Diagnosis clarification by patient-friendly terms, definitions and generalization: implementation study protocol

Authors: Hugo J.T. van Mens[1,2], Gaby Hannen[2], Remko Nienhuis[2],

Roel J. Bolt[3], Nicolette F. de Keizer[1], Ronald Cornet[1]

[1] Department of Medical Informatics, Amsterdam UMC, location AMC, Amsterdam
[2] ChipSoft B.V., Amsterdam
[3] Sint Franciscus Gasthuis & Vlietland

Date: April 11, 2022

Abstract

Introduction Medical data can be difficult to comprehend for patients, but only a limited number of patient-friendly terms and definitions is available to clarify medical concepts. Therefore, we developed a technique of generalizing diagnoses to more general concepts that do have patient-friendly terms and definitions in the SNOMED CT Netherlands Patient-Friendly Extension, by employing the SNOMED CT hierarchy. This increases the number of diagnoses that can be clarified significantly. Additionally, we validated the generalizations and found that the majority of generalizations were considered correct and acceptable to use in practice. The generalizations and diagnosis clarifications with synonyms and definitions that were already available will be implemented into a hospital patient portal.

Objectives In this study we aim to evaluate the implementation of these clarifications into the diagnosis list of a patient portal with actual users to assess up to what extent it meets their information needs and to obtain feedback for further improvement of the clarification functionality.

Methods We measure the usage of the clarification functionality with aggregated, routinely available EHR and log file data. Feedback will be collected about the quality of the clarifications for quality improvement and the type of questions that arise about the diagnoses and the clarifications will be analyzed. Additionally, we will explore differences in personal characteristics of users that log in with users that view their diagnosis list, view the clarifications, and those that provide feedback.

Discussion This study will provide insight into patient portal user information needs by analyzing the actual use of a clarification functionality for the diagnosis list. The feedback on the clarification quality can be used to improve the clarifications. Additionally, this study will show up to what extent the patient-friendly extension and generalization can be useful to generate quality clarifications from the perspective of actual patient portal users.

Versioning

Date	Description
7-5-2021	Version submitted to AMC METC and Franciscus research desk
1-11-2021	Added evaluation of PFE (next to generalization, there are clarifications with synonyms and definitions), removed collections of questions that patients have, provided more details about the research database and data extraction.

11-11-2021	Add aggregation by registration type and codelist.
5-4-2022	Removed part of sentence about patient service desk data that was supposed to be removed already. Changed title to final.
11-4-2022	Adapted title and added header

Contents

Abstract 1
Versioning 1
1 Introduction 5
1.1 Scientific background 5
1.2 Rationale for the study 5
1.3 Objectives of the study 6
2 Study context 6
2.1 Organizational setting 6
2.2 System details and system in use 6
3 Methods 6
3.1 Study design 6
3.2 Participants 6
3.3 Study flow 7
3.4 Outcome measures or evaluation criteria 7
3.4.1 Overall 7
3.4.2 User level 7
3.4.3 Diagnosis clarification level 7
3.5 Methods for data acquisition and measurement 8
3.5.1 Audit trail and EHR data reuse 8
3.5.2 Feedback 8
3.6 Methods for data analysis 9
4 Expected results 9
4.1 Demographic and other study coverage data 9
4.2 Unexpected events during the study 10
4.3 Study findings and outcome data 10
4.4 Unexpected observations 10
5 Discussion 10
5.1 Strengths and limitations 10
5.2 Meaning and generalizability of the study 10
5.3 Conclusions 10
6 Declarations 10
6.1 Ethics approval and consent to participate 10
6.2 Consent for publication 10
6.3 Availability of data and materials 11
6.4 Competing interests 11

6.5 Funding 11
6.6 Authors' contributions 11
6.7 Acknowledgements 11
7 Attachments 11
7.1 Attachment A. Feedback question 11
7.2 Attachment B. Data extraction from database and reports from EHR Report Builder functionality (to be developed) 11
7.2.1 Research database 11
7.2.2 Extracted data 11
8 References 13

1 Introduction

1.1 Scientific background

Medical data can be difficult to comprehend for patients, but only a limited number of patient-friendly terms and definitions is available to clarify medical concepts. Laymen understand information and knowledge on a more general level, in less detail, than patients and clinicians. Specialists tend to register medical data with even more details. To clarify what a medical term means we hypothesize a short description in more general terms might thus be sufficient if it is at the right level of detail. This of course does not replace the need to inform patients thoroughly during consultations and to provide patient information resources, but we believe this can help patients understand data in their medical record. Patients can also read back what was discussed during consultations and find some clarification of medical concepts.

Therefore, we developed a method to generalize diagnoses to more general concepts that do have patient-friendly terms and definitions in the SNOMED CT Netherlands Patient-Friendly Extension (PFE), by employing the SNOMED CT hierarchy [1]. We showed that this method increases the number of diagnoses that can be clarified significantly. Additionally, two raters with a medical and terminological background validated a representative sample of 1200 clarifications and found more than 85% of clarifications were considered to be correct and acceptable to use in practice [2].

We further improve the clarifications based on the input from the validation study and update the clarifications with the latest version of the PFE. The final set of clarifications will contain clarifications consisting of direct synonyms and definitions available in the PFE (e.g. 'phlebitis' in Table 1) and clarifications that are generated by the generalization to concepts that do have PFE synonyms and definitions (e.g. 'pulmonic valve regurgitation' in Table 1).

Table 1 Examples of diagnoses registered in Dutch diagnosis lists of medical records and their corresponding clarifications that can be displayed when clicked on the diagnosis description or info button.

Medical diagnosis description	Clarification
Phlebitis ⓘ	Another word for "phlebitis" is inflammation of vein: Inflammation of a vein which makes it red, swollen, and painful.
Pulmonic valve regurgitation ⓘ	A type of leaky heart valve. Leaky heart valve: This is a heart valve that closes poorly so that oxygen-rich blood no longer flows properly through the body. This causes complaints such as shortness of breath, fatigue after exertion and dizziness.
Congenital cyst of adrenal gland ⓘ	A type of inborn abnormality and hormonal disorder. Cyst: cavities in the body filled with liquid.
Lowe syndrome ⓘ	A type of inborn abnormality, mental disorder, and disorder of brain, kidney, eye, and metabolism. It is hereditary.

1.2 Rationale for the study

The clarifications have not been evaluated by actual patient portal users. In the current study we want to evaluate the implementation of these diagnosis clarifications in a patient portal diagnosis list.

1.3 Objectives of the study

First, we aim to evaluate patient portal users' information needs by analyzing to what extent they actually use the clarification functionality when they view their diagnosis list and for which diagnoses. Second, we will evaluate the quality of the clarifications from the perspective of the users. Third, we will explore differences in user characteristics between users that view their diagnosis list, those that use the clarification functionality, and those that provide feedback.

2 Study context

2.1 Organizational setting

The study will be carried at one or more hospitals that register diagnoses with the Diagnosethesaurus, use the health information system HiX (ChipSoft B.V., Amsterdam, The Netherlands) version 6.2, and have implemented the web-based HiX patient portal. Currently, we are planning to carry the study with the hospital Sint Franciscus Gasthuis & Vlietland (Franciscus).

Franciscus had two hospital locations, five out-patient locations and a catchment area of about 400,000 inhabitants. It had over 4,614 employees (3,285 full-time equivalent employees) and 318 medical specialists. The hospital had 568,035 outpatient clinic visits, 38,140 surgeries, 31,525 day treatments, 36,910 clinical admissions, 48,439 emergency visits, and 4,109 new-borns in 2019. (Franciscus, 2019) Between November 1, 2019 and October 31, 2020, 45,951 patients had logged in to their portal and 22,140 patients had viewed their diagnosis list, which thus were about 61 views per day. This was a relatively higher number than other hospitals that ran HiX 6.2, from a selection of hospitals that were contacted to gain insight into diagnosis list views.

2.2 System details and system in use

Patients, or their significant others that have been authorized by patients, use the patient portal, for instance, to view their medical data, schedule appointments, message their health care provider securely, and complete questionnaires. The clarifications will be implemented in the diagnosis list of the patient portal. The description of the diagnosis will be highlighted, underlined and provided with an info icon. When clicked, the diagnosis description and a clarification of the diagnosis will be displayed. For the quality improvement of the clarifications, users can provide feedback about the clarification. A warning will be displayed for the clarifications with supertypes, stating that the clarification was generated automatically and might contain mistakes. For questions about their diagnosis, patients will be referred to their doctors.

3 Methods

3.1 Study design

Post implementation study with reuse of routinely collected data about diagnoses, routinely logged data about system usage, quality improvement feedback.

3.2 Participants

Aggregated data on all patient portal users during the study period will be exported from the EHR. Patient portal users can be the patients themselves or others that are authorized to use the portal to access the patient's EHR functionality through portal. They can be authorized by the patient themselves or be the parents of a child that is a patient at the hospital for instance. Additionally, users will be able to provide feedback on the clarifications when they view them. This will thus result in a convenience sample with those users that viewed their problem list, clicked on the info buttons to view the clarifications, and took the effort to provide feedback on the clarifications.

3.3 Study flow

We will analyze usage data about the logins on the patient portal, diagnosis list views, which diagnoses are displayed when users view their diagnosis list, the number of diagnoses with clarifications and which info buttons were clicked by users. Secondly, a feedback question will be asked about the quality of the clarifications if users view a clarification. They are free to choose to provide feedback or not. Therefore, we expect less users to go further in the process of completing these steps, and some users to provide more input about different clarifications. Users may log in, view their diagnosis list, display clarifications and provide feedback multiple times. If users have any questions about the diagnoses or about the clarifications, they can address them towards their doctor.

Steps:

1. Login into patient portal
2. Display diagnosis list
3. Click on info button to display clarification
4. Provide feedback on the clarification displayed

3.4 Outcome measures or evaluation criteria

3.4.1 Overall

The coverage of the clarifications will be measured as the percentage of diagnoses with a clarification that was displayed (number of diagnoses with a clarification divided by the number of diagnoses). The use of the clarification functionality will be measured as the percentage of unique info-buttons clicked compared to total number of info-buttons (number of diagnosis with clarifications) displayed on the diagnosis list. For both measures we will take the median per patient.

3.4.2 User level

We will aggregate user characteristics of users for each step, to compare differences between subgroups that complete each of the four steps described in 3.3. We will distinguish unique user logins, users that view their diagnosis list, users that clicked on any info-button, user ratings, from the number of logins, views, clicks and ratings per user. User characteristics are user type (authorized or patient user), age group and gender.

3.4.3 Diagnosis clarification level

For each diagnosis we will aggregate the number of times they were viewed, the info button was clicked, and collect feedback. Users can rate the quality of the clarifications on a seven-point scale from very bad to very good and optionally describe in a free-text input field why they provided that rating.

Table 2 Outcome measures collected on each aggregation level

Level	Outcome measure	Data source
Overall	Logins	Log data
	Users displaying their diagnosis list	Log data
	Users who clicked on info-buttons	Log data
	Users who provided feedback	Feedback questions
User	User type	Log data
	Age group	Medical record
	Gender	Medical record

	Number of diagnoses registered	Medical record
	Number of logins	Log data
	Number of diagnosis list views	Log data
	Number of problems displayed	Log data
	Number of info-buttons displayed	Log data
	Number of info-buttons clicked	Log data
	Number of clarifications rated	Feedback questions
Diagnosis	Diagnosethesaurus id	Medical record
	Start date	Medical record
	End date	Medical record
	Specialty	Medical record
	Additional diagnosis data: - Status: null, confirmed, rejected, suspected? - Active?	Medical record
	Clarification available? (whether the diagnosis has an info-button or not)	
Diagnosis clarification	Displayed on portal	Log data
	Info-button clicked	Log data
	Clarification quality	Feedback questions
	Feedback	Feedback questions

3.5 Methods for data acquisition and measurement

3.5.1 Audit trail and EHR data reuse

The audit trail contains data on which data were viewed, including what diagnoses were displayed on the patient portal diagnosis list and whether other actions were taken on the patient portal by which user. We will reuse these audit trail data to derive which diagnoses were viewed by patients and for which diagnoses the info button was used. We will also reuse age and gender already registered in the EHR to explore differences in user characteristics. These data will be made available by the participating hospital(s) in aggregated form without any directly identifying personal information.

3.5.2 Feedback

For quality improvement each user will be asked two simple and minimally invasive feedback questions: (1) how bad or good do you find this explanation (very bad 1 – very good 7)? (2) why/can you motivate your score? The questionnaire functionality of the EHR will be used for this purpose. The results can be viewed by the hospital staff using the report functionality of the EHR. It will be ensured that the feedback data will not contain any directly identifying personal details before extracting it for this research.

We expect users to provide this feedback from their personal experience, and thus users might enter personal information or ask personal questions. Therefore, the input will be analyzed by the hospital itself, to assess whether the feedback contains questions that need to be addressed by someone. Moreover, this feedback will be monitored to assess if any issues arise. The hospital can contact the patients to address their question and where necessary a clarification can be corrected or removed, or the functionality may be turned off completely if necessary.

3.6 Methods for data analysis

Results from audit trail, EHRs and feedback data will be made available by the hospital for the purpose of this research in aggregated and anonymized form. This poses some limitations to the statistical data analysis. Consent would otherwise be required to carry out the analysis on individual patient record level. Free text from the questions asked and feedback provided will be anonymized by an authorized functionary from Franciscus. Anonymization will be carried out by removing directly identifying data, such as dates, names of patients, clinicians or others. We will analyze the feedback and questions thematically and summarize them narratively.

Data will be analyzed using R programming language in R studio.

4 Expected results

4.1 Demographic and other study coverage data

Coverage of diagnoses by clarifications. Percentages of patients that viewed their diagnosis list, clicked on info buttons, and provided feedback.

Table 3 Total, median and IQR of logins, views, clicks and ratings

	Total number of unique users Median and IQR	Total number of actions Median and IQR
Logins		
Viewed their diagnosis list		
Clicked on one or more info-buttons		
Provided feedback on one or more clarifications		

Table 4 Characteristics of patients that logged in, viewed problems, used the info-button and provided feedback

Statistic	Category	n (% + CI) logged in	n (% + CI) viewed problems	n (% + CI) used info button	n (% + CI) feedback
User type	Patient				
	Authorized user				
Gender	Male				
	Female				
	Other				
Age group	0 years				
	1-11 years				
	12-15 years				
	16-18 years				
	19 – 29 years				
	30 – 39 years				
	40 – 49 years				
	Etc.				
Number of diagnoses	0				
	1				
	Etc.				
Specialties					
Totals					

4.2 Unexpected events during the study

Any events occurring during the study that may influence the design or results will be reported.

4.3 Study findings and outcome data

Median and IQR of info buttons clicked after viewing diagnosis list. Median and IQR of the quality of diagnosis clarifications. Thematic analysis of user feedback and quality ratings of diagnosis clarifications.

4.4 Unexpected observations

Any observations that were not expected will be reported, such as unintended effects or additional insights. User feedback will be monitored weekly by an authorized functionary of Franciscus to be able to intervene if unexpected issues arise with the clarification functionality.

5 Discussion

5.1 Strengths and limitations

Without permission of the users, we cannot obtain individual patient data to run a mixed effects model. Therefore, this research is limited to aggregate data from reports of the EHR. However, the aggregate data will provide insight into different user groups. The approach of reusing existing log and EHR data provides a more representative picture of users than making patients or laymen fill out (long) surveys. The brief quality ratings are minimally invasive for end users. Sometimes another person might use the account of some person to access the patient portal, for example a parent using their child's account, instead of logging in as an authorized user. We cannot verify this from the data itself.

5.2 Meaning and generalizability of the study

This study will provide insight into patient portal user information needs by measuring the actual use of a clarification functionality for the diagnosis list. The feedback on the clarification quality can be used to improve the clarifications. Additionally, this study will show up to what extent generalization can be useful to generate quality clarifications from the perspective of actual patient portal users. Furthermore, the type of questions asked and the feedback provided will provide more insight into the information needs of the patients.

5.3 Conclusions

PFE: Patient-friendly Extension

6 Declarations

6.1 Ethics approval and consent to participate

A waiver from the Medical Research Ethics Committee of Amsterdam UMC, location AMC was obtained and filed under reference number W21_259. It confirmed that the Medical Research Involving Human Subjects Act (in Dutch: WMO) does not apply to the study and that therefore an official approval of this study by the ethics committee was not required under Dutch law. Approval from the Data Protection Officers of the participating organizations will be obtained.

6.2 Consent for publication

No consent for publication will be required, because no individual person's data will be published.

6.3 Availability of data and materials

Data and materials will be available upon request.

6.4 Competing interests

HM and GH are employed by ChipSoft. ChipSoft is a software vendor that develops the health information system HiX. The authors declare that they have no further competing interests.

6.5 Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

6.6 Authors' contributions

6.7 Acknowledgements

7 Attachments

7.1 Attachment A. Feedback question

English translation

- What do you think of this explanation?
  - 1 2 3 4 5 6 7
  - Very bad O O O O O O O Very good
- Why? Can you elaborate on your score?

Original Dutch

- Hoe vindt u deze uitleg?
  - 1 2 3 4 5 6 7
  - Heel slecht O O O O O O O Heel goed
- Waarom? Kunt u uw score toelichten?

7.2 Attachment B. Data extraction from database and reports from EHR Report Builder functionality (to be developed)

7.2.1 Research database

On the acceptance environment of the hospital a database will be placed with the required tables for the data extraction. The tables come from the audit trail and the production environment and are updated weekly. Unnecessary columns will be left out, to limit the personal data. From these tables, the data-extraction will be carried out with SQL queries. The results will be verified by an authorized functionary of the hospital. Data are shared through SURF Filesender.

7.2.2 Extracted data

7.2.2.1 Logins, views, clicks and ratings

Login data will only be one column with the number of logins per patient. Similar for number of views, clicks and feedback. For example, raw data will be: 2,2,1,22,1,1,5,2,1,1,2,7,1,13,6

7.2.2.2 Aggregate personal data

A table with variable, value and numbers to create Table 4, e.g. variable, value, logins, views, clicks, ratings; gender, male, 1000, 100, 10, 1; gender female; 1500, 150, 15, 2.

7.2.2.3 Medical specialties

Medical specialities will be aggregated for each step with the number of patients that is treated by that speciality, e.g. cardiology,500,50,5,0;neurology,100,10,1,1. This data will be used to create Table 4.

7.2.2.4 Diagnosis data

We will retrieve aggregate diagnosis data per diagnosis id about the number of patients with the diagnosis that logged in, number of views, number of clicks and number of ratings per diagnosis.

For example, Table 5 with the number of diagnoses, views, clicks and ratings. Each diagnosis has a Diagnosethesaurus ID and a description.

We will also aggregate diagnoses viewed by registration type and codelist, such as Table 6 Fictitious data about patients, diagnoses, views, clicks and ratings per registration type

Table 5 Fictitious diagnosis data with number of patients, diagnoses, views, clicks and ratings per diagnosis

id	Description	patients	diagnoses	views	clicks	ratings
7195	Gonartrose	60	65	5	2	1
11563	multipele orgaanstoornissen	55	60	4	2	1
11338	globusgevoel	54	55	4	1	1
…	…	…	…	…	…	…
6023	zwarte haartong	2	2	0	0	0
11502	aangezichtspijn	1	1	1	1	1
83801	aanpassingsstoornissen - met angst	1	1	1	1	0
37203	ataxia teleangiectasia	1	1	1	1	1

Table 6 Fictitious data about patients, diagnoses, views, clicks and ratings per registration type

RegistrationTypeId	RegistrationType	patients	diagnoses	views	clicks	ratings
MA	General points of attention	1000	1500	5000	-	-
C	Complication	2000	3000	10000	-	-
D	Diagnosis	4000	8000	30000	400	40

7.2.2.5 Feedback data

Ratings and feedback will be obtained per diagnosis. Directly identifying data such as names (between brackets), dates and places will be removed from the free-text feedback by an authorized functionary of Franciscus. To distinguish feedback from unique users, a userId will be generated with an auto-number (1, 2, 3 and so on).

Table 7 Fictitious feedback data. Text between brackets are directly identifying data, such as the dates, names and places that will be removed for anonimization

id	userId	description	rating	feedback
37203	1	ataxia teleangiectasia	1	Dit is veel te vaag en te algemeen er staat in de uitleg helemaal niet wat

				juist belangrijk bij deze aandoening is. Ik loop er al mijn hele leven mee rond, en het werd gediagnosticeerd werd in [1992] in [Rotterdam]… Anders had ik het ook gewoon kunnen vragen aan mijn arts of googlen.
7195	1	gonartrose	3	-
11563	1	multipele orgaanstoornissen	4	het klopt wel, maar het zegt niets, dat moet toch beter kunnen… ik heb wel wat anders aan mijn hoofd
6023	2	zwarte haartong	7	Goed, moeten ze overal doen!
7195	2	gonartrose	2	wat betekent dat eigenlijk voor mij? Ik ben helemaal niet oud, waarom heb ik het dan? Dat had [dokter Jansen] mij niet verteld.
11338	3	globusgevoel	7	Waarom schrijven ze dan niet gewoon ook brok in je keel
5047	4	presbyacusis	6	Wat voor soort slechthorendheid is het dan? Er staat ‘een soort’, is het dan geen slechthorendheid, maar slechthorendheid ofzo?
11563	1	multipele orgaanstoornissen	2	-

8 References

1. van Mens HJT, de Keizer NF, Nienhuis R, Cornet R. Clarifying Diagnoses to Laymen by Employing the SNOMED CT Hierarchy. Stud Health Technol Inform. 2018;247:900-4. PMID:29678091.
2. van Mens HJT, Martens S, Paiman M, Mertens S, Nienhuis R, de Keizer NF, et al. Diagnosis clarification by generalization to patient-friendly terms and definitions: A validation study. Forthcoming.

3. Franciscus. Jaarverslag 2019 (Annual Report). 2019. Retrieved from https://franciscus.maglr.com/franciscus-jaarverslag-2019/jaarverslag on April 5, 2021.