International Journal of Medical Informatics 108 (2017) 78–84

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Research Paper

Need of informatics in designing interoperable clinical registries

Majid Rastegar-Mojarad[a,b,*], Sunghwan Sohn[a], Liwei Wang[a], Feichen Shen[a], Troy C. Bleeker[c], William A. Cliby[c], Hongfang Liu[a]

[a] *Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA*
[b] *Dep. of Health Informatics and Administration, UW-Milwaukee, Milwaukee, WI, USA*
[c] *Mayo Clinic, Rochester, MN, USA*

ARTICLE INFO

Keywords:
Informatics
Clinical registry
Machine learning
Natural language processing

ABSTRACT

Clinical registries are designed to collect information relating to a particular condition for research or quality improvement. Intuitively, informatics in the area of data management and extraction plays a central role in clinical registries. Due to various reasons such as lack of informatics awareness or expertise, there may be little informatics involvement in designing clinical registries. In this paper, we studied a clinical registry from two critical perspectives, data quality and interoperability, where informatics can play a role. We evaluated these two aspects of an existing registry, Gynecology Surgery Registry, by mapping data elements and value sets, used in the registry, to a standardized terminology, SNOMED-CT. The results showed that majority of the values are ad-hoc and only 6 of 91 procedures in the registry could be mapped to the SNOMED-CT. To tackle this issue, we assessed the feasibility of automated data abstraction process, by training machine learning classifiers, based on existing manually extracted data. These classifiers achieved a reasonable average F-measure of 0.94. We concluded that more informatics engagement is needed to improve the interoperability, reusability, and quality of the registry.

1. Introduction

National Institute of Health (NIH) defines registry as "*a collection of information about individuals, usually focused around a specific diagnosis or condition*" utilized for research and quality improvement. Another definition of clinical registry is "*an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes*." [1]. Clinical registries have been designed for various purposes such as: studying the natural history of disease [2], analyzing clinical outcome of surgery/treatment [3], comparing different treatment methods [4], and measuring quality of care [5]. However many clinical registries have been designed successfully and there are user guides, aimed to assist with designing registries [1], but still there are some caveats on designing clinical registries such as: interoperability. Recently the United States congress approved a bill [6], which requires the U.S. Department of Health and Human Services to make recommendations regarding the structure and scope of clinical data registries. This bill mostly focuses on a set of standards to aid interoperable exchange of information between clinical notes and registries [7] and contains some recommendations about design and structure of clinical registries.

Besides interoperability, we faced some other challenges in designing successful and cost-effective clinical registry, while we were developing an enterprise-wide clinical registry infrastructure at Mayo clinic. We studied several existing clinical registries and noticed that 1) data quality 2) cost of human abstraction 3) lack of interoperability with EMRs and 4) lack of a master data resource are some of challenges in designing a clinical registry.

In this study, we hypothesized that effective informatics engagement in designing clinical registries can lead to cost effective, reusable, and interoperable clinical registries. Informatics '*studies the representation, processing, and communication of information in natural and artificial systems*' [8] and in healthcare domain, informatics defines as "*applying information science, computer technology, and statistical modeling techniques to develop decision support systems for improving both health service organizations' performance and patient care outcomes*" [9]. In the process of designing and implementing a clinical registry, informatics can contribute significantly, at least, in two tasks:

1) Defining data elements and determining the corresponding value sets
2) Collecting data and populating the registry.

The first task is critical for designing a reusable and interoperable

* Corresponding author at: Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA.

http://dx.doi.org/10.1016/j.ijmedinf.2017.10.004
Received 18 May 2016; Received in revised form 29 September 2017; Accepted 2 October 2017
1386-5056/

clinical registry. To ensure interoperability of registry, data elements and value sets should come from a standardized and universal health care terminology [10,11] and the value sets should be comprehensive and cover all possible values for the associated data elements. In biomedical informatics domain, there is a valuable resource called, Unified Medical Language System (UMLS) [12], which integrates and distributes key terminologies and coding standards to assist with creating effective and interoperable systems. One of the common and popular clinical terminologies in the UMLS is Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) [13], which can be used in the first task.

The second task is another area that informatics could play an important role and affect cost of human abstraction and more importantly quality of data. Data collection tool (DCT) [14] and Computer-Assisted Coding (CAC) [15–17] are two informatics tools which can assist abstractors in chart abstraction process and make the process automatic or semi-automatic. Using these tools in the second task, not only reduces the coding burden, but decreases some human errors and inconsistency between resources by following a simple rule in informatics "*one entry of a piece of data, many uses*" [18].

In this paper, we studied a clinical registry, Gynecology Surgery Registry, used by the Gynecologic Surgery practice at Mayo Clinic in Rochester, Minnesota. It contains basic encounter information, patient demographics, and various surgical related data elements such as procedures, diagnoses, and co-morbidities. In this study, we focused on one data element, *procedure,* of the registry, which captures and codifies the list of procedures performed in gynecologic surgeries. The study contains two parts. To assess interoperability of the clinical registry, we investigated data elements and their value sets and cross-referenced with a standardized terminology, SNOMED-CT. In the second part, we focused on chart abstraction process and making the process more automatic and error-proof. As CDC, we trained multiple binary classifiers, for each procedure in the registry, to identify whether procedures are reported in clinical notes or not. To find the best set of features and learning model for the classifiers, three classification methods (i.e., Naïve Bayes, Random Forest, and Support Vector Machine (SVM)) and three sets of features (i.e., unigrams, bi-grams, and topics retrieved by Latent Dirichlet Allocation [19]) were evaluated. To obtain more insights, the classifiers are analyzed and reasons for low performance in some of the classifiers are discussed.

In the following sections, we first discuss related work. Then the case study is presented. The results of our analysis are presented next followed by error analysis of the classifiers. Finally, the learned lessons, limitations, and future work are discussed.

2. Related work

Many clinical registries have been developed and studied for different conditions and diseases such as: "Alzheimer's Prevention Registry" [20], "Genome Connect" [21], and "Cancer Genetics Network" [22]. Shahian et al. [23] developed a clinical registry to study readmission measure for coronary artery bypass grafting surgery. McCombs et al. [24] studied and analyzed data from a department of veterans affairs clinical registry to evaluate the risk of long-term morbidity in patients with chronic hepatitis C. Sites et al. [25] illustrated the use of international clinical registry in quality improvement. Nwomeh et al. [26] studied trauma registry that as one of components in trauma care systems. Megan Quinn [27] has studied characteristics of cancer in adolescents using Tennessee cancer registry from 2004 to 2008. This type of publications mentioned or presented the importance role that clinical registries can play in various types of researches [28], but there are not much about how to design a successful clinical registry [29], what main concerns are and how to address those concerns. A publication supported by Robert Wood Johnson Foundation [28] is one of few publication which highlighted shortcomings in designing registry and noted these flaws can limit the role of registry. Gliklich et al. [29] provided a comprehensive user guide to design and develop a clinical registry. Silva et al. developed a standard framework for developing a device registry [30]. In this study, we emphasized the role of informatics in designing clinical registry.

Clinical registries value depends on the quality of their data [31–33]. Data in clinical registries have been compared with administrative claims data in several studies [34–37]. However, none of these studies focused on accuracy of clinical registries [32] or accuracy in data population process. There are three main factors impacting data accuracy 1) errors in original resources [38] 2) missing data [39–41] and 3) human errors [42]. The first one could be fixed to some extent with cross-referencing different resources such as clinical notes, surgery notes, structured data, and lab tests. Missing data issue has been addressed in several studies. Mendelsohn at el. [39] studied and characterized missing data in clinical registries and associated factors. Norris at el. [40] developed a method for handling missing data in a cardiac registry. They merged registry data with administrative data to fill missing data. In this study, we only addressed the third one, human errors in populating process. After assessing the accuracy of data in our case study, we discussed how informatics could improve the accuracy and decrease human involvement in populating process.

To improve the quality of data in clinical registry and decrease the ratio of errors (especially when subjective judgment is involved [15]) in the process of collecting information from medical records (chart abstraction process), CAC could be a useful tool. In general, CAC systems utilize natural language processing and machine learning algorithms to facilitate coding process. Predicting procedure codes from clinical notes or other type of text data has been studied in several domains. Hersh et al. [16] developed a machine learning system to assess the accuracy of predicting procedures codes from emergency room dictations. Using available data in trauma registry data, they trained the logistic regression classifiers with words appeared in the notes as features. Resnik at el. [15] implemented a CAC system that performed strongly relative to human performance. Morris at el. [17] developed an automated coding system called LifeCode which could be accurate as human coders. However, because of ambiguity in some of medical coding rules and guidelines, involving a human abstractor besides CAC system, seems necessary and will improve the accuracy. In addition to increase consistency in coding, a CAC system decreases needed labor and time for the process. In our case study, we implemented a CAC system using natural language processing and machine learning algorithms to investigate the potential use of assisting the human abstractors in the populating process.

3. Case study: gynecology surgery registry

The original database used in Gynecology Surgery Registry was derived from a professional society database in gynecologic cancer, and started to collect Gynecology Surgery data in 1990s at Mayo clinic. The primary goal was to tracking data retrospectively rather than focusing quality of elements. There was no electronic medical record available when the database was started and several modifications have occurred over the ensuing years. These limitations have made the database difficult to systematically collect data and hindered interoperability. Hence, we performed a case study applying informatics to the current Gynecology Surgery Registry at Mayo Clinic with respect to interoperability and data quality. We also evaluated the feasibility of using the current registry data to automatically codify procedures.

Fig. 1 shows a snapshot of one surgical encounter in the registry. The current database contains 10,160 visits from 1/20/1998 to 12/11/2014. For visits (7123 visits) with surgical notes, a human abstractor extracted procedures from the surgical notes. In Fig. 1, for the surgical note appeared in "*Procedure From SIRS*" textbox, a human abstractor identified 6 procedures (each row is one procedure). In the current design of the Gynecology Surgery Registry, each procedure is a combination of three fields: "*Anatomic location*", "*Procedure*", and "*Method*

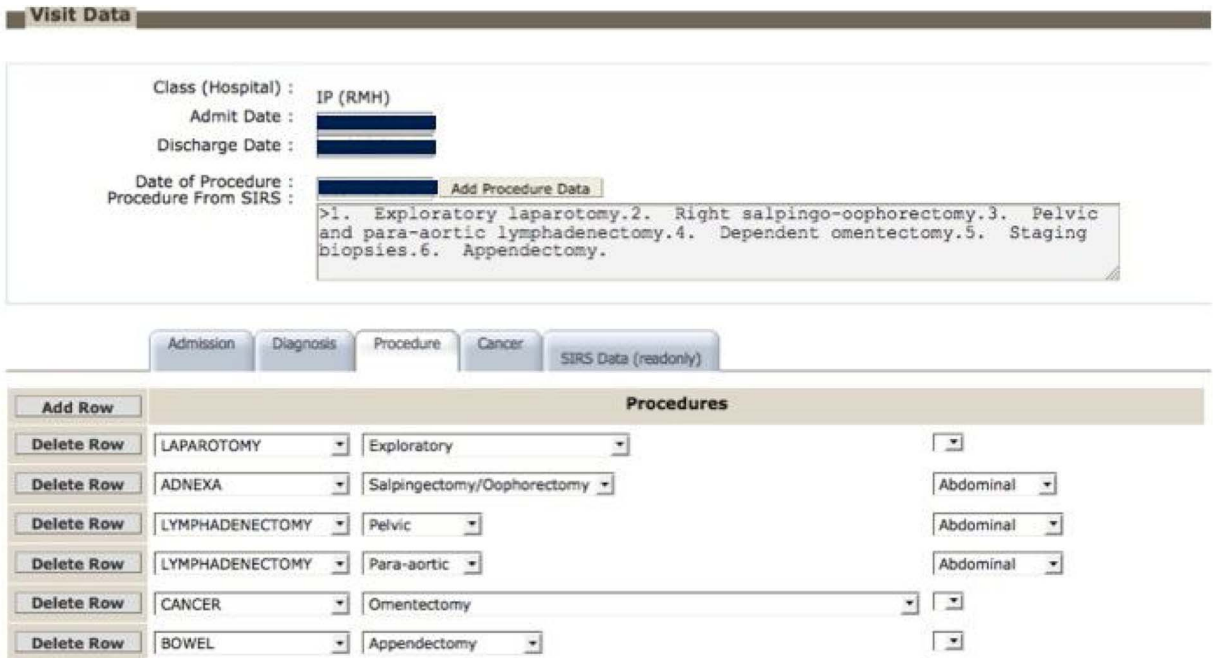

Fig. 1. A snapshot of a surgical encounter in Gynecology Surgery Registry.

or Approach”. For some procedures, only one or two of these fields are filled. For example in Fig. 1, the second procedure has value for all the categories, while the fifth one does not have any value for “*Method or Approach*”.

4. Methods

In this section, first we explain the method used to investigate interoperability and reusability of the gynecology registry. To do so, we assessed the coverage of data elements and value sets, used in the registry, in a standardized terminology. Second, we describe the classifiers that trained and implemented to automate chart abstraction process and improve data quality.

4.1. Assessing the registry value sets in a standardized terminology

First we investigated the existence of values stored in our registry in a standardized terminology. Second we compared their semantic type (category) in the registry and the terminology. As a standardized terminology, we used the UMLS, one of the common terminologies in biomedical informatics. The UMLS provides a set of broad subject categories, called semantic types (such as: procedure, anatomy, disorder, etc.), to categorize biomedical concepts. These semantic types are used in the second step of our investigation.

In our case study (Fig. 1), a procedure contains three elements “*Anatomic Location*”, “*Procedure*”, and “*Method or Approach*”. We retrieved all values for these elements, which were stored in the registry during 1/20/1998 to 12/11/2014. In the first step, using MetaMap [43], a tool to identify biomedical concepts, we investigated whether these values are available in the UMLS or not. For the values, which appeared in the UMLS, we compared their UMLS semantic types with their category in our case study.

In addition to investigating the values in the individual elements, we created a list of procedures in the registry and searched those in the UMLS. A procedure is a combination of “*Anatomic Location*”, “*Procedure*”, and “*Method or Approach*” fields in the registry. For example, for the first and second rows in Fig. 1, “*Laparotomy exploratory”* and “*Adnexa Salpingectomy/oophorectomy abdominal*” are created as procedures. We retrieved all combinations of these elements from the registry and created a list of values for procedure. These procedures are only searched in the SNOMED-CT, because it contains clinical terms. A clinical expert carried out this search manually and for the procedures that she did not find any exact match, she retrieved the closest (semantically) match from the SNOMED-CT.

4.2. Developing binary classifiers

In order to illustrate the use of informatics in CDC process, we developed a binary classifier for each of the combinations (procedures). For a combination with more than 100 occurrences in our registry, we defined the task of procedure extraction as a binary document classification task where all corresponding surgical notes of the procedure are treated as positive instances and all other surgical notes are treated as negative ones. In order to identify the best feature set for these classifiers, we generated multiple feature sets containing 1) unigrams, 2) bi-grams, 3) topics generated by topic modeling, and evaluated different combinations of those. Unigrams (single terms excluding stop words) and bi-grams (two neighbor words) feature sets are coming from n-gram model, a probabilistic language model. Unigrams are treated as the main features for the classifiers. As surgical notes usually contain several items, in order to generate bi-grams each item is processed separately. For example, the following text is from one of surgical notes in the registry:

> “*1. Bilateral pelvic lymphadenectomy. 2.Bilateral para-aortic lymphadenectomy. 3.Appendectomy. 4. Complete Omentectomy*”

Each item is treated like an individual section, and the following bi-grams are generated for this note:

1. “Bilateral pelvic”
2. “Pelvic lymphadenectomy”
3. “Bilateral para-aortic”
4. “para-aortic lymphadenectomy”
5. “Complete Omentectomy”

The last set of features comes from topic modeling, a statistical model for discovering hidden topics in documents. As a document can cover multiple topics, topic modeling calculates probability of documents belonging to each topic. We ran topic modeling on the surgical notes (as documents) and utilized the identified hidden topics as features in our classifiers. In fact, we utilized topic modeling as clustering method. LDA generates the probability of each document belonging in the different categories (topic). These probabilities are used as feature for the classifiers. For topic modeling, we used Latent Dirichlet Allocation (LDA), a generative statistical model. LingPipe [44], a suite of Java libraries, implemented LDA and is used in this project.

In addition to evaluating multiple feature sets, we assessed three learning methods for this task, Naïve Bayes, Random Forest, and Support Vector Machine (SVM). Using Weka [45] and LibSVM [46], we implemented these classifiers in Java.

We should mention that in our dataset the number of negative instances is relatively higher than positive instances (for most of the procedures). To smooth the effect of the unbalanced distribution of positive and negative instances, we assigned different weights to each class [47], using the distribution of the classes.

5. Evaluation Metrics

To evaluate performance of the classifiers, we used three common metrics, Precision, Recall, and F-Measure.

$$Precision = \frac{Number\ of\ True\ Positive}{(Number\ of\ True\ Positive + Number\ of\ False\ Positive)}$$

$$Recall = \frac{Number\ of\ True\ Positive}{(Number\ of\ True\ Positive + Number\ of\ False\ Negative)}$$

$$F - Measure = \frac{2*(Precision*Recall)}{(Precision + Recall)}$$

These metrics could be calculated for either positive or negative class. More common way is to involve the distribution of instances and calculate weighted average of the precisions, recalls, and F-measures. The weighted F-measure can be calculated by the following equation:

$$WeightedF - Measure = \frac{\left(\frac{NumberofPositiveinstances}{NumberofNegativeinstances}*F - Measure_{positive}\right) + F - Measure_{negative}}{\frac{NumberofPositiveinstances}{NumberofNegativeinstances} + 1}$$

Similar equations calculate weighted average precision and recall.

6. Results

There are 7,123 surgical encounters retrieved from the registry.

6.1. Basic statistics of value sets

The value set of three fields: *anatomic location*, *procedure*, and *method/approach*, contains 90 unique values. Among those, 67 (74%) values were found in the UMLS. Table 1 shows the top 10 most frequent values in the registry and how many times they occurred in each fields. One value appearing in different fields shows inconsistency in the registry. We found out that these 67 values belong to 30 different semantic categories in the UMLS. Table 2 shows top 10 frequent semantic types. The most frequent one is "*procedure*" occurring 21,026 times.

6.2. Mapping the procedure combinations to SNOMED-CT

The database contains 91 unique (*location, procedure, method*) combinations: 9 combinations appeared more than 500 times, 26 appeared more than 100 and less than 500 times, and 56 appeared less than 100 times. An expert was able to find only 6 (0.065%) combinations in the SNOMED-CT. This low coverage translates to a low interoperability of the system. Table 3 shows the top 10 frequent ones and the closest match found in the SNOMED-CT. Finding the exact match happened so rarely and the main reason for that are the procedures modifiers.

Table 1
Top 10 frequent values in the registry.

Term	Total occurrence	Anatomic location	Procedure	Method/ Approach	Found In UMLS
Uterus	4763	4762	0	1	Yes
Lymphadenectomy	3801	3801	0	0	Yes
Cancer	3583	3583	0	0	Yes
Adnexa	3449	3448	0	1	Yes
Salpingectomy/ oophorectomy	3342	1	3341	0	Yes
Hysterectomy	3026	1	3025	0	Yes
Abdominal	2862	0	0	2862	No
Urinary	2337	2337	0	0	Yes
Pelvic	2131	0	2131	0	Yes
Laparotomy	2052	2052	0	0	Yes

Table 2
Top 10 frequent semantic types in the value sets.

Semantic Type	Count
Procedures (Therapeutic or Preventive Procedure)	21026
Anatomy (Body Part, Organ, or Organ Component)	9736
Anatomy (Tissue)	5153
Concepts & Ideas (Qualitative Concept)	4377
Disorders (Finding)	3613
Procedures (Health Care Activity)	3080
Anatomy (Body Location or Region)	2569
Concepts & Ideas (Spatial Concept)	2510
Anatomy (Body Space or Junction)	2131
Occupations (Occupation or Discipline)	1719

Table 3
Top 10 frequent procedure combinations in the registry and their closest match in the SNOMED-CT.

Procedure Combination	Count	Closest match in SNOMED-CT
Adnexa; Salpingectomy/ Oophorectomy	2158	Bilateral salpingectomy with oophorectomy
Exploratory; Laparotomy	1739	Exploratory laparotomy
Pelvic; Lymphadenectomy; Abdominal	1374	Pelvic lymphadenectomy
Cancer; Omentectomy	1427	Omentectomy
Uterus; Hysterectomy; Abdominal radical	1194	Radical abdominal hysterectomy
Para-aortic; Lymphadenectomy; Abdominal	1266	Excision of periaortic lymph nodes
Cancer; debulking	928	Debulking of pelvic tumor
Uterus; Hysterectomy; Robotic	493	Hysterectomy
Bowel; Appendectomy	627	Appendectomy
Pelvic; Lymphadenectomy; robotic	347	Pelvic lymphadenectomy

6.3. Classification results

Table 4 illustrates the average of precision, recall, and F-measure (10 fold-cross validation) of classifiers for top 10 most frequent combinations. We provided results for different feature sets and learning models.

The results in Table 4 demonstrates that using SVM as training method and the combination of unigram, bi-gram, and topic modeling features, obtained the best performance. This classifier achieved average F-measure of 0.864 for the positive class and weighted f-measure of 0.94%. We believe that SVM obtained the best performance because it handles unbalanced distribution of the classes. Table 5 presents the performance of this classifier for top 10 combinations.

Table 4
Average of precision, recall, and F-measure of classifiers.

Method	Unigram	Bi-gram	LDA	Positive class			Weighted F-Measure
				Precision	Recall	F-Measure	
Naïve Bayes	X			0.671	0.911	0.768	0.897
Naïve Bayes	X	X		0.646	0.921	0.752	0.889
Naïve Bayes	X	X	X	0.641	0.923	0.750	0.888
Random Forest	X			0.887	0.733	0.797	0.924
Random Forest	X	X		0.893	0.738	0.800	0.927
Random Forest	X	X	X	0.892	0.737	0.800	0.927
SVM	X			0.769	0.943	0.841	0.931
SVM	X	X		0.799	0.947	0.862	0.942
SVM	**X**	**X**	**X**	**0.802**	**0.946**	**0.864**	**0.943**

Table 5
Performance of the best classifier for top 10 most frequent combinations.

Combination (Location; Name of procedure; Method)	Positive class			Weighted F-Measure
	Precision	Recall	F-Measure	
Adnexa; Salpingectomy/ Oophorectomy	0.706	0.886	0.786	0.820
Exploratory; Laparotomy	0.916	0.962	0.938	0.961
Pelvic; Lymphadenectomy; Abdonimal	0.828	0.967	0.892	0.945
Cancer; Omentectomy	0.971	0.978	0.975	0.987
Uterus; Hysterectomy; Abdominal radical	0.772	0.946	0.850	0.933
Para-aortic; Lymphadenectomy; Abdonimal	0.858	0.954	0.903	0.955
Cancer; debulking	0.628	0.885	0.735	0.902
Uterus; Hysterectomy; Robotic	0.854	0.973	0.909	0.983
Bowel; Appendectomy	0.901	0.931	0.916	0.981
Pelvic; Lymphadenectomy; robotic	0.589	0.976	0.735	0.962

7. Error analysis

After training the classifiers, we analyzed the results to identify reasons for misclassification of surgical notes. In this section, we reviewed two examples of false positive and two of false negative cases.

A) False Positives: (Meaning our system identified a procedure in a note, but human abstractor did not)

Example 1)

Surgical note: *"1. Robotic-assisted hysterectomy. 2. Bilateral salpingo-oophorectomy.3. Robotic-assisted bilateral pelvic lymphadenectomy"*

Our system identified "*Adnexa; Salpingectomy/oophorectomy*" procedure in the note, but human abstractor did not assign the procedure to this note. After reviewing the procedures assigned by the human abstractor to the note, we found out that the abstractor assigned the procedure to the note, but he/she also mentioned a method of the procedure. The abstractor assigned "*Adnexa; Salpingectomy/ Oophorectomy;* ***Robotic***" to the note. The abstractor used additional information to determine the method for the procedure. Basically, all Salpingectomy/oophorectomy cases have been done robotically in the practice since several years ago. This is not reflected in the documentation of surgical notes. Therefore, the same text can be mapped to multiple combinations. From data quality point of view, when a procedure can be done using multiple methods, the abstractor should capture the exact method used during the surgery from diverse sources or indicate not obtainable. The value sets should be clearly defined ontologically rather than a flat list since clearly, the combination assigned by the classifier is related to the combination entered by the abstractor.

Example 2)

Surgical Note: *"Exploratory laparotomy. Total abdominal hysterectomy. Bilateral salpingo-oophorectomy."*

The abstractor did not assigned any procedure to this note. The system made the correct decision and identified the procedure "*Adnexa; Salpingectomy/oophorectomy*" in the note. We considered this as human error in chart abstraction process.

B) False Negatives: (Meaning human abstractor assigned a procedure in a note, but our system did not)

Example 1)

Surgical Note: *"Abdominal exploration. Suturing of right hemi-diaphragm times two. Cauterization liver capsule for hemostasis."*

The abstractor entered "*Adnexa; Salpingectomy/oophorectomy"* based on this note. It is not obvious from the surgical note itself to obtain this procedure information, and we believe that abstractor used another piece of information besides the note.

Example 2)

Surgical Note: *"1. Exploratory laparotomy. 2. Abdominal sacrocolpopexy (Pelvitex). 3. Left salpingo-oophorectomy.4. Right salpingectomy. 5.Burch retropubic urethropexy. 6. Intentional cystotomy with placement of suprapubiccatheter.7. Posterior colpoperineorrhaphy."*

The system missed assigning *"Adnexa; Salpingectomy/oophorectomy*" procedure to the note.

8. Discussion

In this paper, we hypothesized that informatics can significantly contribute in designing a reusable, cost-effective, and interoperable clinical registries. As case study, we reviewed an existing clinical registry, Gynecology Surgery Registry. After studying the procedure section of the registry, we discovered that 74% of values, used for anatomic location, procedure name, and method/approach fields, exist in a standardized terminology (Table 1), but only 6 (less than 1 percent) out of 91 combinations of these fields matched to SNOMED-CT concepts (Table 3). The main reason of the low percentage is that the procedure combination in the registry is an ad-hoc data element. This data element has more modifiers comparing to procedures in the SNOMED-CT where some modifiers can be inferred from the ontological relationships and some can be defined using post-coordination SNOMED-CT expressions. For example, "*Adnexa; Hysterectomy; Abdominal radical*" combination in the registry is mapped to "*Radical abdominal hysterectomy*" procedure in the SNOMED-CT, but the combination contains a modifier "*Adnexa*" indicating tissues surrounding the organ which by default, the procedure will remove those tissues. The same thing for "*Uterus; Hysterectomy; Vaginal*" combination that is mapped to "*Vaginal hysterectomy*" in the SNOMED-CT. The use of standard terminologies with rich relationships can partially resolve the issues by supporting inference and increasing interoperability. Engaging informatics experts, familiar with standardized terminologies, in defining data elements and value sets could resolve this issue and promote interoperability of registry. Using standardized terminology not only solves the previous issue, but also improves data quality. Our investigation revealed that one of top 10 frequent terms in the registry is misspelled: "*Abdonimal*". Surprisingly, it occurred eight times more than the correct spelling "*Abdominal"*. Another error was assigning "*cancer*" to "*Anatomic location*" category. Using standardized terminologies can prevent these types of errors occurring. We should highlight that the registry has been used since 1998 and several abstractors (in different time period) had been involved in the data entry process, which caused some of these errors. Another reason for these errors and inconsistencies is that since 1998 some procedure names or methods have been modified or discarded. Considering the fact that terminologies keep changing, some inconsistencies in the registry are inevitable.

In the second part of the study, we discussed another task in designing clinical registries, collecting data for populating registry. For our case study, we showed that informatics tools such as: CDC, can assist abstractors in chart abstraction process. For each procedure (combination) in our case study, we trained a binary classifier using different feature sets and learning models. Table 4 shows that the classifiers performed reasonable well and obtained an average of 0.94 F-measure for top 10 frequent procedures. Table 5 shows the performance of our best classifier for top 10 most occurred combinations in more details. Our main limitation in training the classifiers was gold standard. As mentioned in the method section, we used existing registry data to train and evaluate the classifiers, but the registry has been populated by a single human abstractor (in different time period) and it is not ideal to use for training classifiers. In general, the performance of binary classifiers looks acceptable. However, it is not clear how well two human abstractors agree with each other. We consider a tool acceptable if it benchmarks with a gold standard, created through a judification process, and behaves just like a human abstractor when evaluated against the gold standard. However, it is quite expensive to derive such gold standard. The current resultant classifiers can assist human abstraction aiming for reducing effort rather than replace human abstraction. One of the objectives in our enterprise-wide clinical registry project is to decrease cost of human abstraction by deploying advanced informatics approaches and our experiment demonstrates it is feasible to reduce effort through secondary use of existing registry data.

9. Conclusion

In this study, we investigated the role of informatics in designing a reusable and interoperable clinical registry. We targeted two tasks in the design process: 1) defining data elements and value sets 2) populating registry. As a case study, we considered a Gynecology Surgery Registry that has been used since 1998. We cross-referenced data elements in the registry with a standardized terminology. Our investigation revealed some data quality issues in the registry: 1) misspelling 2) non-standardized definitions of value sets or data elements 3) inconsistency in the process of manual chart abstraction. We discussed how engaging informatics experts could solve these issues to some extent, and make the registry more interoperable. In addition, we presented that informatics tools are able to assist human abstractors in chart abstraction process and improve data quality. Using surgical notes and features such as: unigram, bi-gram, and topic modeling categories, we trained multiple binary classifiers to identify 91 different procedures from notes. Our best classifier obtained an acceptable F-measure of 0.94 using a noisy data.

Contributions

Liu and Cliby defined the problem and Liu, Cliby, Sohn, and Rastegar-Mojarad contributed in designing the solution. Bleeker prepared the dataset and Rastegar-Mojarad, Wang, and Shen implemented the classifier. All the authors contributed in preparing the manuscript.

Conflict of interest

We have no conflicts of interest to disclose.

Summary points

We evaluated the data quality and interoperability of an existing registry by mapping data elements and their value sets in the registry to a standardized terminology, SNOMED-CT. As case study, we used Gynecology Surgery Registry, used by the Gynecologic Surgery practice at Mayo Clinic in Rochester, Minnesota. To automate data abstraction process, we trained binary classifiers, for each procedure in the registry, based on existing manually extracted data. The study showed that only 13 % of 91 unique procedures in the registry could be mapped to SNOMED-CT concepts. The binary classifiers obtained an average F-measure of 0.864.

Acknowledgments

The authors would like to thank Joshua J. Pankratz, Jay B. Doughty, Nate Smrekar, Annette Kuball, and Paula Boos for providing the dataset. This study was made possible by National Institute of HealthR01 GM102282-01.

References

[1] R.E. Gliklich, N.A. Dreyer, M.B. Leavy (Eds.), Registries for Evaluating Patient Outcomes: A User's Guide [Internet], 3rd ed., Agency for Healthcare Research and Quality (US), Rockville (MD), 2014[cited 17 October 2016]. (AHRQ Methods for Effective Health Care). Available from: http://www.ncbi.nlm.nih.gov/books/NBK208616/.
[2] H. Khalili, M. Neovius, A. Ekbom, J.F. Ludvigsson, J. Askling, A.T. Chan, et al., Oral contraceptive use and risk of ulcerative colitis progression: a nationwide study, Am. J. Gastroenterol. (October (11)) (2016).
[3] A.J. Luikku, A. Hall, O. Nerg, A.M. Koivisto, M. Hiltunen, S. Helisalmi, et al., Multimodal analysis to predict shunt surgery outcome of 284 patients with suspected idiopathic normal pressure hydrocephalus, Acta Neurochir. (Wien) (October (14)) (2016).
[4] A.J.M. Watson, J. Hudson, J. Wood, M. Kilonzo, S.R. Brown, A. McDonald, et al., Comparison of stapled haemorrhoidopexy with traditional excisional surgery for haemorrhoidal disease (eTHoS): a pragmatic, multicentre, randomised controlled trial, Lancet (October 7) (2016).
[5] J.A. Holmes, J.T. Bensen, J.L. Mohler, L. Song, M.H. Mishel, R.C. Chen, Quality of care received and patient-reported regret in prostate cancer: analysis of a population-based prospective cohort, Cancer (September (13)) (2016).
[6] M.D. Hirsch, Legislation Would Spur Growth, Interoperability of Clinical Data Registries [Internet], FierceEMR, 2017 ([cited 12 Frebruary 2015]. Available from:), http://www.fierceemr.com/story/legislation-spurs-growth-interoperability-clinical-data-registries/2014-07-29.
[7] A.A. Sinaci, Laleci Erturkmen GB: A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains, J. Biomed. Inform. 46 (October (5)) (2013) 784–794.
[8] Ž. Turk, Construction informatics: definition and ontology, Adv. Eng. Inf. 20 (April (2)) (2006) 187–199.
[9] T.T.H. Wan, Healthcare informatics research: from data to evidence-based management, J. Med. Syst. 30 (1) (2006) 3–7 (Feb 1).
[10] N.A. Dreyer, S. Garner, Registries for robust evidence, JAMA 302 (7) (2009) 790–791 (Aug 19).
[11] B. McCourt, R.A. Harrington, K. Fox, C.D. Hamilton, K. Booher, W.E. Hammond, et al., Data standards: At the intersection of sites, clinical research networks, and standards development initiatives, Drug Inf. J. 41 (3) (2007) 393–404 (May 1).
[12] O. Bodenreider, The unified medical language system (UMLS): integrating biomedical terminology, Nucl. Acids Res. 32 (suppl 1) (2004) D267–70 (Jan 1).
[13] SNOMED Clinical Terms [Internet]. Available from: http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html.
[14] A.P. Moore, E.C. Bryant, G.W.J. Olivier, Development and use of standardised data collection tools to support and inform musculoskeletal practice, Man.Ther 17 (Decemver (6)) (2012) 489–496.
[15] P. Resnik, M. Niv, M. Nossal, G. Schnitzer, J. Stoner, A. Kapit, et al. Using Intrinsic and Extrinsic Metrics to Evaluate Accuracy and Facilitation in Computer Assisted Coding | Perspectives. In: Perspectives in Health Information Management Computer Assisted Coding Conference Proceedings [Internet]. [cited 12 February 2015]. Available from: http://perspectives.ahima.org/using-intrinsic-and-extrinsic-metrics-to-evaluate-accuracy-and-facilitation-in-computer-assisted-coding/#.

VNzHIuf0MjE.
[16] W.R. Hersh, T.K. Leen, P.S. Rehfuss, S. Malveau, Automatic Prediction of Trauma Registry Procedure Codes from Emergency Room Dictations, (1998).
[17] W.C. Morris, D.T. Heinze, H.R. Warner Jr., A. Primack, A.E. Morsch, R.E. Sheffer, et al., Assessing the accuracy of an automated coding system in emergency medicine, Proc. Amia Symp. 59 (2000) 5–9.
[18] L.Q. Thede, Informatics: where is it? J. Issues Nurs. 17 (January (1)) (2012) 10.
[19] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach Learn Res. (March (3)) (2003) 993–1022.
[20] Alzheimer's Prevention Registry [Internet]. Available from: http://www.endalznow.org/.
[21] Genome Connect [Internet]. Available from: http://www.clinicalgenome.org/.
[22] Cancer Genetics Network [Internet]. Available from: http://www.cancergen.org/.
[23] D.M. Shahian, X. He, S.M. O'Brien, F.L. Grover, J.P. Jacobs, F.H. Edwards, et al., Development of a clinical registry-based 30-day readmission measure for coronary artery bypass grafting surgery, Circulation 130 (5) (2014) 399–409 (Jul 29).
[24] J. McCombs, T. Matsuda, I. Tonnu-Mihara, S. Saab, P. Hines, G. L'italien, et al., The risk of long-term morbidity and mortality in patients with chronic hepatitis C: results from an analysis of data from a Department of Veterans Affairs Clinical Registry, JAMA Intern. Med. 174 (2) (2014) 204–212 (Feb 1).
[25] B.D. Sites, M.J. Barrington, M. Davis, Using an international clinical registry of regional anesthesia to identify targets for quality improvement, Reg. Anesth. Pain Med. 39 (December (6)) (2014) 487–495.
[26] B.C. Nwomeh, W. Lowell, R. Kable, K. Haley, E.A. Ameh, History and development of trauma registry: lessons from developed to developing countries, World J. Emergency Surg. 1 (1) (2006) 32 (Oct 31).
[27] M. Quinn, Exploring the Unique Characteristics of Cancer in Adolescents and Young Adults in Tennessee. Electronic Theses and Dissertations [Internet], (2012) (May 5; Available from:), http://dc.etsu.edu/etd/1438.
[28] R.W.J. Foundation, How Registries Can Help Performance Measurement Improve Care, (2010) (Jun [cited 5 February 2015]; Available from:), http://folio.iupui.edu/handle/10244/898.
[29] R. Gliklich, N. Dreyer, M. Leavy, Registries for Evaluating Patient Outcomes: A User's Guide [Internet], third edition, (2014) (Available from:), http://www.effectivehealthcare.ahrq.gov/.
[30] K.R. da Silva, R. Costa, E.S. Crevelari, M.S. Lacerda, C.M. de Moraes Albertini, M.M. Filho, et al., Glocal clinical registries: pacemaker registry design and implementation for global and local integration − methodology and case study, PLoS One 8 (7) (2013) (Jul 25 [cited 2015 Mar 3]; Available from:), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3723676/.
[31] D.G.T. Arts, N.F. de Keizer, G.-J. Scheffer, Defining and improving data quality in medical registries: a literature review, case study, and generic framework, J. Am. Med. Inf. Assoc. 9 (6) (2002) 600–611.
[32] F. Bray, D.M. Parkin, Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness, Eur. J. Cancer 45 (March (5)) (2009) 747–755.
[33] L. Kamell, S.S. Hoyler, J.E. Seiffert, J.L. Young, D.E. Henson, D.P. Winchester, et al., A national quality improvement effort: cancer registry data, J. Surg. Oncol. 58 (3) (1995) 155–161.
[34] M.I. Awad, A.G. Shuman, P.H. Montero, F.L. Palmer, J.P. Shah, S.G. Patel, Accuracy of administrative and clinical registry data in reporting postoperative complications after surgery for oral cavity squamous cell carcinoma, Head Neck (March (12)) (2014).
[35] S.K. Pasquali, X. He, J.P. Jacobs, M.L. Jacobs, M.G. Gaies, S.S. Shah, et al., Measuring hospital performance in congenital heart surgery: administrative versus clinical registry data, Ann. Thorac. Surg. (January (24)) (2015).
[36] E.H. Lawson, D.S. Zingmond, B.L. Hall, R. Louie, R.H. Brook, C.Y. Ko, Comparison between clinical registry and medicare claims data on the classification of hospital quality of surgical care, Ann. Surg. 261 (February (2)) (2015) 290–296.
[37] D.W. Jantzen, X. He, J.P. Jacobs, M.L. Jacobs, M.G. Gaies, M. Hall, et al., The impact of differential case ascertainment in clinical registry versus administrative data on assessment of resource utilization in pediatric heart surgery, World J. Pediatr. Congenit. Heart Surg. 5 (3) (2014) 398–405 (Jun 23).
[38] S. Hasan, R. Padman, Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach, AMIA Annu. Symp. Proc. 2006 (2006) 324–328.
[39] A.B. Mendelsohn, N.A. Dreyer, P.W. Mattox, Z. Su, A. Swenson, R. Li, et al., Characterization of missing data in clinical registry studies, Ther. Innov. Regul. Sci. 49 (1) (2015) 146–154 (Jan 1).
[40] C.M. Norris, W.A. Ghali, M.L. Knudtson, C.D. Naylor, L.D. Saunders, Dealing with missing data in observational health care outcome analyses, J. Clin. Epidemiol. 53 (April (4)) (2000) 377–383.
[41] D.A. Southern, C.M. Norris, H. Quan, F.M. Shrive, P.D. Galbraith, K. Humphries, et al., An administrative data merging solution for dealing with missing data in a clinical registry: adaptation from ICD-9 to ICD-10, BMC Med. Res. Methodol. 8 (Janury (1)) (2008) 1.
[42] M. Nahm, Data Accuracy in Medical Record Abstraction. UT SBMI Dissertations, (2010) (Open Access) [Internet]. 2010 May 6; Available from: http://digitalcommons.library.tmc.edu/uthshis_dissertations/15.
[43] A.R. Aronson, Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program, Proc. Amia Symp. 1 (2001) 7–21.
[44] Alias-i. LingPipe 4.1.0. 2008.
[45] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, SIGKDD Explor. Newsl. 11 (November (1)) (2009) 10–18.
[46] C.-C. Chang, C.-J. Lin, LIBSVM: a Library for Support Vector Machines, ACM Transactions on Intelligent Systems and Technology, 2011.
[47] M. Rastegar-Mojarad, Extraction and Classification of Drug–Drug Interaction from Biomedical Text Using a Two-Stage Classifier. Theses and Dissertations, (2013) (Dec 1).