

Making the Best Use of Textual ED Data for Syndromic Surveillance



Debbie Travers, PhD, RN, FAEN, CEN

Associate Professor, Health Care Systems & Emergency Medicine

Faculty, Carolina Health Informatics Program

University of North Carolina at Chapel Hill

Text Processing Tools

- What the tools do
- How to obtain tools
- How to implement them

Chief Complaints

- Challenges
- Tool developed at UNC

A Visit to the ED

Chief Complaints- Influenza

Cough	Congestion	Sob
Prod cough	Congest	Short of brth
Uri sx	Fever	Dyspnea
URI symptoms	Febrile	Diff breath
URI?	Temp	Difficulty breathing
Flu?	High temp	Resp distress
Flu sx	Temp 102.3	Sat low
Flu-like illness	Fvr up to 102.3	Sat 89%

Chief Complaint: *Primary reason for visit*

Overwriting the CC



N/V/dizzy

Overwriting the CC



Syncope

Overwriting the CC



MI

Overwriting the CC



410.9

(myocardial infarction, acute)

Documenting Fever in CC: Varies

- Multiple complaints
 - CC fields- some ED systems limited to one
 - Prioritization- **CHIEF** complaint (not entire list)
- What is a fever?
 - Peds: 38C (100.4), high fever 39C (102.2)
 - Clinicians
 - “Did you measure it at home?”
 - “What route? How long did you measure it?”
 - Patient
 - “My norm is 97 so when temp is 99 I’m febrile”

Using Chief Complaints (CC) for Secondary Purposes

such as Syndromic Surveillance

- No standard terminology
 - Until standard adapted/developed & widely adopted, must use messy text

“fever”

Simple pattern matching

Fever for two days	→	fever
Denies fever	→	fever
feb sz	→	no fever
fevet	→	no fever
temp 102	→	no fever

ED Chief Complaint Conference

- **Consensus Conference, 2006**
 - Baltimore, before ISDS annual conference
- **40 stakeholders**
 - Emergency Medicine/Nursing
 - Clinicians, researchers, administrators, quality/surveillance
 - ED system vendors
 - ISDS community
- **10 recommendations**
 - Controlled vocabulary needed
 - All user communities' needs should be addressed
 - Capture patients' words and clinicians' terms

Haas et al., 2008 Toward Vocabulary Control for Chief Complaint, Academic Emergency Medicine

National Standards- ED Data

- DEEDS-
 - **Data Elements for Emergency Department**
- DEEDS Version 3 creation
 - *McClay et al, 2015, JAMIA, Standard for improving emergency information operability: the HL7 data elements for emergency department systems*
- HL7 Emergency Care Workgroup
- 525 concepts, 200 lab tests (LOINC)
- Foundation for the adoption of standards for ED data

DEEDS Update

- **HL7 V3 DEEDS specification**

[https://www.hl7.org/documentcenter/public temp E8692EE7-1C23-BA17-0CE4E039544CD500/wg/emergencycare/HL7%20V3%20Specification%20DEEDS%20Ballot%20Document%20-%20Abbreviated.pdf](https://www.hl7.org/documentcenter/public_temp_E8692EE7-1C23-BA17-0CE4E039544CD500/wg/emergencycare/HL7%20V3%20Specification%20DEEDS%20Ballot%20Document%20-%20Abbreviated.pdf)

- **Data elements**

- Reason for Visit

- The basis, or purpose, for an encounter in an ED setting (may be a process or procedure)

- Presenting problem

- Clinically relevant concept describing the healthcare professionals' interpretation of the cause creating the health issues promoting the patient to seek care.

- Review of symptoms, Physical examination/findings

- Multiple data elements

Approaches to Using CC Data

Natural language processing*

- Accessing data in the form of narratives or free text
- Creating machine-understandable interpretations
- NLP methods in use for CC data for surveillance#
 - Keyword
 - Statistical
 - Linguistic processing

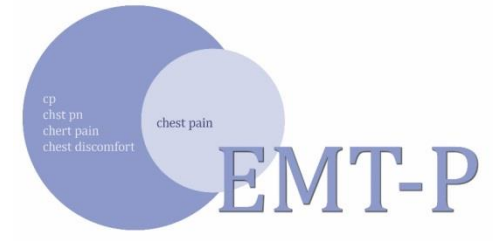
* *van Bommel & Musen, Handbook of Medical Informatics, 1997.*

Conway, Dowling & Chapman, JBI, 2013

Emergency Medical Text Processor (EMT-P)

- Natural language/text processing system
- Extracts standardized terms from ED CC text
- Utilizes Unified Medical Language System® (UMLS®)
- <http://nursing.unc.edu/research/emtp/>

Development, Testing



- Developed* & Validated#
 - Sublanguage analysis, 39,038 ED visits
 - Validated on 203,509 ED visits
 - Expert panel review of output, 96% accurate
- Refined for surveillance
 - Biosense-funded project
 - Review of statewide ED data (NC DETECT)
 - Validated for syndromic surveillance

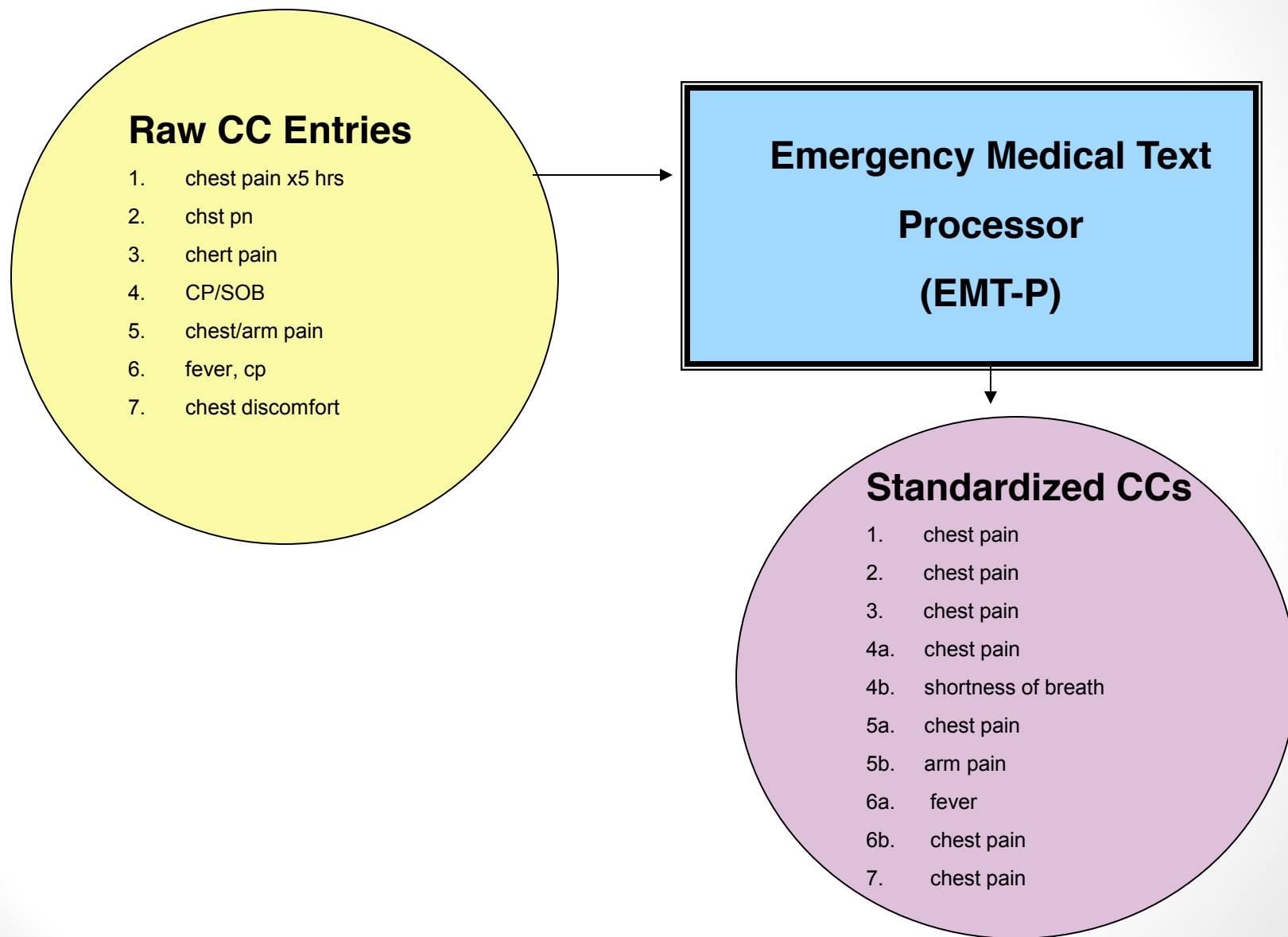
*Travers & Haas, J Biomed Inform, 2003

Travers & Haas, Acad Emerg Med, 2004

Emergency Medical Text Processor (EMT-P)

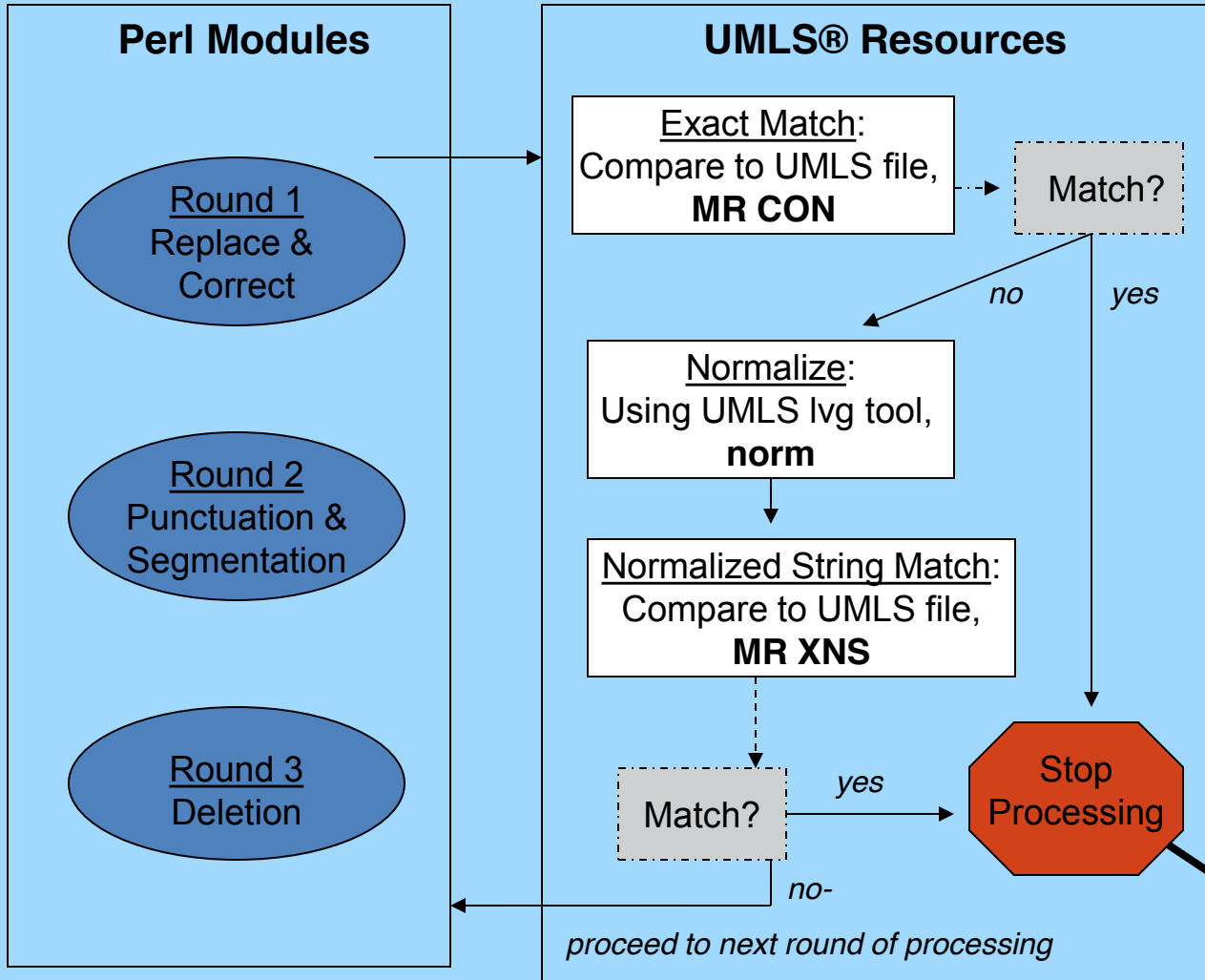
- Used as a pre-processor
 - For syndromic surveillance
 - After pre-processing, classifier applied to data
- NLP Tools
 - Linguistic processing
- Modules
 - Simple → complex
 - Stops processing when CC matches standardized term
 - Clean CCs
 - Replace acronyms, abbreviations, misspellings, truncations
 - Expand coordinate constructions & other syntactic structures
 - Map cleaned CCs to standard concepts
 - from Unified Medical Language System® (UMLS®)

Figure 1- Overview, Cleaning ED Text Entries

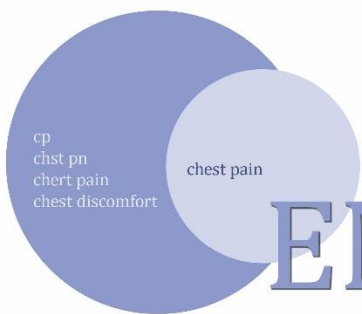


Raw
CC
entries

Emergency Medical Text Processor (EMT-P)



Standardized
CCs



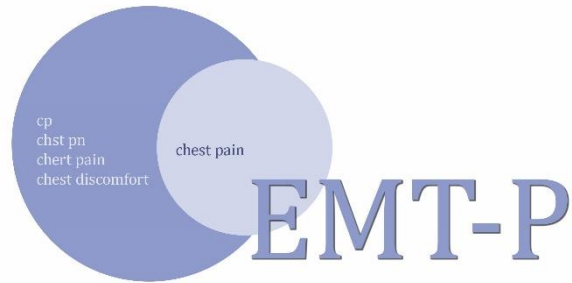
EMT-P

System Inputs/Outputs

Original CC entry	Processed CC(s)	UMLS Output Concept ID (CUI)
fevet feb sz fvr Temp 102	fever	C0015967
SOB SHOB dyspnea	shortness of breath	C0392680
sore throat/chest	sore throat sore chest	C0242429 C0008031
aches all night	myalgia	C0027121

EMT-P Users

- Boston Public Health
- NC DETECT
- Arizona Department of Health
- Researchers
- Incorporated into EMT-C triage tool



To obtain

**iBridge Network
Kauffman Foundation**

[http://www.ibridgenetwork.org/unc/
emergency-medical-text-processor-
version-2-3](http://www.ibridgenetwork.org/unc/emergency-medical-text-processor-version-2-3)

Triage Notes

- Challenges
 - For secondary use
- Tool developed at UNC

A Visit to the ED

CC (100% visits)	Triage Note* (27% visits & increasing)	Syndrome
Fever	amb c/o yest fever 102, n/v.today just general aches	Gastrointestinal
Fever	27 yo male, c/o cough and SOB x1 wk. Denies wheez. Yellow sputum.	Respiratory
Fever	Rash on chest, spread to back, abd & neck. Fever today with back pain. Aches.	Fever Rash

Triage Note: *More detailed description of history of present illness*

EMT-C System

- Developed at UNC in 2011-13
- Pre-processor & classifier
 - Pre-processing with EMT-P modules (NLP)
 - Machine learning system
 - Vector space methodology
 - Key component: Master Term List

Emergency Medical Text Classifier (EMT-C)*

- Extraction of Gastrointestinal (GI) symptoms
 - “Finding the needle in the haystack”
 - 1,000 ED visits per day
 - ~2% with GI symptoms
- Methods
 - Uses NLP and machine learning tools
 - To identify visits with high likelihood of acute GI infection

Machine Learning

- Train system to identify patterns
- Requires gold standard training set
 - Syndromic surveillance- gold standard
 - Record is syndrome negative or positive
 - Ways to identify gold standard rating
 - Manual review of sample
 - Diagnosis, laboratory results
- Relevance feedback
 - Manual or automated
 - Refines the system to improve accuracy

Emergency Medical Text Classifier (EMT-C)

Before EMT-C	After EMT-C
amb c/o yest fever 102, n/v.today just general aches	Ambulatory Fever, nausea vomiting yesterday Today general aches
27 yo male, c/o vom,diarea x2 days. Denies temp. h/a too.	27 year-old male Vomiting, diarrhea for 2 days Denies fever Headache

EMT-C Evaluation- Informatics Lab

- Evaluation on 3353 records
 - Weighted to N=2,418,167
 - Gold standard- manual review by experts

	Sensitivity	Specificity
Baseline	0.28	0.97
EMT-C v.1.1	0.89	0.81

- Measured accuracy
 - **Sensitivity**- *If EMT-C rates the record syndrome negative, it's probably accurate (but there will be false pos ratings)*
 - **Specificity**- *If EMT-C rates the record as syndrome positive, it's probably accurate (but there will be false neg ratings)*

EMT-C Evaluation- Production

- Focus of grant
 - Real-world implementation, not just in lab
- Challenges- production system
 - small set sizes
 - missing and sparse data
- User input into system
 - Burden of false positives --too great
 - Increased workload not feasible
 - EMT-C is adjustable- false positives, negatives
 - Favored **Positive Predictive Value (PPV)** over Sensitivity
 - PPV: *what is the probability that the patient meets the syndrome criteria?*

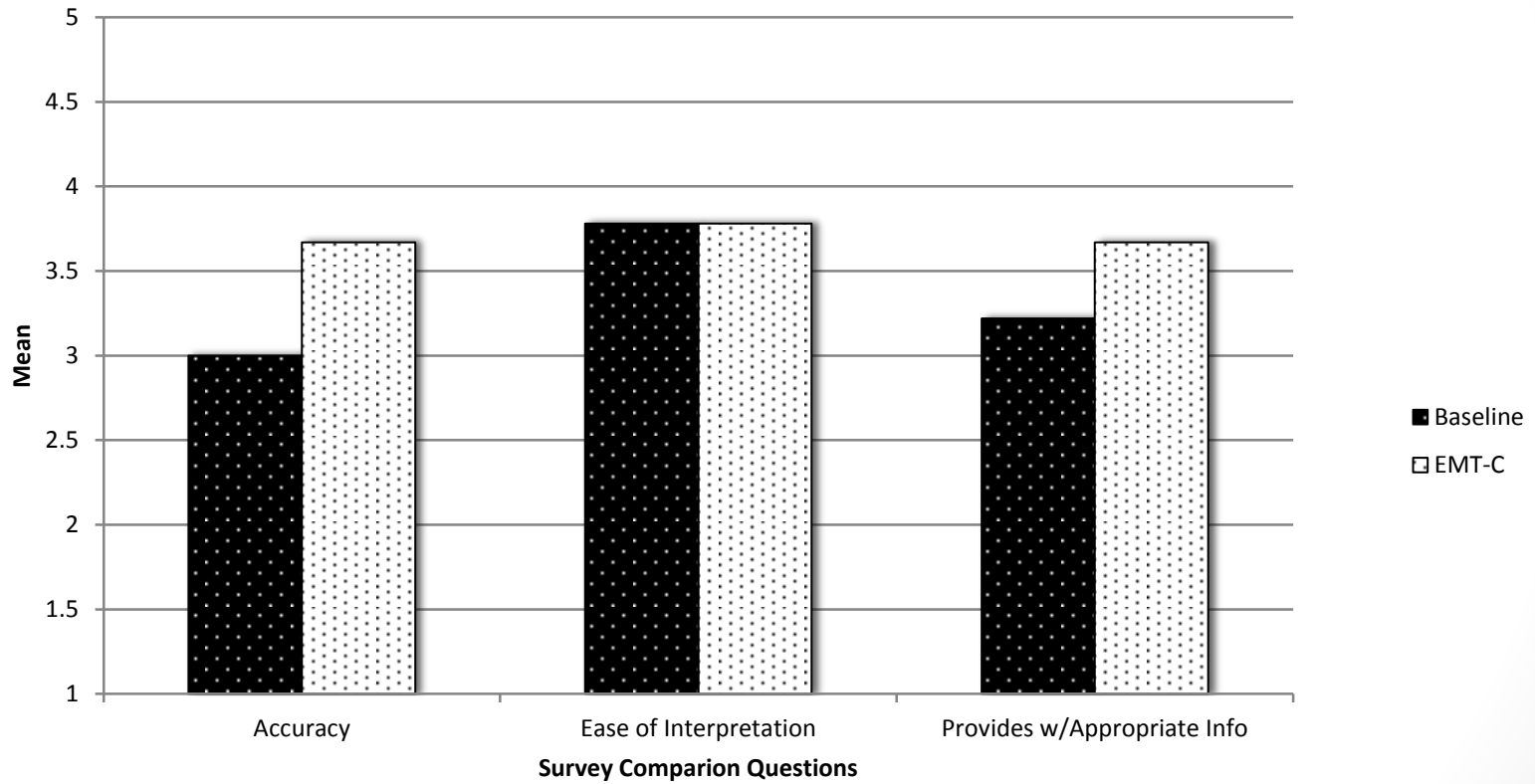
EMT-C Evaluation- **Production**

- N=500 records, weighted to N=273,409
- Gold standard- manual review by expert (with access to diagnoses)

	Sensitivity	Specificity	Positive Predictive Value
Baseline	0.69	0.80	0.46
EMT-C v.1.1	0.79	0.83	0.53

User Evaluation

Pre-Post Survey Comparisons



False Positives

CC	Temperature	Triage Note	Diagnoses
N/V/D; FEVER	<i>None listed</i>	<i>None listed</i>	172.4 - MALIGNANT MELANOMA OF SKIN OF SCALP AND NECK 787.01 - NAUSEA WITH VOMITING
FEVER HEADACH VOMITING	<i>None listed</i>	<i>None listed</i>	311 - DEPRESSIVE DISORDER NOT ELSEWHERE CLASSIFIED 616.4 - OTHER ABSCESS
FEVER	38	2yo boy w/ broviac, J-tube significant gastrointestinal history who presents with fever for 3 days responsive to Ibuprofen and Broviac that has problems drawing and flushing for 3 days. pt is on chronic TPN. Has been sleeping more the last few days. Central line repaired 5 days previously.	<i>Not available at time of manuscript submission</i>

Take-Home Messages

- No perfect syndromic classifier
 - Some false positives unavoidable
 - Sparse/missing data a problem
- Machine learning techniques
 - Show promise for syndromic surveillance
 - Master term list vital

Take-Home Messages

- Users don't want a large percentage of false positives
 - Samoff* & colleagues
 - <25% of signals required response
 - Users need syndromic surveillance system oriented toward
 - **Efficiency (not sensitivity)**

**Samoff E, Waller AE, Fleischauer A, et al. Integration of syndromic surveillance data into public health practice at state and local levels. Public Health Reports. 2012 May-June;127(3):310-317.*