

Characterizing Patient-Friendly “Micro-Explanations” of Medical Events

Lauren Wilcox^{1,2}, Dan Morris¹, Desney Tan¹, Justin Gatewood³, Eric Horvitz¹

¹Microsoft Research
{dan,desney,horvitz}
@microsoft.com

²Dept of Computer Science
Columbia University
wilcox@cs.columbia.edu

³MedStar Institute for Innovation
justin.s.gatewood@medstar.net

ABSTRACT

Patients’ basic understanding of clinical events has been shown to dramatically improve patient care. We propose that the automatic generation of very short *micro-explanations*, suitable for real-time delivery in clinical settings, can transform patient care by giving patients greater awareness of key events in their electronic medical record. We present results of a survey study indicating that it may be possible to automatically generate such explanations by extracting individual sentences from consumer-facing Web pages. We further inform future work by characterizing physician and non-physician responses to a variety of Web-extracted explanations of medical lab tests.

Author Keywords

Electronic Medical Records, Personal Health Records

ACM Classification Keywords

J.3 [Computer Applications] Life Sciences – Health.

General Terms

Human Factors

INTRODUCTION

Active involvement by patients in their own care has been shown to increase patient satisfaction, which in turn results in greater compliance, reduced stress and complications, and several additional benefits to quality of care [7].

Patient involvement depends on patients being well-informed about their care. Fortunately, today’s patients have several information sources at their disposal. For example, physicians routinely explain symptoms, diagnoses, conditions, and treatments to patients verbally. Unfortunately, such in-person explanations are not available to meet all patient information needs, nor are they as readily accessible as other materials. Educational materials about tests, procedures, and medications can provide useful information for patients. However, such materials can be overwhelming amidst the distress common in hospital environments. Furthermore, a great deal of medical information is inaccessible to those lacking health literacy; information packets are

rarely tailored to a specific patient, and traditional delivery media such as paper make dynamic delivery difficult [3].

Online health resources (e.g. webmd.com) provide comprehensible, patient-friendly health literature, but target patients who are cognitively and logistically able to formulate a search query, then comprehend detailed results. Results typically contain much more information than is necessary to provide basic understanding, and often include facts that are irrelevant or even needlessly worrying to a specific patient. Consequently, much online content is often ill-suited for delivering information in the small chunks demanded by patients who are in a cognitive state to receive only limited information or who are using space-constrained media such as in-room displays and mobile devices.

In addition to online information sources, the increasing availability of electronic medical records (EMRs) promises to provide patients with unprecedented levels of access to *their own* medical data. Despite this availability, patients often remain uninformed about the tests and procedures conducted during clinical visits, leaving them ill-equipped to participate in medical decisions. This discrepancy is largely a consequence of the structure and terminology used in EMRs, which are designed for clinicians and administrators, rather than for patients.

We therefore propose that leveraging multiple information sources – specifically existing online health resources *and* the EMR – can provide concise, patient-friendly, and personalized explanations of medical events. Such medical *micro-explanations* would be suitable for real-time presentation to patients on mobile devices or in-room displays, or for offline presentation through a personal health record.

Fully-automatic generation of explanations remains technically difficult. However, we hypothesize that current consumer-facing Web resources, although not designed for this purpose, contain sentences that are suitable for presentation as standalone explanations. Furthermore, situation-specific extraction of appropriate sentences from these resources, based on information already available in the EMR, can circumvent the need to create explanations “from scratch”.

As a first step toward supporting automated, Web-based micro-explanation extraction, we present a survey study that characterizes the properties of “good” Web-derived explanations, according to both physicians and non-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05....\$10.00.

physicians. The specific contributions of this note are:

- 1) Preliminary validation of our hypothesis that the Web is a source of concise, patient-friendly explanations of lab tests, suitable for automated extraction.
- 2) A survey study characterizing the properties of “good” explanations, targeted at defining features for automated Web content extraction.
- 3) Exploration of the similarities and differences between how doctors and patients view these explanations.

RELATED WORK

Recent work has demonstrated the benefits of computer-based presentation of medical data to hospital patients, through kiosks [5], in-room displays [9] or agents [1]. This work motivates ours, which opens a new approach for automatic content generation for these patient-facing media.

Parallel work has explored natural-language generation and extraction techniques for summarizing medical content. Delegér [2] used parallel medical and lay-person corpora to map specific phrases between “patient-speak” and “doctor-speak”; a similar approach is taken by [4]. Portet et al [6] attempt to create textual summaries of continuous signals (e.g. EKG traces) in hospital settings. DiMarco [3] addresses the problem of semi-automatic personalized educational materials, using physician input to tailor generic materials to specific patients. This and other work in medical summarization complements our proposed approach, but no work to date has explored the unique properties of short, real-time explanation delivery, or the opportunity to base such explanations on existing consumer-friendly sources.

METHODOLOGY

To pursue insights about the value of consumer-facing Web pages serving as a source of explanations for medical events, we conducted a survey to evaluate physicians’ and non-physicians’ opinions of Web-extracted sentences, presented as explanations for diagnostic lab tests. We scope the present work to lab tests, given the finding by [8] of patient interest and appreciation about diagnostic test results in a personal health record. The survey took about an hour and participants were given a software gratuity.

Physician Participants

Seven physicians (six male) participated. All were actively practicing in U.S. hospitals and indicated that explaining tests to patients was a part of their responsibilities. Participants’ specialties were emergency medicine, oncology, cardiology, pediatrics, surgery, and internal medicine (2).

Non-Physician Participants

Twenty-three non-physician participants (9 male) from the U.S. (14 states) were recruited based on having had recent hospital experiences as patients or family members of patients. Sixteen participants had hospital experiences in the past two years; all had hospital experiences in the past ten years. Participants’ ages ranged from 21 to 63 (mean 41), and education ranged from high school to post-graduate.

Survey Design

We prepared four “patient profiles” based on common Emergency Department presentations, chosen via review of patient records in a large, urban hospital. Each profile (approximately one paragraph) included the sample patient’s current symptoms, current medications, and a summary of their medical history and was reviewed by two non-participating experts for content validity. For each profile, experts provided a list of four lab tests that would likely be ordered. For each of these lab tests (16 total), we collected nine alternative one-sentence explanations (144 total).

To select the sample explanations for our survey, we performed a Web query for each lab test and chose the first three HON-certified¹ Web pages appearing in the results. From these pages, we hand-selected sentences that would fit the micro-explanation format. A large set (hundreds) of candidate sentences existed, so we selected sentences that spanned the space of possible explanations to ensure a representative sample of explanatory content (i.e., removing redundancy across sites while diversifying information).

Each participant responded to two of the four patient profiles. Each participant was thus presented with 72 explanations total (9 explanations for each lab test, 4 lab tests per profile, 2 profiles). Participants rated each explanation according to how appropriate they found the explanation to be for the patient described (physicians) or how helpful they would find the explanation if delivered to them in a hospital (non-physicians).

Participants used a rating scale of 1-5, and were instructed to use scores above 3 to indicate that the benefit of the explanation outweighs any concerns about its flaws. For each rating given, participants elaborated on the rationale for their choice using free text. Additionally, after all lab explanations had been examined for both sample patients in the study, both participant groups were asked to reflect on their preferences and describe characteristics of meaningful explanations, again in free text.

RESULTS AND DISCUSSION

Overview

Not surprisingly, the most consistent theme in all responses was a desire for “simplicity”. In fact, 23/30 participants mentioned this explicitly in their end-of-survey reflections, including 6/7 physicians. Despite this common high-level goal, however, a deeper analysis of our results reveals that participants have widely-varying concepts of what makes a “simple”, patient-friendly explanation. This section will explore this complexity, along with several additional themes that emerged in survey responses.

Web-extracted sentences as explanations

One of our core hypotheses is that consumer-facing Web pages can serve as the basis for extracting single sentences,

¹ The “Health on the Net” (HON) Foundation certifies the credibility of health-related sites through physician review.

suitable for delivery as micro-explanations. This is not obvious, however; Web sites are composed as a whole, and single sentences may lose meaning out of context.

Our results suggest that even sampling only 9 candidate sentences per lab test (an automated system could examine many more) is sufficient to extract at least one “good” explanation. For both physicians and non-physicians, every lab test had at least one explanation whose mean score was higher than 3.0; participants were instructed that “A rating of 3 or higher means that you consider the explanation of enough benefit to display”. This suggests that extracting individual sentences from Web pages, despite the loss of context, is promising for micro-explanation generation.

Features of “good” explanations

Another goal of our study was to understand the relationship between participant preferences and characteristics of explanation text. To investigate correlations between measurable features and survey responses, we manually coded several features of the survey explanations, including:

1. Mention of *symptoms*
2. Mention of *organs, systems, or body parts*
3. Mention of *problems* or *diagnoses*
4. Sentence length in words
5. Use of *acronyms* and *medical terms*
6. Number of *medical terms* or *acronyms defined*

These features were hypothesized to be consistent with good explanations, based on discussions with physicians.

The mention of a specific symptom and the mention of an organ or system in an explanation correlated with participant preference: the top five most-preferred explanations across all labs (i.e., the explanations with the highest mean scores) *all* contained both a symptom and an organ/system reference, and *none* of the five least-preferred explanations contained either of these features. The most-preferred explanation overall is illustrative of these properties:

The troponin test is used to help diagnose a heart attack, to detect and evaluate mild to severe heart injury, and to separate it from chest pain that may be due to other causes.

Interestingly, despite participants’ overall desire for simplicity, longer explanations were preferred overall by both physicians and non-physicians. The five most-preferred explanations overall had a mean length of 30.0 words (± 8.0), while the five least-preferred explanations had a mean length of 15.8 words (± 9.2).

This preference for somewhat long sentences was explained differently by physicians, who explained high ratings for more complex sentences based on *accuracy*, and patients, who frequently explained high ratings for more complex sentences based on “sounding like a professional answer” and “sounding like something a doctor would say”.

The preliminary correlations observed between encoded features and participant preference support our hypothesis

that a feature-driven approach to scoring and ranking explanations could underlie a micro-explanation system.

Sources of divergence between patients and physicians

We also aimed to understand differences between physician and non-physician responses to Web-extracted explanations. To explore this, we identified explanations whose mean scores differed by more than 0.8 between physician and non-physician participants (16 explanations). Examining these explanations yields insights into how physicians and non-physicians approach these explanations differently.

14 of these 16 explanations were cases in which physicians rated explanations *lower* than non-physicians, and in fact all 14 had means above 3.0 (“acceptable”) for non-physicians but below 3.0 (“unacceptable”) for physicians. Most of this divergence was caused by explanations that physicians felt were inaccurate in this specific case: although the sentence was true, it reflected a narrow application of the test that was not appropriate for the particular patient profile being examined. For example, regarding a test that is sometimes used to diagnose clotting disorders (easy to understand) but was really being used here to assess liver function, P3 (physician) responds to an explanation about clotting by saying: “This is a reason to order this test, but not necessarily in this patient.”, and P1 (physician) states: “True, but not applicable to someone presenting in the ER.”

This presents a challenge for automated extraction techniques: even given credible information, it is critical to target explanation selection to a specific patient. This need for patient-specificity re-affirms our hypothesis that drawing from multiple information sources – specifically the Web (for candidate content) and the EMR (for patient-specific selection among candidates) – is necessary for automatic generation of explanations.

Another surprising reason emerged for physician preference of several explanations relative to non-physicians: physicians thought the explanation was too complicated for non-physicians to understand, but non-physicians appreciated the technical detail and saw it as “professional”. For example, the following liver panel explanation was rated significantly higher by non-physicians:

A liver panel or one or more of its component tests may be used to help detect liver disease if a person has symptoms that indicate possible liver dysfunction or if a person is being monitored or treated for a known condition or liver disease.

A physician described this explanation as “too wordy”, while non-physicians indicated that “it’s detailed” and “something a doctor would say”.

The two explanations that physicians rated highly but non-physicians did not were both explanations for the “lipase level” test, and both used the term “pancreatitis”. While physicians were generally concerned about conforming to a patient-friendly level of technical detail, they did not expect patients to be unfamiliar with this term. This presents an-

other interesting challenge for automated extraction: understanding patient vocabularies and finding the boundary between layman and medical terminology.

The danger of too much information

Both patients and physicians noted that while detail is critical, too much information not only limits clarity, but creates risk and confusion. This particularly applied to the mention of diseases in explanations. Responding to tests that clearly and concisely stated these goals, referring specifically to diseases, non-physicians in particular expressed potential fear. P17, for example, states: “The mention of all of these diseases while already under duress would cause more stress.” P20 similarly expresses: “You are using medical terms that as a confused and disoriented individual are just scaring me. The longer the words are [...], the scarier it sounds.” These findings suggest other challenges for automated extraction, including balancing *specificity* and *patient comfort*, and minimizing inappropriate anxieties that may come with mention of low-probability risks.

Need for personalization

In many cases, even among non-physicians, participants disagreed significantly on the quality of an explanation, suggesting a need for automated systems to recognize various aspects of personal preference.

In fact, 49 of 135 explanations were “polarizing”, meaning that at least 25% of non-physician participants assigned score ≥ 4.0 (“good” to “excellent”), and at least 25% assigned scores ≤ 2.0 (“poor” to “unacceptable”). Exploring the rationales for these ratings highlights several sources of variation in non-physician responses.

Many lab tests are used as broad initial screenings, but non-physicians varied significantly in their willingness to accept “broad screening” explanations as useful. An explanation for one such test referred to “determining general health status”, yielding scores of 1 (“I would hope any lab test would provide information on general health status”) and 5 (“I understand that this kind of test is used to rule out many medical issues”) from non-physician participants.

Another common explanation pattern that polarized non-physician respondents was the use of *indications for a test* that did not refer specifically to diagnostic outcomes. The following explanation is illustrative:

If a patient is having symptoms such as fatigue or weakness or has an infection, inflammation, bruising, or bleeding, then the doctor may order a CBC to help diagnose the cause.

P14 (who assigned a 4.0 rating) specifically likes “that this answer tells me why the test might have been ordered for me”, but P7 (who assigned a 2.0 rating) responds “Only states causes of the test, does not say what the test does”.

A third source of significant disagreement among non-physicians emerged from varying health literacy: some explanations were concise and complete but used a term or

acronym not familiar to some participants. The explanation “to determine if your blood glucose level is within healthy ranges” prompted responses of “This is straightforward and easy to understand” (P17, rating 5.0) and “If you don't know what blood glucose is, this explanation is useless.” (P27, rating 2.0).

Finally, some respondents expressed a strong dislike for explanations that were not grammatically complete, even if they were conceptually correct (e.g. an explanation beginning with “To determine if your lungs are functioning well enough...”). This need for personalization opens a further challenge for automated extraction systems: personalization based not only on EMR data but on user preference as well.

CONCLUSION AND FUTURE WORK

We studied opportunities for generating medical *micro-explanations* from consumer-facing Web content. We explored physician and non-physician responses to sentences extracted from Web pages and presented as micro-explanations for medical lab tests. We aim to develop algorithms for presenting real-time explanations to hospital patients. Although the present work focused on explaining lab tests, we expect that insights and results presented here will generalize to other EMR content, e.g. medications. Further work is required to validate this hypothesis.

REFERENCES

1. Bickmore, T., Pfeifer, L. and Jack, B. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. *Proc ACM CHI 2009*.
2. Deléger, L. and Zweigenbaum, P. Extracting lay paraphrases of specialized expressions from monolingual comparable medical corpora. *Proc 2009 Workshop on Building and Using Comparable Corpora*.
3. DiMarco, et al. Authoring and generation of individualised patient education materials. *Proc AMIA 2006*.
4. Elhadad, N. and Sutaria, K. 2007. Mining a lexicon of technical terms and lay equivalents. *Proc 2007 Workshop on Bio, Translational, and Clin Lang Processing*.
5. Jones, R. The role of health kiosks in 2009: literature and informant review. *Int J Env. Res Public Health*, 6.6 (2009).
6. Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripada, S., Freer, Y., and Sykes, C. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, 173 (2009).
7. Pratt, W., Unruh, K., Civan, A., and Skeels, M. Personal Health Information Management. *Communications of the ACM*, 49.1 (2006).
8. Tang, P.C. and Lansky, D. The missing link: bridging the patient-provider health information gap. *Health Affairs* 24 (2005).
9. Wilcox, L., Morris, D., Tan, D., and Gatewood, J. Designing patient-centric information displays for hospitals. *Proc ACM CHI 2010*.