Is the problem list in the eye of the beholder? An exploration of consistency across physicians

John C Krauss,[1] Philip S Boonstra,[2] Anna V Vantsevich,[3] and Charles P Friedman[4]

ABSTRACT

Objective Quantify the variability of patients' problem lists – in terms of the number, type, and ordering of problems – across multiple physicians and assess physicians' criteria for organizing and ranking diagnoses.

Materials and Methods In an experimental setting, 32 primary care physicians generated and ordered problem lists for three identical complex internal medicine cases expressed as detailed 2- to 4-page abstracts and subsequently expressed their criteria for ordering items in the list. We studied variability in problem list length. We modified a previously validated rank-based similarity measure, with range of zero to one, to quantify agreement between pairs of lists and calculate a single consensus problem list that maximizes agreement with each physician. Physicians' reasoning for the ordering of the problem lists was recorded.

Results Subjects' problem lists were highly variable. The median problem list length was 8 (range: 3–14) for Case A, 10 (range: 4–20) for Case B, and 7 (range: 3–13) for Case C. The median indices of agreement – taking into account the length, content, and order of lists – over all possible physician pairings was 0.479, 0.371, 0.509, for Cases A, B, and C, respectively. The median agreements between the physicians' lists and the consensus list for each case were 0.683, 0.581, and 0.697 (for Cases A, B, and C, respectively).Out of a possible 1488 pairings, 2 lists were identical. Physicians most frequently ranked problem list items based on their acuity and immediate threat to health.

Conclusions The problem list is a physician's mental model of a patient's health status. These mental models were found to vary significantly between physicians, raising questions about whether problem lists created by individual physicians can serve their intended purpose to improve care coordination.

RESEARCH AND APPLICATIONS

INTRODUCTION

In 1968, Lawrence Weed proposed the concept of the "problem list," a transformative idea with the goal of helping physicians treat their patients more effectively and efficiently. Dr. Weed called the problem list a "dynamic 'table of contents' of the patient's chart, which can be updated at any time . . . The essential combination of clarifying single problems and integrating multiple problems is greatly facilitated by a medical record that is structured around a total problem list."[1] In today's world, in which healthcare is much more fragmented than in Dr. Weed's time, the problem list has taken on an even greater significance. In particular, it holds the potential to greatly improve care coordination by supporting the sharing of patient statuses during transitions of care between providers and between care environments.

In testimony to their perceived value, problem lists are central to the accreditation criteria of the Joint Commission[2] and were a key objective of Meaningful Use policy.[3] In the context of care coordination, the problem list is a communication vehicle that, ideally, would ensure that providers assuming the care of a patient inherit the referring providers' understanding or "mental model" of the patient's condition. However, there is little reason to assume that a free-text problem list, expressed in each provider's own words, would be a conduit for such shared understanding of a patient's health status. A recent report by Adler-Milstein et al. illustrated how sharing problem lists does not automatically create coordination of care between providers: "practices differed in whether or not they included acute problems on the problem list and short-term medications on the medication list."[4]

In general recognition of these challenges, other studies have addressed the institutionalization of problem lists,[5] their utilization by outpatient providers in academic and community healthcare settings,[6] physician attitudes towards problem lists,[7–9] physician acceptance of approaches designed to improve the accuracy of electronic problem lists,[10] as well as the variance in problem list utility,[11] accuracy,[12] and encoding[13] by different types of physicians. In aggregate, these studies support the contention that the problem list's utility cannot be assumed. Other authors have proposed mechanisms to augment the problem list's utility: organizational policies[14,15] and interventions,[16] coded terminology,[17–19] automated inference from free text,[20,21] clinical decision support[22,23] and Natural Language Processing,[24–26] improving the design and visualization of problem lists,[27,28] and using a Wiki-model for problem lists.[29] However, these ideas have largely been realized only in isolated implementations and preliminary tests.

The point of departure for this study is the proposition that the existing literature has overlooked a fundamental set of questions that must be addressed before any approaches can be successfully implemented to realize the problem list's potential as a care coordination tool. The applicability of the problem list for patient care planning and coordination tacitly assumes that problem lists are objective (ie, that all physicians will agree, for a given patient, on what that patient's problems are, on their order of importance, and on what criteria they should be ranked). Recognizing that the problem list is fundamentally an abstract representation of the patient, constructed in the mind of each provider, we must understand to what extent these subjective appreciations vary naturally across providers, in terms of their content

Correspondence to John C Krauss, Division of Hematology/Oncology, Department of Internal Medicine, University of Michigan Medical School, 3-219 Cancer Center, 1500 E Medical Center Dr, Ann Arbor, MI, 48109, USA; jkrauss@med.umich.edu For numbered affiliations see end of article.

and organization. Does it make sense to talk about *THE* problem list for each patient, or, at the opposite extreme, is the number of valid problem lists for a patient equal to the number of providers who care for that patient? Some authors have argued that standardizing content terminology and the granularity of problem expression are the key strategies to unleash the problem list's potential to support care coordination.[30] However, if physicians appreciate their patients in fundamentally different ways, merely standardizing the way these perceptions are expressed will not achieve this purpose, making a deeper approach to this challenge necessary.

One approach to address this question is to explore through experimentation how different physicians construct the problem list of one patient encountered at the same point in time. By describing how physicians intuitively construct and organize problem lists and quantifying the extent of the differences between them, we can begin to identify approaches and build tools that will help the problem list realize its potential to improve care.

Accordingly, this study addresses four interconnected research questions: (1) How various are the lengths of physician-subjects' problem lists and to what extent does problem list length vary by subject and case?; (2) Taking into account the length, content, and ordering of problem lists, how similar are the physician subjects' problem lists overall?; (3) How much consensus exists within a statistically inferred "consensus problem list"?; and (4) What are the predominant criteria used by physicians to rank items in the problem list?

MATERIALS AND METHODS
Study Design
The study engaged a sample of primary care physicians ($n = 32$) in constructing problem lists for three identical, diagnostically challenging clinical cases that were presented to the physician subjects in 2- to 4-page written narratives. The study was approved by the institutional review board (IRB) of The University of Michigan. As described below, the task of constructing the problem lists was minimally constrained, to create the most valid representation of each subject's mental model of the case while still allowing for comparisons of the problem lists across physicians. After identifying and ordering the problems in lists for all three cases, the physician subjects were asked to describe the criteria they used to order the problems they listed.

Case Material
Three cases were selected for this study from a larger set developed for research on physicians' clinical reasoning that have been employed in prior published research studies.[31,32] The cases were selected based on their relevance to outpatient medicine as well as for varying complexity and clinical domain. The cases were described in 2- to 4-page written narratives.

Case A involved a patient presenting with a chief complaint of shortness of breath. The patient was found on initial evaluation to have a fever, rales, and splenomegaly. The patient's laboratory examinations were significant for hypoxia, anemia without iron deficiency, thrombocytopenia, and sputum positive for gram-positive diplococci, and a chest X-ray showed bibasilar infiltrates. The patient's past medical history included diet-controlled diabetes mellitus, anxiety with depression, and post-menopausal symptoms requiring hormone replacement therapy.

Case B involved a patient presenting with a chief complaint of fever, sore throat, and facial pain. The patient was found on initial evaluation to appear acutely ill and to have encephalopathy, proptosis, cranial nerve palsies, hepatomegaly, and ascites. The patient's laboratory examinations were significant for hyperglycemia with acidosis, renal failure with hyponatremia and hyperkalemia, cerebrospinal fluid with increased

neutrophils, and ascites with increased neutrophils, and a computed tomography (CT) scan showed a right maxillary sinus mass and a right orbital fracture. The patient's past medical history included disabling schizophrenia, alcohol and intravenous drug use in remission, active smoker, alcoholic cirrhosis, and a laparotomy for a gunshot wound.

Case C involved a patient presenting with a chief complaint of shortness of breath. The patient was found on initial evaluation to have hypertension, a systolic ejection murmur, and a blind right eye. The patient's laboratory examinations revealed iron deficiency anemia and a urinary-tract infection, and a chest X-ray showed a pleural effusion and a pericardial effusion, which was bloody on pericardiocentesis. The patient developed a diastolic murmur after pericardiocentesis. The patient's past medical history included hypertension as well as a hysterectomy and bilateral oophorectomy.

Subjects
Physician subjects were recruited by e-mail invitations. All members of the general internal medicine faculty and family medicine faculty at the University of Michigan who primarily see outpatients ($n \sim 150$) were invited to participate in the study. A total of 38 physicians agreed to participate in the study, and informed consent was subsequently obtained per the IRB-approved protocol.

Pilot Study and Standardized Representation of Common Problem List Items
A pilot study employing the first six physician subjects, whose data were not included in the final analyses, served several purposes. It refined the experimental procedure and verified that the subjects were challenged by the selected cases. It also identified, for each case, the problems that were likely to be included on many of the subjects' problem lists in the subsequent experiment. This allowed for the construction of a standardized representation, using the Systematized Nomenclature of Medicine (SNOMED) and International Classification of Diseases – 9 (ICD-9) terminology, of the commonly listed items on problem lists, to allow for making valid comparisons of the lists across subjects. Without structured language representation of the commonly listed items, there would be no objective basis for deciding whether problems listed by the subjects in their own words should be considered indicative of the same underlying clinical concept.

Experimental Method
The remaining 32 subjects were engaged in the study protocol that resulted from the pilot study. Study sessions with each physician (all conducted by the first author – J.C.K.) lasted from 30–90 min. The cases were presented in a fixed order (A, B, C), with the most complex case (Case B), as determined from previous studies, being second in the order.[32] For each of the three patient cases, physicians were first asked to review the case and then generate a written problem list in their own words. Each physician subject's "natural language" problem list for each case was then transferred to cards, with each problem represented on a separate card. For each item in a subject's natural language list that, in the investigator's opinion, potentially corresponded to an item for which a standardized representation already existed, the investigator offered the standardized representation, on a pre-printed card, as a substitute. This was then followed by asking clarifying questions, such as: "You listed 'estrogen replacement therapy,' would you agree that this is equivalent to 'post-menopausal symptoms requiring hormone replacement therapy'?" If the subject accepted the substitution, a card with the standardized representation was employed. Otherwise, the item was listed on a card in the subject's own words. Once all of a physician subject's items for a case

were listed on cards, the subject was asked to organize the cards into an ordered list. In the resulting semi-standardized lists, items listed on the pre-printed cards were considered to represent the same clinical concept for analytical purposes, and items retained in natural language were considered to represent different concepts.

To explore the reasoning that physicians used to order their problem lists, they were asked: "What criteria did you use to order the diagnosis in the problem lists that you generated?" Some physicians requested prompting as to organizational schemes. When this occurred, the investigator would give the example, "One way of ordering the problems would be alphabetical." Responses were recorded by dense notes taken during the session and annotated immediately afterward.

Analysis

To address the first research question, we calculated and descriptively analyzed the lengths of the problem lists generated by the physician subjects. The impact of each physician and case on the number of problems in each list was estimated using a log-linear regression model for the number of problems listed in excess of one as a function of random physician and case effects.

To address the second research question descriptively, we initially visually summarized all of the subjects' problem lists for each case with a histogram, capturing both the frequency with which a given item was listed across all subjects and the rank order in which that item appeared in the subjects' lists.

We then created an index of similarity for any pair of problem lists for a given case. This index is a novel extension of the rank-biased overlap (RBO) measure described previously in the literature.[33–35] Values of RBO lie in the interval [0,1], with RBO = 1 corresponding to two identical lists and RBO = 0 corresponding to two completely nonoverlapping lists. In the intermediate, RBO is an average of agreements over many depths, where "agreement at depth d" is defined as the proportion of the first d items that two lists have in common. For example, at depth d = 2, agreement can be 0, 1/2, or 2/2, corresponding to two lists having zero, one, or two of their first two items in common. The RBO measure as originally described in the information retrieval literature assumes each list's length to be infinite and thus does not exactly suit the purposes of this study. Our modification of the RBO index, the length-dependent RBO (LDRBO), is still an average of agreements and ranges in value from zero to one, and its measures are interpreted the same as those of the RBO. However, it is tailored to lists that are deliberately finite and vary in length. For a pair of lists, the LDRBO is the average of agreements for all values of d between 1 and the length of the longer of the two lists being compared. (At the point that d exceeds the length of the shorter list, all the items from the shorter list are considered.) Like its predecessor, the LDRBO is "rank-biased," meaning that agreement at a high rank propagates to all subsequent agreements, and agreement on Rank 1 disproportionately increases LDRBO, followed by Rank 2, etc.

To address the third research question, we used the LDRBO to calculate a "consensus problem list" for each case, which is a consolidation of the 32 physicians' lists into a single hypothetical list. We first looked at every possible list (considering any problem listed by at least one physician) of any length and calculated its LDRBO with each of the 32 physicians' lists. The consensus problem list is defined as the list having the maximum median LDRBO across all subjects, and the corresponding maximum value of the median LDRBO may be interpreted as portraying how much agreement exists between the physicians' problem lists for a case. The Supplementary Appendix provides the technical details of our approach.

To address the fourth research question, we qualitatively summarized the physician subjects' descriptions of their rationales for ordering their problem lists. At the completion of the generation of the

problem lists, an interview was conducted about the criteria the physicians used to organize items in the problem lists. The three problem lists were reviewed, and the physicians were asked to name the most important concept for listing the problems. These concepts were subsequently tabulated and displayed graphically.

Figures 1, 2, and all the statistical analyses were done in R,[36] using the R package RColorBrewer.[37]

RESULTS

Research Question 1 – Exploring Problem List Length

The median problem list length and interquartile range for Cases A, B, and C were 8 (5–9), 10 (8–12.5), and 7 (6–9), respectively. The log-linear model estimates the extent to which the number of problems in the lists varied by physician and case. Across all the cases, the overall mean list length was 8.2 problems. Case-specific deviations from this mean, across all physicians, were -0.7 (Case A), 2.4 (Case B), and -1.0 (Case C). Physician-specific deviations, across all cases, varied from -2.3 to 4.2, with an interquartile range of 1.6 on these deviations. This analysis revealed that, for the samples employed in this study, the variances in problem list length from case to case and from physician to physician are of comparable magnitude.

Research Question 2 – Exploring the Overall Similarity of the Problem Lists

Figure 1 plots all the items listed for each case across all subjects, ordered according to frequency of appearance and shaded to indicate where in the subjects' rankings that item was listed. Supplementary Table S1 lists the full names of all the problems in the same order. This figure is visually suggestive of the high degree of variability in the subjects' lists. For example, the proportion of highly unique items – those listed by only one or two physicians – was 12/28, 24/47, and 16/30 for Cases A, B, and C, respectively. The median LDRBO over all possible physician pairings (32 × 31 ÷ 2 = 496) was 0.479, 0.371, and 0.509 for the three cases, respectively, pointing to greater agreement between Cases A and C. Out of all pairings across all three cases (496 × 3 = 1488), in only one instance (occurring in Case A) were identical lists generated by a pair of physicians.

Research Question 3 – Exploring the Consensus Problem List

The computed consensus problem lists for Cases A, B, and C, which contained 8, 15, and 7 problems, respectively, are given in Table 1. Figure 2 compares each physician's problem list for each case with the underlying consensus list for each case, ordered by increasing similarity to the consensus list, and starting with the physician lists that are least similar to the consensus lists. The median LDRBO (ie, the median average overlap) between the consensus list and the 32 physicians' lists for Cases A, B, and C is 0.683, 0.581, and 0.697, respectively.

Research Question 4 – Exploring the Ordering of Items in the Problem List

The creation of any list requires a list ordering structure, which is not always explicitly stated. For example, the problem list could be ordered alphabetically, by time of occurrence of the problem, or by the acuity of the problem. The rationales offered by physicians for ordering their problem lists employed somewhat different language but clustered around the concepts of acuity and immediacy (Figure 3). Sixteen (50%) physicians specifically used the term "acuity" to describe their ranking, and ten (31%) used the phrases "life-threatening," "needs immediate action," or "immediate danger," connoting the general concept of immediacy. The remaining physicians provided "root

Figure 1: Frequency of all the problems listed by the physician subjects, shaded according to ranking within a physician's problem list.
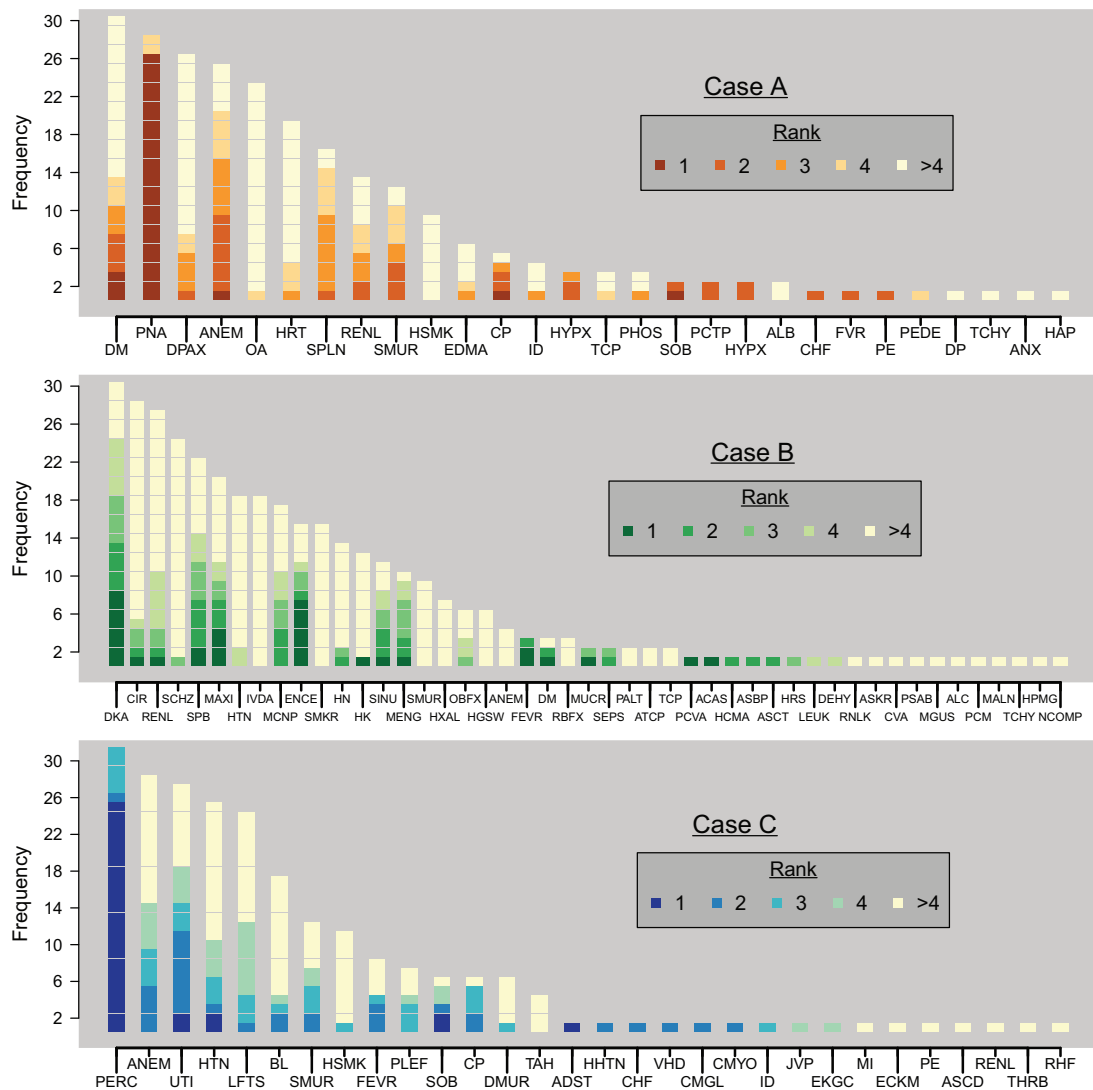
cause" (n = 2), "chief complaint" (n = 2), ongoing management (n = 1), and mortality (n = 1) as their rationale for ordering their lists.

DISCUSSION

Through an experimental protocol, we have formally quantified the variability in physicians' mental models of three complex cases as reflected in ordered problem lists. We have portrayed this variability through an analysis of the length, content, and ordering of the problem lists. Problem list lengths varied by comparable degrees by physician and case. Based on the small sample of cases employed in this study, we did not find a strong tendency for an individual physician to consistently produce either long or short lists. The combined effects of all three of these sources of variability can be captured by the LDRBO index, which can also be used to mathematically combine multiple problem lists into a single consensus list.

For Cases A and C, there was widespread agreement on the top-ranked problem. Beyond the top-ranked item, many problem lists contained the same items but not necessarily in the same order, and, as reflected in the range of list lengths, there was great variability in the

tacit criteria used by physicians to include a problem in the list. Case B is more complex, due to the patient having multiple problems with competing priorities in terms of both acuity and potential mortality. There was greater variability in both the number and nature of the problems listed by each physician in the problem list for Case B.

Overall, the naturally occurring variation in physicians' mental models of a complex clinical case – as reflected in the length, content, and ordering of their problem lists – is sufficient to engender concern about the utility of a natural language problem list as a tool for care coordination. The results of this study can allow us to extrapolate that the problem list that a physician receiving a new patient would be cognitively "expecting," as reflected by the problem list he or she would hypothetically generate for that patient, would likely be very different from the problem list that the referring physician actually transmits. The variability in the problem lists generated in this study by a substantial sample of practicing physicians suggests that, on average, the referring physician's list will contain about 50% of the information – taking into account the length, content, and order of the problem lists, reflected in the LDRBO – that the receiving physician would expect to get. Even though a large

Figure 2: All problem lists, ordered according to agreement with the consensus problem list. Full problem names are given in Table 1. Shaded problems are those also listed in the consensus list, and "x"-ed problems do not appear on the consensus list. The connected line plots LDRBO between each problem list and the consensus list.

RESEARCH AND APPLICATIONS
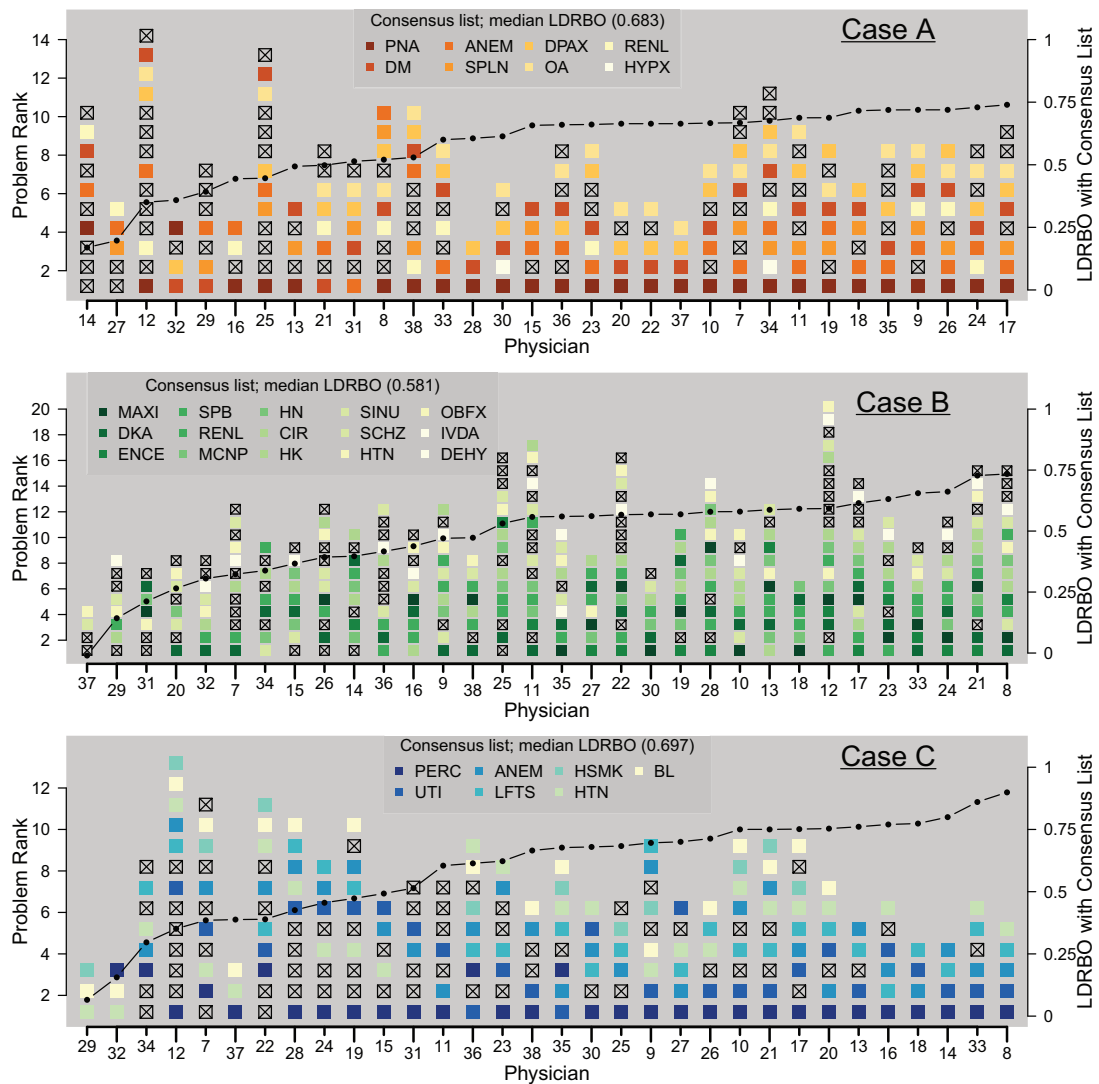
proportion of physician subjects reported using the same general criteria for ordering their problem lists, there was significant variability in where an item (when it was included in multiple lists) actually appeared in the lists. Although these findings confirm assertions that interventions to standardize the problem list are needed to improve their utility for care coordination, our study also suggests that these interventions may need to transcend terminological and format standardization to more cognitively oriented interventions that prompt physicians to think about problem lists in more uniform ways.

The criteria that the physician subjects reported using for ranking items in the problem list, which were focused on acuity and immediacy, were relatively homogenous. However, the observation that items listed by many physicians often appeared in different locations in the problem lists they generated suggests that subjects may have been applying or interpreting these common criteria very differently when constructing their lists.

The problem list can provide a synoptic perspective on the patient's current health status, but the divergent views of physicians on what items to include in the problem list often may not present a clear picture. This uncertainty in the items to be included in the problem list extends to the highest levels of healthcare, with the Office of the National Coordinator defining the problem list as the "current and active diagnosis, as well as the past diagnosis relevant to the current care of the patient," and the Joint Commission naming the problem list the "problem summary list" and defining it as "any significant medical diagnoses and conditions, any significant operative and invasive procedures, and adverse or allergic drug reactions, any current medications, over-the-counter medications, and herbal preparations," and the American Health Information Management Association defining the problem list as "a list of illnesses, injuries, and other factors that affect the health of an individual patient, usually identifying the time of occurrence or identification and resolution . . . including items from organization defined national standards, known significant medical diagnoses and conditions, known

RESEARCH AND APPLICATIONS

Table 1: Consensus Problem Lists

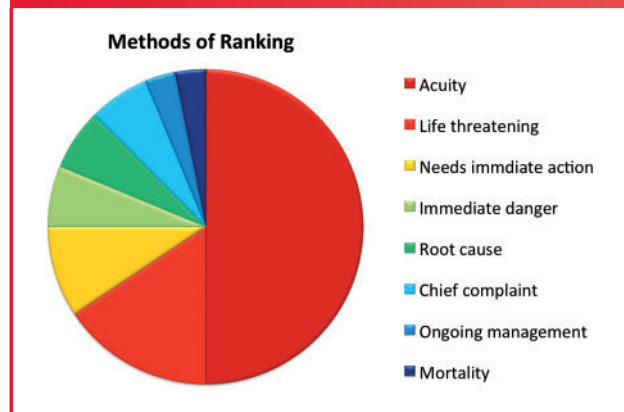| Rank	Case A	Case B	Case C
1	Pneumonia	Maxillary sinus mass	Pericardial effusion
2	Diabetes mellitus	Diabetic ketoacidosis	Urinary tract infection
3	Anemia	Encephalopathy	Anemia
4	Splenomegaly	Spontaneous bacterial peritonitis	Elevated liver function tests
5	Depression with anxiety	Renal failure	History of smoking
6	Osteoarthritis	Multiple cranial nerve palsies	Hypertension
7	Renal failure	Hyponatremia	Blind in right eye
8	Hypoxia	Cirrhosis due to alcohol	
9		Hyperkalemia	
10		Sinusitis	
11		Schizophrenia	
12		Hypertension	
13		Orbit fracture	
14		History of IV drug use	
15		Dehydration	

IV, intravenous.

significant operative and invasive procedures affecting current health, and known adverse and allergic reactions."[14] The lack of agreement on the definition of a problem list leaves each institution to determine its own problem list policy, which makes the interoperability of problem lists between institutions more difficult.

Even with a unified problem list policy in place, there is little incentive in many documentation workflows to curate problem lists. For example, a cardiologist may need to resolve a patient's previously unexplained symptoms of peripheral edema, shortness of breath, and systolic murmur into congestive heart failure caused by mitral regurgitation, but may rely on the patient's primary care provider to update the problem list. The problem list is an essential part of handing off the patient to the next provider, so implementing the suite of ideas that have proven successful for the I-PASS hand-off (Illness severity, Patient summary, Action list, Situational awareness, Synthesis by receiver) may serve as a model for improving the quality of problem lists.[38] Ultimately, each care provider must represent his/her understanding of the overall health status of the patient in the patient's chart and use the problem list to convey that understanding to all the providers involved in the patient's care.

Our study has several limitations. Most important, although the objectives of this study required a constructed experiment, a written synopsis of a case is never the same as seeing and examining an actual patient, and the effects of this experimental intervention are difficult to estimate. Second, our relatively small sample of cases raises

Figure 3: Rationale for ranking of problem lists.

Methods of Ranking

- Acuity
- Life threatening
- Needs immdiate action
- Immediate danger
- Root cause
- Chief complaint
- Ongoing management
- Mortality

important questions about whether the variability of problem lists might vary between different disease domains. We intentionally selected complex cases to reflect those clinical situations in which well-coordinated hand-offs in care transitions might be particularly important. Partially standardizing the terminology used to express items in a problem list obviated the need for investigators to judge whether two related items were clinically synonymous. Offering the physician subjects standardized substitutes for terms expressed in their own words likely biased the results, but this would be a conservative, homogenizing bias. To the extent that this bias existed, the physicians' actual problem lists would be more variable than those reported on here. Third, the manual system we used is different from the workflow that happens in most electronic health records, in which the problem list items are chosen from a large list of symptoms and diagnosis provided in the electronic health record. We plan on replicating this experiment in one or more commercial systems to determine the amount of variability this workflow introduces.

CONCLUSION

Problem lists are physicians' subjective mental models of their patients' health status. We have formally quantified the variability in physician agreement on problem lists. Further study of the generation and organization of the problem list may allow for the creation of problem lists that present a more consistent abstraction of patients' current health status and care needs.

CONTRIBUTORS

J.C.K., P.S.B., A.V., and C.P.F. participated in and are responsible for the data analysis.

COMPETING INTERESTS

None.

SUPPLEMENTARY MATERIAL

Supplementary material is available online at http://jamia.oxfordjournals.org/.

REFERENCES

1. Weed LL. Medical records that guide and teach. *N Engl J Med.* 1968;278(11):593–599.

2. Joint Commission Primary Care Medical Home (PCMH). *Certification for Accredited Ambulatory Health Care Organizations Question & Answer Guide.* http://www.jointcommission.org/assets/1/18/PCMH_Q_and_A_Guide_. Accessed July 4, 2015.

3. Maintain Problem List. *HealthIT.gov.* http://www.healthit.gov/providers-professionals/achieve-meaningful-use/core-measures/patient-problem-list. Accessed July 4, 2015.

4. Adler-Milstein J, Cohen G, Cross D, et al. *Assessing Stage 3 MU Coordination Criteria: Policy Report.* https://www.si.umich.edu/people/julia-adler-milstein. Accessed July 8, 2015.

5. Zhou X, Zheng K, Ackerman M, Hanauer D. Cooperative documentation: the patient problem list as a nexus in electronic health records. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. ACM*; 2012:911–920.

6. Wright A, Maloney F. Understanding problem list utilization through observational techniques. *J Gen Intern Med.* 2010;25 (Suppl 3):S205–S567.

7. Holmes C, Brown M, Hilaire DS, Wright A. Healthcare provider attitudes towards the problem list in an electronic health record: a mixed-methods qualitative study. *BMC Med Inform Decis Mak.* 2012;12:127.

8. Jao C, Hier D, Galanter W, Valenta A. Assessing physician comprehension of and attitudes toward problem list documentation. *AMIA Annu Symp Proc.* 2008;990.

9. Wright A, Maloney FL, Feblowitz JC. Clinician attitudes toward and use of electronic problem lists: a thematic analysis. *BMC Med Inform Decis Mak.* 2011;11:36.

10. Feblowitz J, Henkin S, Pang J, et al. Provider use of and attitudes towards an active clinical alert: a case study in decision support. *Appl Clin Inform.* 2013;4(1):144–152.

11. Wright A, Feblowitz J, Maloney FL, Henkin S, Bates DW. Use of an electronic problem list by primary care providers and specialists. *J Gen Intern Med.* 2012;27(8):968–973.

12. Luna D, Franco M, Plaza C, et al. Accuracy of an electronic problem list from primary care providers and specialists. *Stud Health Technol Inform.* 2013;192:417–421.

13. Rothschild AS, Lehmann HP, Hripcsak G. Inter-rater agreement in physician-coded problem lists. *AMIA Annu Symp Proc.* 2005;644–648.

14. Acker B, Bronnert J, Brown T, et al. Problem list guidance in the EHR. *J AHIMA.* 2011;82(9):52–58.

15. Kadlec L. Resolving problem list problems. HIM's role in maintaining an effective EHR problem list. *J AHIMA.* 2013;84(11):58–59.

16. Bakel LA, Wilson K, Tyler A, et al. A quality improvement study to improve inpatient problem list use. *Hosp Pediatr.* 2014;4(4):205–210.

17. Campbell JR, Xu J, Fung KW. Can SNOMED CT fulfill the vision of a compositional terminology? Analyzing the use case for problem list. *AMIA Annu Symp Proc.* 2011;181–188.

18. Wright A, Feblowitz J, Mccoy AB, Sittig DF. Comparative analysis of the VA/Kaiser and NLM CORE problem subsets: an empirical study based on problem frequency. *AMIA Annu Symp Proc.* 2011;1532–1540.

19. Fung KW, Xu J, Rosenbloom ST, Mohr D, Maram N, Suther T. Testing three problem list terminologies in a simulated data entry environment. *AMIA Annu Symp Proc.* 2011;445–454.

20. Wright A, Pang J, Feblowitz JC, et al. A method and knowledge base for automated inference of patient problems from structured data in an electronic medical record. *J Am Med Inform Assoc.* 2011;18(6):859–867.

21. McCoy JA, McCoy AB, Wright A, Sittig DF. Automated Inference of Patient Problems from Medications using NDF-RT and the SNOMED-CT CORE Problem List Subset. *Poster Session Presented at: AMIA 2011 Annual Symposium*; 2011 October 22–26; Washington D.C.

22. Galanter WL, Hier DB, Jao C, Sarne D. Computerized physician order entry of medications and clinical decision support can improve problem list documentation compliance. *Int J Med Inform.* 2010;79(5):332–338.

23. Wright A, Pang J, Feblowitz JC, et al. Improving completeness of electronic problem lists through clinical decision support: a randomized, controlled trial. *J Am Med Inform Assoc.* 2012;19(4):555–561.

24. Huser V, Fung KW, Cimino JJ. Natural language processing of free-text problem list sections in structured clinical documents: a case study at NIH clinical center. *AMIA Summits Transl Sci Proc.* 2014;2014:178.

25. Solti I, Aaronson B, Fletcher G, et al. Building an automated problem list based on natural language processing: lessons learned in the early phase of development. *AMIA Annu Symp Proc.* 2008;687–691.

26. Meystre S, Haug P. Improving the sensitivity of the problem list in an intensive care unit by using natural language processing. *AMIA Annu Symp Proc.* 2006;554–558.

27. Spry KC. An infographical approach to designing the problem list. *2nd ACM SIGHIT Symp. Int. Health Informatics Symp Proc.* 2012;791–794.

28. Bui AA, Aberle DR, Kangarloo H. TimeLine: visualizing integrated patient records. *IEEE Trans Inf Technol Biomed.* 2007;11(4):462–473.

29. Mehta N, Vakharia N, Wright A. EHRs in a web 2.0 world: time to embrace a problem-list Wiki. *J Gen Intern Med.* 2014;29(3):434–436.

30. Holmes Casey. The problem list beyond meaningful use: part 2: fixing the problem list. *J AHIMA.* 2011;82 (3):32–35.

31. Meyer AN, Payne VL, Meeks DW, Rao R, Singh H. Physicians diagnostic accuracy, confidence, and resource requests, A vignette study. *JAMA Int Med.* 2013;173(21):1952–1958.

32. Friedman CP, Elstein AS, et al. Enhancement of clinicians diagnostic reasoning by computer based consultation, a multisite study of 2 systems. *JAMA.* 1999;282:1851–1856.

33. Fagin R, Lotem A, Naor M. Optimal aggregation algorithms for middleware. *J Comput Syst Sci.* 2003;66(4):614–656.

34. Wu S, Crestani F. Methods for ranking information retrieval systems without relevance judgments In : Proceedings of the 2003 ACM Symposium on Applied Computing, 2003 Mar 9 (pp. 811-816). *ACM Trans Inf Syst.* 2003:6–11.

35. Webber W, Moffat A, Zobel J. A similarity measure for indefinite rankings. *ACM Trans Inf Syst.* 2010;28(4):1–38.

36. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

37. Neuwirth Erich. RColorBrewer: ColorBrewer Palettes. R. package version 1.1-2. 2014. http://CRAN.R-project.org/package=RColorBrewer. Accessed Aug. 1, 2015.

38. Starmer AJ, Spector ND, Srivastava R, et al. Changes in medical errors after implementation of a handoff program. *N Engl J Med.* 2014;371 (19):1803–1812.

AUTHOR AFFILIATIONS

[1]Division of Hematology/Oncology, Department of Internal Medicine, University of Michigan Medical School, 3-219 Cancer Center, 1500 E Medical Center Dr, Ann Arbor, MI, 48109, USA. jkrauss@med.umich.edu

[2]Department of Biostatistics, School of Public Health, II, 1415 Washington Heights, University of Michigan, Ann Arbor, MI, 48109-2029, USA

[3]San Francisco VA Medical Center, Mental Health Service, 4150 Clement St, San Francisco CA, 94121, USA

[4]Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, 48109, USA.