Health Information Technology Department
Mashhad University of Medical Sciences

In the name of God

Mashhad University of Medical Sciences

Need of informatics in designing interoperable clinical registries

Najmeh Shariati

Dr. Khalil Kimiafar

Shariatin951@mums.ac.ir

International Journal of Medical Informatics

- the official journal of the **European Federation for Medical informatics (EFMI)** and **IMIA**
- **Impact Factor: 3.210**

Majid Rastegar-Mojarada[a,b], Sunghwan Sohna[a], Liwei Wanga[a], Feichen Shena[a],

Troy C. Bleekerc[c], William A. Clibyc[c], Hongfang Liua[a]

h-index: 8

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

CONTENTS

1. Introduction
2. Related work
3. Case study: gynecology surgery registry
4. Methods
5. Evaluation Metrics
6. Results
7. Discussion
8. Conclusion

1- INTRODUCTION

Registry

- **NIH**: “a collection of information about individuals, usually focused around a specific diagnosis or condition”
- “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- INTRODUCTION

Challenges: (design)

1) data quality
2) cost of human abstraction
3) lack of interoperability with EMRs
4) lack of a master data resource

1- INTRODUCTION

informatics:

- ‘studies the representation, processing, and communication of information in natural and artificial systems’ **(IT)**

- “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” **(health)**

1- INTRODUCTION

informatics: (design & implementation)

1) Defining data elements and determining the corresponding value sets

2) Collecting data and populating the registry.

2- RELATED WORK

This type of publications mentioned or presented the importance role that clinical registries can play in various types of researches, but **there are not much about how to design a successful clinical registry**, what main concerns are and how to address those concerns.

2- RELATED WORK

Clinical registries value depends on the quality of their data. (accuracy)

3 main factors impacting data accuracy:

1) errors in original resources
2) missing data
3) human errors

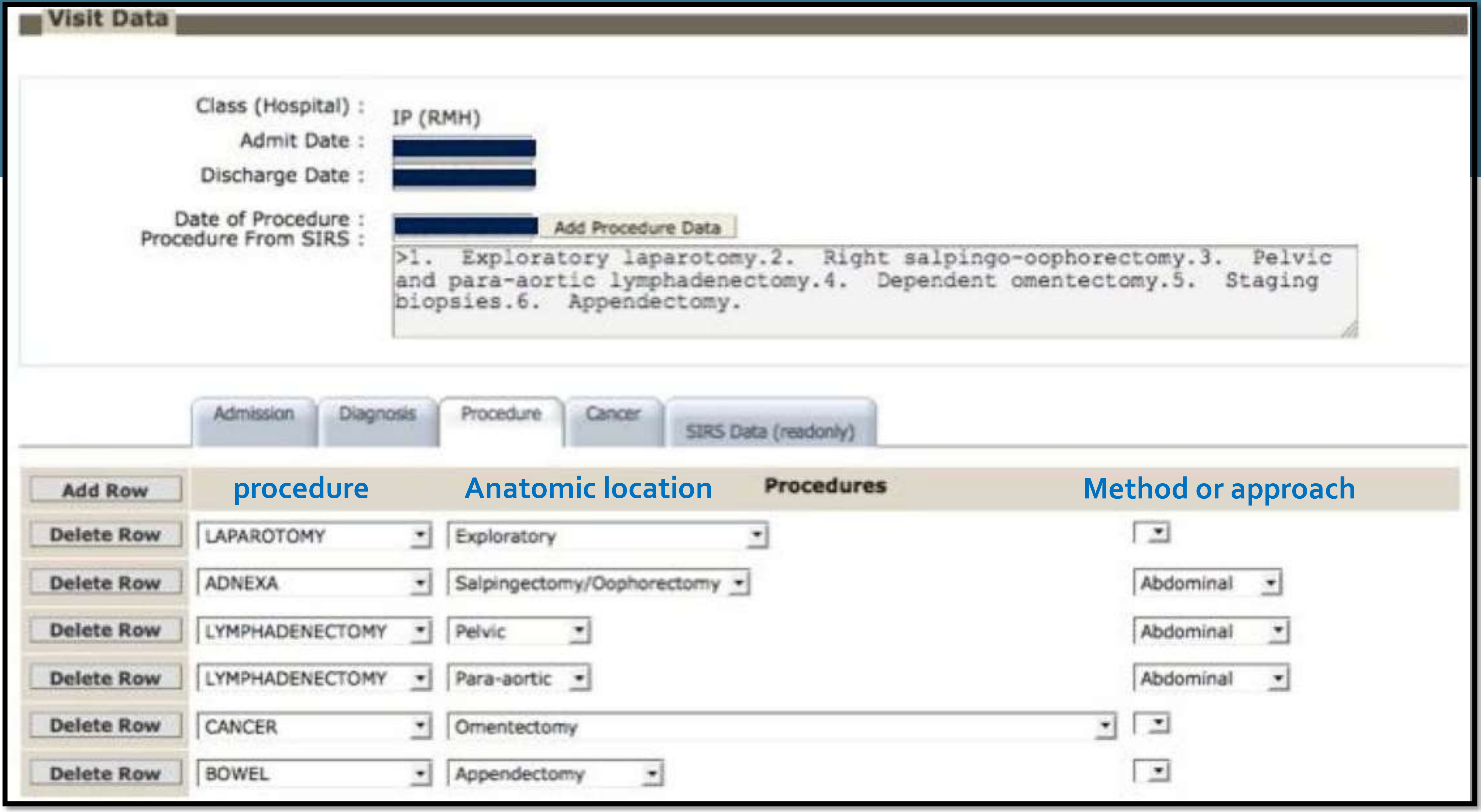
2- RELATED WORK

CAC: Computer-Assisted Coding

- To improve the quality of data in clinical registry and decrease the ratio of errors
- Natural Language Processing
- Machine Learning algorithms

3- CASE STUDY: GYNECOLOGY SURGERY REGISTRY

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.


Visit Data
Class (Hospital) : IP (RMH)
Admit Date : [redacted]
Discharge Date : [redacted]
Date of Procedure : [redacted]
Add Procedure Data
Procedure From SIRS :
>1. Exploratory laparotomy.2. Right salpingo-oophorectomy.3. Pelvic and para-aortic lymphadenectomy.4. Dependent omentectomy.5. Staging biopsies.6. Appendectomy.
Admission
Diagnosis
Procedure
Cancer
SIRS Data (readonly)
Add Row
procedure
Anatomic location
Procedures
Method or approach
Delete Row
LAPAROTOMY
Exploratory
ADNEXA
Salpingectomy/Oophorectomy
Abdominal
LYMPHADENECTOMY
Pelvic
Para-aortic
CANCER
Omentectomy
BOWEL
Appendectomy

4- METHODS

4.1. Assessing the registry value sets in a standardized terminology

4.2. Developing binary classifiers

5- EVALUATION METRICS

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

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

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

5- EVALUATION METRICS

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

6- RESULTS

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
Abdonimal	2862	0	0	2862	No
Urinary	2337	2337	0	0	Yes
Pelvic	2131	0	2131	0	Yes
Laparotomy	2052	2052	0	0	Yes

6- RESULTS

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

6- RESULTS

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; Abdonimal	1374	Pelvic lymphadenectomy
Cancer; Omentectomy	1427	Omentectomy
Uterus; Hysterectomy; Abdominal radical	1194	Radical abdominal hysterectomy
Para-aortic; Lymphadenectomy; Abdonimal	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- RESULTS

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**

6- RESULTS

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- DISCUSSION

- 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)

7- DISCUSSION

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
Abdonimal	2862	0	0	2862	No
Urinary	2337	2337	0	0	Yes
Pelvic	2131	0	2131	0	Yes
Laparotomy	2052	2052	0	0	Yes

Need of informatics in designing interoperable clinical registries

8- CONCLUSION

1) Defining data elements and determining the corresponding value sets

2) Collecting data and populating the registry.

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.

8- CONCLUSION

- Our best classifier obtained an acceptable F-measure of 0.94 using a noisy data.

Thanks for Your Attention

Health Information Technology Department
Mashhad University of Medical Sciences

NAJMEH SHARIATI

Shariatin951@mums.ac.ir