Decision Support Systems and Education
J. Mantas et al. (Eds.)

doi:10.3233/978-1-61499-921-8-10

Caveats for the Use of the Active Problem List as Ground Truth for Decision Support

Werner CEUSTERS [a b,1] and Jonathan BLAISURE [a b]
[a] *Department of Biomedical Informatics, University at Buffalo, Buffalo, NY, USA.*
[b] *Institute for Healthcare Informatics, University at Buffalo, Buffalo, NY, USA.*

Abstract. Diagnoses recorded on the problem list are increasingly being used for decision support applications. To obtain insight in the adequacy of the clinical user interface to capture what the clinician has in mind, and to reconstruct the clinical reality of the patient, we analyzed in the database of an EHR system the transactions that resulted from managing the problem list. Our findings indicate (1) that caution is required when using the evolution of the problem list for determining comorbidity or ongoing disease, and (2) that similarities or differences in problem list annotation sequences do not always correspond with similarities resp. differences in disease courses. It is to be investigated whether automatically identifiable subsets of problem list evolution patterns exist from which ground truth reliably can be inferred or whether clinicians need more education in how problem list user interfaces should be used to avoid erroneous interpretations by clinical decision support applications.

Keywords. Medical Records, Problem-oriented; Diagnosis, Computer-Assisted

1. Introduction

Data generated in clinical practice and originally stored and managed in electronic healthcare records (EHR) with the primary goal to support direct patient care become more and more combined in multi-center repositories that adhere to common data models (CDM) [1]. However, the many biases and idiosyncrasies that EHR data suffer from [2] form a barrier to use these data for EHR-integrated decision support applications at the point of care

The problem list is the most utilized resource in point of care decision support applications for suggesting treatment, generating medication alerts or differential diagnoses, and also for adequate maintenance of the list itself [3]. Maintaining a problem list is in the USA a 'meaningful use' criterion, but keeping the list complete and correct is a major problem despite the narrow definition as '*a list of current and active diagnoses as well as past diagnoses relevant to the current care of the patient*' [4,5]. This is far removed from Weed's original conception as '*a complete list of all the patient's problems, including both clearly established diagnoses and all other unexplained findings that are not yet clear manifestations of a specific diagnosis, such as abnormal physical findings or symptoms*' [6]. Also adequate automatic interpretation of what the list contains so that it could be used as a reliable resource for decision support, is still a major challenge [4].

[1] Werner Ceusters, Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, 77 Goodell street, Suite 540, Buffalo NY 14203, USA; E-mail: ceusters@buffalo.edu.

2. Background

The University at Buffalo's Institute for Healthcare Informatics' (IHI) aggregates fully identified healthcare data sets from distinct sources into a centralized environment. The long-term goal is to have this data repository maximally explicit and self-explanatory [7]. The strategy involves: (1) to identify, and where possible remove, ambiguities, (2) to represent explicitly what is implied in certain assertions in the source EHRs, and (3) to identify and mark inconsistencies and incoherencies. Although it has been claimed rightfully that from assertions in the medical record one can only infer by approximation what is, or has been '*the true state of the patient*' [8] and that diagnoses are specifically prone to biases and inconsistencies [2], it is possible to construct a uniform representation which constrains the diagnostic interpretations about the ground truth by keeping track of the data and what the data are about [9]. Obstacles encountered thus far include: (1) lack of documentation about the data models of commercial EHR systems [10]; (2) doubts about the completeness and accuracy of turnkey data extraction tools as witnessed by observations that using different tools on the same EHR yields different results [11], and (3) information distortion resulting from oversimplified CDMs [12].

The goal of the work reported on here was to obtain a first impression of the extent to which the transactions registered in the EHR's database resulting from managing the problem list provide insight in the adequacy of the clinical user interface (1) to capture what the clinician has in mind, and (2) to reconstruct the clinical reality of the patient.

3. Methodology

The user-interface of the EHR system studied allows users to create items on the problem list and manipulate existing ones such as activating or deactivating problems, updating the chronicity, declaring a problem to be resolved and 'transitioning' a problem into another one. Any such manipulation cannot be undone, but can be marked as having been entered in error. Adding a problem creates for that patient a new record in both the *problem header* (PH) and *problem instance* (PI) tables of the EHR's database. Any further manipulation creates a new PI record, and updates the *last modified date* in the corresponding PH record. When a clinician transitions a problem into another one – e.g. angor pectoris into myocardial infarction – a new record is created in a *problem transition* (PT) table indicating the direction of the transition.

We collected all the records from these tables. We identified in the PI records the data fields for which a change in value from one record to the next would qualify as a change in the corresponding problem; e.g. a problem first marked as *active* and then *inactive* would count as a problem change, while a mere change in documenting clinician or date of observation would not. Whenever 2 subsequent PI-records exhibited a value change in any of these data fields, a *PH-change record* was created. We then computed tallies of the various sorts of problem changes that occurred over time. We looked for odd change patterns, e.g. repetitive activations and de-activations of the same problem, transitions of problems into existing older problems, …, or lack of change where expected (e.g. problems initiated as chronic and inactive without any further change). We investigated whether problems were over time further documented – e.g. whether clinical notes were added – without being marked as changed. We selected a few examples for a complete reconstruction of (1) the change history as documented, and (2) what might be inferred to have happened on the side of the patient, modulo ambiguities.

4. Results

Out of the 6 million PI records covering in total 550,000 patients, 764,333 PI records (13%) covering 79,881 patients (15%) were found to mark at least one problem change. From these records, 394,722 PH-change records were created, documenting for 369,611 individual problems one or more problem changes for a total count of 663,791. 94.52% of these problems underwent only one change, 4.42% two, and 0.93% three. The 499 remaining problems (0.13%) underwent four or more changes, maximum 34. Inspection of the PI-records revealed that problem changes can happen along 3 dimensions simultaneously: (1) a '*problem status*' dimension with 7 possible values accounted for 299,122 changes (45%), (2) a '*problem type*' dimension with 6 possible values which in 99.94% of PI-change records was either set to 'chronic' or 'acute' accounted for 138,800 changes (21%) and (3) a '*problem category*' dimension with 29 possible modifiers such as 'history of', 'risk for', etc., accounted for 225,879 changes (34%). Table 1 shows the distribution of problem changes along the problem-status dimension.

Table 2 summarizes the change history of 6 cases (involving 3 different types of disorders) selected for exhibiting odd problem change patterns along the problem-status dimension: (1) A1 and A2 for the high number of transitions, percentages for 1, 2 and more than 2 transitions being 94.2%, 5.0% and 0.8% respectively for a total of 23,162 problems (3.9% of all problems); (2) C1, C2 and B2 – B1 being selected for comparison with B2 – for a repetitive 'active/resolved' pattern which occurred in 2,910 problems (.008%). Another odd pattern, multiple repetitions of 'active/inactive' was found in only 40 problems and is not further discussed here.

Table 1. Changes in problems along the problem-status dimension.

To: / **From:**	**Active**	**Resolved**	**Entered in error**	**Transitioned**	**Denied**	**Inactive**	**Assumed resolved**	***FROM* TOTAL**
Active	-	**80.9%**	10.5%	7.1%	0.4%	1.1%	0.0%	**237,160 (79.3%)**
Resolved	38.8%	-	**56.8%**	1.1%	3.3%	0.0%	0.0%	**59,713 (20.0%)**
Denied	10.3%	4.1%	**85.4%**	0.1%	-	0.1%	0.0%	**2,009 (0.7%)**
Assumed resolved	**60.0%**	0.0%	40.0%	0.0%	0.0%	0.0%	-	**5 (< 0.1%)**
Entered in error	**41.5%**	31.7%	-	0.0%	17.1%	9.8%	0.0%	**41 (< 0.1%)**
Inactive	**96.7%**	0.0%	3.3%	0.0%	0.0%	-	0.0%	**184 (0.1%)**
***INTO* TOTAL**	**23,543 (7.9%)**	**191,945 (64.2%)**	**60,642 (20.3%)**	**17,418 (5.8%)**	**2,959 (1.0%)**	**2,605 (0.9%)**	**0 (0.0%)**	**299,112 (100%)**

Legend: From 'X' \To 'Y' percentages are relative to the FROM 'X' TOTAL; *FROM* TOTAL and *INTO* TOTAL percentages are relative to the overall total.

5. Discussion

Despite the lack of documentation concerning the database structure of the EHR under scrutiny, it turned out to be possible to reconstruct the sequence of events in which items on the problem list are manipulated. However, some of these sequences raise questions about how clinicians organize the list. What is presented here are only a few illustrative examples selected from a set of cases in which in the sequence of documented events none was ever marked as having been entered in error, annotated as being a working hypothesis, or modified with an uncertainty code.

The evolutions of the problem lists of cases A1 and A2 (Table 2) clearly indicate disease progressions consistent with diabetes. However, at certain points in time, the accuracy of the problem list with respect to the ground truth can be questioned.

Table 2. Evolution of the active problem list of six patients.

PtID	PrID	Problem	EHR update event 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A1	1	Type II DM - Uncomplicated, Controlled	N	D			D											
	2	Type 2 DM - Uncomplicated, Uncontrolled			N													
	3	Type 2 DM With Manifestations - Uncontrolled					N	D										
	4	Type 2 DM, uncontrolled, with neuropathy							N	D								
	5	Controlled type 2 DM with neurological manifestations									N	D						
	6	DM type 2, uncontrolled w/neurologic complication											N	D	D			
	7	DM type II, controlled, with no complications															N	D
A2	1	Prediabetes	N						D		D			D				
	2	Impaired Fasting Glucose		N			D		D		D		D					
	3	Diabetes Mellitus			N	D		R								A		
	4	Type II diabetes mellitus, well controlled																N
B1	1	Cerumen Impaction In Both Ears	N	D			D	D	D	D	R	A						
	2	Impacted cerumen of right ear										N						
B2	1	Cerumen Impaction	N	D	R	A	R	A	R	A	D	D						
C1	1	Acute Sinusitis	N	D	D	R	A	R	A	D	R	A						
C2	1	Acute Sinusitis	N	D	D	D	D	D	R	A	D	D	R	A	R	A	D	

Legend: 'PtID': patient identifier; 'PrID': problem header identifier; 'N': Problem header added and marked active; 'D': problem documentation added; 'A': problem header marked active; 'R': problem resolved; Arrow: problem 'transitioned' into the problem pointed at by the arrow. Solid background: problem 'active'.

During update event 2 and 3 for patient A1, the list states that the patient has two problems simultaneously: controlled and uncontrolled diabetes. This is rectified at event 4 by transitioning the former into the latter. In similar vein, patient A2 is stated to have at the same time diabetes and prediabetes, the former, oddly, being documented as resolved at event 5. Also alternatingly documenting the problems 'prediabetes' and 'impaired fasting glucose' (IFG) is odd since IFG is one of the two alternative defining criteria for prediabetes, thus only one disease entity is involved. Our hypothesis that the annotations were created by two distinct providers turned out to be wrong. The situation is rectified at event 11 but in a strange manner: by transitioning one into the other, rather than by stating that one was in error as it was never the case that the patient had prediabetes and impaired fasting glucose as two comorbidities.

Also the problem list entries and subsequent documentations of the patients with impacted cerumen (B1 and B2) seem to divert from what would be good practice, be it in slightly different ways. Clearly, cerumen impaction is an entity which is typically treated when diagnosed and after treatment ceases to exist. The logical approach would thus be to mark the disorder as resolved and to deactivate it as a problem. Yet, in case of B1, the problem is kept active despite several treatments over the years, until it is finally declared resolved at event 7, but reactivated at event 8. In case of B2, the problem is repeatedly declared to be resolved and then reactivated whenever a new instance of this type of disorder comes into existence. The problem is here, so we believe, that the clinicians in these cases don't do justice to what it means for a problem to be 'active'. Active problems are those that must be referenced on the problem list either because (1) the patient is currently suffering from them or (2) they existed in the past, but were of a type that impacted the future health of the patient. In both cases here the problems stay on the problem list without all the time satisfying either of the two conditions. Granted, overly frequent cerumen impaction can itself be viewed as an active problem, but individual impactions are, once resolved, clearly no threat for a patient's future health. Similar considerations can be made for the two patients with recurrent episodes of acute sinusitis (C1 and C2) in which case one can also wonder why 'recurrent sinusitis' was not added as a chronic problem with several exacerbations in the form of acute sinusitis.

6. Conclusion

Clearly, whenever a diagnosis is added to the problem list – except for entries marked later as entered in error or bearing uncertainty – some disorder related entity of a type suggested by the diagnostic label must exist or have existed on the side of the patient; such fact can be used for simple decision support applications. But caution is required if one seeks to use the evolution of the problem list as a source to infer the number of such entities that exist as comorbidities, or whether disease entities once referenced still exist when subsequent annotations are added to existing problems. Our findings also suggest that providers use in different ways the facilities offered by the clinical user interface to further qualify problem list entries and the changes thereof. The consequence is that such differences cannot be taken to reflect distinct types of disease courses, nor that similarities in such documentation sequences reflect similar disease courses. Therefore, caution is also required when attempting to use the problem list to compute such disease courses. It is further to be investigated (1) whether there exist automatically identifiable subsets of problem list evolution patterns from which ground truth reliably can be inferred, (2) whether clinicians need to be more thoroughly educated in how problem list user interfaces should be used to *avoid* mistakes of the sort discussed, and (3) whether displaying the complete history of the problem list may help in *detecting* mistakes.

Acknowledgement

This work was supported in part by Clinical and Translational Science Award NIH 1 UL1 TR001412-01 from the National Institutes of Health.

References

[1] V. Huser, J.J Cimino, Desiderata for healthcare integrated data repositories based on architectural comparison of three public repositories, *AMIA Annual Symposium proceedings* (2013), 648-656.
[2] W.R Hersh, M.G. Weiner, P.J. Embi, J.R. Logan, P.R. Payne, E.V. Bernstam, H.P. Lehmann, G. Hripcsak, T.H. Hartzog, J.J. Cimino, J.H. Saltz, Caveats for the use of operational electronic health record data in comparative effectiveness research, *Medical care* **51** (2013), 30-37.
[3] R.A. Greenes, *Clinical decision support : the road to broad adoption*, Elsevier/Academic Press, London UK/Waltham MA, USA, 2014.
[4] A. Wright, A.B. McCoy, T.T. Hickman, D.S. Hilaire, D. Borbolla, W.A. Bowes, W.G. Dixon, D.A. Dorr, M. Krall, S. Malholtra, D.W. Bates, D.F. Sittig, Problem list completeness in electronic health records: A multi-site study and assessment of success factors, *Int J Med Inform* **84** (2015),784-790.
[5] C.M. Hodge, S.P. Narus, Electronic problem lists: a thematic analysis of a systematic literature review to identify aspects critical to success, *J Am Med Inform Assoc* **205** (2018), 603-613.
[6] L. Weed, Medical records that guide and teach. *New England Journal of Medicine* **278** (1968), 593-600.
[7] W. Ceusters, B. Smith, Y.H. Chiun, Clinical Data Wrangling using Ontological Realism and Referent Tracking, *CEUR Workshop Proceedings* **1237** (2014), 27-32.
[8] A.L. Rector, W.A. Nowlan, S. Kay, Foundations for an electronic medical record, *Methods Inf Med* **30** (1991), 179-186.
[9] W. Hogan, Representing the aboutness of a diagnosis, *Stud Health Technol Inform* **247** (2018), 805-809.
[10] J.P. Bona, W. Ceusters, Replacing EHR structured data with explicit representations, *CEUR Workshop Proceedings* **1515** (2015), 85-86.
[11] S.T. Liaw, J. Taggart, H. Yu, S. de Lusignan, Data extraction from electronic health records - existing tools may be unreliable and potentially unsafe, *Aust Fam Physician* **42** (2013), 820-823.
[12] W. Ceusters, J. Blaisure, A Realism-Based View on Counts in OMOP's Common Data Model, *Stud Health Technol Inform* **237** (2017), 55-62.