Improving the Sensitivity of the Problem List in an Intensive Care Unit by Using Natural Language Processing

Stéphane Meystre MD PhD\textsuperscript{a,b}, Peter Haug MD\textsuperscript{a}.
\textsuperscript{a} Department of Medical Informatics, University of Utah, Salt Lake City, Utah, U.S.
\textsuperscript{b} Office Informatique, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland.

Abstract
To improve the completeness of an electronic problem list, we have developed a system using Natural Language Processing to automatically extract potential medical problems from clinical free-text documents; these problems are then proposed for inclusion in an electronic problem list management application.

A prospective randomized controlled evaluation of this system in an intensive care unit is reported here. A total of 105 patients were randomly assigned to a control or an intervention group. In the latter, patients had their documents analyzed by the system and medical problems discovered were proposed for inclusion into their problem list. In this population, our system significantly increased the sensitivity of the problem lists, from 8.9\% to 41\%, and to 77.4\% if problems automatically proposed but not acknowledged by users were also considered.

Introduction
The problem-oriented Electronic Health Record (EHR), centered on the problem list, is seen by many as a possible answer to the quality of healthcare and medical errors reduction challenges. At Intermountain Health Care (IHC; Salt Lake City, Utah), the problem list is an important piece of the medical record, and a central component of HELP\textsubscript{2}, a new clinical information system in development\textsuperscript{1}. To enable its potential benefits, the problem list has to be as accurate, complete and timely as possible. Unfortunately, problem lists are usually incomplete and inaccurate, and are often totally unused. To address this deficiency, we have created an application using Natural Language Processing (NLP) to harvest potential problem list entries from the multiple free-text electronic documents available in a patient’s EHR. The medical problems identified are then proposed to the physicians for addition to the official problem list\textsuperscript{2}. We hypothesize that the use of NLP to automatically provide potential medical problems will improve the completeness of this Automated Problem List.

Background
More than three decades after Larry Weed proposed the Problem-Oriented Medical Record (POMR)\textsuperscript{3,4} as an answer to the complexity of the medical knowledge and clinical data, and to address weaknesses in the documentation of medical care, the problem-oriented EHR and the problem list have seen renewed interest as an organizational tool\textsuperscript{5-10}. Advantages to this approach are that the problem list provides a central place for clinicians to obtain a concise view of all patients’ medical problems, that it facilitates associating clinical information in the record to a specific problem, and that it can encourage an orderly process of clinical problem solving and clinical judgment. The problem list in a problem-oriented patient record also provides a context in which continuity of care is supported, preventing redundant actions\textsuperscript{5}. The Institute of Medicine\textsuperscript{11} recommends that the Computer-based Patient Record contain a problem list that specifies the patient’s clinical problems and the status of each. Also, convinced of the benefits of the problem list, the Joint Commission for the Accreditation of Hospitals (JCAHO\textsuperscript{12}) has established it as a required feature of hospital records.

The patient record contains a large amount of information captured as narrative text. These free-text documents represent the majority of the information used for medical care\textsuperscript{13} but decision-support, research, and quality improvement create a need for structured and coded data instead. As a possible answer to this issue, NLP can be used to convert free-text into coded data\textsuperscript{14}.

Several groups have evaluated NLP techniques with medical free-text. Examples are MedLEE (Medical Language Extraction and Encoding system)\textsuperscript{15}, and applications developed by our medical informatics group at the University of Utah: SymText\textsuperscript{16} and MPLUS\textsuperscript{17}. Other systems that automatically map clinical text concepts to standardized vocabularies have been reported, notably MetaMap\textsuperscript{18,19}. MetaMap, and its Java\textsuperscript{TM} version called MMTx (MetaMap Transfers), were developed by the U.S. National Library of Medicine (NLM). They map concepts in the analyzed text with UMLS concepts. The mapped
concepts are ranked, but no negation detection is performed.

Negation detection is a required feature when analyzing clinical narrative text where findings and diseases are often described as absent. To this end, negation detection algorithms have been developed, like NegEx20.

At Intermountain Health Care (IHC), the web-based clinical information system called HELP21 offers secured access to clinical data through specialized modules like “Patient search”, “Labs”, “Medications”, and “Problems”. The “Problems” module allows viewing, modifying, and adding medical problems along with their status (active, inactive, resolved, or error) and other information. This electronic problem list was already in use in the outpatient setting, but was not commonly used for hospitalized patients. The medical and surgical ICU at the LDS Hospital in Salt Lake City was piloting the use of this problem list. This ward participated in our study. All other inpatient wards were using a paper-based problem list, or no problem list at all.

Materials and Methods
As mentioned earlier, the Automated Problem List system uses NLP to extract potential medical problems from free-text medical documents. The two main components of the system are a background application and the problem list management application. The background application does text processing and analysis and stores extracted problems in the central clinical database. These problems can then be accessed by the problem list management application integrated into HELP2. The background application has already been evaluated and has shown good performance.21 During the study described here, the background application looked for 80 different diagnosis problems, which were selected based on their frequency in the clinical environments chosen for our evaluation (a medical and surgical intensive care unit). The NLP tools used in this experiment were based on MMTx with a customized data subset. The reviewers also examined the patients’ electronic problem list. They mapped problems that were entered as free-text (not coded) with the corresponding coded problem, and also mapped children of our targeted problems to the relevant parent problem (such as Fallot’s triad mapped with atrial septal defects).

Problem list completeness: Three different problem lists were considered for each patient: the reference standard (i.e. what should have been in the problem list), the “official” problem list (i.e. problems recorded in the clinical database as active, inactive, or resolved), and the “potential” problem list (i.e. problems included in the “official” list, with the addition of problems that had not been reviewed and changed from a proposed status). The content of each patient’s official and potential problem list was compared with the reference standard, and each
problem list entry was counted and categorized as true positive (TP; problem present in the patient’s documents and in the problem list), false negative (FN; problem documented in the patient’s record but absent from the problem list), false positive (FP; problem absent but listed in the problem list), or true negative (TN; problem absent and not listed in the problem list).

We then calculated the sensitivity (TP / TP+FN) for each patient, and averaged across the group of patients analyzed.

**Results**

Ten different reviewers, all physicians, reviewed clinical documents to create the reference standard. Reviewers’ overall agreement was very good, with a Finn’s R of 0.897 when reviewing documents and 0.995 when reviewing problem lists. Finn’s R was used instead of Cohen’s kappa, because the agreement table was strongly skewed, with far more true negatives than true positives.

Because of lack of usage of the electronic problem list in the cardiovascular surgery ward during the study, this experiment and its analysis were focused on the ICU patients. In this ward, 105 patients were enrolled, with 54 patients randomly assigned to the intervention group, and 51 to the control group. A total of 719 medical problems were automatically extracted and proposed during the study, and about half were added to the “official” problem list or rejected as errors (Fig. 1).

**Problem list sensitivity:** Mean and 0.95 confidence intervals were computed (Table 1). Analysis of all patients from the ICU showed significant differences between control and intervention groups (Fig. 2). Patients in the intervention group had a more complete problem list, with a significantly higher sensitivity. When analyzing the potential problem list (i.e. including problems that had remained proposed), the sensitivity was again significantly higher, reaching about 77%.

When analyzing the subset of patients with a problem list that had been edited during their stay in the ICU (43 of the 105 patients), the sensitivity of their problem list was also significantly higher in the intervention group.

![Figure 1](image1.png)  
*Figure 1: Number and proportion of proposed medical problems and their subsequent status modifications.*

Statistical analysis was executed using a non-parametric test (Mann-Whitney test) for non-normality reasons.

**Discussion**

This evaluation of our Automated Problem List system suggests that the addition of NLP to support completeness was successful. The sensitivity we measured was significantly increased. Enhancing the problem list management application with NLP made the problem list more complete.

The excellent inter-reviewer agreement in this study allowed us a high quality reference standard. This was made possible by the use of explicit review techniques; the list of targeted problems was always provided beside the document or problem list to review.

Our results are difficult to compare to other published results because only very few such studies have been published. A rare example is an evaluation by Szeto et al., measuring the accuracy of an outpatient problem list for 9 different diagnoses.

A sensitivity of 49% was measured. Our study considered 80 different diagnoses, and gave very similar specificity results, but the sensitivity without intervention was much lower. The effect of our Automated Problem List system increased the sensitivity to a similar degree.
The medical problem list figures prominently in our plans for computerized physician order entry and medical documentation in the new HELP2 system currently under development at IHC. A well-maintained problem list will significantly enhance these applications. The Automated Problem List could be beneficial for many reasons: a better problem list could potentially improve patient outcomes and reduce costs by reducing omissions and delays, improving the organization of care, and reducing adverse events. It could enhance decision-support for applications requiring knowledge of patient medical problems.

Several limitations of our system and of this study need to be discussed. A first important issue is the use of the problem list by physicians. The application suite in which these tools were embedded originated in the outpatient setting and is in the process of moving into the hospital environment. As mentioned in the background section of this paper, the problem list currently used in this environment is paper-based and is usually incomplete and not timely. In fact, it is often totally unused. There is good reason for this. None of the therapeutic or documentation functions that are done by the physicians are currently tied, in any way to the problem list. In the electronic record that is evolving at our institution, the problem list will be integrated into the care process. Electronic order entry, documentation, and a variety of decision support tools will be tied to it. Unfortunately, these applications either do not currently exist or exist as limited prototypes. The single function currently mediated through the problem list is the “infobutton” a tool that provides problem-specific medical knowledge to the user. Therefore, this study is best seen as an effort to explore the possibilities offered by NLP to support the problem list. Today, encouraging clinicians to enter data in the problem list and convincing them of the benefits of the problem list remains a challenge.

Another issue with this system is the scalability of the list of targeted medical problems. Our system is currently designed to extract only 80 different medical problems. Those 80 problems represented about 64% of all coded medical problem instances in our EHR in 2003; however, many more problems will be needed to allow this system to be used in other settings. A very simple solution is to use the default full UMLS data set provided with MMTx instead of a custom data subset, but this reduces the NLP module performances: it decreases the recall and makes it slower.

Acknowledgments
This work is supported by a Deseret Foundation Grant (Salt Lake City, Utah). We would like to thank Min Bowman for her help with the modified Problems module. We would also like to thank Greg Gurr for his advice and his help. Scott Narus and Stan Huff also gave us helpful advice and guidance for which we are grateful. Finally, we are especially grateful to Terry Clemmer whose enthusiasm for the problem list made this study possible.

References


