# Leveraging user's performance in reporting patient safety events by utilizing text prediction in narrative data entry

Yang Gong [a,*], Lei Hua [a,b], Shen Wang [c]

[a] School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX, USA
[b] Informatics Institute, University of Missouri, Columbia, MO, USA
[c] Department of Nursing, Tianjin First Central Hospital, Tianjin, China

## ARTICLE INFO

## ABSTRACT

*Background:* Narrative data entry pervades computerized health information systems and serves as a key component in collecting patient-related information in electronic health records and patient safety event reporting systems. The quality and efficiency of clinical data entry are critical in arriving at an optimal diagnosis and treatment. The application of text prediction holds potential for enhancing human performance of data entry in reporting patient safety events.

*Objective:* This study examined two functions of text prediction intended for increasing efficiency and data quality of text data entry reporting patient safety events.

*Methods:* The study employed a two-group randomized design with 52 nurses. The nurses were randomly assigned into a treatment group or a control group with a task of reporting five patient fall cases in Chinese using a web-based test system, with or without the prediction functions. T-test, Chi-square and linear regression model were applied to evaluating the outcome differences in free-text data entry between the groups.

*Results:* While both groups of participants exhibited a good capacity for accomplishing the assigned task of reporting patient falls, the results from the treatment group showed an overall increase of 70.5% in text generation rate, an increase of 34.1% in reporting comprehensiveness score and a reduction of 14.5% in the non-adherence of the comment fields. The treatment group also showed an increasing text generation rate over time, whereas no such an effect was observed in the control group.

*Conclusion:* As an attempt investigating the effectiveness of text prediction functions in reporting patient safety events, the study findings proved an effective strategy for assisting reporters in generating complementary free text when reporting a patient safety event. The application of the strategy may be effective in other clinical areas when free text entries are required.

Published by Elsevier Ireland Ltd.

* Corresponding author. 7000 Fannin St. Suite 165, Houston, TX 77030, USA. Tel.: +1 713 500 3547; fax +1 713 486 0117.
E-mail addresses: Yang.Gong@uth.tmc.edu; gongyang@gmail.com (Y. Gong).

## 1.　Introduction

Many attempts have been made to investigate the difficulties in clinical data entry in order to promote the acceptance and quality-in-use of clinical information systems [1–4]. With the advance of non-narrative entry templates and natural language processing techniques, the application of structured entries in clinical information systems is increasing because of the merit of interoperability and reuse. However, given the complexity of healthcare, failure to include essential fields and lack of options in structure entries are still common. Structured entries also received critiques due to lack of flexibility and expressiveness in clinical communication. Therefore, unstructured narrative data entry plays an indispensable role in clinical data entry.

For patient safety event reporting, narrative entry has been a prevalent and dominant format. However, it is now in a transition of becoming a supplementary format to non-narrative forms released by national and international organizations [5,6]. Narrative comment field is intended to collect case details beyond structured entries. Previous studies showed that voluntary reporters usually ignored this field or described events with inaccurate and incomplete terms and sentences [7,8]. Two major barriers are identified due to the multitasking and busy nature of healthcare and the lack of languages or knowledge for reporters to describe patient safety events in detail [9]. To help remove the barriers and achieve quality-in-use, it requires a user-centered design process which embraces cultural, strategic and technical considerations [10]. We proposed to examine the utilization of text prediction applied in the narrative fields in reporting patient safety events.

Text prediction, also known as word, sentence or context prediction originated in augmentative and alternative communication (AAC) to increase text generation rates for people with the disabilities of motor or speech impairment [11,12]. The advance of natural language processing techniques has brought text prediction into a broader scope of daily computing activities, such as mobile computing [13] and radiography reports [14]. Studies on optimization [15] and impact [16] of text prediction have received much attention. This study focusing on the impact evaluation, would respond three basic concerns regarding text prediction in reporting patient safety events. These concerns include whether text prediction would increase (1) reporter's engagement in narrative comment field; (2) quality of narrative entry, which is highly valuable in generating actionable knowledge yet received little attention; (3) efficiency of reporting patient safety events, which remains unclear based on mixed results across the fields [16]. In this study, we employed a two-group randomized experiment to examine the impact of text prediction in the three aspects.

## 2.　Background

This study is grounded in a user-centered design of patient safety event reporting systems. Reporting systems in use show the problems of underreporting [17] and low quality of reports [8,18], though efforts from all levels are made to improve the systems [19,20]. Prior to this study, we have consecutively conducted heuristic evaluation, cognitive task analysis and think-aloud user testing [4,7,21,22] which revealed interface representational issues and identified that the introduction of text prediction holds promise in increasing efficiency and quality of event reporting. Therefore, text prediction functions on data entry were proposed to bridge the information gaps induced by competing tasks in the work domain and languages or knowledge needed for quality reporting [23].

In this study, we prototyped and examined two prediction functions aiming at facilitating data entry in free text format. The process may be helpful for the introduction and customization of domain specific, predefined drop-down lists as well as the optimization of text prediction accuracy and interface representation.

## 3.　Methods and materials

### 3.1.　Participants

Potential candidates who were nurses and experienced in reporting and analyzing patient safety events in Tianjin First Central Hospital (TFCH) in Tianjin, China were identified and invited to participate in the study. Two candidates were on leave of absence during the study period, and three candidates felt not confident with operating computers. As a result, the remaining 52 nurses from 21 clinical departments enrolled in the study. All nurses were female between the ages of 30 and 52 years. On average, they had approximately 20 years of nursing experience and reported patient safety events for at least four years since the implementation of a citywide computerized reporting system in 2009. None of them used the interfaces of this study before. Each participant signed an informed consent form approved by the Ethics Committee at the TFCH (No.E2012022K). This study was also approved by the Institutional Review Board at the University of Texas Health Science Center at Houston (No. HSC-SBMI-12-0767).

### 3.2.　Interfaces

Two experimental interfaces were developed for the purpose of control over the configurations and data collection. The interfaces have two language versions in English and Chinese. The contents and layouts of two interfaces were identical, carrying the same task of the 13 multiple-choice questions (MCQs) [24] and a narrative comment field for the collection of case details. The MCQs and narrative comment field were developed based on the Common Formats released by AHRQ [24]. A sample of the MCQs is shown in Fig. 2. A single exception between the two experimental interfaces was the provision of text prediction functions to the narrative field. During the experiment, participants had to complete the MCQs prior to the narrative comment field. We added branching logic to the MCQs to ensure that only the relevant questions were displayed and irrelevant questions were skipped based on the response to a branching question. For example, when "assisted fall" was selected, all questions associated with "unassisted fall" would be skipped. The branching also helped prepare prediction functions.
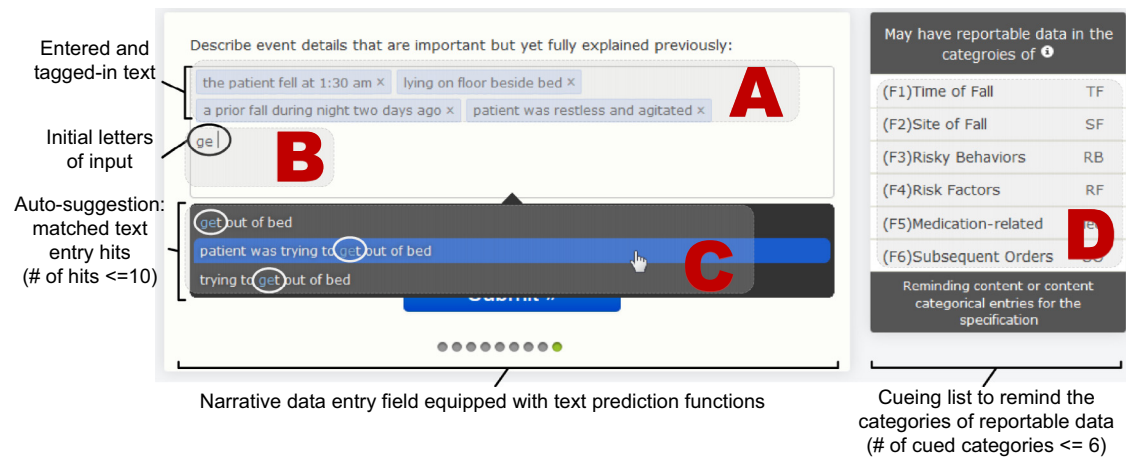
**Fig. 1 – The layout of interface elements associated with the narrative comment field and text prediction functions, cueing list (CL) and autosuggestion (AS). CL (shown in box D) served as a reminder of the key characteristics of reportable data associated with the event. Matched AS was triggered and listed as entry options when initial letters (shown in box B & C) of description were typed in. Matched letters and cursor line were highlighted in blue as shown in the box C. The reporter was expected to select one of them and make any changes when deemed necessary.**

Fig. 1 illustrates the interface in English and Fig. 2 illustrates the one in Chinese where the narrative comment field resides with text prediction functions. The interfaces with prediction functions were developed using PHP 5.2.6, JavaScript, MySQL 5.0.51b plus a JavaScript library (JQuery 1.7 [25]) and two open source modules (SlidesJS [26] and Tag-it [27]).

Following the experimental designs in peer reviewed publications [16,28], the predicted text items were manually selected per case from a pool of entries which was automatically generated based upon a collection of fall cases. Two domain experts (X.L. & Y.S.) reviewed all testing cases, transcribed and categorized the key elements with colloquial language in both Chinese



**Fig. 2 – The layout in the Chinese equivalent of Fig. 1 based on domain experts' review. This version of interface was used in Tianjin First Central Hospital where the randomized tests were conducted.**

characters and Pinyin Input. These contents were populated into a text entry dictionary, a pool of AS and CL items. A Soundex-based phonetic matching function of MySQL was employed for AS to retrieve and present matched entries in the dictionary. The number of listing items for AS and CL was not greater than ten, a trade-off number balancing inspecting efforts against predicting sensitivity [15]. On the treatment interface where AS and CL were offered, participants were allowed to choose from the predicted entries and modify the entries. On the control interface where no prediction was offered, participants had to key in narratives by using a standard QWERTY keyboard.

### 3.3. Testing cases

Each participant was asked to report five cases in a randomized sequence. The cases were selected from two sources – a repository of de-identified 346 fall cases [18] and a publicly accessible online database of Morbidity and Mortality (Web M&M) [29]. The cases were translated into Chinese, rephrased and reviewed by two domain experts for the purpose of quality, readability, and consistent complexity. The narrative excerpted from one of cases is shown as an instance below.

> "… patient was alert and oriented X3 (person, time and location) upon assessment, and instructed upon admittance not to getting up without assistance. He had been sleeping and attempted to get up to go to the bathroom. He forgot to call the staff to have plexipulses (a device) undone, and tripped on plexi tubing and attempted to catch himself on the overhead bars. He landed on the floor …"

### 3.4. Experimental design

With a permuted-block algorithm and random block [30], fifty-two participants were randomly assigned to two groups. Twenty-five participants were allocated into the control group and twenty-seven into the treatment group. During each participant's testing, the presenting sequence of five cases was randomly determined at the time of allocation. According to stratified randomization, one case was fixed at the first position and the other four were fully randomized. Prior to each testing, there was a mandatory training session combined with verbal instructions and hands-on practice. All participants were trained on both control and treatment interfaces until they felt comfortable with the interfaces. Each participant was briefly surveyed after completing the experiment for collecting any comments or suggestions on the reporting.

In practice, when a patient safety event occurred, a reporter would initiate a report upon witness's word-of-mouth information. Similarly, the interfaces presented a case on the first page. Participants were expected to read the case descriptions and report it based on recall. Pauses and questions during the timed sessions were discouraged except when participants were in transition between reports. Keystroke level operations including mouse clicks and keystrokes were timed and logged into a MySQL database. All reporting sessions were recorded by using Camtasia Studio® 7 for data reconciliation.

### 3.5. The processing of data

We first analyzed the completion time and response accuracy on the MCQs through SQL queries of logged keystrokes. A t-test was conducted to examine whether or not the identical part prior to text data entry was of significant difference between the groups. Then, we analyzed ten dependent measures (shown in Table 1) in the text data entry to investigate each participant's performance and variation in terms of efficiency, effectiveness, and engagement. Apart from the nine measures retrieved from the logged data, report comprehensiveness was measured based on a blind review of two domain experts. The experts used de-identified narrative comments of both groups for rating. As illustrated in Section (A) of Fig. 1, the narrative comments were segmented by a pipeline symbol "|" that was added as participants pressed <Enter> key to conclude the description of an event. The experts independently reviewed the text chunks by grading them in a binary score either one or zero. Score one indicates a chunk was event-related and not a duplicate, the chunk thus contributed one point to an overall comprehensiveness score of the report. The range of overall comprehensiveness scores was between one and twelve in the study. Any rating differences greater than one point were resolved by the team discussion. The mean of comprehensiveness scores was used to represent the report comprehensiveness.

To examine the impact of text prediction (CL and AS) functions on participants' performance, a t-test and Chi-squared test were conducted between the two groups. Kernel density was applied to examine the distributions of text generation rate and the reporting comprehensiveness of narrative comments between the groups. A linear regression model was also used in examining the interactions between the measures. All statistical computing were executed by using MySQL embedded functions and R Studio ver 0.97.

## 4. Results

All participants, with each reported five cases, successfully completed the experiment and thus generated 260 reports. There were 25 participants allocated in the control and 27 in the in the treatment group, accounting for 125 and 135 reports, respectively. The means of participants' ages were $43.6 \pm 5.8$ versus $41.1 \pm 6.6$ years in the control and treatment group. The 260 reports contained 2849 MCQ answers and 238 unstructured narrative comments. The completion times of MCQs were $131.0 \pm 50.0$ versus $129.5 \pm 45.7$, and the response accuracy were 79.4% versus 79.2% in the control and treatment group. No significant differences between the groups were found in the abovementioned variables according to the t-test at significance level of 0.05. Of the 238 unstructured narrative comments, 105 comments were contributed by the control group and the rest 133 comments by the treatment group. The unanswered comment fields were 20 versus 2 between the two groups. The detailed findings on the narrative comment fields are shown in Table 2.

**Table 1 – Dependent measures on efficiency, effectiveness, engagement and autosuggestion.**

| Measures | Data collection | Evaluating dimensions | Methods |
|---|---|---|---|
| Efficiency-related | | | |
| Completion time | Recorded at the millisecond level by interfaces | Time length of completing a narrative comment | Descriptive statistics, and t-test |
| Keystrokes | Recorded by interfaces | Keystroke counts of completing a narrative comment | Descriptive statistics, and t-test |
| Text generation rate | Text length divided by completion time | The speed of text generation measured by "letters/second" | Descriptive statistics, and t-test |
| Effectiveness-related | | | |
| Text length | Recorded and calculated at the unit of letter | Text length (in letters) of a narrative comment | Descriptive statistics, and t-test |
| Text chunks | As demonstrated in Fig. 1, pressing <enter> resulted in a tag in the text segment, i.e. text chunk | Number of text chunks in a narrative comment field | Descriptive statistics |
| Chunk length | Text length divided by number of text chunks | Mean length of text chunks in a narrative comment field | Descriptive statistics |
| Reporting comprehensiveness | A blind review by two experts | Number of event characteristics described in the text | Expert review, descriptive statistics and t-test |
| Engagement-related | | | |
| Non-adherence rate | Amount of unanswered narrative comment fields divided by amount of narrative comment fields | Proportion of unanswered narrative comment fields | Descriptive statistics, and Chi-squared test |
| AutoSuggestion (AS)-related | | | |
| Influenced chunks | Identified text chunks contained in original AS | Number of text chunks accepted as suggested by AS | Descriptive statistics |
| Influential rate | Number of influenced chunks divided by the number of total text chunks created in a narrative comment | Percentage of text chunks created via AS rather than keyed in | Descriptive statistics |

## 4.1. Efficiency–text generation rate

According to Table 2, the completion time of the narrative comments did not show a significant effect ($p = 0.782$) between the groups. However, the participants in the treatment group contributed 44.7% more text with 28.2% fewer keystrokes than those in the control group, accounting for a 70.5 % increase in the text generation rate, which was a significant improvement in reporting efficiency at significant level of 0.01. Fig. 3 shows a comparison of distributions of text generation rates between the two groups, which reveals the text prediction helped the participants in the treatment group reach a higher rate in generating text ($p < 0.01$).

## 4.2. Effectiveness–reporting comprehensiveness

The number of text chunks ($5.4 \pm 2.5$) and the mean length of the chunks ($37.7 \pm 18.6$) in the treatment group are significantly ($p < 0.01$) greater than those ($4.1 \pm 2.5$ and $30.3 \pm 13.1$) in the control group. Review scores for reporting comprehensiveness range from $5.1 \pm 2.4$ in the treatment group to $3.8 \pm 2.3$ in the control group. Fig. 4 illustrates the distributions of

**Table 2 – Participants' performance on the narrative comment fields between groups.**

| Measures | Samples excluding blank spaces | | | |
|---|---|---|---|---|
| | Control (N = 105) | Treatment (N = 133) | Variation | p-value |
| Efficiency-related | | | | |
| Completion time (seconds) | $139.6 \pm 99.6$ | $142.9 \pm 82.2$ | ↑ 2.3% | 0.782 |
| Keystrokes | $144.9 \pm 110.7$ | $104.0 \pm 86.9$ | ↓ 28.2% | 0.002 |
| Text generation rate (letters/second) | $0.95 \pm 0.35$ | $1.62 \pm 0.99$ | ↑ 70.5% | 0.000 |
| Effectiveness-related | | | | |
| Text length (letters) | $127.9 \pm 96.6$ | $185.1 \pm 86.4$ | ↑ 44.7% | 0.000 |
| Text chunks | $4.1 \pm 2.5$ | $5.4 \pm 2.5$ | ↑ 31.7% | 0.000 |
| Chunk length (letters) | $30.3 \pm 13.1$ | $37.7 \pm 18.6$ | ↑ 24.4% | 0.000 |
| Reporting comprehensiveness | $3.8 \pm 2.3$ | $5.1 \pm 2.4$ | ↑ 34.2% | 0.000 |
| Engagement-related | | | | |
| Non-adherence rate | 20/125(16.0%) | 2/135(1.5%) | ↓ 14.5% | 0.000 |
| AS-related | | | | |
| Influenced chunks (N = 120) | – | $3.8 \pm 1.9$ | – | – |
| Influential rate | – | 66.9% | – | – |
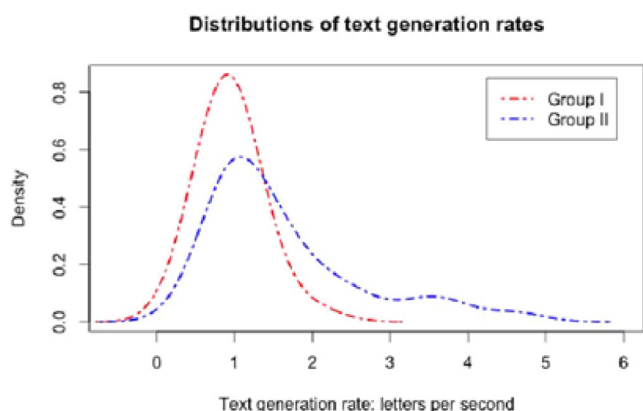
Fig. 3 – Text generation rates in the control group (I) and treatment group (II).

comprehensiveness scores between the two groups. The difference of comprehensiveness scores is statistically significant which exhibits an effective intervention contributed by the prediction functions.

### 4.3. Participant's engagement–non-adherence rate

The *Chi*-squared test identified a significant difference in the non-adherence rates of the narrative comment field between the groups. There were 20 comment fields out of 125 reports left blank in the control group; whereas 2 out of 135 in the treatment group. The lower non-adherence rate indicates that participants were more actively engaged in describing the event details. Such an effect was expected because of the provision of CL in the treatment group.

### 4.4. AS–utility and influence rate

Among 133 narrative comments generated in the treatment group, AS was used 460 times during the text input on 120 comments, accounting for an AS utility rate of 90.2%. On average, AS influenced $3.8 \pm 1.9$ text chunks with an overall influen-

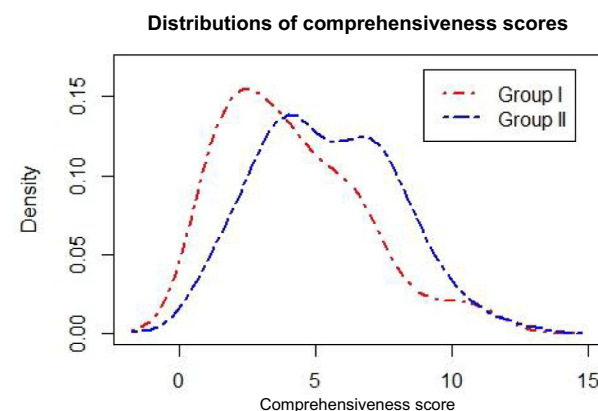**Effectiveness – reporting comprehensiveness**



Fig. 4 – Comprehensiveness scores in the control group (I) and the treatment group (II).
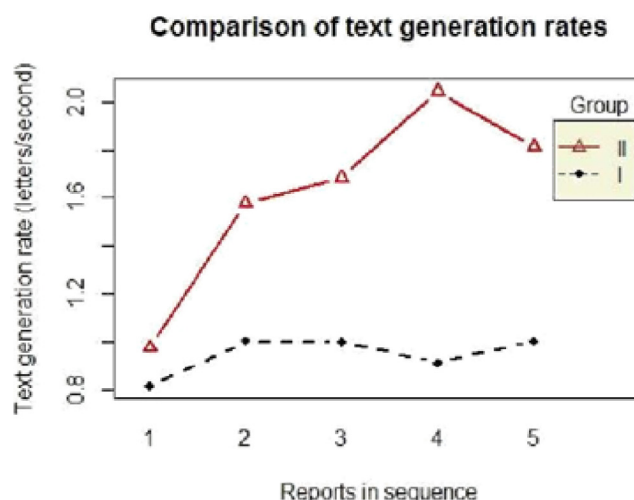


Fig. 5 – Text generation rates as the experiment proceeded in the control group (I) and treatment group (II).

tial rate of 66.9%. The regression analysis showed this influential rate increased as the experiment proceeded ($p = 0.029$). Meanwhile, the text generation rate in the treatment group increased over time as shown in Fig. 5. In contrast, the text generation rates in the control group showed certain flatness along the process of reporting.

## 5. Discussion

Clinicians working under time constraints are usually expected to enter data for documentation in a timely and efficient manner [31,32]. The comprehensiveness of entered data is critical to clinicians' decision-making and creation of actionable knowledge. To accommodate the expectation in patient safety event reporting, we introduced two text prediction functions of CL and AS, which are attached to narrative comment field pervasively used in patient safety reporting systems. A two-group randomized design examined the impact of the predictions on participant performance in terms of efficiency, effectiveness and engagement. The results are positive and may be of guidance toward designing and optimizing reporting systems toward a safer healthcare system.

### 5.1. Reporting efficiency

One of the major findings explains that the text predictions can improve participant's efficiency, which is critical to busy clinicians. The study scrutinized three measures associated with the efficiency in terms of completion time, keystrokes and text generation rates. During a similar period of time, the treatment group produced much more text, translated into a higher text generation rate, than the control group. The upward trend of text generation rate in the treatment group implies the prediction functions can be manifest when users are adapted to them.

In the treatment group, participants generated more text with 28.2% fewer keystrokes. This is consistent with the study

results in a variety of fields [14,33]. Nevertheless, the introduction of the predictions, accounted for increased cognitive loads, eye gaze movements and the total number of mouse clicks, presents mixed results because keystroke savings alone may not adequately explain the increased efficiency in this study [34–36]. With the trend of health information systems migrating from desktops to mobile terminals, the core value of keystroke savings with on-screen keyboards would be amplified. Usually, keystrokes with on-screen keyboards have a greater time cost than those with regular computer keyboards.

### 5.2. Reporting comprehensiveness

The prediction function CL and AS contributed to the comprehensiveness score of narrative comments by cueing frequent characteristic categories, sentences and terms associated with the event, and leveraging the breadth and depth of comments. The functions served as mnemonic tips which transform a pure recall process into a mixed effort comprising of recall and recognition. Consequentially, the participants in the treatment group generated more information in terms of text length, number of text chunks and comprehensiveness scores.

We noted that typographical errors or imprecise phrases are inevitable within the text generated from both groups. However, the treatment group had a tendency of fewer issues than the control group. The cueing list utilized in the treatment group implies a supportive application for event reporting systems.

### 5.3. Participant's engagement

In contrast to our previous study which showed a 73.3% non-adherence rate among inexperienced users [7], the experienced users of both groups in this study were much more engaged in generating comments. We believe that experienced participants possess better knowledge and reporting skills which help them identify the essence of reports and complete the case descriptions in depth.

As shown in Table 1, the non-adherence rate of control group was 14.5% and treatment group was 1.5% as a result of the introduction of CL in the narrative comment field. Through the post-experiment survey, some participants in the control group explained their non-adherence as (1) a slip of skipping the field unintentionally; (2) having no clues on further describing event characteristics and (3) memory fade. As a remedy, the dynamic display of CL may have altered the optional nature of the comment field or may be more inviting to participant's conscious attention [37] to the interface. This dynamic CL rendered a compelling message to the participants signaling the importance of filling the field.

### 5.4. Limitations and future research

We noticed that the investigated narrative comment fields served as a complementary component in a single field functioning as a catch-up after the MCQs regarding an identical topic. As a result, the findings may be not totally applicable to pure free text fields without an introduction of structured entries. The topics on patient safety can be more complex when beyond one single event type.

The high quality CL and AS entries prepared by a manual process may not be fully scalable when a completely automatic approach is employed based on case similarity and event frequency. Expert review of CL is highly suggested for reaching a similar quality and completeness effect presented in this study. The AS carefully prepared by domain experts in the study demonstrates high quality and may not fully represent the AS potentially created by an automatic approach. Further studies developing domain specific AS must include an expert review process so as to ensure the performance and user experience.

The number of predicted listings may vary from one setting to another. In general, an extended list would cost more time for a reporter to skim or inspect, and thus there would be a greater chance of missing correct responses when time is limited. We are interested in investigating further if there is a tradeoff between accuracy and length of the predicted listings and how the tradeoff would impact participant's performance. For this reason, the increased comprehensiveness can only be partially credited to the introduced AS and CL functions.

Although the text generation rate and accuracy increased, AS function may become a constraint which limits the participant's breadth of recall and thinking at certain points. This is because busy users tend to recognize or ignore choices rather than specify answers [8,38,39] when predefined options are provided in MCQs, event classification, or error types. Thus, it is worthwhile in future research paying attention to such an open question between speed and quality.

## 6. Conclusion

Text data entry, as an indispensable format in clinical information systems, was demonstrated to be enhanced by the text prediction functions in terms of efficiency, data quality and engagement in a two-group randomized experiment. This study disclosed the necessity of the text prediction functions in facilitating experienced domain users in text data entry when reporting patient safety events. The quality of event reporting plays a key role in learning from the events. Simply counting numbers of the reports without an insight into the quality of the contents beyond the structured data fields may undermine the learning value of the reports.

## Authors' contributions

Yang Gong: Responsible for study design, data collection and analysis, and coauthored the manuscript with review and revision responsibilities.

Lei Hua: Responsible for the conceptualization, study design, prototype development, testing session observation, data collection and analysis, coauthored the manuscript.

Shen Wang: Organized and coordinated with the study participants, assisted in preparing testing cases, data analysis and had review responsibilities.

## Conflicts of interest

## Acknowledgments

## Summary table

Known on the topic

- Text data entry is indispensable toward the quality care delivery but suffered from low quality and user's engagement.
- Text prediction is widely used in mobile computing.

  Our contribution

- Text prediction can increase the efficiency, data quality and user's engagement of text data entry if the functions were properly designed.
- Text prediction would be still helpful for experienced users to enhance their performance.
- An initial investigation about the text prediction functions in support of documentation for reporting patient safety events.

REFERENCES

[1] C.J. McDonald, The barriers to electronic medical record systems and how to overcome them, J. Am. Med. Inform. Assoc. 4 (3) (1997) 213–221.

[2] B. Kaplan, Reducing barriers to physician data entry for computer-based patient records, Top. Health Inf. Manage. 15 (1) (1994) 24–34 (eng). [Epub 1994/07/07]. PubMed PMID: 10135720.

[3] S.H. Walsh, The clinician's perspective on electronic health records and how they can affect patient care, BMJ 328 (7449) (2004) 1184–1187.

[4] L. Hua, S. Wang, Y. Gong, Text prediction on structured data entry in healthcare: a two-group randomized usability study measuring the prediction impact on user performance, Appl. Clin. Inform. 5 (1) (2014) 249–263 PubMed PMID: 24734137. Pubmed Central PMCID: 3974259.

[5] AHRQ, Users' Guide AHRQ Common Formats for Patient Safety Organizations, 2008, Agency for Healthcare Research and Quality, Rockville, MD.

[6] Towards an international classification for patient safety: the conceptual framework 2009. <http://library.tue.nl/csp/dare/LinkToRepository.csp?recordnumber=657653>.

[7] L. Hua, Y. Gong, Design of a user-centered voluntary patient safety reporting system: understanding the time and response variances by retrospective think-aloud protocols, Stud. Health Technol. Inform. 192 (2013) 729–733 (eng). [Epub 2013/08/08]. PubMed PMID: 23920653.

[8] Y. Gong, Data consistency in a voluntary medical incident reporting system, J. Med. Syst. 35 (4) (2009) 609–615 (eng). [Epub 2010/08/13]. PubMed PMID: 20703528.

[9] R.J. Holden, B.T. Karsh, A review of medical error reporting system design considerations and a proposed cross-level systems research framework, Hum. Factors 49 (2) (2007) 257–276 (eng). [Epub 2007/04/24]. PubMed PMID: 17447667.

[10] G. Klein, Sources of Power: How People Make Decisions, revised ed, The MIT Press, 1999.

[11] D. Beukelman, P. Mirenda, Augmentative and alternative communication, 2005.

[12] N. Garay-Vitoria, J. Abascal, Text prediction systems: a survey, Univ. Access Inf. Soc. 4 (3) (2006) 188–203.

[13] S.I. Mackenzie, W.R. Soukoreff, Text entry for mobile computing: models and methods, theory and practice, Human-Computer Interaction 17 (2 & 3) (2002) 147–198.

[14] J. Eng, J.M. Eisner, Informatics in radiology (infoRAD): radiology report entry with automatic phrase completion driven by language modeling, Radiographics 24 (5) (2004) 1493–1501.

[15] S. Hunnicutt, J. Carlberger, Improving word prediction using Markov models and Heuristic methods, Augment. Altern. Commun. 17 (4) (2001) 255–264.

[16] H.H. Koester, S.P. Levine, Learning and performance of able-bodied individuals using scanning systems with and without word prediction, Assist. Technol. 6 (1) (1994) 42–53 (Eng.). [Epub 1993/12/09]. PubMed PMID: 10147209.

[17] J. Kim, D.W. Bates, Results of a survey on medical error reporting systems in Korean hospitals, Int. J. Med. Inform. 75 (2) (2006) 148–155 (eng). [Epub 2005/08/13]. PubMed PMID: 16095963.

[18] Y. Gong, Terminology in a voluntary medical incident reporting system: a human-centered perspective, in: Proceedings of the 1st ACM International Health Informatics Symposium, ACM, Arlington, Virginia, USA, 2010, pp. 2–7.

[19] J. Rosenthal, M. Takach, Guide to State Adverse Event Reporting Systems, 2007, National Academy for State Health Policy, Portland, ME.

[20] L.T. Kohn, J.M. Corrigan, M.S. Donaldson, To Err Is Human: Building a Safer Health System. Report of Committee on Quality of Healthcare in America, 1999, Institute of Medicine, National Academy of Science, Washington, D.C.

[21] L. Hua, Y. Gong, Developing a user-centered voluntary medical incident reporting system, Stud. Health Technol. Inform. 160 (Pt 1) (2010) 203–207.

[22] L. Hua, Y. Gong, Usability evaluation of a voluntary patient safety reporting system: understanding the difference between predicted and observed time values by retrospective think-aloud protocols, in: M. Kurosu (Ed.), Human–Computer Interaction Applications and Services. Lecture Notes in Computer Science, Springer Berlin, Heidelberg, 2013, pp. 94–100. 8005.

[23] L. Hua, Y. Gong, Information gaps in reporting patient falls: the challenges and technical solutions, Stud. Health Technol. Inform. 194 (2013) 113–118 (eng). [Epub 2013/08/15]. PubMed PMID: 23941941.

[24] AHRQ, Common formats. https://pso.ahrq.gov/common, 2011.

[25] JQuery, <http://jquery.com/>.

[26] J.S. Slides, <http://www.slidesjs.com>.

[27] A. Ehlke, S. Challand, T. Schmidt, L. Carneiro, Tag-it. <http://aehlke.github.io/tag-it/>.

[28] D.J. Higginbotham, A.M. Bisantz, M. Sunm, K. Adams, F. Yik, The effect of context priming and task type on augmentative communication performance, Augment. Altern. Commun. 25 (1) (2009) 19–31.

[29] AHRQ, Web M&M morbidity and mortality rounds on the web: cases & commentaries. https://psnet.ahrq.gov/search?Site2Search=PSNet&q=web+m%26m (accessed 04.12.11).

[30] J.P. Matts, J.M. Lachin, Properties of permuted-block randomization in clinical trials, Control. Clin. Trials 9 (4) (1988) 327–344.

[31] J. Allan, J. Englebright, Patient-centered documentation: an effective and efficient use of clinical information systems, J. Nurs. Adm. 30 (2) (2000) 90–95.

[32] L. Poissant, J. Pereira, R. Tamblyn, Y. Kawasumi, The impact of electronic health records on time efficiency of physicians and nurses: a systematic review, J. Am. Med. Inform. Assoc. 12 (5) (2005) 505–516.

[33] M.S. Tuttle, N.E. Olson, K.D. Keck, W.G. Cole, M.S. Erlbaum, D.D. Sherertz, et al., Metaphrase: an aid to the clinical conceptualization and formalization of patient problems in healthcare enterprises, Methods Inf. Med. 37 (4–5) (1998) 373–383.

[34] J. Light, P. Lindsay, L. Siegel, P. Parnes, The effects of message encoding techniques on recall by literate adults using AAC systems, Augment. Altern. Commun. 6 (3) (1990) 184–201.

[35] H.H. Koester, S. Levine, Effect of a word prediction feature on user performance, Augment. Altern. Commun. 12 (3) (1996) 155–168.

[36] C. Goodenough-Trepagnier, M.J. Rosen, Predicitive assessmet for communication aid prescription: motor-determined maximum communication rate, in: L. Bernstein (Ed.), The Vocally Impaired: Clinical Practice and Research, Grune & Stratton, Philadelphia, 1988.

[37] D.A. Norman, Cognitive Artifacts. Designing Interaction, Cambridge University Press, 1991, pp. 17–38.

[38] M. Tamuz, M.I. Harrison, Improving patient safety in hospitals: contributions of high-reliability theory and normal accident theory, Health Serv. Res. 41 (4 p2) (2006) 1654–1676.

[39] M. Tamuz, E. Thomas, K. Franchois, Defining and classifying medical error: lessons for patient safety reporting systems, Qual. Saf. Health Care 13 (1) (2004) 13–20.