contain metaphors, wordplay, and so on. This absence is of course characteristic for expert language, though metaphors and analogies could be used to make difficult content more accessible to the layman reader. The text is not characterised by euphemisms with the exception of the mention of bodily functions. In the cases where the euphemism is created by means of Latin": faeces" [faeces]. It may not be understood by the average patient, whereas the euphemisms which make use of ev- eryday expressions are much easier for the patient to understand (Jensen 2013: 61-62)": der er gang i maven" [the stomach is working]	euphemism	Style	1	Lexis	informality
22	Stylistic markers	metaphor/euphemism/ analogies	In connection with the e-records, stylistic markers are notable through their absence. There are a few metaphors like: "Vi giver alt hvad vi har i skuffen imod encefalopati." [we give everything we have in the drawer against encefalopati] but generally the text is very formal and does not contain metaphors, wordplay, and so on. This absence is of course characteristic for expert language, though metaphors and analogies could be used to make difficult content more accessible to the layman reader. The text is not characterised by euphemisms with the exception of the mention of bodily functions. In the cases where the euphemism is created by means of Latin": faeces" [faeces]. It may not be understood by the average patient, whereas the euphemisms which make use of ev- eryday expressions are much easier for the patient to understand (Jensen 2013: 61-62)": der er gang i maven" [the stomach is working]	analogy	Style	1	Lexis	informality
23	Clinical concepts	incorrect explanation	Errors in this category involved incorrect explanations of some clinical concepts from the original texts. Many incorrect explanations concerned the nature of diabetes and the function of the liver, and the pertinence of liver functioning to dia-betes (glycogen metabolism). For example, with respect to explaining the nature of diabetes, two participants stated that it was "a disorder of glucose production," rather than of glucose metabolism. Another participant wrote that "some diabetics have a liver disease," while another suggested that diabetes damages the liver. Several stated that the function of the liver is to produce either glucose or insulin, and that it is this function that is impaired in diabetes. A related, and probably underlying, error involved explanation of various chemical substances involved in glucose metabolism. Clinical concept, incorrectly explained (CC) - Incorrect explanation of a disease mechanism or biological process or concept: "insulin is an enzyme"; "diabetes is a disease where the liver can't produce a mechanism or biological process or concept"	confusion	Semantics	4	Semantics	confusion

24	Clinical research	misunderstandings	Errors in this category involved incorrect explanations of some clinical concepts from the original texts. Many incorrect explanations concerned the nature of diabetes and the function of the liver, and the pertinence of liver functioning to diabetes (glycogen metabolism). For example, with respect to explaining the nature of diabetes, two participants stated that it was “a disorder of glucose production,” rather than of glucose metabolism. Another participant wrote that “some diabetics have a liver disease,” while another suggested that diabetes damages the liver. Several stated that the function of the liver is to produce either glucose or insulin, and that it is this function that is impaired in diabetes. A related, and probably underlying, error involved explanation of various chemical substances involved in glucose metabolism. Several participants seemed to confuse glucose and glycogen with insulin and did not understand that glycogen was a form of glucose, rather than a separate substance involved in glucose metabolism. This is exemplified by the following statement from one of the participants, “A normal body will after a meal store extra insulin in the muscles and liver as glycogen.” Another participant wrote that “glucose breaks down consumed calories.” Yet another wrote that in diabetics “sugar drops after eating.” Other errors involved nuanced variations on the ones described above, such as stating that conversion of glucose to glycogen is done by “blood cells.”	confusion	Semantics	4	Semantics	confusion
25	Medication	incorrect medication nature	Medication Nature, Incorrect (MedNI) – Ascribing an incorrect function to a medication; calling a non-medicinal substance a medicine “Coumadin is a sleep medication”; “2H2O is a medicine”	confusion	semantics	4	Semantics	confusion
26	Medication	generic name misspelling	Medication, generic name misspelling (MedGMiss) – Misspelling of a recognizable generic drug name “acetaminophin”	confusion	semantics	4	Semantics	confusion
27	Medication	non-prescribed medication type	Non-prescribed medication type, generic name, or form, (2) a medication purpose not corresponding to any medications in the documents or (3) a name that looks like a brand name, but isn't, and cannot be related to anything in the text “antibiotic,” “aspirin,” “some nasal spray,” “medication for liver functioning,” “Devton”	confusion	semantics	4	Semantics	confusion
28	Medication	brand name misspelling	A spelling error in a brand name that can be corrected by replacing or deleting one letter or switching two letters with each other “ambion” or “ambient” for “ambien”; “Coudamin” for “Coumadin”	confusion	semantics	4	Semantics	confusion
29	Medication	brand name confusion	“Swapping” two existing medications with each other; brand not mentioned in the document, but it does exist “Flomax” for “Flovent”	confusion	semantics	4	Semantics	confusion
30	Medication	partial memory (MedPM)	More than 1 character is incorrect, but the original brand name mentioned in the document is generally recognizable or can be inferred “Courdin” for “Coumadin,” “Landin” for “Lantis”	ambiguity	Lexis	4	Lexis	ambiguity
31	Medication units	incorrect	Test results or medication dosage reported in the wrong units “Respiration 18 beats per minute” “Respiration 18 beats per minute”	ambiguity	Content	4	not applicable	not applicable
32	Medication regimen	incorrect	Medication regimen/schedule/ dosage reported incorrectly: “1000 mg” for “10 mg” - ‘as needed’ instead of “daily”	ambiguity	Content	4	not applicable	not applicable
33	Devices	incorrect explanation (DevIE)	Using incorrect device name or misrepresenting its general purpose “pacemaker” for “ICD implant,” “spectroscopy machine for taking X-Rays”	terminology	Lexis	4	Lexis	confusing terminology
34	Procedures	incorrect explanation	Incorrect name or purpose, specific steps, or mechanism of a procedure “detects low level of blood glucose” instead of “measures glycogen metabolism in the liver;” “this method involves pairing a complex device with a painkiller;”	inaccuracy	Content	4	not applicable	not applicable
35	Terminology	clinical term, misspelling	Misspelling of medical and health-related terms “atherosclerosis” for “atherosclerosis”	misspelling	Lexis	4	Lexis	misspelling

36	Terminology	terminology confusion	Substituting a specialized medical term with another medical or general term or a non-word that is similar to another medical or general term, confusing a medical term with a medication name, or making an error in an abbreviation or acronym; applying an inappropriate medical term to a contextually relevant description: "diabetic phrenopathy" for "diabetic nephropathy," "Lasik" for "Lasix", ICD implant" for "ICD implant", "too much sugar in the blood is called "hypoglycemia"	confusion	Lexis	4	Lexis	confusing terminology
37	Findings	inaccurately reported	A finding is a clinically significant observation or measure (or set of observations or measures), potentially indicative of an underlying medical problem. Misrepresenting some characteristic of a finding reported in the Visit Note, but clinically possible and the original source is clear "hard abdomen" instead of "soft abdomen"	confusion	semantics	4	not applicable	not applicable
38	Findings	non-existent	Clinically possible, but no clear source in the Visit Note: "frequent pain"	confusion	semantics	4	not applicable	not applicable
39	Findings	non-sensical	Clinically impossible finding: "trouble breathing a green expectorant"	confusion	semantics	4	not applicable	not applicable
40	Diagnosis	wrong	Attributing as a diagnosis a disease or a condition, not mentioned in the original document: "patient suffers from neuropathy"	confusion	semantics	4	not applicable	not applicable
41	Diagnosis	inaccurately reported	Misrepresenting some characteristics of a diagnosis mentioned in the document: "acute diabetes" instead of "Diabetes Mellitus"	terminology	Lexis	4	not applicable	not applicable
42	Diagnosis	non-sensical	Clinically impossible/meaningless diagnosis: "ischemic lung disease"	terminology	Lexis	4		not applicable
43	Other	non-existent direction	Inaccurate recall of instructions regarding self-care and follow up visit: "return to the office in a few months"	ambiguity	Content	4	Semantics	confusion
44	Other	demographics, inaccurately reported	Incorrect report of patient's age/race: "60-years-old man"	ambiguity	Content	4	not applicable	not applicable
45	Other	patient's circumstances, inaccurately reported	Inaccuracy in reporting the details of the patient's knowledge or time of treatment and visits: "patient was on a medication, but was not sure what it was treating"	inaccuracy	Content	4	not applicable	not applicable
46	Abbreviations	general non-standardized	Not classified in any of the specific categories below, non-standardized but common in everyday Swedish	abbreviation	Lexis	2	Lexis	abbreviations
47	Abbreviations	general standardized	Found in the Swedish Writing Rules	abbreviation	Lexis	2	Lexis	abbreviations
48	Abbreviations	general standardized, punctuation mistakes	Found in the SWR, with minor punctual deviations	abbreviation	Lexis	2	Lexis	abbreviations
49	Abbreviations	food/meals	Meals, food and cooking	abbreviation	Lexis	2	Lexis	abbreviations
50	Abbreviations	medical assessment and intervention	Assessment methods and scales, treatments, medicines	abbreviation	Lexis	2	Lexis	abbreviations
51	Abbreviations	diseases	Diseases	abbreviation	Lexis	2	Lexis	abbreviations
52	Abbreviations	units	Length, volume and other units	abbreviation	Lexis	2	Lexis	abbreviations
53	Abbreviations	institutions/hospitals	Names of institutions	abbreviation	Lexis	2	Lexis	abbreviations
54	Abbreviations	laboratory values	Blood lipids, iron status etc.,	abbreviation	Lexis	2	Lexis	abbreviations
55	Abbreviations	health care professionals	Professions within health care	abbreviation	Lexis	2	Lexis	abbreviations
56	Abbreviations	body parts	Organs and other body parts	abbreviation	Lexis	2	Lexis	abbreviations
57	Abbreviations	names of persons or products	Names of persons or products	abbreviation	Lexis	2	Lexis	abbreviations
58	Abbreviations	probable spelling mistakes	No identified meaning, interpreted as spelling mistake	abbreviation	Lexis	2	Lexis	abbreviations
59	Terminology	foreign words	Historically, medical terminology origins from both Latin and Greek. The Latin terms generally denote body parts while Greek, the language of pathology, give rise to diagnostic terms as well as names for different medical specialties [12]. Nowadays, English expressions are gaining influence on Swedish health records since this is the vocabulary used in textbooks and medical journals. Many foreign words have been assimilated to the Swedish medical language by the addition of Swedish inflections, but will seem "Latin" to the layman reader and hence incomprehensible. Previous studies have shown differences in professional and layman vocabulary in the Swedish medical domain [23]. Also, the close resemblance of expressions in different languages, combined with Swedish inflections, results in diverging spellings which complicates automatic processing of clinical text.	foreign words	Style	3	Lexis	confusing terminology



60	Lexis	negations	Many negations were probably due to mandatory reporting of certain aspects, e.g., for X-ray of lungs to negate pulmonary infiltrates, for CT of brain to negate tumors or bleeding.	negation	Syntax	3	Syntax	negation
61	Coherence	findings non standardized	If the report conveys a new finding, this is not written in a standardized way. Instead, for each such situation, it is described in more varied ways. This has implications for a future text simplification system, as these more varied formulations probably convey more details that affect the patient directly. Observations in this study confirm earlier findings that different medical terms can be used for the same pathology when excluding and reporting normality, or reporting specific findings.	non standardized findings	Coherence	3	Lexis	confusing terminology
62	Coherence	incomplete sentences	Only 23 of the 100 most frequent sentences were complete, containing both a subject and a predicate. However, the majority of the short sentences contain an implicit subject and predicate, e.g. Ingen stas (No stasis) could be rewritten as a sentence such as "The radiology image (subject) shows (predicate) no stasis"	incomplete sentence	Syntax	3	Syntax	sentence
64	Lexis	speculation	An important part of the radiology report is the concluding remarks with diagnostic speculation and reasoning, often intertwined with expressions for hedging and uncertainty. For text simplification, these parts need to be considered with great care since this poses special problems for layman comprehension [27].	speculation	Style	3	Lexis	describing phrase
65	Coherence	argument overlap	Principles for improving local coherence typically involve strategies such as the addition of argument overlap (making each sentence repeat the linking word from a previous sentence). Local coherence deals with sentence overlaps, often mentioning a concept from the previous sentence at the beginning of a new one; this makes it clear and unambiguous what various pronouns refer to (e.g., does "it" refer to the heart or the procedure performed on it?). When text consists largely of bulleted or numbered lists, it is hard to do this kind of local coherence correction. For example, in a medications list made up of numbered sentence fragments, concepts mentioned in new sentences cannot be clearly linked to earlier sentences.	argument overlap	Coherence	5	Coherence	microstructure
66	Coherence	sentence connectives	the use of sentence connectives, and the rearrangement of clauses so that sentences repeat old ideas before introducing new ones.	sentence	Coherence	5	Syntax	sentence connectives
67	Coherence	introducing of background concepts	Improving global coherence involves introducing background concepts	introducing of background concepts	Coherence	5	not applicable	not applicable
68	Coherence	making important references explicit	making important references explicit;	making important references explicit	Coherence	5	not applicable	not applicable
69	Coherence	explaining causal connections	explaining causal connections between events	explaining causal connections	Coherence	5	Syntax	explicit causality
70	Coherence	headers	adding headers and topic sentences	headers	Coherence	5	Coherence	macrostructure
71	Coherence	clear links	clearly linking subtopics to the main topic	clear links	Coherence	5	Coherence	macrostructure
72	Quality of information	abbreviations	Approximately 33% of (n = 56/169) patients reported difficulties with understanding the content of their medical records. The most commonly reported difficulties were understanding abbreviations, medical words and terminology and the meaning and significance of test results	abbreviation	Lexis	6	Lexis	abbreviations
73	Quality of information	terminology	Approximately 33% of (n = 56/169) patients reported difficulties with understanding the content of their medical records. The most commonly reported difficulties were understanding abbreviations, medical words and terminology and the meaning and significance of test results	terminology	Lexis	6	Lexis	expert terminology

74	Quality of information	meaning/significance of test results	Approximately 33% of (n = 56/169) patients reported difficulties with understanding the content of their medical records. The most commonly reported difficulties were understanding abbreviations, medical words and terminology and the meaning and significance of test results	meaning	Semantics	6	Semantics	meaning
75	Coherence macro	headings	Insert a heading above every subject change in the text. Make sure that the level of each section created in this way is clear. Most respondents added headings to the text. Respondents preferred to separate sections with those headings at points in the text where the topic changed and where this transition was not regarded as very obvious at first glance. Respondent: It seems as if this piece of text links up with this piece of text, but nevertheless this is a new narrative. Because this is all about function and how the lungs work and now it suddenly is about how someone gets asthma, and how you can recognize it. So, this is an entirely new subject, therefore a new heading must be added in between. Also, at points in the text where the subject change was already indicated by a white line, many respondents chose to add a heading above the paragraph. For instance, where white lines separated enumerations of factors concerning asthma, one-word headings were often placed above these very short sections. Headings that were already in the text were changed if respondents thought that the vocabulary was too difficult or did not cover the section adequately. Some respondents changed headings to attract the reader's attention or to address the reader more directly. This was done by adding words like you and your child into the headings.	headings	Coherence	7	Coherence	macrostructure
76	Coherence macro	macro signals	Start every section that is preceded by a heading with a brief explanation of the relation between that section and the previous text content. Most respondents who added macro signals did this in the form of short introductions at the beginning of paragraphs, introducing the theme and explaining how it related to the previous information. Respondent: Then the fourth section is about treatment and then you should actually refer to the causes of asthma, so to the content of previous pages, that is, to the body and the environmental influences. Well, then I could say in the section on treatment "Hereditary causes cannot be influenced by ourselves. But environmental factors can." Macro signals were sometimes also placed in the middle of a section, specifically to add explanatory information. Relevant concepts that had been mentioned in other sections were repeated in the explanation of the new concept if this was thought to contribute to a proper understanding of a new concept. Respondent: Here, in the paragraph about medication, you could also say something about how important it is that the bronchi stay open as widely as possible. And some explanations should be provided about the mechanisms of the muscles, so how the medicine works on those muscles. For that purpose, the information from previous paragraphs can be used.	macro signals	Coherence	7	Coherence	macrostructure

77	Coherence micro	argument overlap	<p>In a new sentence, repeat from the preceding sentence the word to which the reader should relate the information in the current sentence. For important concepts in a text, consistently use the same words, that is, avoid synonyms. The experimental text already had considerable argument overlap in the sense that nouns were regularly repeated in succeeding sentences but often rather implicitly in the respondents used the same words again, mostly without further comments. Also, when adding information, they seemingly automatically repeated the noun or the subject about which they added text.</p> <p>Original text: [at the beginning of the section about the lungs:] When we breathe, we provide our bodies with oxygen. The inhaled air flows <i>via</i> the nose and mouth. Respondent: [adds two sentences in the front:] I would start with "Our bodies need oxygen. Oxygen is the fuel with which we walk and move". In such instances, none of the respondents came up with synonyms for important nouns but used the same words instead. While rewriting, some respondents consciously commented on this issue: Original text: When we breathe, we supply our body with oxygen. The breathed air flows via the nose or mouth past the throat and the vocal cords through the wind- pipe to the left and right lung. Respondent: And what happens very often is that people don't like to use the same word twice, then they use other words. Here it says "When we breathe, we supply our body with oxygen," and in the next sentence it says "breathed air." Then I start thinking "where did the oxygen go here? And is oxygen the same as air?" Most people don't recognize that. So, usually I choose the same word so that people consistently get the same understanding.</p>	argument	7	Coherence	microstructure
----	-----------------	------------------	--	----------	---	-----------	----------------

78	Coherence micro	ambiguous pronouns	<p>Only choose a demonstrative pronoun if there is only one possible referent in the previous sentence. When a demonstrative pronoun could refer to two concepts in the previous text, repeat the referent word or phrase itself. Pronouns that ambiguously refer to a noun or concept are detrimental to a straightforward understanding of text. Pronouns in the text that are potentially ambiguous for the reader were recognized as such by almost half of the respondents and replaced by the nouns that the ambiguous pronoun referred to. Original text: A narrowing of the airways occurs because of the contraction. On the inside of the airways extra slime is formed to remove the stimulus. This blocks the airway-passage even more. Respondent: Here is again a reference about which I wonder where it refers to. To the slime or to the contraction? This way, the process is explained less exactly, the information is hidden to some extent. I would repeat the exact word that "this" refers to, so "The slime blocks . . ." etc. Typically, as respondents read the text, they sometimes paused, indicating that they were confused, and asked aloud what certain words referred to exactly. In such instances, respondents always replaced the ambiguity with the noun itself. Related to the first principle, the existing argument overlap was made more explicit by replacing the pronoun with the noun it stood for. Even when respondents understood pronouns without hesitation, they replaced them with nouns if they thought that there was a chance of misunderstanding in the reader. Sometimes respondents replaced ambiguous pronouns seemingly automatically while actually tackling another problem. Specifically, in rewriting sentences that they considered too long into two separate sentences, they replaced the pronoun by the noun-subject of the second sentence: <b>Original text:</b> The oxygen is delivered to the blood and this transports it through the body. Respondent: I think that for a brochure a sentence is already too long when it contains more than one statement. Here for instance. I would make this "Oxygen is delivered to the blood. The blood transports the oxygen through the body." Because splitting a sentence into two does not necessarily require that demonstrative pronouns are replaced by nouns, the fact that it was done here suggests that it was not just an outcome of the splitting procedure but an improvement in itself.</p>	ambiguity	Lexis	7	Lexis	ambiguity
----	-----------------	--------------------	---	-----------	-------	---	-------	-----------

79	Coherence micro	descriptive elaborations	<p>When descriptive well-known terms are added to difficult new concepts, text coherence and thus text comprehension is improved. Respondents added only a few of these descriptive elaborations when they found a concept too difficult or uncommon for the reader. In such instances, respondents mainly chose to add examples to provide a more concrete explanation for the concept. To a lesser extent they chose to add a more common synonym to the concept. Original text: Children whose symptoms can every time be related to repeated contact with certain organic substances, are probably allergic to those. Respondent: "Certain organic substances," that's an unknown term. I don't know an alternative for that, but in any case, I would write that totally differently. With an example or so, adding one or two examples of organic substances. Adding explanations to difficult concepts, however, was generally not regarded as the best remedy to repair breakdowns in comprehensibility of concepts in the text. Whenever respondents did not regard a difficult concept as crucial for the reader, they preferred to delete it altogether. Important concepts that were rated as too difficult preferably were described in another way, usually with some more words than in the original sentence. Original text: [header:] Various stimuli that can cause hyper reactivity. Respondent: I wonder if I find the term "hyper reactivity" suitable. I think for a lot of people that don't read very much; such a word does not mean anything to them. Therefore, I would throw out such terms that I don't find very accessible. And almost always you can describe in plain language what's going on. Along with unavoidable jargon or unusual words, add descriptive elaborations of generally known concepts.</p>	descriptive terms	Coherence	7	Lexis	defining phrase
80	Coherence micro	sentence connectives	<p>Use sentence connectives where possible to clarify the relations between sentences. Especially causal relations among concepts should be made explicit this way. At points in the text where respondents thought sentences could be related more explicitly, they added words to connect those sentences with each other, resulting in one longer sentence. Words like but, like, for instance, except, moreover, and conversely were added to the existing text to clarify relations. <b>Original text:</b> The oxygen is given to the blood and subsequently this transports it through the whole body. Our body also produces waste. Respondent: I would not want to see these things separate from each other. I would link them more, ... by using a word like "conversely": "Conversely, our body also produces waste". Whereas some connectives were inserted really deliberately, others seemed to be inserted rather automatically, meaning that respondents used explicit connective words in pieces of text that they wanted to add in. For instance, along with every factor concerning asthma attacks, many respondents added practical advice on how to act upon them. In those cases, they used words like therefore or thus to make the connection clear.</p>	sentence connectives	Syntax	7	Syntax	sentence

81	Coherence micro	given-new order	<p>Within sentences, maintain an information order of 'given-new': start a sentence with information that has been mentioned earlier in the text and insert after that the 'new' (i.e., to the reader unknown) information, which the reader has to relate to the previous information. At hardly any instance did some respondents change the order of the information in a sentence from the suboptimal "new-given" order to the more optimal given-new order. In the few cases in which respondents did improve sentence order into given-new, they did this either without further remarks or as the result of another, strongly related problem that they tackled. For instance, some respondents found that the subject of a sentence should be placed in the beginning of a sentence, from which an unintended given-new order arose. Original text: An important part of the airways are the smallest branches, the bronchi. [Where the bronchi had been mentioned in the previous paragraph]. Respondent: What I see here ...is that the subject is placed far to the back of the sentence. "The bronchi, the smallest branches of the lungs, are an important part of the airways." It is actually very simple, it all sounds very boring but it reads so much easier when you immediately encounter the subject, the verb and then the information. In Dutch it is increasingly often turned around. Then you first encounter a lot of text and after that the subject. Although respondents made no remarks about changing sentences to a given-new order, it should be noted that when they added sentences, in many instances they employed a given-new order, seemingly without conscious deliberation.</p>	given-new order	Coherence	7	Coherence	microstructure
82	Coherence micro	Explicit actor	<p>Some respondents applied the cognitive coherence principle of the "explicit actor", that is, they added the actor in sentences in which it was not explicitly mentioned. Typically, however, most respondents made the actor of a verb more explicit by rewriting passive sentences into active ones (see the Other Textual Coherence Changes section, below). Furthermore, some respondents repeated the actor when they thought the text was unclear in who did what. State the actor of every important verb explicitly in the text. Similarly, every concept that is important to a proper comprehension by the reader should be named explicitly in the text. <b>Original text:</b> The finest branches end up in many little alveoli. Around those there are veins. The oxygen is given to the blood and this transports it through the entire body. Respondent: In any case I would want to make clear who does what and here it is "what does what?" But "the oxygen is given," then it suddenly is much less concrete and then I would say, "Those alveoli give the oxygen to the blood". In addition, many respondents commented on the distance that is created between the text and the reader by using general sentences without actors such as you and your child. Consequently, many respondents put these terms in as the actors of sentences. Original text: The slime is then swallowed or spitted out. Respondent: If it's possible I would make it active. You should address the reader directly. It makes the text really awkward if you write a lot in the passive voice. Because it almost seems as if it is not related to yourself When I read this brochure I think, "it is me, or a person who swallows." The reader might not get that. So: "Then you swallow the slime or you spit it out".</p>	explicit actor	Syntax	7	Syntax	explicit actor

83	Coherence micro	explicit causality	<p>Make causality explicit within and between sentences by means of explicit descriptions of causal processes and an appropriate word-ordering of cause followed by effect. In general, only some changes and hardly any remarks were made with regard to making causal relations more explicit. Even though the sentences themselves contained very few explicit markers of causality, words or phrases such as because, as a consequence, if . . . then, and therefore were only infrequently added in the text for reasons of explicating causal relations between concepts. Original text: People with an allergy react differently to certain substances than people without an allergy. Tussocks of plants for instance: these harmless substances do not bother most people, but do harm people with an allergic predisposition. Respondent: I would say here "If you come in contact with tussocks of plants, then you will get complaints in your airways. That is a consequence of an allergy, so in that case there is a chance that you have an allergy". As already mentioned in the section on sentence connectives, in cases where respondents wanted to add text on practical actions that could be taken to deal with asthma, they did use causality indicators in the form of connective words. However, respondents only rarely explicitly mentioned causality as a coherence improvement. Causality was more often made explicit with regard to the macrostructure of the text. Specifically, some respondents changed the order of sentences into one fitting a "cause-and-effect sequence," starting with a situation and subsequently elaborating the effect(s). Respondents mentioned "plain logic" or "taking the reader by the hand" as a basis for such information order changes.</p>	explicit causality	Syntax	7	Syntax	explicit causality
84	Coherence additional	sentence length	<p>Changes were made in surface characteristics of the text, which in the minds of the respondents specifically concerned the length of sentences. In this vein, many respondents shortened or split up sentences they considered too long. Respondents said they wished to restrict sentence length to a maximum of approximately 12 words. As a result, sentences contained only a limited number of different concepts that the reader should keep active in working memory. Thus, the information quantity is kept well within the limits of the working memory span, so that no breaks occur in information processing.</p>	sentence	Syntax	7	Syntax	sentence
85	Coherence additional	passive to active sentence structure	<p>A considerable number of respondents explicitly preferred sentences constructed in the active voice instead of the passive. As a side effect, a subject is put into the sentence, which in turn increases textual coherence.</p>	passive voice	Syntax	7	Syntax	passive voice
87	Coherence additional	main verb to front	<p>Third, a small minority of respondents put the main verb in the beginning of sentences, so that the reader would not have to keep too much information in working memory before encountering the verb to which it should be "related."</p>	main verb to front	Syntax	7	Syntax	main verb to front
88	Coherence additional	replace difficult words	<p>All respondents frequently replaced difficult or unusual words and terms with more commonly known synonyms. When they found a word or term superfluous or confusing, they deleted it altogether. This change improves text coherence in the sense that the incoming information matches well with the existing knowledge networks of the readers so their working memory is not overloaded with unnecessary information.</p>	terminology	Lexis	7	Lexis	expert terminology

89	Coherence additional	add/delete content	A fifth kind of textual change was used by all respondents; namely, they intended to increase text comprehension by adding or deleting whole pieces of information to explain more and reduce redundant information. For instance, respondents frequently proposed to insert practical advice on feasible countermeasures against causes of asthma attacks throughout the brochure. By adding this information, the writers wanted to meet readers' wishes and expectations about the health education material. In a similar vein, detailed information that was not thought of as necessary to trigger correct behavior was deleted to prevent confusion.	added content	coherence	7	not applicable	not applicable
90	Non regular expressions meaning	context dependency	ex. cut out (1. restriction - 2. excision)	context	Lexis	8	Lexis	context
91	Non regular expressions meaning	qualifier	ex. slightly more	qualifier	Lexis	8	Lexis	unclear qualifier
92	Non regular expressions meaning	misspelling	ex. gaulbladder (gallbladder)	misspelling	Lexis	8	Lexis	misspelling (linguistic error)
93	Non regular expressions meaning	shortened form	ex. alz (Alzheimer Disease)	abbreviation	Lexis	8	Lexis	abbreviations
94	Non regular expressions meaning	defining phrase	ex. breast removal (mastectomy)	defining phrase	Lexis	8	Lexis	defining phrase
95	Non regular expressions meaning	exemplar	ex. aspirin (analgesic)	exemplar	Lexis	8	Lexis	exemplar
96	Non regular expressions meaning	describing phrase	ex. ticking bomb (occurs suddenly)	describing	Lexis	8	Lexis	describing phrase
97	Non regular expressions meaning	colloquialism	ex. pee (urination)	colloquialism	Lexis	8	Lexis	colloquialism
98	Non regular expressions meaning	neologism	ex. ADD-lets (ADD patients)	neologism	Lexis	8	Lexis	neologism
99	Non regular expressions meaning	transformation	ex. break (fracture)	terminology	Lexis	8	Lexis	expert terminology
100	Non regular expressions meaning	doctor-ese	ex. postop	terminology	Lexis	8	Lexis	expert terminology
101	Concepts	general ambiguity	not mappable to UMLS fuzzies	general	Content	8	not applicable	not applicable
102	Concepts	lack of specificity	Not mappable to UMLS ex. private area	lack	Content	8	not applicable	not applicable
103	Concepts	non-credible concept	Not mappable to UMLS ex. mucoid plaque	concept	Content	8	not applicable	not applicable
104	Change of Skopos	unidiomatic expressions	It goes without saying that factual errors and wrong or unidiomatic translations, do not lead to user-friendly inserts. Similarly, caiques, word-for-word translations and unidiomatic expressions have a negative impact on the readability of a text. A higher level of formality, stylistic errors and inconsistent terminology are more intriguing, for whereas a complex syntactic structure, formal expressions, expert terms, etc. may not be 'wrong' from a linguistic point of view, such features are certainly not adequate in view of the skopos of the inserts.	terminology	Lexis	9	Lexis	expert terminology
105	Change of Skopos	level of formality in terms of syntax	A higher level of formality, stylistic errors and inconsistent terminology are more intriguing, for whereas a complex syntactic structure, formal expressions, expert terms, etc. may not be 'wrong' from a linguistic point of view, such features are certainly not adequate in view of the skopos of the inserts. A translator who renders a simple syntactic structure in the source text by a complex one or introduces specialist terms in the target text, does not convey the user-friendliness of the original insert adequately.	formal syntax	Syntax	9	Lexis	high formality



106	Change of Skopos	level of formality in terms of lexis	<p>The lexis in Danish translations is more formal. In the following examples the Danish translators have rendered informal phrases by more formal expressions: EXAMPLE 4: Source text: after breakfast Target text: After dagensforste maltid [Literally: after the first meal of the day] EXAMPLE 5: Source text: How this medicine is to be used: to relieve symptoms that may occur with hay fever or other year-round allergies Target text: Anvendelsesomrade: Symptomerpa scesonbetinget allergisk hosnue eller helarssnue [Literally: Application area: symptoms of seasonal allergic hay fever or year-round allergies] EXAMPLE 6: Source text: if you drink alcohol Target text: ved alkoholindtagelse [Literally: in connection with alcohol consumption]. This formalisation of lexis is often combined with agent displacement measures, such as the above-mentioned introduction of the passive voice. Latin medical terms are characteristic of English, which has virtually no medical expressions deriving from Anglo-Saxon, whereas everyday Danish medical vocabulary is characterised by ancient, simple and immediately understandable words (often compounds) derived from Danish roots ('uterus' for instance is a 'livmoder' [literally: a 'life mother']). This means that English has many Latin medical terms which are familiar to English consumers such as 'hepatitis', 'teta- nus' and 'ulcer' all of which have common Danish names (for analogous examples involving German, French and English see Kirkness 1997). It is a major problem when these common English terms are transferred into Danish where they constitute expert terminology unknown to Danish consumers. Such renditions heighten the level of formality in Danish inserts and sometimes make inserts incomprehensible to the average Danish consumer. Two examples: EXAMPLE 7: Source text: for oral use Target text: for oral anvendelse Instead of the common Danish expression = indtages gennem munden [Literally: to be taken through the mouth] EXAMPLE 8: Source text: injections Target text: injektioner Instead of the common Danish word = indsprojtning [Literally: in spraying]. This phenomenon also applies to non-medical Latin-based words which are common in English, but which in Danish are used in medical LSP.</p>	formality	Lexis	9	Lexis	high formality
-----	------------------	--------------------------------------	--	-----------	-------	---	-------	----------------

107	Change of Skopos	stylistic errors	As mentioned above, the phrasing of some English inserts may be infelicitous because the consumer aspect was disregarded in the inter-generic translation. Naturally, problems arise when the Danish translator subsequently transfers the unfortunate solutions. Obvious examples are renditions of English specialist terms (which are unknown to the average British consumer) by Danish specialist terms (with which average Danes are also unfamiliar). This does not improve the readability of the insert: EXAMPLE 9: Source text: Interferon alfa-2bfor subcutaneous use - Target text: Interferon alfa-2bfor subkutan anvendelse EXAMPLE 10: Source text: The usual dose is 3 million International Units (IU) three times a week - Target text: Den scedvanlig dosis er 3 millioner 'Internationale enheder (IU) tre gange om uge. When translators uncritically transfer expert terms from source texts and fail to take into account inadequate style, let alone blatant errors in the English texts, problems that arose in the first stage shown in figure 1 may make their way into the Danish translation. In these cases, the translators fail to fulfil their obligation to act as linguistic experts who should intervene and improve the text according to the new skopos necessitated by the fact that the audience are now Danish consumers.	stylistic errors	Style	9	Lexis	high formality
108	Change of Skopos	inconsistent terminology	Our last category comprises cases where the inserts become more complex and ambiguous because the translator uses synonyms to refer to the same phenomenon and thus switches between expert and non-expert terms. Sometimes the inconsistencies derive from the source text, but quite often the synonyms are introduced in the course of the translation process. EXAMPLE 11: Source text: Women of childbearing age Synonyms used in one and the same Danish - Target text: Fertile kvinder [literally: fertile women]... Kvinder i den fødedygtige alder [literally: Women of childbearing age] EXAMPLE 12: Source text: level of glucose in your blood Synonyms used in one and the same Danish - Target text: blodets glukosekoncentration [literally: the glucose concentration in the blood]... blodsukkeret [literally: the blood sugar] An expert in medicine will recognise these terms as synonyms. But to the layman such 'stylistic variation' is confusing. We could go on listing other sub-groups of problematic translations from the inserts we have analysed. From the point of view of user-friendliness, the inserts abound in translation mistakes and errors, and it is important to stress that it is the sum of these features rather than the single, isolated example which makes the Danish inserts near-incomprehensible to many people. But the infelicitously translated inserts are only symptoms. We shall therefore now address the causes of these problems, and we will try to diagnose what goes wrong in the chain of translation.	terminology -inconsistent	Lexis	9	Lexis	expert terminology
109	Lexical simplification	abbreviation	Abbreviation identification in English biomedical and clinical texts has been studied extensively (e.g. Xu et al. (2007), Liu et al. (2001)). For detection of Swedish medical abbreviations, there are fewer studies. Dannélls (2006) reports detection of acronyms in medical journal text with 98% recall and 94% precision by using part of speech information and heuristic rules. Clinical Swedish presents greater problems than medical texts, because of ad hoc abbreviations and noisier text. By using lexicons and a few heuristic rules, Isenius et al. (2012) report the best F- score of 79% for abbreviation detection in clinical Swedish.	abbreviations	Lexis	10	Lexis	abbreviations

110	Lexical simplification	compound words	Good compound analysis is critical especially for languages whose orthographies concatenate compound components. Swedish is among those languages, in which every such concatenation thus corresponds to a word. The most common approach to compound splitting is to base it on a lexicon providing restrictions on how different word forms can be used for generating compounds. For example, Sjöbergh and Kann (2006) used a lexicon derived from SAOL (the Swedish Academy word list), and Östling and Wirén (2013) used the SALDO lexicon of Swedish morphology (Borin and Forsberg, 2009). With this kind of approach, compound splitting is usually very reliable for genres like newspaper text, with typical accuracies for Swedish around 97%, but performs poorer in domain specific genres.	compound words	Lexis	10	Lexis	compound words
111	Lexical simplification	terminology	The detection of English medical terminology is a widely researched area. An example of term detection in English clinical texts is Wang and Patrick (2009) work based on rule-based and machine learning methods, reporting 84% precision. For Swedish clinical text, Kokkinakis and Thurin (2007) have employed domain terminology matching and reached 98% precision and 87% recall in detecting terms of disorders. Using similar approaches, Skeppstedt et al. (2012), reached 75% precision and 55% recall in detecting terms of disorders. With a machine learning based approach, improved results were obtained: 80% precision, 82% recall (Skeppstedt et al., 2014). Skeppstedt et al. (2012) have also demonstrated the negative influence of abbreviations and multiword expressions in their findings.	terminology	Lexis	10	Lexis	expert terminology
112	Lexical simplification	spelling	A system for general spelling correction of Swedish is described by Kann et al. (1998), but we are not aware of any previous work related to spelling correction of Swedish clinical text. An example of spelling correction of clinical text for other languages is Tolentino et al. (2007), who use several algorithms for word similarity detection, including phonological homonym lookup and n-grams for contextual disambiguation. They report a precision of 64% on English medical texts. Another example is Patrick et al. (2010) and Patrick and Nguyen (2011), who combine a mixture of generation of spelling candidates based on orthographic and phonological edit distance, and a 2-word window of contextual information for ranking the spelling candidates resulting in an accuracy of 84% on English patient records. Siklósi et al. (2013) use a statistical machine translation model (with 3-grams) for spelling correction, achieving 88% accuracy on Hungarian medical texts.	misspelling	Lexis	10	Lexis	misspelling (linguistic error)
114	Understanding	abbreviations	Many patients requested explanations of medical terms (42%), abbreviations and acronyms (13%), and information on tests or results (17%) and metric weight measurements (5%).	abbreviation	Lexis	11	Lexis	abbreviations
115	Understanding	acronyms	Many patients requested explanations of medical terms (42%), abbreviations and acronyms (13%), and information on tests or results (17%) and metric weight measurements (5%).	abbreviation	Lexis	11	Lexis	abbreviations

116	Understanding	explanation of medical terms	Table 1. Level of patient agreement with questions about accessing their electronic patient record (n = 100). Question %	terminology	Lexis	11	lexis	expert
			Found it easy to find your way around the record 73					
			Found registration section useful 85					
			Found record summary useful 94 Found consultation details useful 90 Found medication details useful 59 Found referrals section useful 42					
			Found at least one section useful 99					
			Found record summary useful 94					
			Found consultation details useful 90					
			Found medication details useful 59					
			Found referrals section useful 42					
			Found at least one section useful 99					
			Found registration section easy to understand 94					
			Found record summary easy to understand 84					
			Found consultation details easy to understand 80					
			Found medication details easy to understand 61					
			Found referrals section easy to understand 41					
			Found record easy to understand overall 73					
Found record difficult to understand overall 5								
Worried about security - before seeing record 48								
Confident of security in use - after seeing record 61								
Most patients found it easy to understand their records. Where problems arose, it was with the record summaries or consultation details (Table 1). Many patients requested of medical terms (42%), abbreviations and acronyms (13%), and information on tests or results (17%) and metric weight measurements (5%).								
117	Understanding	information on tests and results	Many patients requested explanations of medical terms (42%), abbreviations and acronyms (13%), and information on tests or results (17%) and metric weight measurements (5%).	explanation on	Content	11	Semantics	meaning
118	Understanding	explanation of medical terms	Most patients found it easy to understand their records. Where problems arose, it was with the record summaries or consultation details (Table 1). Many patients requested explanations of medical terms (42%), abbreviations and acronyms (13%), and information on tests or results (17%) and metric weight measurements (5%).	terminology	Lexis	11	Lexis	expert terminology
119	Understanding	abbreviations and acronyms	Most patients found it easy to understand their records. Where problems arose, it was with the record summaries or consultation details (Table 1). Many patients requested explanations of medical terms (42%), abbreviations and acronyms (13%), and information on tests or results (17%) and metric weight measurements (5%).	abbreviation	Lexis	11	lexis	abbreviations
120	Understanding	results	Most patients found it easy to understand their records. Where problems arose, it was with the record summaries or consultation details (Table 1). Many patients requested explanations of medical terms (42%), abbreviations and acronyms (13%), and information on tests or results (17%) and metric weight measurements (5%).	results	Lexis	11	Semantics	meaning
121	Understanding	explanation on metric weight measurements	Most patients found it easy to understand their records. Where problems arose, it was with the record summaries or consultation details (Table 1). Many patients requested explanations of medical terms (42%), abbreviations and acronyms (13%), and information on tests or results (17%) and metric weight measurements (5%).	measurements	Content	11	Semantics	meaning
122	Additional requirements	explanation of medical terms	Improved understanding: explanation of medical terms, use of easy-to-understand language, glossary of acronyms, imperial conversion for weight and height, normal ranges (for example, body mass index) for test results	explanation of medical terms	Lexis	11	Lexis	expert terminology
123	Additional requirements	easy-to-understand language	Improved understanding: explanation of medical terms, use of easy-to-understand language, glossary of acronyms, imperial conversion for weight and height, normal ranges (for example, body mass index) for test results.	terminology	Lexis	11	Lexis	expert terminology

124	Additional requirements	glossary of acronyms	Improved understanding: explanation of medical terms, use of easy-to-understand language, glossary of acronyms, imperial conversion for weight and height, normal ranges (for example, body mass index) for test results.	abbreviations and acronyms	Lexis	11	lexis	abbreviations
125	Additional requirements	imperial conversion for weight and height	Improved understanding: explanation of medical terms, use of easy-to-understand language, glossary of acronyms, imperial conversion for weight and height, normal ranges (for example, body mass index) for test results.	explanation on metric measurements	Lexis	11	Lexis	unexplained numbers
126	Additional requirements	normal ranges for test results	Improved understanding: explanation of medical terms, use of easy-to-understand language, glossary of acronyms, imperial conversion for weight and height, normal ranges (for example, body mass index) for test results.	results	Content	11	Semantics	meaning
127	Comprehension barrier	conceptual	Forty-four participants made comments that pointed to records comprehension barriers that were related to insufficient conceptual knowledge.	conceptual knowledge	Content	12	not applicable	not applicable
128	Comprehension barrier	reference support	Twenty-seven of these were related to problems that could be ameliorated by pointers to general reference type support while 17 would require individualized decision support.	medical knowledge	Support	12	not applicable	not applicable
129	Comprehension barrier	terminology support	Thirty-eight participants commented on professional language as a barrier to record comprehension. Across all coding categories, medical terminology was the most frequently cited difficulty area (comments by 24 participants). The need for terminology support was often expressed as a preference for "simpler words" or "laymen's terms" in place of "medical terms", "as if they were written for non-medical degreed person". Some participants also noted that electronic record format would make providing terminology resources (e.g., online dictionaries) easier.	terminology	Lexis	12	Lexis	expert terminology
130	Comprehension barrier	language	A significant proportion of language-related comments (9) also had to do with the use of abbreviations and special codes in the records (#1.2.3 Abbreviation and Codes Understanding Support). One of the participants expressed her frustration by writing, "The abbreviations, acronyms and symbols doctors use are a mini-foreign- language to most of us".	abbreviation	Lexis	12	Lexis	abbreviations
131	Comprehension barrier	structure	Twenty participants commented on the record structure as a factor contributing to comprehension difficulty. Eleven of these comments concerned data ordering (#3.5 Data Ordering): some participants felt that topical organization was preferable to temporal, or vice versa, others asked for topical categories arranged within temporal ones (eg, "Put them in a structured document so I can compare apples to apples over the course of the longitudinal record").	structure	Coherence	12	Coherence	macrostructure
132	Comprehension barrier	understanding	Finally, fifteen participants commented on the role that health professionals and advocates should play in the process of record comprehension. These participants often felt that no matter how much comprehension support is provided within the record, interpretation of this professional document requires medical expertise. They, therefore, felt that records should be jointly reviewed and discussed by patients and health professionals (e.g., "Review and discuss reports (especially lab) with a healthcare provider.")	joint record review	Support	12	not applicable	not applicable

133	Results	results	The prevalence of categories and subdivisions are provided in Table 1. The most prevalent note characteristics across categories were General Medical Acronyms (99.1%), Medical Jargon (96.7%), Typographical Errors (49.0%), and Lab Tests & Infectious Disease (46.6%). As seen in Table 2, the average resident note was significantly longer than the average hospitalist note (359.9 and 223.8 words, respectively). Furthermore, residents displayed significantly higher usage of General Medical Acronyms, Lab Tests & Infectious Disease, Pharmaceutical, Medication Delivery & Timing, and Vitals. Table 3 exhibits the most frequently coded words from each category or subdivision.	explanation on results	Support	13	not applicable	not applicable																																																																																
134	Results	results	<table border="1"> <thead> <tr> <th colspan="4">Table 1</th> </tr> <tr> <th colspan="4">Prevalence of medical Shorthand, Jargon, subjective descriptors, health status and Typographical errors within medical notes</th> </tr> <tr> <th>Note statistic</th> <th>Hospitalist</th> <th>Resident</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Notes by Author type (N)</td> <td>231</td> <td>106</td> <td>337</td> </tr> <tr> <td colspan="4"><b>Acronyms &amp; Abberivations</b></td> </tr> <tr> <td>General medical acronym</td> <td>98.70%</td> <td>100%</td> <td>99.10%</td> </tr> <tr> <td>Lab tests &amp; infectious</td> <td>35.90%</td> <td>69.80%</td> <td>46.60%</td> </tr> <tr> <td>Pharmaceuticals</td> <td>16.90%</td> <td>34%</td> <td>22.30%</td> </tr> <tr> <td>Medication delivery &amp;</td> <td>25.10%</td> <td>66%</td> <td>38%</td> </tr> <tr> <td>Vitals</td> <td>25.50%</td> <td>55.70%</td> <td>35%</td> </tr> <tr> <td>Medical jargon</td> <td>95.20%</td> <td>100%</td> <td>96.70%</td> </tr> <tr> <td colspan="4"><b>Subjective descriptors</b></td> </tr> <tr> <td>Age-specific Phrases</td> <td>0%</td> <td>0%</td> <td>0%</td> </tr> <tr> <td>Affect &amp; Mentation</td> <td>10%</td> <td>17.90%</td> <td>12.50%</td> </tr> <tr> <td>Weight</td> <td>5.20%</td> <td>9.40%</td> <td>6.50%</td> </tr> <tr> <td colspan="4"><b>Mental and personal health</b></td> </tr> <tr> <td>Psychiatric history</td> <td>9.50%</td> <td>9.40%</td> <td>9.50%</td> </tr> <tr> <td>Substance use history</td> <td>15.20%</td> <td>19.80%</td> <td>16.60%</td> </tr> <tr> <td>Sexual history</td> <td>0.40%</td> <td>0.90%</td> <td>0.60%</td> </tr> <tr> <td>Typographical errors</td> <td>45.90%</td> <td>54.70%</td> <td>49%</td> </tr> </tbody> </table>	Table 1				Prevalence of medical Shorthand, Jargon, subjective descriptors, health status and Typographical errors within medical notes				Note statistic	Hospitalist	Resident	Total	Notes by Author type (N)	231	106	337	<b>Acronyms &amp; Abberivations</b>				General medical acronym	98.70%	100%	99.10%	Lab tests & infectious	35.90%	69.80%	46.60%	Pharmaceuticals	16.90%	34%	22.30%	Medication delivery &	25.10%	66%	38%	Vitals	25.50%	55.70%	35%	Medical jargon	95.20%	100%	96.70%	<b>Subjective descriptors</b>				Age-specific Phrases	0%	0%	0%	Affect & Mentation	10%	17.90%	12.50%	Weight	5.20%	9.40%	6.50%	<b>Mental and personal health</b>				Psychiatric history	9.50%	9.40%	9.50%	Substance use history	15.20%	19.80%	16.60%	Sexual history	0.40%	0.90%	0.60%	Typographical errors	45.90%	54.70%	49%	acronyms and abbreviations	Lexis	13	Lexis	abbreviations
Table 1																																																																																								
Prevalence of medical Shorthand, Jargon, subjective descriptors, health status and Typographical errors within medical notes																																																																																								
Note statistic	Hospitalist	Resident	Total																																																																																					
Notes by Author type (N)	231	106	337																																																																																					
<b>Acronyms &amp; Abberivations</b>																																																																																								
General medical acronym	98.70%	100%	99.10%																																																																																					
Lab tests & infectious	35.90%	69.80%	46.60%																																																																																					
Pharmaceuticals	16.90%	34%	22.30%																																																																																					
Medication delivery &	25.10%	66%	38%																																																																																					
Vitals	25.50%	55.70%	35%																																																																																					
Medical jargon	95.20%	100%	96.70%																																																																																					
<b>Subjective descriptors</b>																																																																																								
Age-specific Phrases	0%	0%	0%																																																																																					
Affect & Mentation	10%	17.90%	12.50%																																																																																					
Weight	5.20%	9.40%	6.50%																																																																																					
<b>Mental and personal health</b>																																																																																								
Psychiatric history	9.50%	9.40%	9.50%																																																																																					
Substance use history	15.20%	19.80%	16.60%																																																																																					
Sexual history	0.40%	0.90%	0.60%																																																																																					
Typographical errors	45.90%	54.70%	49%																																																																																					
135	Results	results		terminology	Lexis	13	lexis	expert terminology																																																																																
136	Results	results		subjective describing phrases	Lexis	13	Lexis	describing phrase																																																																																
137	Methods	typographical errors	We defined typographical errors as the following: 1) spelling errors (chills); 2) capitalization errors (Pain); 3) repeated words (he was going). We did not code the omission of capitalization of medical proper nouns as errors (hodgkin's).	spelling	Lexis	13	lexis	spelling																																																																																
138	Results	results		spelling	Lexis	13	Semantics	meaning																																																																																