Annotation of pain and anesthesia events for surgery-related processes and outcomes extraction

Wen-wai Yim  
Palo Alto Veterans Affairs  
Stanford University  
Palo Alto, CA 94305, USA  
wwyim@stanford.edu

Dario Tedesco  
University of Bologna  
Bologna, 40126, Italy  
Stanford University  
Palo Alto, CA 94305, USA  
dariot@stanford.edu

Catherine Curtin  
Palo Alto Veterans Affairs  
Stanford University  
Palo Alto, CA 94305, USA  
ccurtin@stanford.edu

Tina Hernandez-Boussard  
Stanford University  
Palo Alto, CA 94305, USA  
boussard@stanford.edu

Abstract

Pain and anesthesia information are crucial elements to identifying surgery-related processes and outcomes. However, pain is not consistently recorded in the electronic medical record. Even when recorded, the rich complex granularity of the pain experience may be lost. Similarly, anesthesia information is recorded using local electronic collection systems; though the accuracy and completeness of the information is unknown. We propose an annotation schema to capture pain, pain management, and anesthesia event information.

1 Introduction

Post surgical pain continues to be a challenging problem for the health system. Firstly, continued pain after surgery, or chronic persistent postsurgical pain, is common with about 20% of patients having pain long after the wounds have healed (Neil and Macrae, 2009; Kehlet et al., 2006). Secondly, inadequate acute post operative pain control contributes to adverse events such as impaired pulmonary function and impaired immune function (White and Kehlet, 2010). Finally, post surgical pain can be a gateway to addiction, which has taken on increased urgency with the current opioid crisis (Waljee et al., 2017). To improve these problems, it is crucial to have a clear understanding of the patients’ pain and its treatments.

There is some evidence that different interventions such as the use of multi-modal pain management and different anesthesia types, e.g. use of regional anesthesia and nonsteroidal anti-inflammatory drugs, can improve pain management (Baratta et al., 2014). However, different analgesic treatments have different side-effect profiles; moreover, some treatment combinations are not appropriate for certain populations. Furthermore, genetics, age, prior exposure to surgery, and social norms influence the experience of pain. Therefore, there is a clear need to capture anesthesia and pain information and relate them to individual history, social, and genetic factors to improve surgical outcomes.

Even with mandated collection, pain is not always recorded (Lorenz et al., 2009). Even when recorded as structured data, there are a variety of scales that are institution-dependent, e.g. a site-specific 0-10 numeric rating scale or a multi-dimensional questionnaire such as the Brief Pain Inventory. Additionally, it is difficult to capture the rich complex characteristics of pain in structured ways. Anesthesia type, on the other hand, may be recorded or inferred from procedures, medications, or structured input as part of surgery documentation. However, such recording practices differ by institution and local software.

In this work, we present annotation schemas for pain, pain treatment, and anesthesia events for text extraction, as well as report on inter-annotator agreement and corpus statistics. The ultimate goal is to build a new system or adapt an existing system, using this annotated corpus, to automatically extract such information from clinical free text. The extracted data could then be used to complement missing structured information, facilitating greater opportunities for longitudinal study of patients’ pain experience long after initial surgery.
2 Related work

To our knowledge, there is no systematic creation of a pain annotation schema for text extraction, however we reference two extraction systems that identify pain information based on their own targeted needs. (Heintzelman et al., 2013) created a system that extracted pain mentions, severity, start date, end date. Their annotation was based on a created 4-value severity of pain created by the development team. Items were identified using the Unified Medical Language System (UMLS) vocabularies for dictionary look-up (Bodenreider, 2004). Dates and locations were extracted by developed contextual rules. In another work, (Redd et al., 2016) used a series of regular expressions to extract pain score in intensive care unit notes. In contrast to previous works, our work provides a more detailed set of annotations that include different clinical aspects of pain, as well as two other event types (treatment and anesthesia) important for studying outcomes. Similarly, there has not been any work on anesthesia-specific annotation and extraction.

Relating this work to a larger context, our pain, treatment, and anesthesia event annotations can be thought of as more specific reincarnations of the CLEF corpus and i2b2 event annotations (Roberts et al., 2008; Uzuner et al., 2011). For example, under the CLEF annotation schema, pain would fall under the condition entity, with the pain’s location aligning to CLEF’s locus/sub-location/locality schema. Drug, intervention, and negation for conditions are also elements we capture in our annotation schema. Under the i2b2/N/A 2010 concepts, assertions, and relations challenge schema, pain would be considered a medical problem and pain treatments or anesthesia could be identified treatments. Our annotation of status’ are related to assertion and relations between pain and treatment function similarly to their medical problem treatment relations. Pain and treatment annotation can also be compared to medication and adverse drug events, where instead the focus of events are on pain symptoms and treatment concepts (Uzuner et al., 2010; Karimi et al., 2015).

3 Corpus creation

We drew data from two sources (1) Stanford University’s (SU) Clarity electronic medical record database, a component of the Epic Systems software, and (2) MTsamples.com, a online source of anonymized dictated notes. With approval of an institutional review board, we identified a cohort of surgical patients that underwent 5 procedures associated with high pain: distal radius fracture, hernia replacement, knee replacement, mastectomy, and thoracotomy. We focused on three note types: anesthesia, operative, and outpatient clinic visit notes. Anesthesia and operative notes were sampled from the day of surgery, whereas clinic notes were randomly sampled within 3 months prior and 1 year after the surgery. Because of the variation in clinic notes, we performed stratified random sampling per sub-note type and per surgery category.

From MTsamples, we isolated operative (surgery) and clinic visit notes. Clinic notes were considered those not grouped into specialized categories, e.g. surgery, autopsy, discharge. Frequencies by type are shown in Table 1.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Anesthesia</th>
<th>Clinic</th>
<th>Operative</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTsamples</td>
<td>-</td>
<td>90</td>
<td>75</td>
</tr>
<tr>
<td>SU</td>
<td>90</td>
<td>90</td>
<td>75</td>
</tr>
<tr>
<td>TOTAL</td>
<td>90</td>
<td>180</td>
<td>150</td>
</tr>
</tbody>
</table>

Table 1: Breakdown of note types

4 Guideline Creation

Annotation guidelines were created iteratively with a medical general practitioner as well as a biomedical informatics scientist. The initial pain event schema was derived from existing literature (Fink, 2000) and cues from Stanford Health Care’s pain collection practices. Schemas were designed and altered according to feedback from a surgical attendee and an anesthesiologist.

Our annotation focuses on three event types: pain, treatment, and anesthesia events. Below is a description of the entities (in some cases phrasal highlights) for each type of event. Those concepts marked with a * are event heads for which other entities may attach to.

Pain information:

Pain* - indication of pain including signs and symptoms that denote pain or diseases definitionally characterized as pain, e.g. “myalgia”, with attributes Goal:\{binary\} and Status:\{Current, Past, None, Unknown, Not Patient\}

Description - descriptive characteristics of the indicated pain, e.g. “burning”

Frequency - information regarding periodic oc-
Location - location of pain, with attributes Laterality: {Bilateral, Left, Right, Unspecified} and Type: {Abdomen, Ankle, Arm, Back, Back-lower, Back-upper, Breast, Buttocks, ChestArea, Ear, Elbow, Eye, Foot, Generalized, Groin, Hand, Head, Hip, Incisional, Jaw, Knee, Leg, Mouth, Neck, Nose, Pelvis, Shoulder, Throat, Wrist, Other} (This attribute is useful for matching with structured data that pre-specify locations)

Severity - severity of pain, with attributes SeverityAttribute: {0.1...10, mild, moderate, severe}

Temporal - demarkations of time points at which pain occurs, including time relative to events

Treatment - interventions used on patient (see next section for more information)

Trend - trend of pain with attribute TrendAttribute: {Increasing, Decreasing, No change}

Trigger - events that cause some change in pain, with attribute TriggerAttribute: {Increase, Decrease}

Treatment information:

Effectiveness - Effectiveness of treatment with attributes EffectivenessAttribute: {Alleviates, Worsens, No change}
for anesthesia and surgery notes.

An initial set (Set1) included 15 clinic and 15 operative notes from MTSamples; and 30 anesthesia, 15 clinic, and 15 operative notes from SU. Two rounds of revision and agreement were performed on this set. Changes or adjustments to annotation guidelines were made as necessary during annotator agreement cycles. Because clinic notes presented more complexity, we drew another 15 documents from MTSamples and 15 from SU resulting in a new subset (Set2). EffectivenessAttribute and Goal attributes were added from the second set onwards. Two rounds of revisions were performed on this set. Finally, the combined set was revised. The remaining corpus (60 anesthesia, 120 clinic, 120 operative notes) was evenly split and single-annotated by the two annotators. We used brat, a web-based software, for our annotation (Stenetorp et al., 2012).

Inter-annotator agreement (IAA) was evaluated using F1 measure, the harmonic mean of positive predictive value and sensitivity, for entities, relations, and attributes (Hripcsak and Rothschild, 2005). All reported measures are based on partial matches (text spans need only to overlap). For this, relations require that corresponding entity arguments overlap with accurate relation labels.

### 6 Results

Tables 2-6 show final agreement levels for the separate sets of inter-annotator documents and then for the full inter-annotator corpus for the entities, attributes, and relation levels. We also report the frequencies of each field for the full corpus.

For clinic notes, 125 documents had at least one entity, with 19 ± 19 entities, 10 ± 11 relations per non-empty report. Table 7 shows the top 90% of unique co-occurring relation combinations attached to the same pain entity. Most pain entities appeared either without attached relations or with a Location-Arg. For treatment entities not attached to pain entities as an argument (632 entities), 74% had no attachments, 24% were attached to a Temporal-Arg alone, the rest had either an Effectiveness-Arg relation alone or both. Most relations existed within a close context, however a small number did appear at 2 or more sentences away. This included 10% of Trigger-Arg, 7% of Treatment-Arg, 2% of Severity-Arg, and 2% of Temporal-Arg relations. The remaining relations appeared on the same or one sentence away.

Identification of pain and treatment events for clinical notes was relatively challenging. Ten entities with their related attributes, as well as 8 relation types were involved. Moreover, clinical
notes tend to contain unpredictable expressions, e.g. “pain [...] waxing and waning” or “worse with hiking”, and narrative information that spans over several sentences, the conclusion of which could communicate a resolved status. Thirteen out of 613 mentions of pain were attributed as past. Out of 126 marked TriggerAttributes, 114 were aggravating factors (Increase), with only 12 mentions of alleviating factors (Decrease). Interestingly, many severity attributes were qualitative descriptions with 22 for mild, 13 for moderate, and 23 for severe out of 87 total marked. For treatment types, of 654 identified treatment types, 428 were surgical procedures, 116 medication, 82 physical therapy, 12 steroid injection. The remaining had frequencies of 3-5 each.

Ideologically, there were nuances to annotating pain information. While the easiest references to pain were trivial, e.g. “pain”, some required referencing dictionaries, e.g. myalgia, or reading context, e.g. discomfort. Distinguishing between cause of and timing for pain was not always clear. For example, in “pain is worse in the morning” and “pain [...] when running”, both underlines could be considered as either Trigger or Temporal. Our final decision was to mark as a Trigger when believed to be causal of the pain rather than delineating chronology. Some pain attributes had multiple connotations. For example, “chronic pain”, defined as presence of pain for longer than 3 months, has both a duration and frequency context. We decided to assign chronic as a description attribute. Extent of decisions were specified in annotation guidelines. Finally, there are unavoidable limitations in text interpretation. For example, in “patient is very tender to palpation”, very may be normalized to moderate or severe based on annotator subjectivity. Furthermore, pain may be suggested but not explicitly stated, e.g. “woman [...] with [...] debilitating abdominal wall hernias” (most likely painful), and therefore not captured.

Anesthesia and operative note entity agreement was at 0.923 F1 and 0.934 F1. There was a total of 235 and 254 entities for anesthesia and operative notes. For anesthesia reports, 72 had at least one entity, with 4 ± 5 entities each; operative reports, 130 had at least one entity, with 2 ± 1 entities each. 15% of Pre-incisional intervention entities were marked as Planned for anesthesia reports; 1% for operative reports. Agreements for operative and anesthesia entities and attributes were high (Table 5 and 6). This is due to the focused nature of these domains. However, our annotation schema did not include implicit references, e.g. “skin was anesthetized with 1% lidocaine solution” where lidocaine is often used for local anesthesia.

To improve IAA, further annotation would benefit from pre-annotation of entities trained on this starting set. This would increase consistency and throughput. Additional annotation of a larger corpus would provide larger samples sizes to estimate task challenge for less populated classes.

7 Conclusions and Future Work

In this work, we present a rich annotation schema for pain and pain interventions, as well as an annotation categorization for anesthesia types. Although this work was developed in the surgical setting, the pain annotation schema presented here can be adapted for other settings. Future work includes building our extraction system and applying these data to assess important patient outcomes and health services research.

Annotation guidelines and the MTSamples portion of our corpus is available through our group’s website (med.stanford.edu/boussard-lab.html).

Acknowledgments

This project was partially funded from the Veterans Affairs Big Data Scientist Training Enhancement Program. This project was supported by grant number R01HS024096 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality or the Department of Veterans Affairs.
References


