

A Novel Schema to Enhance Data Quality of Patient Safety Event Reports

Hong Kang, PhD, Yang Gong, MD, PhD

School of Biomedical Informatics, the University of Texas Health Science Center at
Houston, Houston, TX, USA

Abstract

The most important knowledge in the field of patient safety is regarding the prevention and reduction of patient safety events (PSEs). It is believed that PSE reporting systems could be a good resource to share and to learn from previous cases. However, the success of such systems in healthcare is yet to be seen. One reason is that the qualities of most PSE reports are unsatisfactory due to the lack of knowledge output from reporting systems which makes reporters report halfheartedly. In this study, we designed a PSE similarity searching model based on semantic similarity measures, and proposed a novel schema of PSE reporting system which can effectively learn from previous experiences and timely inform the subsequent actions. This system will not only help promote the report qualities but also serve as a knowledge base and education tool to guide healthcare providers in terms of preventing the recurrence of PSEs.

Introduction

Retrospective analysis of health data holds promise to expedite scientific discovery in medicine and constitutes a significant part of clinical research¹. Informatics technology has helped improve the efficiency by replacing paper-based systems in many healthcare organizations and garnering rich health data^{2,3}. For example, dozens of patient safety event (PSE) reporting systems have been established to enable safety specialists to analyze events, identify underlying factors, and generate actionable knowledge to mitigate risks⁴, such as PSRS (Patient Safety Reporting System)⁵ and AHRQ Common Formats (Common Definitions and Reporting Formats)⁶. These systems are initiatives to improve patient safety because data supports further learning and actionable knowledge. However, the success of such systems in healthcare is yet to be seen because the data quality could not reach the satisfying level. Therefore, maintaining the quality of healthcare data is widely acknowledged as problematic but critical to effective healthcare⁷.

A PSE is any process, act of omission, or commission that resulted in hazardous health care conditions and/or unintended harm to the patient⁸. Improving the data quality for PSE reporting systems is a challenging task since many factors are involved: inadequate management structures for ensuring complete, timely and accurate reporting of data; inadequate rules, training, and procedural guidelines for those involved in data collection; fragmentation and inconsistencies among the services associated with data collection; and the requirement for new management methods which utilize accurate and relevant data to support the managed care environment⁹. Sometimes reporting systems cannot guarantee qualities because many reporters complete the task halfheartedly. For example, in the AHRQ Common Formats reports, we found more than 80% reporters tend to choose ambiguous options such as “Unknown” which are meaningless for learning rather than explicit “Yes” or “No” options. These systems will become redundant databases as more and more low-quality reports emerge, from which researchers cannot learn any experience from previous cases. Also, there is a lack of connections among different PSE resources due to the diverse report formats. Users have to spend plenty of time manually investigating helpful information from previous cases¹⁰. Therefore, facing the gap that end-users cannot receive any perceived benefit during and after the reporting, there is an urgent need of a new generation PSE reporting system which can continually enrich the PSE knowledge base by annotating new reports and solutions according to their features, and provide all necessary information (e.g., solutions and prevention options) to the reporters’ cases by measuring the similarities of PSE reports. The quality of the reports will be improved consequently since reporters can receive “benefits” for their reporting rather than passively completing the tasks.

The primary challenge of the new reporting system is measuring the similarity of PSEs, specifically, how to calculate the similarity score between two PSE reports. Different fields provide diverse definitions for the term “similarity”, but there is still no an explicit definition in PSE domain. Based on the theory of cognitive science, such a system handles tasks in the case-based reasoning (CBR) cycle including retrieve, reuse, revise, and retain¹¹. Specifically, we firstly measure the similarity of the current problem to previous ones stored in a sort of database with their known solutions, then retrieve one or more similar cases and attempt to reuse the solution of one of the retrieved cases. The solution is assessed and revised before being proposed. At last, the problem description and its

solution is retained as a new case learned by the whole system to solve new problems. Therefore, the measurement should be based on an assumption that similar PSEs have similar solutions.

AHRQ Common Formats⁶ and International Classification of Patient Safety (released by WHO)¹² defined incident types and event categories for PSE, which are widely accepted and commonly used in patient safety community. Recently, researchers are trying to develop new sets of categories in order to better serve the reporting. Nevertheless, the data formats of individual case were all in the form of ontology (a hierarchical data structure to manage terms and their relationships). Therefore, the comparison between two PSEs could be technically processed through semantic similarity measure as a function that, given two ontology terms or two sets of terms annotating two entities, returns a numerical value reflecting the closeness in meaning between the two¹³. The semantic similarity algorithms have been generally applied in many fields, such as bioinformatics¹³⁻¹⁵, geoinformatics¹⁶, linguistic¹⁷ and natural language processing (NLP)^{18, 19}, etc. The Gene Ontology (GO)²⁰ is the main focus of investigation of semantic similarity in molecular biology, because not only it is the ontology most commonly adopted by the life sciences community, but also comparing gene products at the functional level is crucial for a variety of applications. Numerous researches have demonstrated that the functional relatedness between genes products with GO annotations can be well measured by semantic similarity algorithms^{13, 21-25}, which provide major significance for gene function studies. Similarly, the patient safety community also need an approach to compare PSEs, then offer the users potential hints of solutions to current cases. Intuitively, the form of PSE seems similar to that of GO, since a number of taxonomies have been designed for labeling case through ontology annotations. Accordingly, the methods that work effectively to compare GO products might be also feasible when identifying similarities in PSEs. However, to our best knowledge, the semantic similarity algorithms have never been adopted and assessed in patient safety area.

In this study, we hypothesize that 1) the similarity between PSEs can be measured if they are annotated by the same ontology/taxonomy in patient safety domain, and 2) similar PSEs have similar solutions. We utilized and assessed the semantic similarity measures on the PSE datasets of AHRQ WebM&M (Morbidity and Mortality Rounds on the Web)²⁶ and AHRQ Common Formats⁶. Based on this model, we proposed a novel schema which can process the comparison tasks for PSEs and provide the reporters pertinent suggestions about solutions and prevention options for their cases. The schema will help develop a new generation of PSE reporting system which could improve the data quality of PSE reports by arousing the enthusiasm and motivation of reporters, and hold promise in preventing the recurrence and serious consequences of PSEs.

Method

Datasets of PSE Reports

AHRQ WebM&M: As a public assessable resource, AHRQ WebM&M Web site²⁶ represents illustrative cases of confidentially-reported PSEs on the internet, accompanied by straightforward evidence-based expert commentaries. It also provides a taxonomy with 219 terms to describe the features of PSE from six perspectives: safety target, error type, approach to improving safety, clinical area, target audience and setting of care. Each case in the database has an individual set of annotations on this taxonomy labeled by experts. As of January 7, 2016, 366 cases have been posted on the site. The data size cannot hinder WebM&M from becoming an adequate resource for researching purpose because of the high-quality contents and the diversity of event profiles.

AHRQ Common Formats: AHRQ created the Common Formats to help providers uniformly report PSEs and to improve health care providers' efforts to eliminate harm⁶. The Common Formats is a group of standardized questionnaire-based forms with nine subtypes defined by PSO (Patient Safety Organization), including blood or blood product, device or medical/surgical supply including health information technology, fall, healthcare-associated infection, medication or other substance, perinatal, pressure ulcer, surgery or anesthesia, and venous thromboembolism. There is an individual questionnaire for each subtype which contains 10 to 30 single or multiple choice questions, or subjective questions. Missouri Center for Patient Safety (MCPS), a PSO institute and the collaborator of this study, started to collect real PSE reports from several hospitals in Missouri since 2008. So far the database has included more than 41,000 PSE reports. As part of the national level database, the data provided by MCPS can represent all PSE data of PSO program.

Semantic Similarity Measures

Semantic similarity assesses the degree of relatedness between two entities by the similarity in meaning of their annotations. There are two types of semantic similarity approaches when comparing terms, edge-based and node-based (Table 1). Edge-based approaches are based on counting the number of edges in the graph path between two

terms²⁷, for instance, the shortest path or the average of all paths. Correspondingly, node-based approaches focus on comparing the properties of the terms themselves, their ancestors or descendants. Information Content (IC), the most commonly used approach in GO studies, belongs to this category. IC gives a measure of information for every term and makes them comparable. Edge-based and node-based approaches are intended to score the similarity between two terms, and must be extended to compare sets of terms such as gene products and incident cases. Pairwise and groupwise approaches are the two types of strategies for this issue. Every term in the direct annotation set A is compared against every term in the direct annotation set B in pairwise approaches, then the semantic similarity is considered by every pairwise combination of terms from the two sets (average, the maximum, or sum) or only the best-matching pair for each term. Groupwise approaches calculate the similarity directly by set, graph, or vector. Set approaches are not widely used since they only consider the direct annotations that would lose a lot of information; based on set similarity techniques, graph approaches represent entities as the subgraphs of the whole annotations and calculate the similarity using graph matching techniques; vector approaches compact the information in vector space as binary fingerprints which are more convenient for comparison.

Table 1. Pros and Cons of semantic similarity measures

| Approaches | Pros | Cons |
|-----------------------------------|---|---|
| <i>Measures for terms</i> | | |
| Edge-based | Intuitive, easy to perform. | Assume all the nodes and edges are uniformly distributed and treat them who are in the same depth equally, which is not applicable for real data. |
| Node-based | Measure the terms independent of their depth in the ontology. | The common used term would make more contribution when calculating the similarity. |
| <i>Measures for sets of terms</i> | | |
| Pairwise | The contributions from every pair of terms are concerned. | Over-reliance on the quality of data; time-consuming. |
| Groupwise | Compare term combinations from a macro view instead of relying on integrating similarity between individual terms; time-saving. | Excessive choices could be a trouble. |

We applied and compared three typical semantic similarity measures in this study:

Information Content (IC). As a classic node-based approach, IC gives a measure on how specific and informative a term is. Towards PSE reports, it assumes that a term with higher probability of occurrence may contribute less when measuring the similarity. In this study, the pairwise strategy which calculated the similarity for all pairs of terms and assessed them with average score was adopted. And we used Lin's measure of similarity²⁸ which accounts the IC values for each of term t_1 and t_2 in addition to the lowest ancestor shared between the two terms.

Normalized Term Overlap (NTO). NTO²¹ considers the set of all direct annotations and all of their associated parent terms. Theoretically, NTO might be applicable if the ontology of WebM&M data has been well defined and annotated. The only concern is that the depth of its tree structure is not deep enough which may decrease the comparison resolution. However, in order to further study the applicability of this typical graph-based groupwise approach, we also enrolled it in our assessment.

Vector Space (VS). VS compacts the annotations of a set of terms into a binary vector which is more comparable because the model is based on linear algebra with lots of mature algorithms which can measure similarity, such as cosine measure²⁹. Similar to IC, a variation of VS approach has been used in ontology-based similarity. The approach generates a weight for each term based on the frequency of its occurrence in the corpus, and then replaces the non-zero values in the binary vector with these weights. As the WebM&M cases are well annotated in an ontological structure, VS measure, a vector based groupwise approach, may be potentially applicable to measure the similarity between PSEs.

Expert Review Based on 4-Point Likert Scale Measure

To each query case, the semantic similarity algorithms ranked all the other cases in the database according to the similarity scores. A group of cases with the same interval were chosen from the rank for expert review. To assess the result given by the algorithms, we adopted a 4-point Likert scale measure³⁰ which contains 1-irrelevant, 2-somewhat irrelevant, 3-relevant, and 4-highly relevant. Three experts were invited to label every case with one of the four scales according to the degree of similarity between this case and the given query case. After the experts completed the review, two rounds of discussion were opened for them to provide a final review result. If they cannot reach an agreement to certain case, the case would be labeled by a majority. The final expert result was treated as a golden standard. Any case that was labeled by either 1 or 2 in both expert's result and algorithm's result was regarded as "agreement" and judged as being irrelevant to the query case; conversely, the one that was labeled either 3 or 4 in both of the two results was also regarded as "agreement" but classified as being relevant to the query. The agreement ratio between final expert review and algorithm (sample agreement ratio) was calculated by dividing the numbers of agreement cases by the number of total cases. Then we randomly labeled the same group of cases for 10,000 times and calculated the agreement ratios respectively (random agreement ratios). To evaluate the performance of certain algorithm of the three, one sample t-test was adopted to examine the mean difference between the sample agreement ratio and the random agreement ratios mean (power analysis). The main steps of the evaluation are shown in Figure 1.

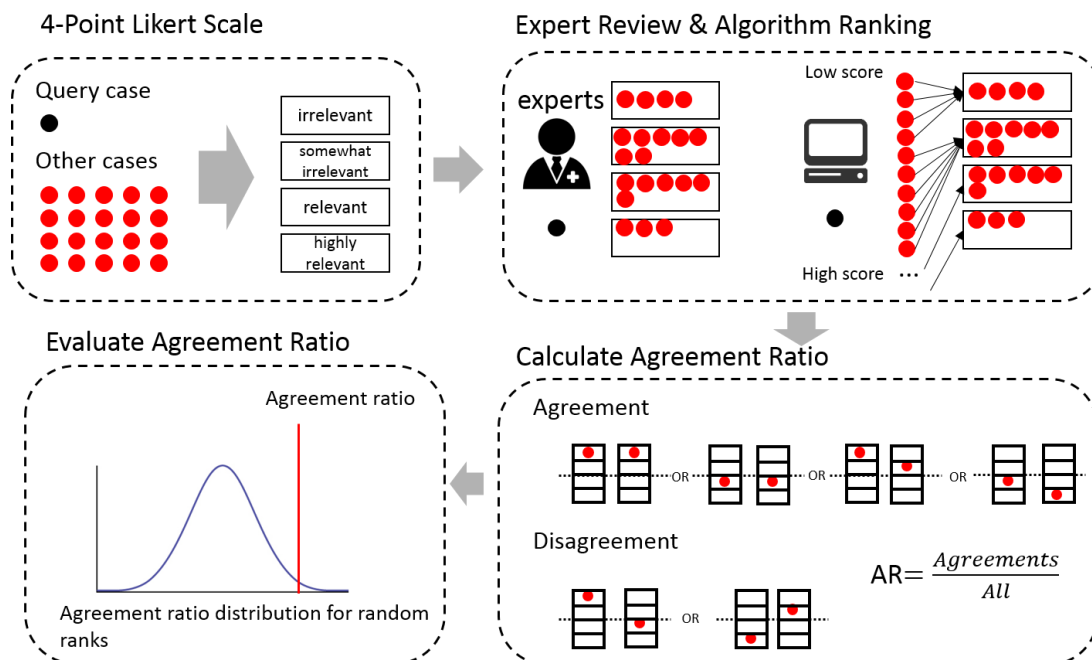


Figure 1. Performance evaluation for the similarity searching model. 1) A 4-point Likert Scale measure which contains 1-irrelevant, 2-somewhat irrelevant, 3-relevant, and 4-highly relevant was adopted; 2) Three experts labeled every case with one of the four scales according to the degree of similarity between this case and the given query case, meanwhile the model ranked the sample cases according to the similarity scores and labeled them with the four scales; 3) The sample agreement ratio was calculated by dividing the numbers of agreement cases by the number of total cases. 4) The performance of the model was evaluated by comparing the sample agreement ratio to the random agreement ratios.

Results

A Similarity Searching Model for PSE Reporting System

To provide solutions for a given PSE report, it should be figured out that which previous cases are similar to the given case. The similarity between PSEs can be measured by comparing their annotations on the same taxonomy. Similar annotations would obtain a high similarity score. AHRQ WebM&M is an appropriate dataset to implement and evaluate the similarity searching model since it provides a taxonomy with 219 terms to describe the features of PSE from six perspectives. In this study, to each pair of the 366 PSE reports, similarity scores were calculated by the three semantic similarity measures (VS, NTO, and IC) separately. As a result, three individual similarity matrices were generated to store all similarity scores.

We developed a local web server to present and evaluate the similarity searching model. Users can either report a new PSE or choose an existing case in the datasets of WebM&M or AHRQ Common Formats as a query case to initiate a similarity searching session. The server will calculate the similarity scores between the query case and every other case, and rank all these cases according to the scores (high to low). Figure 2 shows a screenshot of the similarity searching result when using one of the case in WebM&M named: *Ebola: Are We Ready*³¹ as a query. The similar cases calculated by VS method were presented on the right side with case topics, dates, and authors. When any case is selected by the user, its annotation will be presented on the left side. Any click on these terms will trigger a new searching session by using the clicked term as a new keyword. The system also provides three action buttons on the right side of each similar case including: 1) *ask for details* (the paper icon), to show the details of this case, as well as the commentaries; 2) *choose as a query* (the magnifier icon), to launch a new similarity searching session using this case as a query; 3) *I agree* (the thumb up icon), to receive user feedback and improve the similarity matrices.

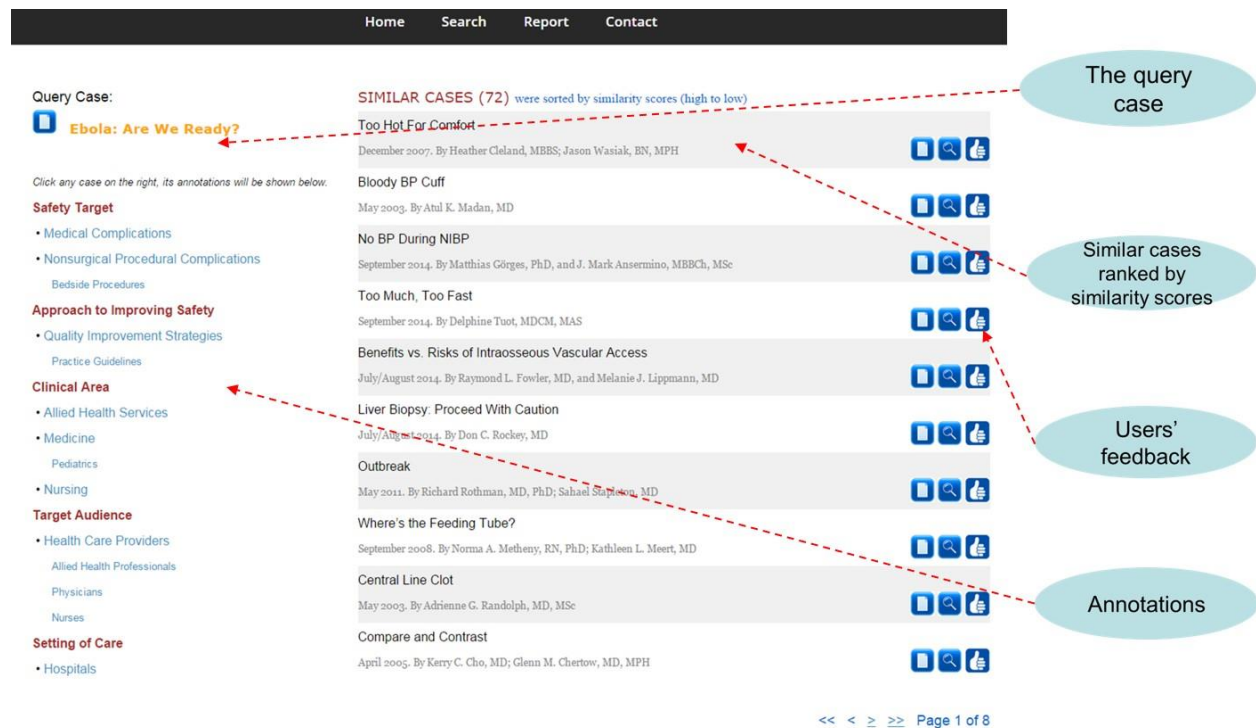


Figure 2. Similar cases of the query case “Ebola: Are We Ready?” calculated by the PSE similarity searching model (Vector Space method)

Model Assessment via Expert Review

We conducted a detailed study of the three semantic similarity algorithms based on AHRQ WebM&M taxonomy. Case 241³², a typical event of nosocomial infections, was chosen as the test query. All the other cases were ranked by our similarity searching model according to the similarity scores, and there were 49 of overall 365 cases carrying nonzero similarity scores in the result list. We randomly extracted 15 cases from the ones with nonzero scores, and 5 cases from the rest, since we supposed the irrelevant cases may occupy 25% of the list according to the natural proportion of the 4-point Likert scale³⁰. Three experts with clinical experience reviewed and rated the 20 cases without implication. The result showed that the agreement of the three experts was 90% before the group discussion, and 100% after the first round of discussion. It was even more encouraging that the only two cases judged as “relevant” to the query by the experts have the highest similarity scores calculated by our model.

Further analysis was then processed to figure out the agreement between algorithms and experts. By comparing with random model, the VS model and NTO model reflect a significantly higher consistency with the experts’ review (Table 2). It is feasible and practical to apply semantic similarity model to measure the similarity of PSEs.

Table 2. The agreements between algorithms and experts

| Model | Agreement | | |
|-------|-----------|----------|----------|
| | Expert 1 | Expert 2 | Expert 3 |
| VS | 80%** | 90%** | 90%** |
| NTO | 90%** | 80%* | 90%** |
| IC | 70% | 70% | 70% |

** P-value < 0.01

* 0.01 ≤ P-value < 0.05

Recommending Solutions to Query Reports

We collected solutions for the most common PSE subtype, patient falls, which occurred about 700,000 to 1,000,000 times each year in U.S. hospitals³³. The solutions were collected and curated from three resources: 1) Pennsylvania Patient Safety Authority Site³⁴; 2) AHRQ WebM&M commentaries; 3) Staff of a PSO institute, Missouri Center for Patient Safety (via interviews). According to the 13 questions in the fall form of AHRQ Common Formats, the solutions of the fall events were divided into two categories: general solutions (applicable to any fall event) and specific solutions (applicable when certain answer options are chosen). Totally 15 general solution entries and 40 specific solution entries were refined as the solution dataset for fall events. In order to provide a succinct solution recommendation to the query case, we further grouped the general solutions into six subtypes: event reporting, education, patient monitoring, risk assessment, communication, and assistive devices.

[Home](#) [Search](#) [Report](#) [Contact](#)

Query Case: SIMILAR CASES SOLUTIONS

ID: 399, Subtype: Patient Fall

Q1: Was the fall unassisted or assisted?
a. Unassisted

Q2: Was the fall observed?
b. No

Q3: Who observed the fall?
N/A

Q4: Did the patient sustain a physical injury as a result of the fall?
b. No

Q5: What type of injury was sustained?
N/A

Q6: Prior to the fall, what was the patient doing or trying to do?
b. Ambulating with assistance and/or with an assistive device or medical equipment

Q7: Prior to the fall, was a fall risk assessment documented?
b. No

Q8: Was the patient determined to be at increased risk for a fall?
N/A

Q9: At the time of the fall, were any of the following risk factors present?
c. Sensory impairment (vision, hearing, balance, etc.)

Q10: Which of the following were in place and being used to prevent falls for this patient?
d. Call light/personal items within reach

Q11: At time of the fall, was the patient on medication known to increase the risk of fall?
c. Unknown

Q12: Was the medication considered to have contributed to the fall?
N/A

Q13: Did restraints, bedrails, or other physical device contribute to the fall (includes tripping over device electrical power cords)?
b. No

Specific Solutions

- Re-evaluate types of assistive devices used by the facility to prevent falls. (Q6.b)
- Provide training to staffs on the use and maintenance of assistive devices. (Q6.b)
- Identify and manage individuals complex resident needs including risk for falls, pressure ulcers, health care-associated infections, and other high risk conditions. (Q7.b)
- Carefully assess risk and evaluate competency before considering involuntary restraints for patients. (Q7.b)
- Deploy monitoring system that can identify high-risk residents. (Q9.c)
- Provide a sitter for agitated patients who appear to be at high fall risk. (Q9.c; Q10.d)
- Educate patients who are at high fall risk to avoid ambulating or getting out of bed without assistance. (Q9.c)
- Require staff to stay with patients who are identified as being at risk to fall while in the bathroom. (Q9.c)
- Move the high fall risk patients to the rooms close to nursing stations. (Q9.c; Q10.d)
- Use of a specific color to identify patients as being at increased risk of fall. (Q9.c)
- Use of falls risk wristbands. (Q9.c)
- Use of a sign inside the room. (Q9.c)
- Use bed alarms to notify staff when a patient attempts to get out of bed. (Q10.d)
- Educate patient to use call button to ask for help getting out of bed if feel unsteady. (Q10.d)
- Educate staff to review and modification of medications, especially psychotropic medications. (Q11.c)

General Solutions

Event Reporting

- Have guidelines for completing patient safety event reports and investigation forms for patients after fall.
- Use of a standardized patient safety event reporting for internal purposes, requiring staff to include environmental factors.

Education

- Address the roles and responsibilities of staff as part of the falls prevention education program.
- Educate patients and their family members regarding the causes of falls and the interventions used to prevent falls.
- Educate and require staff to document the fall in the patient's medical record and request a post-fall investigation.

Patient Monitoring

- Deploy hourly rounds monitoring for patients.

Risk Assessment

- Deploy routine fall risk assessment which were found statistically significant associated with lower rates of falls with injury in 75 facilities.
- Reassess fall risk of patients following a fall.
- Periodic facility review the effectiveness of falls risk assessment tools.
- Modify environmental hazards.

Communication

- Communicate fall risk in the patient's medical record.
- Communicate fall risks when the patient is transferred for testing, therapy, or to another unit.
- Display falls risk indicators on the nurse call system workstation.

Assistive Devices

- Ensure the weight-bearing capacities of assistive devices and aids are known to staff.
- Require patients wear slip-proof socks or shoes.

Figure 3. Solution recommendation for a fall event reported in AHRQ Common Formats. The specific solutions are recommended dynamically according to the report options. E.g., the solution entry “Re-evaluate types of assistive devices used by the facility to prevent falls” was presented because the reporter chose “b. Ambulating with assistance and/or with an assistive device or medical equipment” to answer the Question 6.

As shown in Figure 3, when a user completed a fall report in AHRQ Common Formats (case ID 399), the system provided both similar cases and potential solutions for the user to switch at will. All the specific solution entries have been directly linked to the specific answer option(s), which means an entry will be presented only if the associated option(s) is(are) chosen during the reporting. For example, the solution entry “Re-evaluate types of assistive devices used by the facility to prevent falls” was presented because the reporter chose “b. Ambulating with assistance and/or with an assistive device or medical equipment” to answer the question what was the patient doing or trying to do prior to the fall (Question 6). The general solutions were presented for all fall events. In this way, reporters will spontaneously improve their reports since the quality of the reports relates directly to the quality of the specific solutions they received. It will be a win-win situation that reporters really learn something from the system, and the report quality gets improved.

Discussion

In this study, we proposed a novel workflow to improve the data quality of PSE reporting system, which includes seven key modules: data collection and management, algorithm implementation based on semantic similarity measures, expert review, agreement analysis, statistical test, user interface, and user feedback mechanism (as shown in Figure 4). Using fall events as an example, the proposed schema for PSE reporting system can process the comparison tasks and provide the reporters pertinent suggestions about solutions and prevention options for their reporting cases. The schema will be applied in all PSE subtypes in our future work. It will mend the gap that most current PSE reports are in low quality because the end users cannot receive any perceived benefit during and after the reporting. The task of learning and analyzing from previous experiences will be completed by the intelligent system instead of solely relying on healthcare providers. The more detailed users report, the more accurate suggestions they will receive. Such benefits would greatly enhance the quality of PSE reports. Moreover, the semantic similarity module was embedded into the reporting system which allows users to provide feedback such as whether they agree or disagree with certain similar case or solution by clicking a user feedback button. All the feedback will be returned to the algorithm implementation step and help update the weights of related similarity matrices. In this way, the performance of the similarity searching model will be improved dynamically. The model will be progressively stable and enhanced as the involvement through the users’ feedback.

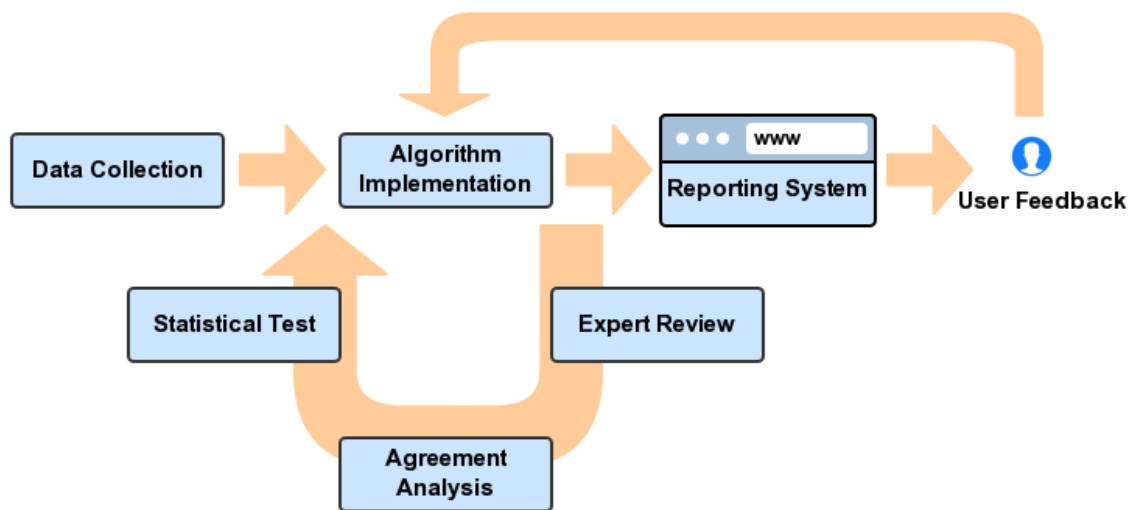


Figure 4. The workflow to improve the data quality of PSE reporting system. Data collection and management, algorithm implementation based on semantic similarity measures, expert review, agreement analysis, statistical test, user interface, and user feedback mechanism are the key modules to improve the data quality of the proposed PSE reporting system.

A challenge of the future work is how to standardize PSE storage and management in order to overcome the lack of interoperability and communication. In this study, we found that PSEs can be compared to each other if they are managed by a hierarchical data structure (e.g., taxonomy) according to their characteristics. Therefore, our future plan is developing a knowledge base in patient safety domain to standardize PSE reports and solutions at the data

level. This standardization will make connections not only among diverse reporting formats but also between solutions and PSE reports. The solutions will be recommended based on the similar cases instead of answer options, which will make the recommendations more targeted. This plan is feasible and low-risk since we are not going to develop a brand-new knowledge base, instead, we will extend part of the terms and relationships defined by International Classification for Patient Safety (ICPS)³⁵ to a knowledge base in patient safety domain. The ICPS is carried out under the World Alliance for Patient Safety of the World Health Organization (WHO). ICPS represents the knowledge related to PSE that occur to patients during hospital stays. For the time being, the representation has focused on falls and pressure ulcer incidents. To construct this knowledge base, we plan to extract the sub-classes including terms and relationships about PSE from ICPS. Then we will use Competency Questions (CQ), a classical method for designing and evaluating knowledge bases³⁶, which are interview questions that require candidates to provide real-life examples as the basis of their answers. Two reviewers with background knowledge in patient safety will participate in the CQ development. To each subtype of PSE, they will review a group of cases and generate the CQs. Each CQ must not be duplicated in its meaning with any of the others to the best of reviewers' judgment. Then we will merge the CQs with the same meanings and break down the subdivisible ones. Thereafter, corresponding classes will be selected and combined with those extracted from ICPS, and a knowledge base describing each PSE subtype will be generated. The developing schema holds promise in improving patient safety and potentially triggering a revolution for data management and analysis in health care industry.

Clinical Decision Support Systems (CDSS) link clinical observations with health knowledge to assist clinical decisions³⁷. The systems influence clinician's decisions and consequently enhance health-care quality. However, most CDSS suffer from common problems in usability, which have received significant attention in the patient safety community³⁸⁻⁴¹. Several subtypes of PSE are associated with clinical decisions, for example, a medication event may lead a drug-drug interaction. Therefore, in the future work, we will design several specific interfaces to CDSS, embed the key function models of our system to some widely used CDSS, and help detect potential risks at the early stage, thus provide better decision support and eventually improve patient safety.

Limitations

Database The main defect of AHRQ WebM&M dataset is the small sample size which comprises only 366 PSE reports by January 2016. The selection criteria are unclear, however, based on our observation, the cases in WebM&M may have been chosen as the most typical ones in each category. Taking patient fall as an instance, as the most common event type, it only has four records in WebM&M, which does not mean patient fall is infrequent, but rather indicates the other cases might be similar to the four typical samples and thus were not included by the experts. This is probably the main reason that the algorithm Information Content (IC) has the worst performance on this dataset. IC treats the frequency of each term as an important parameter that high frequency means low information content. This rationale does not apply to this experiment because the frequencies in WebM&M dataset cannot represent the real PSE distribution.

The limitation of AHRQ Common Formats is the low report quality. Many hospitals are using the Common Formats to report their events because of the requirement of PSO, which helps PSO collect a large sum of PSE reports. However, the quality of the reports is unsatisfactory since not all the reporters reported their cases spontaneously. We found that some reporters tended to choose the obscure options rather than the explicit ones which may cost more time to figure out the details. For example, reporters are more likely to choose the option "Unknown" if provided. Unfortunately, most "Unknown" choices are least meaningful for further analysis, and called missing values. This is consistent with our previous finding⁴².

However, according to the status quo, these two report formats seem to be the best choices for this study since they are by far the only accessible PSE databases with hierarchical feature labels. We believe the quality of reports will be improved if reporters can really learn something from the system during and after their reporting. That is one of the reasons we propose this study.

Assessment Strategy All the assessments in this study were processed with the help of health care experts since there is no golden standard for PSE similarity. Each expert likely brings various biases on his/her different perspectives. For example, a clinician may judge the similarity between PSEs by measuring their severity, while a nurse may judge the similarity based on solution finding. The biases cannot be fully avoided. However, we provided more targeted introductions before every round of expert review, and adopted interviews rather than questionnaires in order to help the experts better understand our purpose.

Conclusion

Voluntary PSE reporting systems have a great potential for improving patient safety through wide adoption and effective use in healthcare. The similarity analysis of PSE is a key to the success of such systems. Focusing on the quality issue of current reporting systems, this study proposed a novel schema to improve the data quality by gathering information from previous experiences and informing subsequent action in a timely manner. Healthcare providers can learn how to avoid hazardous consequences and prevent their recurrence after they report their PSEs to our system. These benefits encourage the reporters to spontaneously provide reports with higher quality, since the more detailed they report, the more accurate suggestions they will receive. Furthermore, the schema holds promise in developing a new generation of self-learning PSE reporting system based on a standard knowledge base in patient safety domain. As the system being used by more organizations for a longer period of time, the internal knowledge base will become more intelligent, eventually provide better services.

Acknowledgement

We thank Drs. Jing Wang and Nnaemeka Okafor for their expertise and participation in expert review. This project is supported by UHealth Innovation for Cancer Prevention Research Training Program Post-Doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant #RP160015), Agency for Healthcare Research & Quality (1R01HS022895), and University of Texas System Grants Program (#156374).

References

1. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med*. 2010;2(57):57cm29.
2. Hersh W. Health care information technology: progress and barriers. *JAMA*. 2004;292(18):2273-4.
3. Walker JM, Carayon P. From tasks to processes: the case for changing health information technology to improve health care. *Health Aff (Millwood)*. 2009;28(2):467-77.
4. Shaw R, Drever F, Hughes H, Osborn S, Williams S. Adverse events and near miss reporting in the NHS. *Qual Saf Health Care*. 2005;14(4):279-83.
5. The Center for Drug Evaluation and Research: U.S. Food and Drug Administration. Available from: <http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/>.
6. AHRQ. Users guide AHRQ Common formats for patient safety organizations: Agency for healthcare Research and Quality; 2008.
7. Orfanidis L, Barnidis PD, Eaglestone B. Data quality issues in electronic health records: an adaptation framework for the Greek health system. *Health Informatics Journal*. 2004;10(1):23-36.
8. Myers SA. Measurements and data integration. *Patient safety and hospital accreditation: A model for ensuring success*. New York: Springer; 2012. p. 250-1.
9. Data quality: mission needs statement for Data Quality Integrated Program Team (IPT). *Military Health Systems*, 2003.
10. Barriers to learning from incidents and accidents 2015. Available from: <http://www.esreda.org/Portals/31/ESReDA-barriers-learning-accidents.pdf>.
11. Aamodt A, Plaza E. Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications*. 1994;7:39-59.
12. WHO. The conceptual framework for the international classification for patient safety. *World Health Organization*, 2007.
13. Pesquita C, Faria D, Falcao AO, Lord P, Couto FM. Semantic similarity in biomedical ontologies. *PLoS Comput Biol*. 2009;5(7):e1000443.
14. Guzzi PH, Mina M, Guerra C, Cannataro M. Semantic similarity analysis of protein data: assessment with biological features and issues. *Brief Bioinform*. 2012;13(5):569-85.
15. Benabderrahmane S, Smail-Tabbone M, Poch O, Napoli A, Devignes MD. IntelliGO: a new vector-based semantic similarity measure including annotation origin. *BMC Bioinformatics*. 2010;11:588.
16. Janowicz K, Raubal M, Kuhn W. The semantics of similarity in geographic information retrieval. *Journal of Spatial Information Science*. 2011;2:29-57.
17. Kaur I, Hornof AJ, editors. A Comparison of LSA, WordNet and PMI for Predicting User Click Behavior. *Proceedings of the Conference on Human Factors in Computing*; 2005: CHI.
18. Gracia J, Mena E, editors. Web-Based Measure of Semantic Relatedness. *Proceedings of the 9th international conference on Web Information Systems Engineering (WISE '08)*; 2008; Springer-Verlag, Berlin, Heidelberg.
19. Pirolli P. Rational analyses of information foraging on the Web. *Cognitive Science*. 2005;29(3):343-73.
20. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32(Database issue):D258-61.

21. Mistry M, Pavlidis P. Gene Ontology term overlap as a measure of gene functional similarity. *BMC Bioinformatics*. 2008;9:327.
22. del Pozo A, Pazos F, Valencia A. Defining functional distances over gene ontology. *BMC Bioinformatics*. 2008;9:50.
23. Lim WK, Wang K, Lefebvre C, Califano A. Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics*. 2007;23(13):i282-8.
24. Schlicker A, Domingues FS, Rahnenfuhrer J, Lengauer T. A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics*. 2006;7:302.
25. Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF. A new method to measure the semantic similarity of GO terms. *Bioinformatics*. 2007;23(10):1274-81.
26. Wachter RM, Shojania KG, Minichiello T, Flanders SA, Hartman EE. AHRQ WebM&M-Online Medical Error Reporting and Analysis Tools, and Products). 2005.
27. Rada R, Mili H, Bicknell E, Blettner M. Development and application of a metric on semantic nets. *IEEE Transaction on Systems, Man, and Cybernetics*. 1989;19:17-30.
28. Lin D, editor An information-theoretic definition of similarity. *Proceedings of the 15th International Conference on Machine Learning*; 1998; San Francisco, CA: Morgan Kaufmann.
29. Popescu M, Keller JM, Mitchell JA. Fuzzy measures on the Gene Ontology for gene product similarity. *IEEE/ACM Trans Comput Biol Bioinform*. 2006;3(3):263-74.
30. Likert R. A technique for the measurement of attitudes. *Archives of Psychology*. 1932;140:1-55.
31. International Classification for Patient Safety: World Health Organization. Available from: http://www.who.int/patientsafety/implementation/taxonomy/development_site/en/.
32. Case 241: Outbreak: AHRQ; 2011 [cited 2016 Jan 6]. Available from: <https://psnet.ahrq.gov/webmm/case/241>.
33. Currie L. Fall and Injury Prevention. In: Hughes RG, editor. *Patient safety and quality: An evidence-based handbook for nurses*. Rockville, MD: AHRQ Publication No. 08-0043; 2008.
34. James JT. A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf*. 2013;9(3):122-8.
35. Patient safety event report: Center for Leadership, Innovation and Research in EMS. Available from: <http://event.clirems.org/Patient-Safety-Event>.
36. Grüninger M, Fox MS. Methodology for the design and evaluation of ontologies. *IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing*1995.
37. Berner ES, Lande TJL. Overview of clinical decision support systems. In: Berner ES, editor. *Clinical Decision Support Systems: Theory and Practice*. 2nd ed. New York, USA: Springer; 2007. p. 3-22.
38. Zhang J. Human-centered computing in health information systems. Part 1: analysis and design. *J Biomed Inform*. 2005;38(1):1-3.
39. Zhang J. Human-centered computing in health information systems part 2: evaluation. *J Biomed Inform*. 2005;38(3):173-5.
40. Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. *J Biomed Inform*. 2005;38(1):75-87.
41. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. *J Biomed Inform*. 2004;37(1):56-76.
42. Gong Y. Data consistency in a voluntary medical incident reporting system. *J Med Syst*. 2011;35(4):609-15.