MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)
© 2015 IMIA and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License.
doi:10.3233/978-1-61499-564-7-188

Toward User-Centered Patient Safety Event Reporting System: A Trial of Text Prediction in Clinical Data Entry

Lei Hua[ab], Yang Gong [a]

[a] *School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX*
[b] *Informatics Institute, University of Missouri, Columbia, MO*

Abstract

As a primary source for learning from lessons, patient safety event reporting systems have been widely adopted. Nevertheless, underreporting and low quality of reports pervade the system. To address these issues, the study proposed two text prediction functions as data entry aids to system users. With 52 subjects, a two-group randomized experiment was conducted to quantify the impacts in terms of the reporting efficiency, quality, and usability attitudes. Consequentially, on structured data entry, the results were an overall 13.0% time reduction and 3.9% increase of response accuracy with the functions; on unstructured data entry, there was an overall 70.5% increase in the text generation rate, a 34.1% increase in the reporting completeness score, and a 14.5% reduction on the amount of text fields ignored by subjects. Subjects' usability attitudes were slightly improved with the proposed functions according to the questionnaire results.

Keywords:

Ppatient Safety Event Reporting; Usability; Text Prediction; Performance Measurement; Randomized Experiment.

Introduction and Background

In 1999, the prestigious report "To Err is Human" released by the Institute of Medicine estimated 44,000 – 98,000 patient deaths each year due to preventable medical errors [1]. In a recently published study in 2013, the estimation was raised to 210,000 – 440,000, which made the medical errors the third-leading cause of death, behind heart disease and cancer in the United States [2]. To learn from the errors in order to improve patient safety and healthcare quality, electronic event reporting (e-Reporting) systems have been proposed and promoted nationwide. However, low-quality reports pervade the systems and undermine the user engagement of e-Reporting.

As one of the most prominent factors associated with low-quality reports, data entry has received much attention in clinical information systems, such as electronic health record and computerized provider order entry system [3,4]. Quality data in patient safety holds promise in prompting system acceptance, which in turn may form a virtuous loop in leveraging system performance, user engagement, and patient safety. In this two-group randomized study, we explored text prediction techniques for facilitating data entry in terms of quality and efficiency.

Method and Material

Subjects

The study enrolled 52 nurses from 21 clinical departments in Tianjin First Central Hospital (TFCH) in China. All nurses were female between the ages of 30 and 52 years. On average, they had 20 years of nursing experience and four years of experience with e-Reporting. All participants 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).

Interfaces

Two experimental interfaces (control and treatment) were developed as a means of configuration control during data collection. The two interfaces were identical in terms of task and representation, requiring the completion of 13 structured multiple-choice questions (MCQs) defined by AHRQ Common Formats [5] and one free text, multiple-line comment field for reporting additional details of patient fall. A single exception between the two interfaces was the provision of text prediction functions in the treatment interface, which included a cueing list (CL) and autosuggestion (AS). As illustrated in Figure 1, part B, the CL was attached to the MCQs (4 out of 13) that had a single text field. For unstructured data entry, the comment field was equipped with both the CL (part C) and AS (part G) in the treatment interface.

All items contained in the CL and AS were manually prepared [6]. Two domain experts reviewed all testing cases, and transcribed and categorized the key elements with the colloquial language in Chinese with characters Hanzi and phonetic typing method Pinyin. Pinyin text input methods (IME) are based on the standard qwert keyboard and are the most popular in Chinese. Pinyin IME does not require additional training for those who recognize English letters.

To prepare the CL and AS, expert reviewed words, phrases, and sentences were populated into a data table serving as reminder items. The number of listed items in any CL and AS was limited to 10, based on a trade-off between inspecting efforts and predicting sensitivity [7]. In the CL, at least one of the items was considered to be more descriptive, accurate, and pertinent to the case being reported than the rest. In the AS, the display of suggested entry options relied on a Soundex algorithm (phonetic matching function) of MySQL and the reporter's initial entries. As illustrated in Figure 1, Part G, the top 10 hits were rendered in the AS list. On the treatment interface, the subjects were allowed to select mixed entries from the AS list and user-generated text input. On the control interface, subjects were not provided any AS or CL; thus, each participant had to generate his or her own text input through a regular qwerty keyboard.

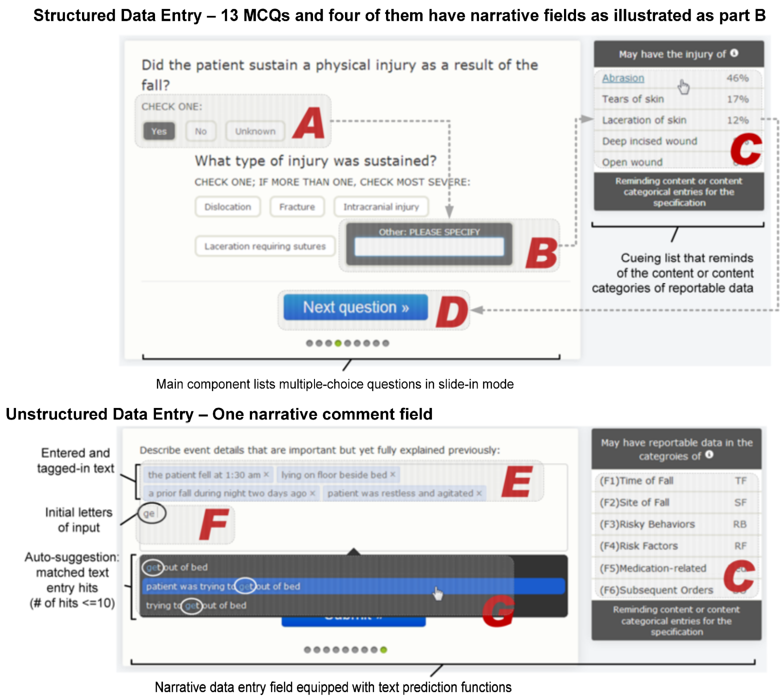

Figure 1 – The layout of interface elements is shown in English for both structured and unstructured data entries. Text prediction of *Cueing Lists* (CL, box C) and *Auto Suggestion* (AS, box G) are triggered with cursors in the corresponding fields. Box A shows the first answer triggering generated by the child's answers to questions B, C, D, etc. Box C is activated only when a single-text field (B) is open. CL reminds reporters of the content or key characteristics of reportable data associated with the event. The number of items listed in CL is <6. When pressing the button (D), a report would show an unstructured page containing boxes E, F and G. When initial letters (shown in box F) of the description are typed in, a matched list provided by AS shows up. The number of items listed in AS is <10. Matched letters and the focused line are highlighted in blue (G). Reporter may select one at a time and modify as necessary. Pressing the "Enter" button (behind G) will tag the current entry in a blue text chunk (as shown in E).

Testing Cases

In the study, each subject reported five patient-fall cases in a randomized sequence. The cases were selected from two sources —a case depository with 346 fall reports from a previous study [8] and a public database of Morbidity and Mortality (M&M) [9]. The five selected cases were translated into Chinese and rephrased by two domain experts for the purpose of quality and readability of text. A similar complexity of the five cases was agreed on by the domain experts. The following English narrative is an excerpt from one of the cases.

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

Experimental Design

With a permuted-block algorithm, the 52 subjects were randomly assigned to two groups. 25 subjects were allocated into the group using the control interface without text prediction; 27 were assigned to the group with the treatment interface. The presenting sequence of five cases for each subject was randomly assigned. The subjects were trained through verbal instructions, and then asked to practice using both interfaces to report a sample case until each felt comfortable with the content and interface interactions. Since the training was conducted prior to grouping and the grouping procedure was blinded to both the subjects and the trainer, this design prevented confounding influences and training bias which might be delivered by the trainer.

Table 1 – The data sources and the units of analysis of key measures in the study

Measures	Data sources	Unit of analysis
Subject		
Age	Hospital nursing office	Years
Proficiency of reporting falls	Graded prior to the experiment	5 points Likert scale (1-low to 5-high)
Reporting efficiency		
Structured data entry	Accumulated time on MCQs	Seconds
Descriptive comments	Completion time on the comment field	Seconds
Text generation rate	Nominator: Letters in length of the comments; Denominator: completion time	Letters/Second*
Quality of reports		
Structured entry accuracy	Nominator: Accumulation of scores on MCQs; Denominator: Maximum of the accumulation	Percentage
Narrative completeness	The number of credited text chunks	Counts
Survey usability attitudes		
User attitudes in four dimensions	Posttest questionnaire	5 points Likert scale (1-low to 5-high)

* To count the length in letters, one UTF-8 encoded Chinese character equals three English letters in length.

In the hospital, a typical scene of event reporting is that a reporter initiates a report upon a witness's word-of-mouth information. This study simulated the natural scene by using the five cases with each having appeared on the first page of the interface. Subjects read the descriptions and answered the questions upon recall.

Pauses and pop-up questions were discouraged, except when a subject was in transition between reports. Keystroke level operations (mouse clicks and keystrokes) for each subject trial were time stamped and logged into a MySQL database. All reporting sessions were recorded using Camtasia Studio® 7 for data reconciliation. Each session was concluded with a questionnaire [10] via SurveyMonkey to reflect user attitudes in e-Reporting. The questionnaire was developed following Nielsen's Attitudes of Usability with a five-point Likert scale, wherein "1" indicated a maximal level of disagreement to the statement and "5" indicated a maximal level of agreement.

Processing of Data

The study collected ordinal and nominal data out of three data sources, including MCQs, narrative comment fields, and questionnaires. Ordinal data are the selected response items in the MCQs and questionnaires, and nominal data are the text entries in the single-line fields of MCQs and the comment field at the end of each report. We examined the ordinal and nominal data in terms of efficiency, effectiveness, and usability attitudes. Other experimental features associated with the CL and AS functions were also investigated. Table 1 lists the data sources and units of analysis of the measures.

Results

All the subjects successfully completed the experimental sessions with each subject having reported five cases, and the group consequentially generated 260 reports and 52 questionnaires. On average, each subject's session took 71 minutes, which consisted of 17 minutes in training and practice, 45 minutes in case-documenting, and 9 minutes in the questionnaire. There were 25 subjects in the control and 27 subjects in the treatment groups, accounting for 125 and 135 reports respectively. The mean ages of the subjects' groupings were 43.6±5.8 (control) versus 41.1±6.7 (treatment). The differences of ages and proficiency scores between the groups were insignificant ($p > 0.05$). The 260 reports contained 2,849 answers to MCQs and 238 unstructured narrative comments. As shown in Table 2, the subjects had eight significant variations between the groups indicated by an arrow ↑ (increase) or ↓ (decrease). Almost all variables, except for mouse clicks, in the treatment group showed an improved performance against the control group.

The analysis showed that the efficiency and quality of reporting can be enhanced by CL and AS for both structured and unstructured data entries. As illustrated in Figure 2, for the structured questions, the completion time was reduced by 13.0%, while the accuracy had a 3.8% increase overall. For unstructured narrative entries, the text generation rate was increased by 70.5%, which significantly increased the generation of free text data volume without an increase of time consumption between the groups. The reports generated in the treatment group were much more complete (34.2%) and richer in text (24.4%) than those in the control group.

As indicated in Table 2, the introduction of CL greatly improved the missing comments in the narrative field where the non-adherence rate dropped from 14.5% to 1.5%. CL functioned as a dynamic display when the page containing narrative comment fields slid in, as shown in Figure 1 (Part E, F & G). CL successfully drew subjects' conscious attention to the interface content. This dynamic CL signaled a compelling message to the subjects about the importance of filling the comment field.

Analytical results out of the questionnaires show potential non-adherence that may occur due to the following reasons: (1) a slip of unconscious skip of the comment field, (2) a lack of ideas of what event characteristics should be described further, and (3) that memory fades when working on the comment field.

Benefits from AS and CL also showed in keystrokes. In the reports, keystrokes were reduced by an average of 48.4%. In contrast, there was a 29.3% increase of mouse clicks. AS, in addition, showed a significant effect in generating 133 narrative comments, wherein AS was selected in 120 (90.2%) comments, totaling 460 times. Meanwhile, AS was inserted into 66.9% text chunks identified in the 133 entries. On average, AS was presented in 3.8±1.9 text chunks in each report. A regression analysis showed that influential rate, defined as AS selection over total text chunks in each report, increased along the experimental progress ($p < 0.05$). Therefore, an observed learning effect amplified the efficiency along the experimental progress, as shown in Figure 4.

Fifty-two questionnaires were completed, which contained 1,300 rated answers. According to the analysis shown in Figure 5, the subjects showed overall good attitudes for the usability of both interfaces. The scores on all four dimensions slightly increased in the treatment group compared to the control group, but were not statistically significant. The text prediction functions implemented in the treatment group did not impact user attitudes in terms of learnability, efficiency, memory, and satisfaction.

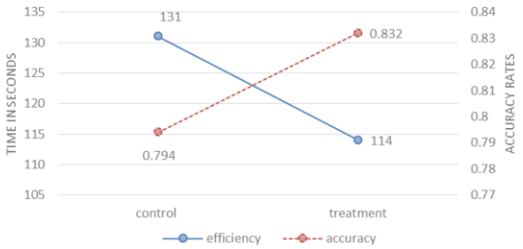

Figure 2 – Reporting efficiency and accuracy of structured data entries increased in the treatment group.

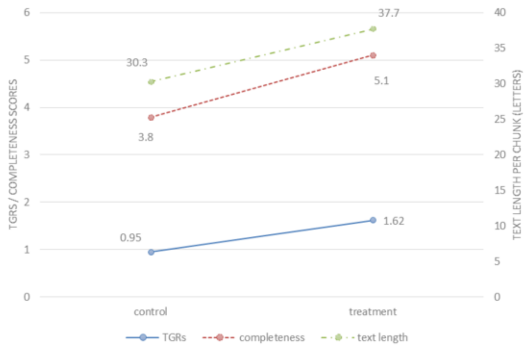

Figure 3 – Text generation rate and data completeness of unstructured data entries increased in the treatment group.

Table 2 – The subjects' performance in the reporting task.

Measure	Control Group	Treatment Group	Variation	*p*-value	Illustration
Subject					
Sample size (subjects/reports)	25/125	27/135			
Age	43.6±5.8	41.1±6.7		0.189	
Proficiency in reporting fall events (Likert scale 1-5)	3.8±0.5	3.9±0.5		0.413	
Reporting efficiency					
Time on case reading (seconds)	84.9±44.8	78.6±43.2		0.307	
Structured data entry (seconds)	131.0±50.0	114.0±41.7	↓ 13.0%	0.004	Figure 2
Descriptive comments (letters)	139.6±99.6	142.9±82.2		0.782	
Text generation rate (letters per second)*	0.95±0.35	1.62±0.99	↑ 70.5%	0.000	Figure 3, 4
Physical operations in a report					
Mouse click	13.3±2.1	17.2±3.7	↑ 29.3%	0.000	
Keystroke	173.2±117.0	89.4±89.4	↓ 48.4%	0.000	
Quality of reports					
Structured entry accuracy	79.4±10.1%	83.2±11.0%	↑ 3.8%	0.000	Figure 2
Missing comments (non-adherence rate)	20/125(16.0%)	2/135(1.5%)	↓ 14.5%	0.000	
Narrative completeness	3.8±2.3	5.1±2.4	↑ 34.2%	0.000	Figure 3
Length of a chunk - richness (letters)*	30.3±13.1	37.7±18.6	↑ 24.4%	0.000	Figure 3
Survey usability attitude (Likert scale 1-5)					
Learnability	4.0±0.3	4.1±0.5		0.545	Figure 5
Efficiency	4.1±0.4	4.3±0.6		0.386	Figure 5
Memory	3.5±0.5	3.8±0.5		0.099	Figure 5
Satisfaction	3.8±0.4	3.9±0.5		0.458	Figure 5

* To count the length in letters, one UTF-8 encoded Chinese character equals three English letters in length.

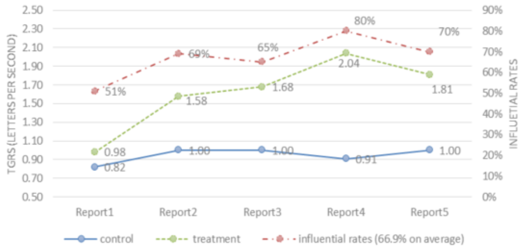

Figure 4. Text generation rates exhibit an overall increase (learning effect) over five reports in the treatment group. In contrast, such an effect does not present in the control group.

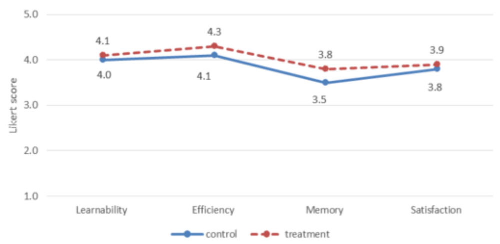

Figure 5. User attitudes of the treatment group are slightly improved than the control group.

Discussion

It is generally agreed that keystrokes could be reduced when clickable on-screen options are readily available on a user-centered interface. However, the keystroke savings alone remain inadequate in explaining the increased efficiency. Earlier research exhibits mixed, inconsistent results regarding improved efficiency of keystrokes in terms of cognitive load, eye gaze movement, and mouse clicking [11,12]. The balance between mouse clicks and keystrokes is an interesting topic in the process of developing an efficient and effective e-Reporting system. As touchscreen handheld devices are becoming pervasive in healthcare, significant keystroke savings on data entry may greatly amplify the benefits of user-centered design. This is because most touchscreen users prefer on-screen menu selection to on-screen keyboard typing.

The two-group randomized experiment proved the effectiveness of AS and CL text prediction in generating quick and high quality data entry in e-Reporting. User attitudes reflected in the questionnaires did not change significantly, but were potentially improved with the introduction of text predictions in e-Reporting. The observed effect may be generalizeable to other data entry interfaces in healthcare. Clinicians working under time pressure are expected to document in a timely manner [13,14]. Meanwhile, the quality of data has to achieve a high standard for decision-making and the creation of actionable knowledge. The solutions such as AS and CL experimented with in this study pave a way for addressing these concerns. The clinical document architecture (CDA) is a markup standard developed by Health Level 7 Internation (HL7) to define the structure of clinical documents, such as discharge summaries, progress notes, etc. Nonetheless, e-Reporting is becoming a routine documentation task where the structure is not yet well defined.

Recently, The Agency for Healthcare Research & Quality (AHRQ) created the Common Formats (CF), which are common definitions and reporting formats, to help providers uniformly report patient safety events and to improve healthcare quality. CFs include paper forms to guide the development of data collection intruments. When transforming the paper forms into electronic interfaces, designers should carefully learn from those lessons based upon the evolvement of electronic medical records.

One of the major purposes of e-Reporting is the collection of information and knowledge about patient safety events. Ironically, the e-Reporting process is not immune to human errors due to time pressure, multi-tasking, or competing priorities.

Recommendations for Interface Development in Healthcare

Guided by user-center design principles, we employed state-of-the-art and open-source techniques to rapidly prototype the dual-interfaces, one exhibited stereotypical functions in prevailing e-Reporting, the other was renovated and embeded with text predictions. The dual interfaces not only met the experimental requirements of randomization and control, but successfully demonstrated the effectiveness and effiency of the innovative design in e-Reporting. Interface design specifically for the purpose of e-Reporting requires further investigation, which would contribute to the purpose of collecting quality events, generating actionable knowledge, and sharing experience in reducing the recurrence of similar incidents.

Our design and development process is instrumental to many clinical information systems. The incremental effect on saving providers' time and improving report quality may be magnified, while evident-based estimates show 440,000 preventable events each year in the U.S. [2].

Limitations

CL offered in the interface presents a high quality contribution by domain experts. In reality, the prediction accuracy based upon an algorithm of event similarity and frequency of appearance may not match the manual effort of experts. In addition, the number of predicted items may differ in other clinical systems depending upon the richness of data entries. We did not provide an extended item list, which may be a burden for reporters' inspection. In the future, the question of whether text prediction with lower accuracy or a longer list would undermine the proven effect is worth further investigation.

It is a noteworthy dilemma that AS, designed for increasing text generation rate and accuracy, may lead to side effects, such as a subject's overdependence on the suggested items, and thus limit constructive thinking [15-17]. This is because, when provided with a predefined list of options, busy clinicians tend to recognize or ignore the options rather than specifying answers. The future development of the AS should carefully keep a balance between the strength and weakness of such protocols, and strategically address the dilemma.

Conclusion

User-centered design aimed at renovating e-Reporting demonstrated the effectiveness of text prediction in clinical data entry. This work may lay the groundwork for further research associated with text prediction in clinical information systems.

Acknowledgments

The authors express heartfelt thanks to Shen Wang, RN, Xiao Liu, RN and Ying Sun, RN at the Office of Nursing at Tianjin First Central Hospital in China who helped review and edit the testing cases, derived proper text prediction responses, scheduled the testing sessions, and graded the subjects' commentary data.

References

1. Kohn LT, Corrigan JM, Donaldson MS. To err is human: building a safer health system. Report of Committee on Quality of Healthcare in America. Institute of Medicine, National Academy of Science, 1999.
2. James JT. A new, evidence-based estimate of patient harms associated with hospital care. Journal of Patient Safety 2013: 9 (3): 122-8.
3. McDonald CJ. The barriers to electronic medical record systems and how to overcome them. Journal of the American Medical Informatics Association 1997: 4 (3): 213-21.
4. Walsh SH. The clinician's perspective on electronic health records and how they can affect patient care. BMJ 2004: 328 (7449): 1184-7.
5. AHRQ. Common Formats 2011 [Internet]. [Place unknown]: AHRQ; 2011 [cited _____give date YYYY MMM DD_____]. Available from: http://www.psoppc.org/web/patientsafety/version-1.2_documents
6. Higginbotham DJ, Bisantz AM, Sunm M, Adams K, Yik F. The effect of context priming and task type on augmentative communication performance. Augmentative and Alternative Communication 2009: 25 (1): 19-31.
7. Hunnicutt S, Carlberger J. Improving word prediction using markov models and heuristic methods. Augmentative and Alternative Communication 2001: 17 (4): 255-64.
8. Gong Y. Terminology in a voluntary medical incident reporting system: a human-centered perspective. Proceedings of the 1st ACM International Health Informatics Symposium (1882996) ACM, 2010: 2-7.
9. AHRQ. Web M&M mobidity and mortality rounds on the web:cases and commentaries [Internet]. [Place unknown]: AHRQ; 2011 [cited 2011 Dec 4]. Available from: http://www.webmm.ahrq.gov/home.aspx
10. Hua L. Posttest questionnaire for user reflections on the system 2013 [Internet]. [Place unknown]: Name Site; date [cited___YYYY_MMM_DD___] Available from: http://www.surveymonkey.com/s/9SLCZ9M
11. Light J, Lindsay P, Siegel L, Parnes P. The effects of message encoding techniques on recall by literate adults using AAC systems. Augmentative and Alternative Communication 1990: 6(3): 184-201.
12. Koester HH, Levine S. Effect of a word prediction feature on user performance. Augmentative and Alternative Communication 1996: 12(3): 155-68.
13. Allan J, Englebright J. Patient-centered documentation: an effective and efficient use of clinical information systems. Journal of Nursing Administration 2000: 30(2): 90-5.
14. Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. Journal of the American Medical Informatics Association 2005: 12(5): 505-16.
15. Gong Y. Data consistency in a voluntary medical incident reporting system. Journal of Medical Systems 2009: 35(4): 609-15.
16. Tamuz M, Harrison MI. Improving patient safety in hospitals: contributions of high reliability theory and normal accident theory. Health Services Research 2006: 41(4p2): 1654-76.
17. Tamuz M, Thomas E, Franchois K. Defining and classifying medical error: lessons for patient safety reporting systems. Quality and Safety in Health Care 2004: 13(1): 13-20.

Address for Correspondence

Dr. Yang Gong. 7000 Fannin St, Houston, Texas, USA 77030.
yang.gong@uth.tmc.edu