MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220095

Annotation Guidelines for Medication Errors in Incident Reports: Validation Through a Mixed Methods Approach

Zoie S.Y. Wong[a], Yuchen Qiao[b], Ryohei Sasano[c], Hongkuan Zhang[c], Kenichiro Taneda[d], Shin Ushiro[e]

[a] Graduate School of Public Health, St. Luke's International University, Tokyo, Japan,
[b] The University of Tokyo, Tokyo, Japan
[c] Nagoya University, Nagoya, Japan
[d] National Institute of Public Health, Saitama, Japan
[e] Kyushu University, Fukuoka, Japan

Abstract

At present no adequate annotation guidelines exists for incident report learning. This study aims at utilizing multiple quantitative and qualitative evidence to validate annotation guidelines for incident reporting of medication errors. Through multiple approaches via annotator training, annotation performance evaluation, exit surveys, and user and expert interviews, a mixed methods explanatory sequential design was utilized to collect 2-stage evidence for validation. We recruited two patient safety experts to participate in piloting, three annotators to receive annotation training and provide user feedback, and two incident report system designers to offer expert comments. Regarding the annotation performance evaluation, the overall accuracy reached 97% and 90% for named entity identification and attribute identification respectively. Participants provided invaluable comments and opinions towards improving the annotation methods. The mixed methods approach created a significant evidential basis for the use of annotation guidelines for incident report of medication errors. Further expansion of the guidelines and external validity present options for future research.

Keywords:

Hospital Incident Reporting, Data Annotation, Medication Errors

Introduction

The WHO Patient Safety Incident Reporting and Learning Systems Technical report and guidelines [27] published in 2020, maintains the importance of seeking effective approaches towards utilizing incident reporting systems for effective patient safety improvement. As a set of widely recognized patient safety guidelines, Minimal Information Model for Patient Safety (MIM PS) [26] provides a minimal information model format and field guidance for incident reporting. One of the key learning challenges remaining is bridging the gap between meaningful patient safety information and narrative text data [34] through information models, taxonomies and classifications suitable for incident reports [27].

Among all the incident types, unsafe medication practices and medication errors can result in severe harm and even death to patients and these are potentially avoidable. In the context of the Global Patient Safety Challenge, WHO announced global goals to reduce medication errors through system strengthening [2]. The challenge of learning incident reports for medication errors is that much of the crucial information is registered in a free text format and is hence not directly analyzable through computerized systems and supervised models.

Annotating clinical documents [18] has been a commonly used means to create gold standard data for automatic information extraction [20]. Recent examples include the 2010 i2b2/VA challenge on concepts, assertions, and relations in the clinical text [22], the n2c2 shared task in 2018 [23], and the i2b2 challenge in Japan [19]. There is a significant body of literature providing instructions for semantic role assignment, argument identification and classification that are crucial for information extraction in general [9; 10; 12; 13] and clinical contexts [8; 21].

Furthermore, existing annotation guidelines for the Adverse Drug Event (ADE) and Medication Extraction [23] exist. For instance, Buchan's guidelines [7] aim at establishing a frame of reference in which different ADE-related information can be registered systematically. The guidelines include two tasks: identifying drug names, dosages, durations and other entities and creating two different relations, including the drug's relation to specific symptoms and diseases; and the drug's relation with ADE. However, the current form of the guide is not suitable for incident report learning purposes.

A fundamental reason why an incident or near miss occurs is due to discrepancies between what is supposed to be delivered from an upstream operation and what is actually delivered to a downstream operation. Medication errors could occur due to such discrepancies across different phases of medication, or associated variables regarding the medication, such as type of drug, dosage, strength, form, and patient identification etc. Often, when reporting medication errors, incident reporters detail aspects of medication and its intention and factuality indications [32]. These unique properties should be captured in incident annotation as they would allow machine learning and AI methods to predict the nature of medication errors and hence meaningful incident information can be captured systematically.

Yet, none of these current schemas are developed in an appropriate way for incident report learning. This motivated us to investigate information extraction approaches for incident report learning through developing gold standard data and investigating named entity recognition (NER) models. As an initial step, we (previously) explained annotation guidelines targeting incident report of medication errors, [32] and in this study, we attempt to evaluate its validity for application. The goals of annotation for medication incidents include:

Identifying different concepts from medication incident reports (named entity identification)

Assigning explanations to the identified concepts based on the narrative text content (attribute identification)

Named entity (NE) refers to a real-world object that can be denoted with a proper name. It could be an abstract concept or a physical referent (e.g. persons, locations, organizations, products, etc.). Assigning an attribute aims to register the properties and characteristics of each NE annotation in order to provide sufficient associated information and, in an overall entry-level sense, to identify what type of incident occurred and pinpoint what kind of errors took place.

Methods

In consideration of a necessary interplay of clinical, information extraction and human factors for annotation, a mixed methods approach allows us to utilize these diverse research lenses to integrate associated data elements [16]. Explanatory sequential mixed method design within a single investigation were employed for validating the use of the guidelines. In this validation exercise, it involves structured annotator training using the developed annotation guidelines, annotator performance evaluation compared with gold standard, and interviews. The research process begins with quantitative data collection and analysis through annotation training and evaluation and completing exit surveys. Then, qualitative information regarding the annotation scheme is synthesized through interviews with annotators and incident report designers/collectors/experts.

Overview of the medication error annotation scheme

The goal for the annotation guidelines is to establish a workable and structured method to extract and retrieve incident registry/information into a machine-readable dictionary of medication error concept terms and indications, which can be used to automatically classify incidents drawn from unstructured incident reports.

When designing the annotation method, we conducted an extensive literature review using state-of-the-art incident reporting guidelines and existing work on medication errors, classification schemes and annotation methods. For instance, the WHO International Classification of Patient Safety [1], WHO Minimal Information Model for Patient Safety (MIMPS) [5], National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) [4], Agency for Health Research and Quality (AHRQ) Common Formats Version 2.0 [31] and European Medicines Agency Good Practice Guide [3], as well as other relevant studies [4; 6; 24; 25; 31; 33], were drawn upon, referenced, and synthesized. The design and technical explanation of the annotation concept is documented in [15; 32]. The annotation task aims to extract a set of NEs and identify the associated properties related to medication errors from incident reports. The annotation consists of two subtasks: (1) Identifying NEs, namely drug, form (form and dose), strength (amount, concentration and rate), timing, duration, frequency, dosage, route and mistaken patient [15] and (2) Applying generic attributes indication to specify types and properties of each annotated NE, including index, status (intention/factuality) and the type of error [32].

Here, we briefly explain the rationale of extracting the explicit properties of medication errors, organizing abstract incident information into meaningful entities associated with incident reporting, and illustrating the methods of annotating targeted entities in incident reports in practice. The key is to capture essential and minimal necessary information of medication errors reporting. Each of the tagged NEs will be associated with certain characteristics. When designing the guidelines, we were concerned with the balance between annotation complexity and implications to patient safety learning.

Study Participants

In the pilot phase, we recruited experienced medical doctors (+15 years' experience, patient safety experts) to examine the feasibility of using the proposed annotation methods. In the evaluation phase, annotators with at least 2 years medical doctor working experience were recruited. They needed to be skilled in incident report writing and/or handling incident reports in their affiliated hospitals. Also, experienced incident report system designers and data owners were invited to observe the evaluation and provide feedback to us.

Quantitative Approaches

There are two phases in this annotation method. Detailed annotation guidelines and training materials are developed for this purpose (and can be made available for the interested researchers to use on request). Sufficient examples are provided in order to demonstrate how to distinguish NEs and their attributes and to illustrate the detailed method of annotation under various practical scenarios.

Brat rapid annotation tool

This annotation guide can be implemented using state-of-the-art text annotation software platforms, such as brat, oxygen xml, prodigy and doccano. For our study, throughout the annotation evaluation, we adopted brat [17] rapid annotation tool (http://brat.nlplab.org) and provide annotation training to participants with illustrative examples using this platform.

Evaluation Data

Prior to the evaluation, our researchers randomly drew on 30 medication incidents from datasets of JQ (46,503) and St. Luke's data (6,628) and created gold standard annotations using the developed guidelines/established methods for annotating medication incidents. From these incident reports, all the taught NEs and attributes are covered.

Training and Evaluation

In both phases, participants receive approximately 3-hours of training to learn the guidelines, and practice how to use brat to annotate with supervision/guidance using case examples, and to conduct the assessments. In part 1 of the training and evaluation, participants learn how to identify NEs for medication errors and in part 2, the participants learn how to identify attributes associated with each tagged NE. Two phases of annotation performance evaluation are conducted. Evaluation metrics in terms of precision, recall, accuracy, F-score are captured.

Surveys

At the end of each phase and at the end of the entire training and evaluation session, annotators carry out annotation experience exit surveys, covering the content of conceptual design, applicability, usefulness, intuition, user-friendliness [10] using a 5-point Likert scale.

Qualitative Approaches

Open-ended questions and expert interviews

Furthermore, we use open-ended questions to collect comments and feedback after each phase and at the end of the evaluation. All the participants and expert members receive a short inter-

view after the training and evaluation phase, allowing us to receive critical comments regarding to the application of annotation guidelines.

Results

Two half-day training sessions with a subsequent testing session were conducted on 27th and 29th Aug 2019 at St. Luke's International University. Three participants joined both phases of training, evaluation and completed all the exit surveys. All of them possessed a medical degree background, were actively practicing medicine, and utilising/reporting incidents. On average the number of years in the practice was 7.8 and typically they handled more than 3 incident reports per year. Patient safety and incident reports experts from the Japan Council for Quality Health Care (JQ) and St. Luke's International Hospital participated and witnessed the training process. Prior to the above exercises, during the pilot phase, two experienced medical officers with significant patient safety training experience refined the training materials and delivery.

Annotation Quality

Annotation evaluation consists of two parts: Identifying NEs and applying attributes indication to specify types and properties of each annotated NE. Table 1 illustrates the annotators' evaluation outcomes at identifying each NE and attributes. The total number of NEs tested was 405 and the number of attributes tested was 200. For NE identification, the overall accuracy reached 97% and overall recall was 83%. In terms of attribute identification, overall accuracy reached 90%.

Table 1– Accuracy, precision, recall, F-measure for annotators compared to the gold standard

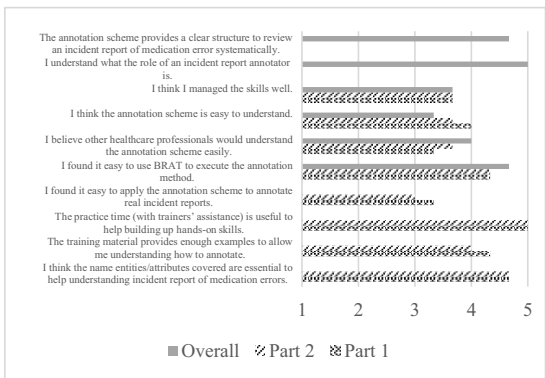
	Accuracy	F1	Recall	Precision
Part 1: Named Entities	0.97	0.81	0.83	0.79
Dosage		0.50	0.93	0.34
Drug		0.89	0.87	0.92
Duration		0.68	0.67	0.70
Form (form)		0.82	0.77	0.88
Form (mode)		0.75	1.00	0.60
Frequency		0.69	0.63	0.76
Route		0.53	0.68	0.43
Strength (amount)		0.88	0.81	0.96
Strength (concentration)		0.85	0.95	0.77
Strength (rate)		0.79	0.87	0.71
Timing		0.79	0.79	0.79
Wrong patient		0.86	0.89	0.83
Part 2: Attributes	0.90			
Index	0.74			
Status	0.89			
Error Occurred	0.95			

Survey Results

Figure 1 presents the survey results collected after each training session and at the end of the exercise. From the final exit survey, all the annotators agreed that the annotation guidelines provided a structured way to systematically review medication error reports (Score 4.6 out of 5). In general, they understood the role of being an incident report annotator. For those questions related to whether the entire annotation scheme is easy to understand, whether they managed to acquire the skill successfully and whether other healthcare professionals would understand the scheme, the average scores lay between 3.33 to 4. The annotators also found it easy to use the brat platform to execute the annotation method (average above 4 out of 5). These questions were also asked at the end of part 1 and part 2 and the results were largely similar.

For both part 1 and part 2 training sessions, the participants found that practice time (with trainers' assistance) was useful to helping build up hands-on skills to achieve the annotation task (average 5 out of 5). The participants agreed that NEs/attributes covered are essential to help understanding incident reports of medication errors and the training material provides enough examples to allow them to comprehend the annotation method. Participants thought that part 1 (average 3.33) was easier than part 2 (average 3) for application to annotate real incident reports. The above results were also confirmed in the participants' interviews.

Figure 1– Survey results (after part 1 and part 2 trainings and Final Exit Surveys)



Discussion

A purposeful mixing of quantitative and qualitative methods was utilized in this study for data collection, data analysis and evidence synthesis. We evaluated the validity of the annotation guidelines for medication errors in incident reporting through providing training to new clinical annotators whilst examining their performance and collecting their feedback.

In general, the annotators accomplished the annotation tasks with accuracy of an 90% or above and the annotators gained a firm appreciation of the role of being an annotator. Participants indicated that the annotation scheme is accessible and manageable and thus comprehensible for other healthcare professionals. Despite the fact that they had never used any annotation platforms before, they felt comfortable to execute the annotation tasks using brat. Feedback from the participants confirmed that practice time with examples (with close supervision under a small training group setting) is important to help to establish annotation skills in practice. Participants felt more confident in identifying NEs than assigning attributes to annotations. This is because NEs are associated with the nature of medication use, which is relatively intuitive, especially since medical doctors have to deal with medication on a daily basis. However, in terms of assigning attributes, it requires good understanding of how the occurrence of incident progresses, through careful reading the incident report.

One participant raised concerns about choosing the appropriate wordings for annotation (i.e. the span). Another indicated that some incident report descriptions were vague and unclear (associated with the selectivity and incompleteness problems [27]), and that led to difficulties in distinguishing the nature of intention and factuality. One suggested a clearly indicated and intuitive pull-down menu view through the annotation platform would be helpful to choose appropriate attributes.

This study validated the use of the annotation method for incident report of medication errors and contributed to development of taxonomy and classification system that caters for systematic registering and capturing meaningful information of incident reports [28]. All participants and experts agreed that uncovering NEs and attributes are useful to systematically capture essential patient safety elements. They also agreed that the guidelines offer a systematic and structured means to review medication error reports.

With the recent advances in AI, such as neural network modelings, natural language processing, and named entity recognition (NER) [11], our annotation guidelines provide a framework for information retrieval directly from medication incident reports and enable forthcoming supervised and semi-supervised learning opportunities to improve the classification of incident reports [14; 29; 30] and register incident report data.

By the time of publication, our team has employed a company specializing linguistic data construction (IR-Advanced Linguistic Technologies Inc.) to create Japanese gold standard annotated incident reports using the public data from the Japan Council for Quality Health Care (JQ) and published the annotated medication incident report data on open domain (https://github.com/HongkuanZhang/IFMIR-Corpus). Through this open source repository effort, we envision that well-validated AI-enabled NER and classification models could revolutionise the way how we collect, utilise, and retrieve information from incident reports in the future. One of the potential benefits is to provide an opportunity to compare incident reports across different institutions/collection system/countries and ultimately capture free-text medication errors in a meaningful and structured manner.

The annotation guidelines (in English) are currently developed for annotating Japanese reports, however, all the subject domain/patient safety-specific designs such as annotation rationale, NE identification and attribute identification would not differ in an English report. Only certain language-specific semantic role assignment and argument identification techniques, (such as how to determine noun phrases) are different and therefore, the entire guidelines can be easily transferred to English incident report usage. In the future, external validation using medication errors incident reports from the world (such as AIRS from Hong Kong [30]) could be possible. Due to the limitation of training capacity and the need for high intensity supervision, we adopted a focused learning group training approach in this instance.

Conclusions

In this study, we validated the use of a newly developed annotation guidelines for incident report of medication errors through a mixed methods approach. Both quantitative and qualitative evidence via annotation performance assessment, questionnaire surveys, and interviews were collected, analyzed and synthesized. Our annotation guidelines provide a practical framework for retrieving patient safety information directly from medication incident reports.

Acknowledgements

This work was supported by the Japan Society for the Promotion of Science KAKENHI grant number

18H03336.

References

[1] Conceptual Framework for the International Classification for Patient Safety, in, WHO, 2009.

[2] Medication Without Harm, in, WHO, 2017.

[3] C.L. Althaus, N. Low, E.O. Musa, F. Shuaib, and S. Gsteiger, Ebola virus disease outbreak in Nigeria: Transmission dynamics and rapid control, *Epidemics* **11** (2015), 80-84.

[4] Anonymous, National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) Taxonomy of Medication Errors, in, National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) US, 1998.

[5] Anonymous, Minimal information model for patient safety incident reporting and learning systems: user guide, in, World Health Organization, Geneva, 2016.

[6] J.K. Aronson, Medication errors: definitions and classification, *Br J Clin Pharmacol* **67** (2009), 599-604.

[7] K. Buchan, Annotation Guidelines for the Adverse Drug Event (ADE) and Medication Extraction Challenge in, 2018.

[8] A. Carson-Stevens, P. Hibbert, H. Williams, H.P. Evans, A. Cooper, P. Rees, A. Deakin, E. Shiels, R. Gibson, A. Butlin, B. Carter, D. Luff, G. Parry, M. Makeham, P. McEnhill, H.O. Ward, R. Samuriwo, A. Avery, A. Chuter, L. Donaldson, S. Mayor, S. Panesar, A. Sheikh, F. Wood, and A. Edwards, Health services and delivery research, in: *Characterising the nature of primary care patient safety incident reports in the England and Wales National Reporting and Learning System: a mixed-methods agenda-setting study for general practice*, Southampton (UK), 2016.

[9] B. Denise, Evaluating classification schema and classification decisions, *Bull Am Soc Inf Sci* **39** (2013), 13-21.

[10] E. Gius, N. Reiter, and M. Willand, Evaluation of Annotation Guidelines, in, 2018.

[11] Z. Liu, M. Yang, X. Wang, Q. Chen, B. Tang, Z. Wang, and H. Xu, Entity recognition from clinical texts via recurrent neural network, *BMC Med Inform Decis Mak* **17** (2017), 67.

[12] D. Maynard and N. Aswani, Annotation and Evaluation, in, 2021.

[13] D. Nadeau and S. Sekine, A survey of named entity recognition and classification, *Lingvisticae Investigationes* **30** (2007), 3-26.

[14] M.S. Ong, F. Magrabi, and E. Coiera, Automated categorisation of clinical incident reports using statistical text classification, *Qual Saf Health Care* **19** (2010), e55.

[15] Y. Shiima and Z.S.-Y. Wong, Classification Scheme for Incident Reports of Medication Errors, *Stud Health Technol Inform* **265** (2019), 113-118.

[16] A. Shorten and J. Smith, Mixed methods research: expanding the evidence base, *Evidence Based Nursing* **20** (2017), 74.

[17] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, and J.i. Tsujii, brat: a Web-based Tool for NLP-Assisted Text Annotation, in, Association

for Computational Linguistics, Avignon, France, 2012, pp. 102-107.

[18] A. Stubbs and Ö. Uzuner, Annotating risk factors for heart disease in clinical narratives for diabetic patients, *J Biomed Inform* **58 Suppl** (2015), S78-91.

[19] T. Takai-Igarashi, R. Akasaka, K. Suzuki, T. Furukawa, M. Yoshida, K. Inoue, T. Maruyama, T. Maejima, M. Bando, M. Takasaki, M. Sakota, M. Eguchi, A. Konagaya, H. Matsuura, T. Suzumura, and H. Tanaka, On experiences of i2b2 (Informatics for integrating biology and the bedside) database with Japanese clinical patients' data, *Bioinformation* **6** (2011), 86-90.

[20] P. Thompson, S.A. Iqbal, J. McNaught, and S. Ananiadou, Construction of an annotated corpus to support biomedical information extraction, *BMC Bioinformatics* **10** (2009), 349.

[21] Ö. Uzuner, B.R. South, S. Shen, and S.L. DuVall, 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text, *J Am Med Inform Assoc* **18** (2011), 552-556.

[22] Ö. Uzuner, B.R. South, S. Shen, and S.L. DuVall, 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text, *Journal of the American Medical Informatics Association : JAMIA* **18** (2011), 552-556.

[23] Ö. Uzuner, A. Stubbs, and L. Lenert, Advancing the state of the art in automatic extraction of adverse drug events from narratives, *Journal of the American Medical Informatics Association* **27** (2019), 1-2.

[24] J. Wang, H. Liang, H. Kang, and Y. Gong, Understanding Health Information Technology Induced Medication Safety Events by Two Conceptual Frameworks, *Appl Clin Inform* **10** (2019), 158-167.

[25] J.I. Westbrook, M. Reckmann, L. Li, W.B. Runciman, R. Burke, C. Lo, M.T. Baysari, J. Braithwaite, and R.O. Day, Effects of two commercial electronic prescribing systems on prescribing error rates in hospital in-patients: a before and after study, *PLoS Med* **9** (2012), e1001164.

[26] WHO, Minimal Information Model User Guide, in, WHO, 2016.

[27] WHO, Patient safety incident reporting and learning systems: technical report and guidance, in, 2020.

[28] WHO, The final draft Global Patient Safety Action Plan 2021-2030, in, 2021.

[29] Z.S. Wong, Statistical classification of drug incidents due to look-alike sound-alike mix-ups, *Health Informatics J* (2014), 1-17.

[30] Z.S. Wong, H.Y. So, B.S. Kwok, M.W. Lai, and D.T. Sun, Medication-rights detection using incident reports: A natural language processing and deep neural network approach, *Health Informatics J* (2019), 1460458219889798.

[31] B. Yao, H. Kang, J. Wang, S. Zhou, and Y. Gong, Toward Reporting Support and Quality Assessment for Learning from Reporting: A Necessary Data Elements Model for Narrative Medication Error Reports, *AMIA Annu Symp Proc* **2018** (2018), 1581-1590.

[32] H.K. Zhang, R. Sasano, K. Takeda, and Z.S.Y. Wong, Development of a Medical Incident Report Corpus with Intention and Factuality Annotation, in: *LERC 2020*, Marseille, 2020, p. pp. 4580-4586.

[33] S. Zhou, H. Kang, B. Yao, and Y. Gong, An automated pipeline for analyzing medication event reports in clinical settings, *BMC Med Inform Decis Mak* **18** (2018), 113.

[34] C. Zins, Conceptual approaches for defining data, information, and knowledge, *Journal of the American Society for Information Science and Technology* **58** (2007), 479-493.

Address for correspondence

Zoie S.Y. Wong, Ph.D., Graduate School of Public Health, St. Luke's International University,

OMURA Susumu & Mieko Memorial St. Luke's Center for Clinical Academia, 5/F, 3-6-2 Tsukiji,

Chuo-ku, Tokyo, 104-0045, Japan (zoiesywong@gmail.com). Phone: +81-3-5550-4101 (Ext. 4397); Fax: +81-3-5550-4114