

Building a Local Surgical Lexicon for Quality Improvement: The Operative

Report Extraction Tool

by

Elizabeth Murphy, MD, MPH

Masters Student Candidate, Biomedical Informatics

Oregon Health Sciences University

Masters Capstone Project Report

## **Background and Significance**

### **Healthcare Costs and Quality**

The US leads all other industrialized nations in health care expenditures, now at almost 18% percent of our gross domestic product (GDP), yet health care quality in the U.S is not significantly better than countries that spend much less.<sup>1</sup> Surgical costs are a major driver of US health care costs. An estimated 29% of the healthcare portion of GDP is devoted to surgical costs and this does not include the costs of surgery–related disability or loss of productivity related to poor surgical outcomes.<sup>2</sup>

In the 2009 American Recovery and Reinvestment legislation, the Health Information Technology for Economic and Clinical Health (HITECH) Act was intended to address healthcare costs and quality, providing financial incentives for the “meaningful use” of Electronic Health Records (EHR) to improve the quality of health care while reducing costs. The Act supported major investments in EHR infrastructure with the hope that better access to healthcare data would produce increased research into improving health outcomes at lower cost. It also required attestation (reporting) of many “Meaningful Use” quality measures so that patient health status could be followed and health care providers could be rewarded (or possibly penalized) for health status improvements in their patients.

Much of the data that supports fulfillment of these measures has proven difficult to extract from clinical text as it is present in unstructured and un-coded free text fields within the EHR. <sup>3</sup> Although natural language processing (NLP) techniques are increasingly efficient at extracting meaningful quality information from electronic health records, NLP has not yet proven successful in extracting surgical quality information.<sup>4,5,6</sup>

### **Quality Models for Surgical Care**

Quality improvement is a major focus of the HITECH Act and the secondary use of EHR data for use in outcomes research has been a goal for both private and public sector health care organizations. <sup>7</sup> The Donabedian model for healthcare quality has persisted for almost 50 years and although it has been used mostly in hospital care and public health settings, it is also applicable to surgical quality improvement. <sup>8,9</sup> According to this model, quality outcomes are often harder to measure, but are obtainable through the use of quality processes and quality structures. In the surgical setting, positive outcomes include restoration of function, repair of injury, elimination of disease, or reduction of disease burden, and increased survival. Negative outcomes can include unnecessary surgery, surgical complications, increased morbidity, disability, decreased function, or death. Surgical processes and structures are often instrumental in determining outcomes. The quality of the operating facility, training of surgeons and staff, auditing of surgeon's statistics, and a culture of safety are key structural determinants of surgical quality. Processes, including verification of correct surgical site and patient, use of evidence-based surgical procedures and selection of appropriate patients for surgery are also central to quality surgical care. Surgical site errors can be

included in the quality continuum with poorer processes and structure contributing to a higher incidence of preventable errors. (Figure 1) Outcomes refer to the impact of the surgical procedure on the patient’s health status. Quality outcomes should be aligned with patient preferences and may be influenced by patient factors that may require additional process modifications.

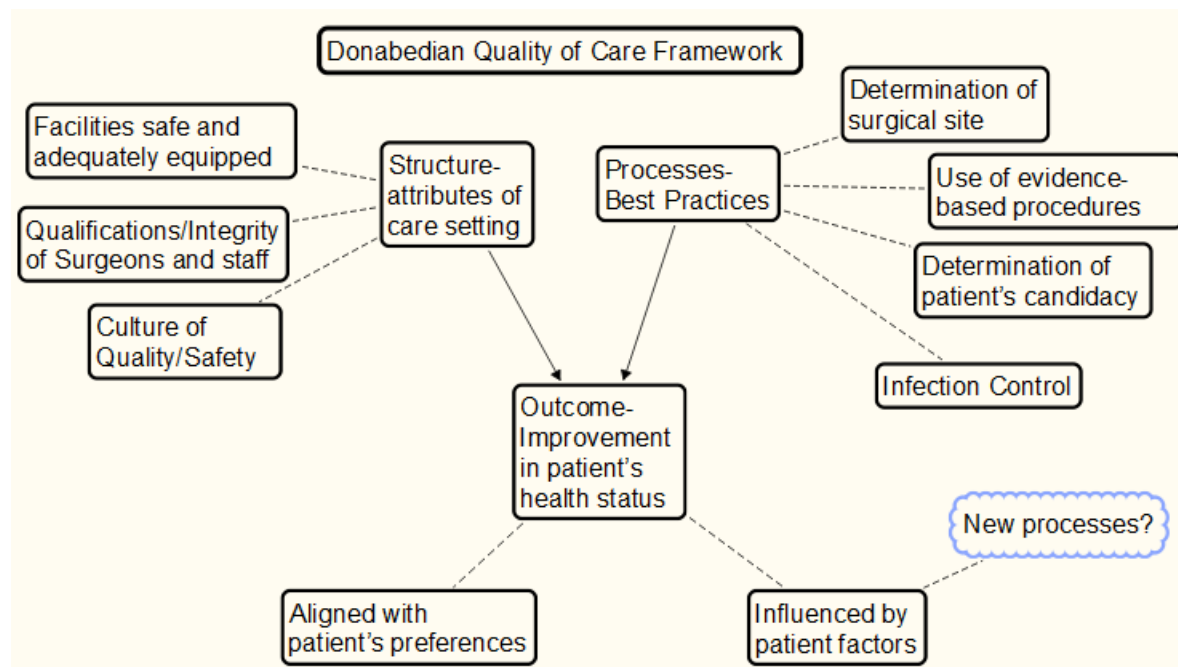


Figure 1. Surgical Quality Model

### Surgical Data Sources

Electronic Health Record data hold the potential for marked improvement in surgical outcomes. Operative reports contain specific information on preoperative and postoperative diagnoses, and specific operative procedures that can be evaluated or compared for effectiveness in treating these diagnoses. Operative reports also contain other potentially useful information including surgical site, indications for surgery,

estimated blood loss and complications of surgery. Unfortunately, much of this information is contained in unstructured free text, and despite the availability of coded data through ICD9 and CPT formats, surgical diagnoses and procedures are inadequately represented by these codes.<sup>10</sup> ICD9 and CPT codes are insufficiently granular to represent many common surgical diagnoses and procedures. They do not include laterality which is important for determination of wrong site surgery or repeated surgeries to the same body part. These systems also include “not elsewhere classified” (NEC) codes that assist with billing, but prevent reliable cohort retrieval for outcome evaluations or clinical research since these categories are akin to a “miscellaneous” category. Since 2004, Medicare has had modifiers for CPT codes that specify laterality and repeat surgery to the same body part but these have not been publically available.<sup>11</sup> In a recent case, Medicare data that was shielded from public disclosure, under the privacy portion of the Freedom of Information Act, was uncovered which exposed a Neurosurgeon performing an average of seven repeat spine surgeries per patient-repeat surgeries and physician owned distributorships of surgical devices have since become a target for legislation.<sup>12</sup>

Although ICD-10 Codes are much more granular than ICD-9 codes and include laterality, they still contain NEC modifiers and ambiguous codes such as “unspecified tear of unspecified meniscus, right knee” or “tear of unspecified meniscus, unspecified knee” making them useful for billing but not for surgical outcomes research.<sup>10,13</sup>

## **The Importance of Ontologies In Understanding Quality Outcomes**

In a classic publication on the need for controlled medical vocabularies and ontologies, “Desiderata for controlled medical vocabularies in the twenty-first century,” Cimino argues that vocabularies must resist temptation to include NEC categories since these categories merely state that the term is not related to existing concepts in a vocabulary and are non-specific.<sup>13</sup> Later works by Cimino and others reiterate the call for standardized terminologies and more specifically the need for biomedical ontologies, where ontology can best be defined as a representation of the conceptual meaning behind each term in a vocabulary and the inclusion of relationships between terms within a vocabulary.<sup>14, 15, 16</sup> The need for terminologies that accurately describe different knowledge domains and clinical text are important for several reasons. More standardized terminologies are needed to communicate medical information between disparate electronic health records systems. Specific domain areas such as neurology or surgery have specialized vocabularies that more accurately describe entities related to medical practice within their domain. Terminology may vary by geographic areas within a country or region, and in a growing global scientific community, medical knowledge must also be transferred across languages. Ontologies allow different terms with the same meaning to map to an underlying concept that defines their meaning. Ontologies are a source for “computable knowledge” allowing data and information retrieval, data integration and interoperability, transfer of knowledge through decision support, and enabling natural language processing applications.<sup>15</sup>

With an explosion of biomedical data, ontologies are increasingly becoming the cornerstone of knowledge management and knowledge representation following the paradigm of data to information (usable data) to knowledge (understood data).<sup>17</sup> Antezana et al define knowledge representation and its dependence on concepts within ontologies in order to represent models of “real-world entities,” allowing understanding of specific domains, and also allowing members of a “domain community” to “efficient(ly) process... information in a computational environment.”<sup>17</sup> Semantic Web technologies are based on ontologies and include standardized knowledge representation languages based on these ontologies, such as the Resource Description Framework (RDF), RDF Schema, the Web Ontology Language (OWL 1), and OWL-DL (description logics.)

Rubin et al describe ontologies from a “functional perspective” since some ontologies not only describe the concepts and relationships between terms, but also the breadth of knowledge within a specific domain and the potential applications for the relationships within the domain.<sup>16</sup> Rubin’s functional view is very similar to Antezana’s knowledge representation model- functional applications for ontologies include “search and query of heterogeneous data,” data exchange between applications and integration, NLP applications, “representation of encyclopedic knowledge,” and “computer reasoning with data.”<sup>16</sup> He describes the Foundational Model of Anatomy (FMA) as an important reference ontology for its function in the formal representation of knowledge.

The FMA is a knowledge source for anatomy which describes anatomic relationships from the cellular to the gross anatomic level. It has a canonical structure that symbolically represents classes and relationships of structures so that they are both human and machine (computer) understandable.<sup>18</sup> It functions as an encyclopedic knowledge source with a comprehensive inclusion of anatomic entities and an extensive representation of relationships to other structures. Since the FMA includes adjacency and lists adjacent structures, it has been useful for predicting penetrating trauma injuries and relationships in radiologic exams.<sup>19</sup> The hierarchical structure of the FMA and comprehensive inclusion of anatomic structures that are often the site of injury or surgical interventions, make it a useful resource in categorizing a surgical diagnosis and procedure terminology by anatomic site.

Data retrieval, integration and knowledge representation are areas that would be helpful in improving quality and outcomes of surgical procedures since we now have the potential to evaluate large cohorts of patients with matching surgical diagnoses that have received specific surgical procedures. Unfortunately, surgical diagnoses and procedures are not captured in sufficient granularity by any of the standard coding methods or by any current ontology. Snomed CT, a comprehensive medical term ontology in the National Library of Medicine's UMLS (unified medical language system) captures diagnoses and procedures, and covers many surgical concepts and terms but it is not complete and contains some errors in concept mapping for surgical terms. For example, total or partial meniscectomies, and meniscal transplant procedures are three very different approaches to meniscal injuries of the knee, but only one generic concept



for meniscectomy exists for these procedures in SNOMED CT. In an ideal system, we would capture the granularity and accuracy of local surgical lexicons and map these terminologies to Snomed CT concepts when possible, eventually adding to and refining the Snomed CT ontology so that it becomes an adequate ontology for surgical concepts. The first step in this process, however, is to develop a comprehensive surgical lexicon that captures the diagnoses and procedures common to a specific health care setting.

Even without an ontology, a surgical lexicon would be useful in categorizing, understanding and improving surgical outcomes in a local environment. Certain types of surgeries, in particular shoulder and spine surgeries, have an inordinately high failure rate and evaluation of different procedures has resulted in improved outcomes.<sup>20, 21, 22</sup> Failed repair of larger rotator cuff tears, particularly in patient over 50, can result in significant morbidity and functional loss and often leads to more significant and riskier surgeries such as reverse total shoulder replacement- detailed analysis of technique effectiveness has been particularly useful in this setting.<sup>23, 24</sup> Having highly specific operative report data would be particularly useful in evaluating the use of spine surgery techniques and in determining the incidence of potentially unnecessary spine surgeries in a care setting. Spinal fusion surgeries are somewhat controversial with high failure and disability rates yet the rate of spinal fusion surgery has increased dramatically in the past twenty years to 122,000 by 2001, 250,000 by 2003 and over 500,000 by 2006, costing 2.5 billion dollars per year just in hardware costs, and the necessity of these surgeries have been brought into question as a major patient safety issue.<sup>25, 26</sup> Highly

experimental surgeries such as artificial disc replacement have taken several years to be disproven due to inadequate methods for cohort identification.<sup>27</sup> Use of costly, ineffective, and potentially unsafe materials by spine surgeons with financial relationships to manufacturers, have also gone undetected due in part to poor outcome and use tracking for surgical procedures.<sup>28</sup>

Surgeons are beginning to understand the importance of a comprehensive surgical lexicon and the development of ontologies in the improvement of surgical care in clinical decision support systems. An orthopedic ontology for the musculoskeletal system of the lower limb has been constructed and used in a computer-aided decision support system for improved classification, and potentially treatment, of club foot deformities.<sup>29, 30</sup>

Surgical ontologies were briefly addressed in the Galen In Use framework almost 20 years ago but strategies to build a comprehensive surgical lexicon and ontology have been sparse since that time.<sup>31</sup> There has been a limited use of the Specialist Lexicon in natural language processing assessments of actions and predicate argument structure frames in operative reports but few studies have addressed methods in operative report data extraction even though these reports are a rich source for data related to surgical practices.<sup>5, 6</sup>

## **Research Question:**

Can an operative report extraction tool efficiently and accurately extract clinically useful data from operative reports creating a surgical lexicon? Can the tool categorize the data and return the data in a format that is useful for surgical quality improvement and research purposes? Can the extracted terms be mapped to an existing ontology in order to standardize surgical terms and build on existing ontologies?

## **Specific Research Aims:**

1. Efficiently and accurately extract information from operative reports in a format that is useful for surgical quality improvement and outcomes research.
2. Extract surgical terms from operative reports towards building a more standardized surgical ontology

## **Research Methods**

### **Pilot Extraction of Diagnosis and Procedure Terms - BTRIS Database**

Although operative reports are most often in free text form, review of multiple of operative reports transcripts from distinct EHR systems and healthcare settings suggested that certain elements of the reports were often organized in a consistent manor. In particular, elements such as preoperative and postoperative diagnoses, and operative procedure, are often framed in a unique manor creating a detectable pattern within the text. Xu et al used a bootstrapping technique to extract free text following

specific phrases in the formation of a disease dictionary using abstracts from Medline.<sup>32</sup> Operative report diagnosis and procedure information appeared to be amenable to extraction through the use of a similar bootlegging technique in combination with the use of regular expression techniques which are available through the Java programming API. A strategy was developed to extract between regular expression terms that flanked preoperative and postoperative diagnosis and operative procedure in order to retrieve these elements from the text.

The Biomedical Translational Research Information System (BTRIS) is a resource providing clinical research data from the NIH Clinical Center and NIH institutes for research use by the NIH intramural community.<sup>33</sup> The data includes de-identified records, including operative reports, from 1976 to present. After obtaining approval from the Office of Human Subject Protection and Research at the NIH, de-identified operative reports from the BTRIS data set were first evaluated by manual inspection. Distinct patterns in format were observed and selected for text extraction using a regular expression technique. A regular expression algorithm was written to extract phrases between the selected regular expressions. A preliminary test extraction was performed on a sample of 98 operative reports and this returned 98 preoperative and postoperative diagnoses and operative procedures.

The extraction code was then modified to remove white space from the captured regular expressions and to allow for tagging of missing diagnoses and procedures in a larger extraction set. A set of 4660 operative reports was then extracted with extraction of

4630 pre-operative diagnoses, 4630 post-operative diagnoses and 4645 operative procedures.

Missing terms were counted and reported by the algorithm. Output work sheets were also reviewed page by page, and all missing values were analyzed. Over 10% (500) of operative reports, were manually reviewed and extracted terms were compared with the terms at the targeted location in the text from each operative report. Precision and recall were calculated using standard formulas

The outputted extraction results were printed to XLSX worksheets. Worksheet data included columns for operative report ID, preoperative diagnosis, postoperative diagnoses, and operative procedure as well as a de-identified subject ID and date of the procedure. Subject ID and dates were removed from the work sheet prior to analysis. A subset of the initial 5000 reports targeted for extraction, were determined to be duplicates and were removed, leaving 4660 unique operative reports for the final analysis.

### **Operative Report Extraction Tool Development**

Creation of an extraction, categorization and ontology matching tool for use in creating a local surgical lexicon was then planned. The first step in this process is the creation of an Operative Report Object with attributes that would include operative report ID, Subject ID, date of surgery, preoperative diagnosis, postoperative diagnosis, surgical procedure, surgical category, and the UMLS CUI (concept unique identifier), if identified.

Preliminary UML structures for the operative report class and the category class that would categorize operative reports by body system, and body part, are presented in Figure 2a and Figure 2b.



Figure 2a. Operative Report Object UML

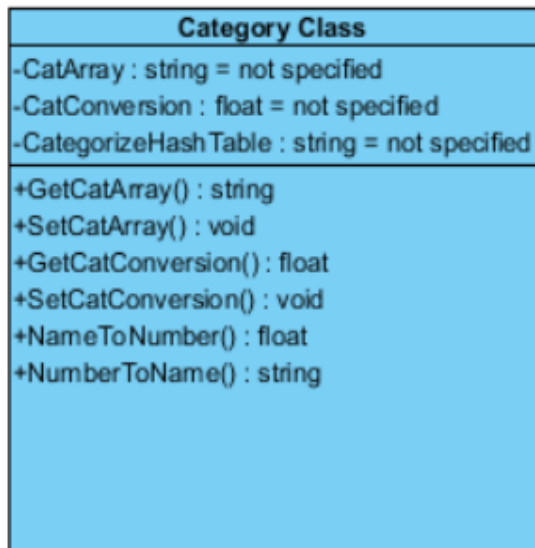


Figure 2b. Category Class UML

An initial model of the “Open Operative Report” tool is shown in Figure 3. The tool includes the creation of an operative report object that contains important operative report attributes including operative report ID, preoperative diagnosis, postoperative diagnosis, operative procedure, operative category and operative diagnosis, operative procedure concept unique identifiers, if available. These are recognized as initial attributes only since multiple other operating report attributes would be mined after further development of the tool. The operative report extraction tool would read free text operative report from text files or XLSX worksheets, and then extract the attributes from each report and output the results to a table that preserves the relationships between diagnosis and procedures.

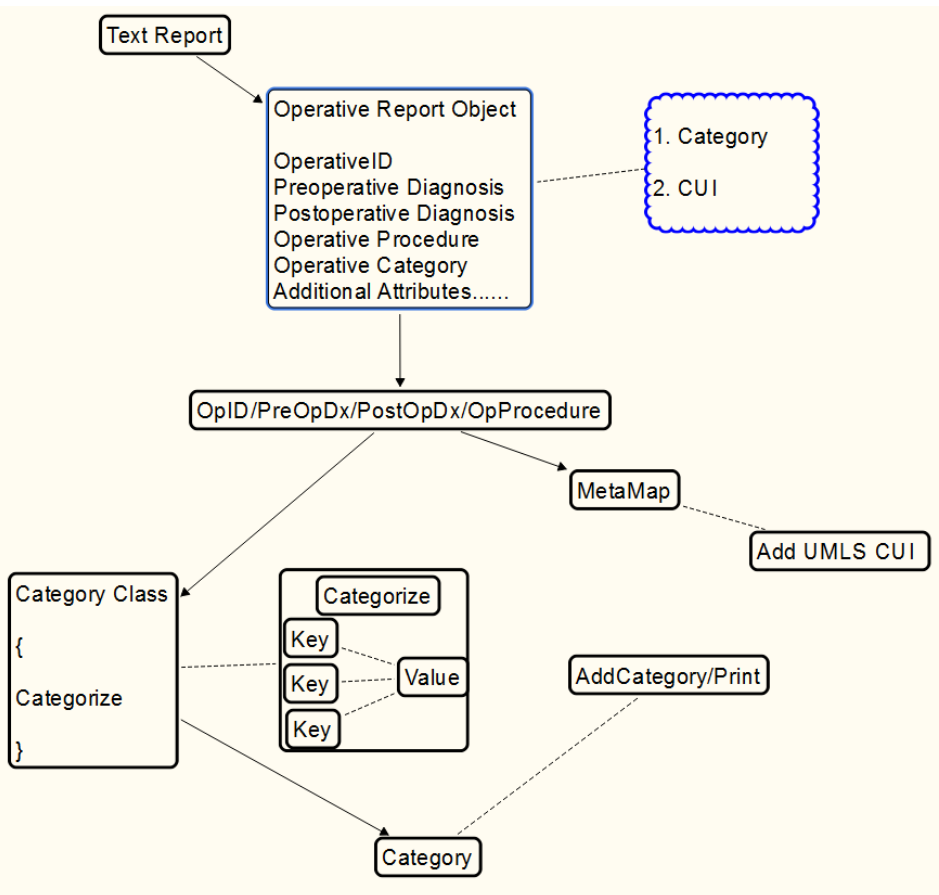


Figure 3. Operative Report Extraction Tool Model

Each row of the table, representing a separate operative report extraction, would then be looped through a categorization hash table, and one or more categories would be assigned to the operative report and added to the table

Extracted terms could also be processed through MetaMap, an NLM tool that allows mapping of terms to UMLS concept unique identifiers, more specifically to the concepts with the standardized ontology, SNOMED CT.<sup>34</sup> If matching concepts are identified, lexical terms would then be assigned specific concept identifiers, consistent with the SNOMED ontology.

### **Categorization Hash Table Development**

A categorization hash table was developed in order to assign each operative report to a specific anatomic category. The Foundational Model of Anatomy (FMA) was consulted through the use of the online FMA Explorer tool which allows in depth anatomic exploration of each body area (Figure 4). Hierarchies within the tool, and subject matter expertise were used to first develop a detailed list of anatomic categories that would become values with the categorization hash table.



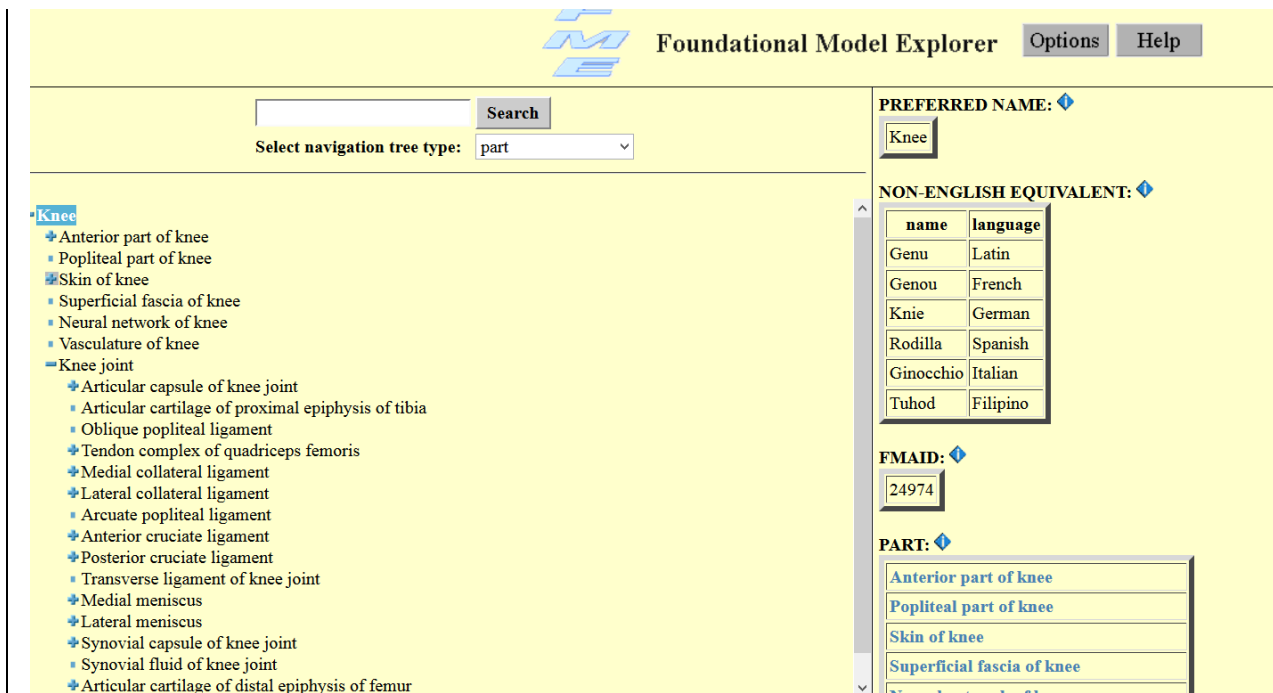


Figure 4. Foundational Model of Anatomy Explorer

Categorization values were chosen for a high degree of granularity including categorization by body system and body part, down to the level of specific digits and spinal levels. Laterality was also included in the categorization. Subject matter expertise, and the FMA was again consulted in the development of hash table keys. A portion of initial lexicon terms were used to verify the keys in the hash table as the keys were being developed. A preliminary hash table was created with over 700 unique keys assigned to 90 anatomic categories based on body system and body part.

A categorization algorithm was then designed in the JAVA API. The output from the initial operative report extraction, including 4660 tuples, each containing extracted terms from a unique operative report were first looped through the categorization hash table, where one or more anatomic categories were assigned. Since operative reports often

included operations to several body parts, more than one category could be assigned to each operative report. The assigned categories were then added to each tuple and again outputted in XLSX format.

### **UMLS/ SNOMED CT Concept Mapping – MetaMap**

The extracted lexical terms from the initial BTRIS extraction were organized and duplicate terms were removed. Using regular expression algorithms and search and replace techniques all modifiers not essential to the term's concept, such as "left", "right", "mild" and "severe", were removed from the lexical terms. All lexical terms were then processed in MetaMap where they were mapped to UMLS concepts, including mapping to concept unique identifiers (CUI), which are the assigned alphanumeric identifier for each concept within the UMLS.<sup>35</sup> In order to capture potential partial matches to an incorrect semantic type, or possible incorrect assignment of semantic type, all semantic types were included in this initial screening.

All matches were manually evaluated with > 5% outputted text reviewed in this initial assessment. Output was evaluated for: complete match with appropriate semantic type, scored as full match; partial number of concepts fully matched, or a single concept with semantically close but imprecise match, partial match; and half of a term matched with wrong semantic type, no match. When an exact concept match was identified, concept unique identifiers (CUIs) were added to each tuple as was the MetaMap preferred CUI name.

## **External Validation of the Operative Report Extraction Tool**

Permission was obtained by the Veteran's Administration IRB for evaluation of patient's operative report data from the Women Veteran's Cohort study (WVCS). The WVCS, established in 2007, collected a cohort of administrative and EHR derived clinical data from the medical records of over 900,000 female and male Veterans from the periods of Operation Enduring Freedom, Operation Iraqi Freedom and Operation New Dawn periods.<sup>35</sup> Over 10,000 operative reports are available within this data collection from VA sites across the U.S..

Prior to analysis, several sample operative reports from multiple VA sites were reviewed, and as predicted, each site appeared to have a highly structured format amenable to regular expression extraction that was specific to that site. It was noted that VA operative report format was not standardized across VA sites, so that regular expression extraction strategy would need to be tailored to each institution.

A single Veteran's Administration site was chosen based on sites with availability of over 1000 operative reports in the database. Less than 10 operative reports were manually reviewed and an extraction algorithm using regular expressions was developed in Java API. An extraction technique similar to the technique used on the BTRIS data<sup>33</sup> was performed on all 1149 operative report records from the site. The extraction algorithm was modified to allow reading of the reports from free text format.

## **Results**

### **Operative Report Extraction- BTRIS Database**

The operative report extraction results contained over 13, 913 diagnoses and procedures. After duplicates were removed, over 6000 unique diagnoses and procedure terms were identified. The results were first examined for missing extraction terms. The rate of missing terms was less than 1% with a total of 67 missed terms for the 13,980 possible terms extracted (three columns times 4660 operative reports). Thirty two of the missed terms were actually missing from the transcribed operative report and these reports were considered to contain true negatives since the extraction result that there was no information between the regular expression terms was correct. Almost all of the remaining missed terms were listed as preoperative “diagnoses” rather than preoperative “diagnosis.” (This could be easily corrected in future extractions using the regular expression “diagnos[ie]s.”)

### **Internal Validation of Extracted Reports**

Internal validation was performed on a randomly selected subset of 500 operative reports, representing over 10% of the extracted records. Preoperative and postoperative diagnoses were all exact matches to the actual diagnoses in the manually reviewed operative reports. Operative procedures were occasionally incorrectly formatted to separate columns for multiple procedures in the XLSX spreadsheet (which will be

resolved by the use of a pipe separated value, free text output). In two cases, an additional sentence was added to the procedure column, starting in “Specimens of” and it was determined that there were two “Specimens of” sentences in each of these reports.

Precision and recall were evaluated by first determining the numbers of true positives, false positives, true negatives and false negatives in each report. Recall is defined as the number of true positives detected / total number of true positives in the targeted operative report fields. Precision is defined as the number of true positives detected/ the number of true positives and false positives detected. Precision and recall were both greater than 99% for this extraction (Table 1).

### INTERNAL VALIDATION BTRIS OPERATIVE NOTES

TOTAL OPERATIVE REPORTS	<b>4660</b>
TOTAL TERMS EXTRACTED	<b>13913</b>
TOTAL UNIQUE TERMS	<b>6042</b>
TOTAL MISSING TERMS	<b>67</b>
MANUAL CHART REVIEW	<b>&gt;10%</b>
PRECISION	<b>99%</b>
RECALL	<b>99%</b>

**Table 1. Internal Validation Operative Report Extraction**

## BTRIS Extraction Categorization Results

Categorization of the BTRIS extraction results produced an XLSX worksheet document with “Operative Category” (site) added to the previous operative report ID, preoperative and postoperative diagnosis, and procedure columns (Figure 5.)

2709	Refractory uveitis and likely indolent intraocular infect	Refractory uveitis and likely indolent intraocular infection, left e	20-gauge pars plana vitrectomy, intraocular lens explantation, and inj	Left Eye
2710	Cataract, left eye, and macular degeneration, left eye.	Cataract, left eye, and macular degeneration, left eye.	Phacoemulsification, cataract extraction, with sulcus intraocular lens	Left Eye
2711	Retained lens fragments, left eye.	Retained lens fragments, left eye.	20-gauge pars plana vitrectomy and lensectomy, left eye.	Left Eye
2712	Primary pigmented nodular adrenocortical disease wit	Primary pigmented nodular adrenocortical disease with Carney	Drainage of paronychia.	Left First Toe
2713	[LAST_NAME i=168] length discrepancy secondary to f	[LAST_NAME i=168] length discrepancy secondary to Proteus s	Removal of "figure of eight" plate, right tibia.	Left First Toe
2714	Gigantism left great toe Proteus syndrome.	Gigantism left great toe Proteus syndrome.	Amputation left 1 toe through the proximal phalanx.	Left First Toe
2715	Synovial cell sarcoma of left foot.	Synovial cell sarcoma of left foot.	Left below knee amputation, midleg, for synovial cell sarcoma of foo	Left Foot

Figure 5. Categorization Sample (operative site highlighted in yellow)

Categories were assigned by the categorization algorithm for each diagnosis and procedure in the operative report with multiple categories assigned if the operation or diagnoses related to multiple body systems or body parts. Each body part category was assigned only once and most operative reports were assigned only one category. The maximum number of categories assigned to a single operative report was five, in a surgery where procedures were performed on five unique body areas. Order of operative categories was not based on order of priority since categories were merely assigned by the order of the category key’s occurrence in free text.

The entire operative report extraction from 4660 unique operative reports, including preoperative and postoperative diagnosis and procedure, was first looped through the

hash table and categories were assigned to each tuple. This was followed by looping the operative procedure only, from each operative report, through the categorization hash table. Results were analyzed by initial analysis of missing categories. Categories were assigned to 97% of reports, using the complete extraction (i.e. all three columns of diagnosis and procedure data). When only the operative procedure for each report was processed through the hash table , 95% of reports were categorized.

Preliminary analysis of categorization data revealed that specificity of categorization was higher in the procedure only categorization results. This was because diagnoses were sometimes not related to the operative procedure (e.g. procedure was insertion of a deep venous access line for a diagnosis of lung cancer). Preliminary manual review analysis of 5% of categorization results, revealed accuracy of categorization in over 95% of categorized reports, though a more complete analysis of categorization accuracy is planned after the hash table algorithm is further refined in future research. A summary of operative report categorization results is provided below.

- > 97% categorization when diagnoses and procedures used
- > 95% categorization with procedures alone
- Categorization allows multiple categories 1-5 categories assigned
- Combination of diagnoses and procedures results in lower specificity
- Preliminary specificity analysis suggests > 95% specificity in categorized reports

## UMLS/ SNOMED CT Concept Mapping Results

Over 6000 unique lexicon terms or phrases were successfully mapped to all semantic types in the UMLS/ SNOMED CT database through MetaMap resulting in 5950 pages of mapping results. As described in the method section, a preliminary analysis of 5% of results was produced and preferred concept terms and concept unique identifiers were added to the operative report tuple when full matches were noted, with highlighting of results fields in which only a partial match or no match was achieved (figure 6).

Lexicon Term	CUI preferred phrase	CUI ID
arytenoid swelling with possible vocal cord paralysis	Vocal Cord paralysis	C0042928
atrial mass	no match	
autoimmune enterocolitis, APECED	APECED	C0085859
autoimmune hemolytic anemia, thrombocytopenia, and splenomegaly	Anemia Hemolytic	C0002878
axilla abscess	axilla abscess	C0263115
axilla incision and drainage	axilla incision	
axillary abscess	axilla abscess	C0263115
axillary abscess in the setting of chronic granulomatous disease	axilla abscess, chronic granulomatous disease	C0263 115/C0018203
axillary adenopathy, rule out T-cell lymphoma	axillary lymphadenopathy/ T-cell Lymphoma	C0578735/C0079772
axillary excisional lymph node biopsy	axillary lymph node biopsy	C0398434

Figure 6. Sample Matches with Concept Unique Identifiers (CUI)

Key:  = partial match  =no match

Preliminary analysis (5% of mapped terms/ phrases) revealed that just over 65% of terms have exact matches to concepts within SNOMED CT with appropriate assignment of semantic type to the matched term (figure 7). Twenty five percent of terms have partial matching to SNOMED.



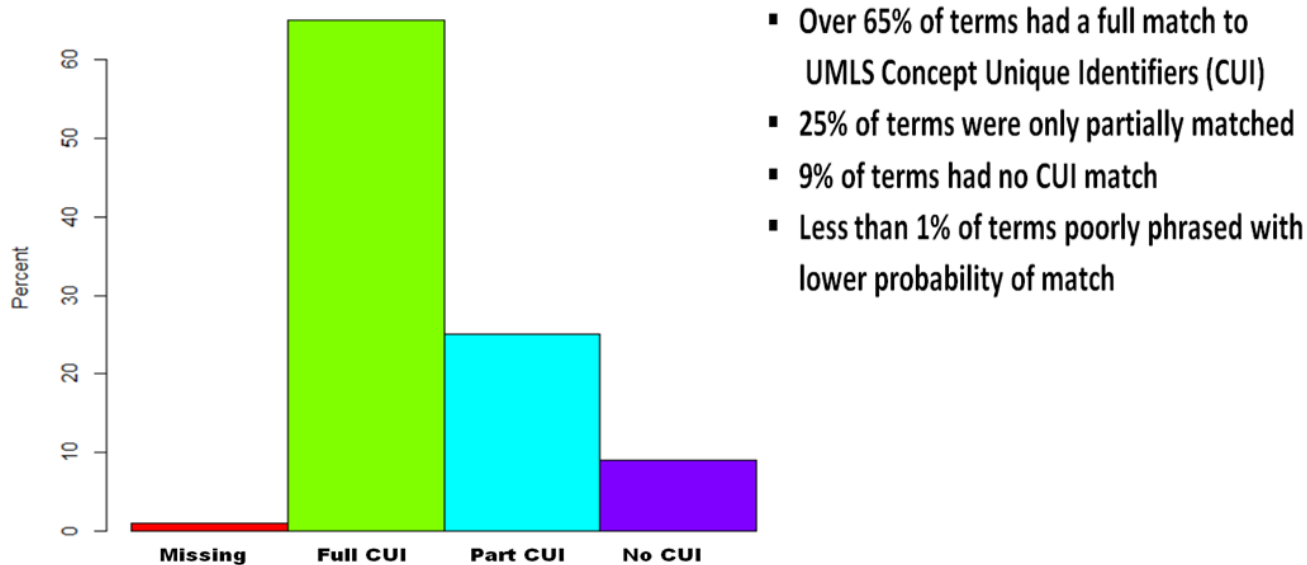


Figure 7. Lexical Terms MetaMap Results

In many of these cases more than one concept was present within a unique lexicon phrase and some but not all concepts were fully matched. In other cases a single concept was partially matched with the correct semantic type, but the match excluded important concept modifiers (e.g. “axillary incision” is the result but an “axillary incision and drainage” was the procedure performed).

Nine percent of terms did not have a SNOMED CT match that accurately described the same concept or semantic type match. For example, “atrial mass” which is a tumor of the atrium of the heart was classified as semantic type anatomic part, “atrium” and another classification for “mass” which is semantic type, finding. Neither of these mappings correctly describe the full concept for this surgical diagnosis so this mapping does not create a useful concept match. Fewer than 1% of terms were incorrectly designated by the surgeon and mapping is considered implausible due to poor choice of

terms (e.g. the surgeon dictated a diagnosis of “acquired absence of breasts” rather than using an accepted term such as bilateral mastectomy).

### **Extraction and External Validation using Veteran Cohort Study Operative Reports**

Results were outputted to an XLSX work sheet with the same format as the BTRIS extraction results. Terms were extracted from 1149 operative reports with 3427 terms retrieved for preoperative diagnosis, postoperative diagnosis and operative procedure. There was again a missing value rate of less than 1% with only 20 empty value fields out of 3447 potential fields. 2024 unique terms were identified and interestingly only 8 of these terms overlapped with the terms harvested from the NIH Clinical Center extraction highlighting the differences in surgical practice type between the two facilities (Table 2).

### **External Validation of Extracted Reports**

External validation was performed by manual review of 120 operative reports (approximately 10%) using the same methods and calculation of precision and recall as that used in the internal validation of BTRIS data. Precision was slightly lower at 98% and recall was again > 99% (Table 1). Although operative procedures were present and accurate in almost all cases, use of “attending surgeon” rather than “surgeon” in the dictated report resulted in addition of the term “attending” to just over 1% of operative procedure terms. This, again is an error that could be easily corrected through a minor modification of the regular expression algorithm.

## EXTERNAL VALIDATION VA OPERATIVE REPORTS

TOTAL OPERATIVE REPORTS	1149
TOTAL TERMS EXTRACTED	3427
TOTAL UNIQUE TERMS	2034
TOTAL MISSING TERMS	20
MANUAL CHART REVIEW	10%
PRECISION	98%
RECALL	99%

**Table 2. External Validation Operative Report Extraction**

### Discussion

Operative reports are a rich source for information on surgical care and the information contained in these reports will be essential to surgical quality improvement and efforts to build a strong evidence base for surgical practices. Almost all operative reports in the U.S. are dictated and transcribed into free text format making data retrieval from operative reports dependent on manual chart review or NLP techniques that have not yet provided comprehensive or high quality operative report data.<sup>5,6,36</sup>

Efforts to introduce structured operative report templates into the electronic health record have been limited and largely based outside the U.S. which may be in part to the

importance of the operative report as a medical legal document in the U.S.<sup>37,38</sup> In a 2014 survey of surgical program directors within the U.S, fewer than 18% of hospitals integrated any structured or “synoptic” operative report templates into the EHR, with the highest use of these templates by Ophthalmology and Obstetrics and Gynecology department where the use was at 30% for very common and routine operations. Operations are very complex and almost every proposal for synoptic or template operative reports leans towards much more customization of templates for specific surgeries, surgical type or by surgeon. This would, in fact, make collection of operative data much more difficult, since customized EHR templates would become more varied than dictation templates or other structures that are currently in use.

Dictation and transcription of operative reports is still the standard practice in the U.S and the 2014 survey of surgical program directors confirmed that this is the method that most surgeons support.<sup>36</sup> Only 19% of program directors supported the use of electronic operative report templates , most often citing a much decreased level of quality in these reports.<sup>36</sup> Directors and surgeons did support the use of templates for dictation of a free text operative report in order to assure that important aspects of the operation are covered in dictated reports and this appears to be the direction of the future.<sup>39</sup>

This research develops a method for extraction of important data related to surgical care from free text operative reports. It suggests the creation of an easy to use tool for the extraction of preoperative, postoperative, and operative procedure information that can be used by local hospitals in order to retrieve information that can be used for

quality improvement, outcomes research and building a local lexicon of surgical diagnoses and procedures. A local lexicon can then be mapped to SNOMED CT in an effort to produce a standardized ontology of surgical concepts that can be used in evaluating outcomes across institutions and in clinical decision support.

### **Significance of the Operative Report Extraction Tool**

Operative reports can be processed by the operative report extraction tool with or without identifiers depending on whether the tool is used for quality improvement or secondary data use. The tool returns data into a table with preservation of the relationships between preoperative and postoperative diagnosis and procedure from the same operative report. This data can be evaluated in a relational data base to evaluate multiple quality outcomes including the outcomes of different surgical procedures for the same surgical diagnosis, the incidence of repeat surgeries by surgeon and patient, and the incidence of outliers in type or numbers of surgical procedures performed. The data is also very useful for cohort retrieval in secondary data use for research. Surgical procedures change and new surgical approaches are often developed with limited evidence for effectiveness. Many surgical approaches that have been in use for some time, particularly surgeries for rarer conditions, also have a limited evidence base to support their use.<sup>21,26,28</sup>

### **Significance of the Operative Report Categorization Tool**

The categorization hash table tool was developed in order to provide actionable (manageable) data from the operative report extractions. The extractions tool is highly

efficient and can produce data from thousands of operative reports within minutes. Surgeries and procedures are most often classified by body system and body part – having a tool that organizes the data in this fashion makes the extracted data more useful by allowing surgical centers to evaluate surgeries to specific areas and for specific types of surgical conditions.

Since this tool categorizes a large percentage ( approximately 95%) of operative reports correctly, the tool may also be very useful in machine learning strategies to predict surgical site. The tool might be used to screen a patient’s clinical records for diagnoses and diagnostic studies and then categorize the results by body part and frequency. Since the tool categorizes with a very high degree of granularity it may even be possible to predict the exact body part that would be the target for surgery. This system may have use in the prevention of wrong site surgeries.<sup>40,41</sup>

### **Significance of Surgical Lexicon Creation**

By creating a local surgical lexicon, this tool also has potential usefulness in improving and expanding existing surgical ontologies. The majority of the lexicon terms mapped to a full match within the SNOMED CT ontology, allowing assignment of UMLS concept unique identifiers (CUI) to local surgical procedures or diagnoses. Expanded use of these CUIs will result in the ability to compare the outcomes of particular surgeries, or the effectiveness of different surgical approaches for the same diagnosis, across institutions creating a much improved evidence base for surgical care. Being able to characterize surgical diagnoses and procedures by their ontologic concept also vastly

expands our ability to use ontologies in surgical clinical decision support and machine learning strategies that can lead to improvement in surgical care.

The MetaMap results from this research also suggest that there still missing concepts with the SNOMED CT ontology. As more of these missing concepts are identified, it may be possible to augment and improve SNOMED CT so that it becomes a more useful and standardized tool with the surgical community. This research demonstrates that the use of MetaMap in mapping surgical lexicon terms is very useful but still quite challenging for use by local surgical departments. Future research will focus on the development of a simplified MetaMap tool for use in identifying CUIs within local surgical data.

### **Limitations of this Research**

1. Although external validation suggests that this tool may be broadly useful, not all institutions may have free text operative reports that are amenable to extraction by these techniques.
2. A proof of concept for the use of the tool in ontology development was explored but in depth ontology development was not pursued in this study.
3. It is not guaranteed that surgical centers will accept this tool especially if it is perceived as a tool to evaluate the quality performance of specific surgeons. Qualitative research will be helpful in exploring surgeon's attitudes towards this tool.

## **Conclusion**

In summary, the operative report extraction tool developed in this research has important potential usefulness in efforts to improve and standardize surgical care. The tool produces a highly efficient extraction of important information from the operative report and preserves the relationships of extracted information within the operative report to further increase its usefulness. The tool also categorizes the extracted data efficiently to enhance usefulness of the data and to make the data results more understandable for potential users. Finally, the tool creates a local surgical lexicon extracting the exact words of the primary surgeon who dictates the report. This captures authentic and high quality data for further analysis. It also allows mapping to existing ontologies so that a more standardized surgical vocabulary may be created in the future, allowing even greater potential for analyzing and improving surgical outcomes.

## **Future Directions**

Future research will be focused on the development of the operative report extraction tool. Research will also focus on the development of a simplified MetaMap mapping tool that may be usable by local surgical centers. Further NLP techniques will be used to extract other important elements from the operative report that are of clinical significance. Finally, qualitative studies should be performed to assure acceptance of this tool by the surgical community.



## Acknowledgements

I would like to thank Dr. Paul Gorman for helpful advice and guidance in the completion of this Capstone Project. I also thank Cynthia Brandt, Chris Lu, Jim Cimino, Stephen Bedrick, Yuli Li and Fabricio Kury, my collaborators in this research.

## References

1. OECD Health Statistics 2013, <http://dx.doi.org/10.1787/health-data-en>. and <http://www.oecd.org/els/health-systems/Health-at-a-Glance-2013.pdf> Accessed 12/01/14.
2. Muñoz E, Muñoz W, Wise L. National and surgical health care expenditures, 2005-2025. *Ann Surg.* 2010 Feb; 251(2):195–200.
3. Elkin PL, Trusko BE, Koppel R, Speroff T, Mohrer D, Sakji S, et al. Secondary use of clinical data. *Stud Health Technol Inform.* 2010; 155:14–29.
4. Garvin JH, DuVall SL, South BR, Bray BE, Bolton D, Heavirland J, et al. Automated extraction of ejection fraction for quality measurement using regular expressions in Unstructured Information Management Architecture (UIMA) for heart failure. *J Am Med Inform Assoc.* 2012 Oct; 19(5):859–66.
5. Wang Y, Pakhomov S, Burkart NE, Ryan JO, Melton GB. A study of actions in operative notes. *AMIA Annu Symp Proc.* 2012; 2012:1431–40.
6. Wang Y, Pakhomov S, Melton GB. Predicate argument structure frames for modeling information in operative notes. *Stud Health Technol Inform.* 2013; 192:783–7.
7. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. *J Am Med Inform Assoc.* 2007; 14(1):1–9.
8. Donabedian A. Evaluating the Quality of Medical Care. *Milbank Q.* Dec 2005; 83(4): 691-729.
9. Donabedian A. The Quality of Care: How Can It Be Assessed? *JAMA.* 1988; 260(12): 1743-1748.
10. Libicki M, Brahmakulam, I. The Cost and Benefit of Moving to the ICD-10 Code Set. RAND TR-132-DHHS. March 2004; ISBN: 0-8338-3585-1. Published 2004

by the RAND Corporation 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138

Available at:  
[http://www.rand.org/content/dam/rand/pubs/technical\\_reports/2004/RAND\\_TR132.pdf](http://www.rand.org/content/dam/rand/pubs/technical_reports/2004/RAND_TR132.pdf)

Accessed 12/4/14

11. Center for Medicare and Medicaid Services Modifier Reference Guide. Available at:<http://www.sccma-mcms.org/Portals/19/assets/docs/Modifier-Reference-Guide.pdf>  
Accessed 12/1/14.
12. Posting on the United States Senate Finance Committee site. Available at:  
<http://www.finance.senate.gov/newsroom/ranking/spotlight/release/?id=fa33d25a-5056-a032-52c2-fc1e7ad33cfe> Accessed 12/2/14
13. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med.* 1998 Nov; 37(4-5):394–403.
14. Cimino JJ, Zhu X. The practical impact of ontologies on biomedical informatics. *Yearb Med Inform.* 2006; 124–35.
15. Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb Med Inform.* 2008; 67–79.
16. Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. *Brief Bioinformatics.* 2008 Jan; 9(1):75–90.
17. Antezana E, Kuiper M, Mironov V. Biological knowledge management: the emerging role of the Semantic Web technologies. *Brief Bioinformatics.* 2009 Jul; 10(4):392–407.
18. Rosse C, Mejino JL Jr. A reference ontology for biomedical informatics: the foundational model of anatomy. *J Biomed Inform* 2003; 36; 478-500.
19. Rubin DL, Dameron O, Musen MA. Use of descriptive logic classification to reason about consequences of penetrating injuries. *AMIA Annu Symp Proc* 2005: 649-53.
20. Burton CV, Kirkaldy-Willis WH, Yong-Hing K, Heithoff KB. Causes of failure of surgery on the lumbar spine. *Clin Orthop Relat Res.* 1981 Jun ;( 157):191–9.
21. Fager CA, Freidberg SR. Analysis of failures and poor results of lumbar spine surgery. *Spine.* 1980 Feb; 5(1):87–94
22. Van der Zwaal P, Thomassen BJW, Nieuwenhuijse MJ, Lindenburg R, Swen J-WA, van Arkel ERA. Clinical outcome in all-arthroscopic versus mini-open rotator cuff repair in small to medium-sized tears: a randomized controlled trial in 100 patients with 1-year follow-up. *Arthroscopy.* 2013 Feb; 29(2):266–73.
23. Bishop J, Klepps S, Lo IK, Bird J, Gladstone JN, Flatow EL. Cuff integrity after arthroscopic versus open rotator cuff repair: a prospective study. *J Shoulder Elbow Surg.* 2006 Jun; 15(3):290–9.

24. Mascarenhas R, Chalmers PN, Sayegh ET, Bhandari M, Verma NN, Cole BJ, et al. Is double-row rotator cuff repair clinically superior to single-row rotator cuff repair: a systematic review of overlapping meta-analyses. *Arthroscopy*. 2014 Sep; 30(9):1156–65.
25. Deyo RA, Gray DT, Kreuter W, Mirza S, Martin BI (Jun 2005). "[United States trends in lumbar fusion surgery for degenerative conditions](#)". *Spine* **30** (12): 1441–5; discussion 1446–7. doi: [10.1097/01.brs.0000166503.37969.8a](#). PMID [15959375](#).
26. Deyo RA, Mirza SK. The case for restraint in spinal surgery: does quality management have a role to play? *Eur Spine J*. 2009 Aug; 18(Suppl 3):331–7.
27. Zindrick MR, Tzermiadianos MN, Voronov LI, Lorenz M, Hadjipavlou A. An evidence-based medicine approach in determining factors that may affect outcome in lumbar total disc replacement. *Spine*. 2008 May 15; 33(11):1262–9.
28. Carragee EJ, Baker RM, Benzel EC, Bigos SJ, Cheng I, Corbin TP, et al. A biologic without guidelines: the YODA project and the future of bone morphogenetic protein-2 research. *Spine J*. 2012 Oct; 22(10):877–80.
29. Pietrobon R, Zaveri A, Cofiel L, Barros J, Shah J. Computational ontologies in orthopaedic surgery. *Clin Orthop Relat Res*. 2010 Oct; 468(10):2612–20.
30. Dao TT, Marin F, and Ho Ba Tho MC. Clinical Validated Computer-Aided Decision System to the Clubfeet Deformities. 31<sup>st</sup> Annual International Conference of the IEEE EMBS. September 2009; 6230-6233.
31. Rossi Mori A, Galeazzi E, Consorti F. An ontological perspective on surgical procedures. *Proc AMIA Annu Fall Symp*. 1996; 115–9.
32. Xu R, Supekar K, Morgan A, Das A, Garber A. Unsupervised method for automatic construction of a disease dictionary from a large free text collection. *AMIA Annu Symp Proc*. 2008; 820–4.
33. Cimino JJ, Ayres EJ. The clinical research data repository of the US National Institutes of Health. *Stud Health Technol Inform*. 2010; 160(part 2): 1299-303.
34. Aronson AR, Lang F-M. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc*. 2010 Jun;17(3):229–36.
35. [http://www.hsrp.research.va.gov/research/abstracts.cfm?Project\\_ID=214169822](http://www.hsrp.research.va.gov/research/abstracts.cfm?Project_ID=214169822) (Accessed 6/17/15)
36. Melton GB, Burkart NE, Frey NG, Chipman JG, Rothenberger DA, Vickers SM. operative report teaching and synoptic operative reports: a national survey of surgical program directors. *J Am Coll Surg*. 2014 Jan;218(1):113–8.
37. Schneider L, Shargall Y, Schieman C, Seely AJ, Srinathan S, Malthaner RA, et al. Design of a consensus-derived synoptic operative report for lung cancer surgery. *Ann Thorac Surg*. 2014 Apr;97(4):1163–8.
38. Spiegle G, Leon-Carlyle M, Schmocker S, Fruitman M, Milot L, Gagliardi AR, et al. Development of a synoptic MRI report for primary rectal cancer. *Implement Sci*. 2009;4:79.

39. Gillman LM, Vergis A, Park J, Minor S, Taylor M. Structured operative reporting: A randomized trial using dictation templates to improve operative reporting. *Am J Surg.* 2010 Jun;199(6):846–50.
40. Kwaan MR, Studdert DM, Zinner MJ, Gawande AA. Incidence, patterns, and prevention of wrong-site surgery. *Arch Surg.* 2006 Apr; 141(4):353–7; discussion 357–8.
41. Van Schoten SM, Kop V, de Blok C, Spreeuwenberg P, Groenewegen PP, Wagner C. Compliance with a time-out procedure intended to prevent wrong surgery in hospitals: results of a national patient safety programme in the Netherlands. *BMJ Open.* 2014; 4(7):e005075.