

Narrative vs structure: Why do we have to choose?

Chris Tackaberry

CEO Clinithink

Inspiration from other verticals....



Healthcare delivery challenge

In pursuit of improved outcomes, HIT systems and data help us....

- Reduce risk, manage uncertainty
- Reduce uncertainty, manage risk
- Live with risk, cope with uncertainty
- None of the above

Quotable quotes

- “My patients are sicker than everyone else’s”
- “We don’t really know what the risk factors are, the answer is in the data but we have no way of pulling it out or even trending it”
- “Symptoms, signs and clinical uncertainty is what I deal with everyday, not codes”
- “If you want me to see the same number of patients [or even more] don’t give me click-itis
- “Physician productivity dropped when we implemented our EMR, but is OK, the administrators re-baselined productivity”
- “At our facility allergies are recorded in a number of different systems as codes or free text – how are we going to meaningfully use that?”
- “Its OK, I spoke to our EMR implementation consultant, she said we can configure a template to capture that stuff – not quite sure how we’ll report on it though”
- “ I want to re-purpose the data we have collected in our EMR and make it work for me”

The problem

- Reporting and coding obligations on providers are becoming more onerous
- Pick something important, measure it, make it better
- In the US, speech-to-text technologies and services account for approx 60% of data input to electronic records (Source: CIDA)
- Narrative style is ingrained in physician practice, culture and thought process: Forms-based transactions are limited and limiting
- Knowledge crisis
- Where is the bar height for precision and certainty?
- Consumer technology is invading the enterprise and driving user expectation

Technology is getting very smart and very plastic (makes the user expectation problem worse)

- Shazam
- Goggle
- Swype for Android
- Tripit
- GPS

All these technologies “guess” through pattern recognition

- Google search experience/Adwords
- Wolfram Alpha, Ask Jeeves, etc
- iPhone contact management

These technologies are highly context-aware, powerful, ubiquitous with simple UX

Bounding the problem

- The Airbus A380
 - has about 4 million parts, with 2.5 million part numbers produced by 1,500 companies from 30 countries (Source: Airbus)
 - There is only one way to describe each part
- “Left fractured radius and ulna”
 - Snomed:
 - *1.4M terms*
 - *compositional grammar*
 - *modifiers, qualifiers and primary concepts*
- **There are probably 20 different legitimate narrative ways of saying “Left fractured radius and ulna”**
 - Rector AL, Nowlan WA, Kay S. Foundations for an electronic medical record. *Methods Inf Med.* 1991 Aug;30(3):179–186

EMR vs Narrative: the wrong question

- Applications which force highly structured data capture
 - produce high value, high quality analytics for specific purposes
 - are expensive to implement
 - may impact front line productivity
 - are often un-popular with clinicians and popular with HIM staff
- Applications which allow free text/narrative data entry
 - are popular with clinicians and a non-starter for HIM staff
 - are less likely to impact frontline productivity
 - are cheaper to implement
 - produce low-value, low quality data for analytics
- Machine-readable free text in correspondence/notes/memos etc is
 - abundant and increasing
 - potentially highly valuable if it could be structured and analysed

Narrative and structure

“slight cough. violent sneezing. very sore eyes both sides. Eczema in childhood. Mother has asthma, no FH atopy otherwise. OE: Moderate conjunctival injection on left. Tympanic membranes pink both sides. Impression: probably hay fever. Possibly has trivial left AOM too.”



History

- ▶ Cough
 - Slightly
- ▶ Sneezing
 - Severe
- ▶ Sore eye
 - Right and left
 - Severe

Past History

- ▶ Eczema
 - Childhood

Family History

- ▶ Asthma
 - Mother
- no** Family history: Atopy

Observations and Findings

- ▶ Conjunctival injection
 - Left
 - Moderate

Assessment

- ▶ Hay fever
 - Probably present
- ▶ Acute left otitis media
 - Known possible
 - Mild

NLP – why now?

- Linguistics – POS recognition
- Horsepower
- Reference terminology
- Brute force pattern matching
- Moons aligning

So what?

- Recognize context – CDS triggers, knowledge management

Outcomes

- Aggregate analysis of unstructured data
 - Care quality
 - Research
 - Clinical trial candidate identification

Quality cycle

- Amplifies the power of other new techniques like machine learning

Surveillance and prevention

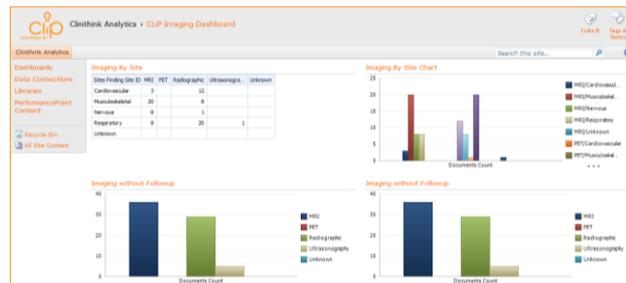
Understanding aggregate data

IMPRESSION:

Essentially normal MR of the lumbosacral spine for a patient of this age. There is no explanation for radicular pain.

Assessment

- Magnetic resonance imaging
 - Lumbosacral spine
 - Normal
 - Radicular pain



NLP creates structured meta-data

Analytics dashboards populated

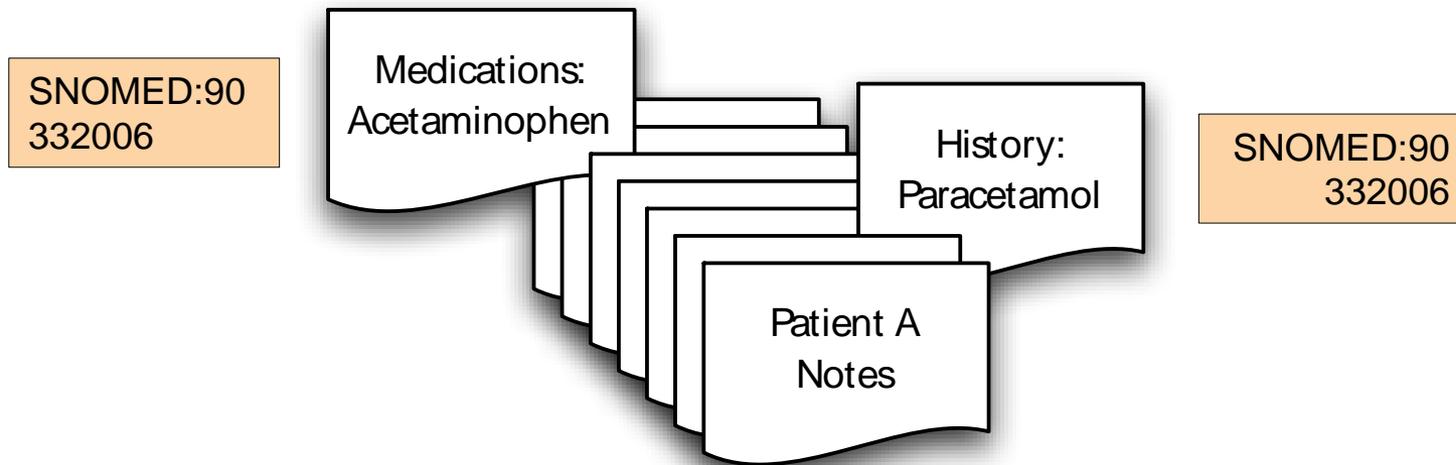
Machine learning primer

- Brute force + a TON of data
- Machine learning **classifier**
- Calibrate/Validate
- Forecast the weather increasingly accurately
- Empirical, data-driven not regression analysis (v hard to swallow)

Adverse Event Triggers

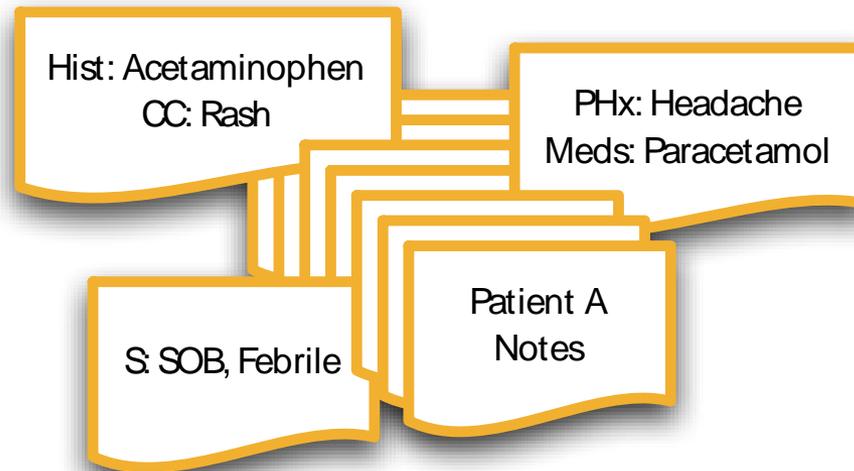
- Current IHI triggers identifying adverse events are generally based on identifying individual instances of coded events, values
 - ADT, laboratory, pharmacy data
- Rudimentary use of clinical notes in existing triggers
 - Information documented in clinical narrative may be very different than that captured structured data, and may indicate warning signs on a patient before lab
 - Some triggers look for exact string matches, but these are unlikely given freedom of expression of medical English
- ADE Triggers by J Rozich et al
 - Many identify only single fact*
 - Diphenhydramine
 - *Many legitimate uses, very low specificity test*
 - Abrupt medication stop
 - *Hard to define “abrupt” programmatically without additional information*
 - *Again, many possible reasons other than ADE*

Opportunities with Concept Identification



- Many ways to express most medical concepts
 - String search vastly limited even for single concepts
 - Virtually impossible for complex, multi-concept arguments
- Coded concepts may be queried regardless of vocabulary chosen

Machine Learning-Based Discovery



- Access to coded concepts facilitates much more accurate automated assessment
- Input: indexed clinical notes to machine learning classifier
- Classifier generation with ADE patient corpus: learn combinations of features indicative of ADEs
- Apply to wider patient populations to identify at-risk individuals

Summary

- Patients and their physicians tell and record stories
- Innovative narrative processing technologies are emerging
- Benefits in outcomes, care quality, regulatory, exchange and surveillance
- Next gen EMR systems **must** support range of structure in input

narrative data is a valuable asset!