A National Web Conference on the Use of Natural Language Processing (NLP) to Improve Quality Management

Presenters:
Brian Hazlehurst, PhD
Alexander Turchin, MD, MS

April 11, 2012
Moderator:
Rebecca Roper, MS, MPH
Agency for Healthcare Research and Quality

Presenters:
Brian Hazlehurst, PhD
Alexander Turchin, MD, MS

There are no financial, personal, or professional conflicts of interest to disclose for the speakers or myself.
Automating Assessment of Asthma Care Quality

Brian Hazlehurst, PhD
Senior Investigator
Kaiser Permanente Northwest Center for Health Research
Quality of care in the U.S. health care system is unacceptably low (IOM, *JAMA* 1998)

“...Serious and widespread quality problems exist throughout American medicine. These problems...occur in small and large communities alike, in all parts of the country, and with approximately equal frequency in managed care and fee-for-service systems of care. Very large numbers of Americans are harmed as a result....”
On average, Americans receive about 55% of recommended medical care processes.

A key component of any solution is the routine availability of information on care delivery performance at all levels.

- Automated, comprehensive, care quality assessments
- The EMR could make possible automated assessment of care, eliminating sampling, surveying, and manual review of charts
A System for Automated, Comprehensive, Quality Measurement

Clinical Guideline

Population selection
Data element selection
Encounter-based extraction

Data standardization
Site-specific data transforms

Terms, Concepts, Rules, Operating Params, for specific application

Application-specific extraction/filter

Measures operationalized in terms of temporarily located numerator and denominator events

EMRAdapter

MediClass

Postprocessor

Measure Implementation

Data Warehouse

CDA (XML)

CDA w/ MC concepts (XML)

EventStream (Flat file)

Quality Measures

Data Extraction

Concept Markup

Quality Measurement
MediClass—A MEDical Record CLASSifier

1. Takes in encounter record (CDA) and marks up each data section with identified clinical concepts.

2. Identifies concepts within text notes (using NLP algorithms) and coded elements of each encounter record.

3. Uses rules defining logical combinations of concepts to infer additional clinical events (classifications) of interest.

<table>
<thead>
<tr>
<th>Quality Measure</th>
<th>Denominator criteria [Index Date]</th>
<th>Numerator criteria [Measure Interval]</th>
<th>Operationalization Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients with the diagnosis of persistent asthma should have a historical evaluation of asthma precipitants.</td>
<td>Patients with persistent asthma [PA Qualification Date]</td>
<td>Patients with a subjective evaluation of precipitants or triggers [observation period]</td>
<td>Probably only found in the text progress notes.</td>
</tr>
<tr>
<td>Patients with the diagnosis of persistent asthma should have spirometry performed annually.</td>
<td>Patients with persistent asthma [PA Qualification Date]</td>
<td>Patients with orders for PFTs or documentation of office spirometry or of PFT results [subsequent 12 months]</td>
<td>Numerator satisfied with documentation of referral to pulmonary specialist if no PFT known available.</td>
</tr>
<tr>
<td>Patients with the diagnosis of persistent asthma should have available short acting beta-2 agonist inhaler for symptomatic relief of exacerbations.</td>
<td>Patients with persistent asthma [PA Qualification Date]</td>
<td>Prescription for a short acting beta-2 agonist to use PRN [subsequent 12 months]</td>
<td>Numerator satisfied if prior/existing active Rx; also combination Rx (i.e., Combivent) or oral/nebulized PRN Rx will count. Exclusion if adverse reaction to b-agonists.</td>
</tr>
<tr>
<td>All patients seen for an acute asthma exacerbation should have current medications reviewed.</td>
<td>Patients with persistent asthma meeting criteria for outpatient exacerbation [Exac. Encounter]</td>
<td>Documentation that medications reviewed by provider [same visit]</td>
<td>Numerator satisfied if provider documents asthma specific medication history in notes or active management of current medication list.</td>
</tr>
</tbody>
</table>
Clinical Events Dataset File (portion)

Patient segment:
- pat1, KPNW, 19xx, F, 'White', '?', 'N', prov1, OBGN, 'loc1', Outpatient, enc1, SmokeAsk, RFV, 20010109
- pat1, KPNW, 19xx, F, 'White', '?', 'N', prov1, OBGN, 'loc2', Outpatient, enc2, Smoker, Smk, 20010109, 'Y'
- pat1, KPNW, 19xx, F, 'White', '?', 'N', prov1, OBGN, 'loc2', Outpatient, enc2, SmokeAsk, Smk, 20010109
- pat2, KPNW, 19xx, F, 'White', '?', 'N', prov2, ONC, 'loc3', Outpatient, enc3, MedsReview, Note, 20010110, 'Main Note,
- pat2, KPNW, 19xx, F, 'White', '?', 'N', prov2, ONC, 'loc3', Outpatient, enc3, ChestExam, Note, 20010110, 'Main Note,

.....

.....
pat3, KPNW, 19xx, M, '?', '?', '?', prov3, FP, 'loc4', Outpatient, enc4, MedOrd, cat2, 20010111, 17, 1, '00172-4390-18', '2P PO Q4-6H PRN', 'ALBUTEROL AER 90MCG', 'INHALATION',
pat3, KPNW, 19xx, M, '?', '?', '?', prov3, FP, 'loc4', Outpatient, enc4, MedOrd, cat6, 20010111, 20, 1, '00179-1228-20', '4P PO BID', 'AZMACORT INHALER', 'INHALATION',
pat3, KPNW, 19xx, M, '?', '?', '?', prov3, FP, 'loc4', Outpatient, enc4, PeakFlow, Note, 20010111, 'Main Note',
pat3, KPNW, 19xx, M, '?', '?', '?', prov3, FP, 'loc4', Outpatient, enc4, AsthmaVisit, Note, 20010111, 'Persistent', 'Main Note',
pat3, KPNW, 19xx, M, '?', '?', '?', prov3, FP, 'loc4', Outpatient, enc4, MedsReview, Note, 20010111, 'Ancillary Note',
pat3, KPNW, 19xx, M, '?', '?', '?', prov3, FP, 'loc4', Outpatient, enc4, AsthmaVisit, Dx, 20010111, 493.90, Pat3, KPNW, 19xx, M, '?', '?', '?', prov3, FP, 'loc4', Outpatient, enc4, SmokeAsk, Note, 20010111, 'Main Note',
pat3, KPNW, 19xx, M, '?', '?', '?', prov3, FP, 'loc4', Outpatient, enc4, Fluvac, Immun, 20010111, 'Done'

.....

.....
The Clinical Events Necessary to Identify “Persistent Asthma”

- Meets any of the following within any 12-month window during qualification period
  - Four “fills” ordered of asthma-specific meds
  - Two “fills” ordered of asthma-specific meds and four outpatient visits coded with asthma Dx
  - Asthma-related ED visit or hospitalization
  - Provider notation that patient has persistent asthma
  - Provider use of “home grown” persistent asthma Dx code
Quality Profile for Patient “X”

- Chest exam performed
- Spirometry ordered or discussed
- Medications reviewed
- Asthma med ordered
- Asked about smoking
- Asthma visit
- Asthma exacerbation
- Persistent asthma notation

Timeline:
- 1/1/01
- 1/1/02
- 1/1/03
- 12/31/03

Persistent Asthma and Exacerbation markers are indicated.
Study populations identified (>12 y.o. with an asthma visit within 3-year observation window)

- Mid-sized HMO (“HMO”)
  - Multiple observation windows in 2001–2008 period
  - Roughly 35,775 study patients per window; 14,000 with persistent asthma

- Consortium of FQHC (“SafetyNet”)
  - Eight orgs with the EMR installed in 2005–2008 period
  - Single observation window (all data available)
  - Roughly 6,880 study patients; 1,800 with persistent asthma
More ACQ Findings

- 22 Outpatient asthma measures identified
  - 18 (80%) were implemented
  - 11 for routine care, 7 for exacerbation care
  - 4 (20%) will require additional effort to implement
    - 2 relied on complex assessment of “control”
    - 2 relied on knowing patients baseline PFT values

- 8 of the 18 (37%) require processing clinician’s text notes, another 7 measures (32%) are enhanced by this processing because the text notes provide an important alternative source for the necessary numerator clinical events
  - In addition, qualification for any measure in the ACQ measure set (as persistent asthma) occurred by text-based assessment in 26% of all patients. Of these, 30% qualified as persistent by text processing alone.
Most ACQ measures performed relatively well in the HMO healthcare system.

- Measure accuracy (agreement with chart review) ranged from 63% to 100% and averaged 88% across all measures (95% CI = 82%, 93%).
- Mean sensitivity was 77% (CI=62%, 92%), and was 60% or greater for 15 of the 18 measures (and 90% or greater for nine of those).
- Mean specificity was 84% (CI=75%, 93%) with 15 measures having specificity of 60% or higher (nine with 90% specificity or greater).
- There were two measures for which specificity was over 90% but which had poor sensitivity.
The automated ACQ analysis was less accurate against the SafetyNet health care system (however, across the evaluable measures at each health care system, specificity was similar with 9 of 16 measures reaching 90% or better)

- Mean overall accuracy was 80% (95% CI=72%, 89%) and ranged from 36% to 99% across all measures
- Mean sensitivity was 52% (95% CI=35%, 69%)
- Mean specificity was 82% (95% CI=69%, 95%)
- Performance was better among the routine measures compared to the exacerbation-related measures
Overall Results of Asthma Care Quality Measurement

- Overall we found that persistent asthma patients received 48.3% (95% C.I. [48.1, 48.5]) of recommended care on average across 166,606 retrospective care evaluations extracted from two electronic medical record systems
  - routine care was higher at 48.8%
  - acute exacerbation care was lower at 26.6%

- Care within SafetyNet system had somewhat lower quality scores compared to the HMO across all groups
  - routine care 42.1% vs. 50.3% of recommended
  - exacerbation care 22.6% vs. 27.1% of recommended
Exacerbations 12 to 24 months post-qualification as “persistent asthma”

Mixed results
- Routine care measures (e.g., evaluation of triggers, flu vaccination, tobacco evaluation) predict WORSE outcomes
- Exacerbation care measures (e.g., meds review, chest exam, spirometry) predict BETTER outcomes

Continue to work to sort out confounding by patient severity
We have generalized this approach and are applying it to assessing obesity treatment (as prescribed by the NHLBI guideline)
- R18 study funded by AHRQ

We are halfway through a 3-year project called the CER HUB, which makes this technology available through a central website hosting research projects that use it
- RO1 project that includes a network of six health systems
- Conducting two CER studies in Asthma Control and Smoking Cessation counseling
CER Hub

Informatics Tools for Evaluating Health and Healthcare

CER Hub is a web-based informatics platform for conducting healthcare research. Research projects using CER Hub technologies are formed as investigator-led communities focused on Comparative Effectiveness Research.

Online collaboration for CER studies
The CER Hub is a web-based mechanism for conducting Comparative Effectiveness Research (CER) where researchers can collaboratively develop protocols to define and operationalize healthcare research questions and methods to answer these from electronic data.

Building clinical data infrastructure
Researchers can develop multi-institutional data sets using CER Hub’s centralized web-based services. These services provide automated tools and support for generating standardized data sets and allow analyses to answer CER questions.

Extracting EMR data
Standardized data processors built on the CER Hub make available natural language processing and knowledge-based systems technologies to automatically identify clinical events in all types of clinical data. Because the CER Hub uses an emerging standard for representing the complete medical record, data from any EMR implementation can be uniformly processed.

Learn More »  View CER Projects »
Asthma Care Quality (ACQ) Study

Contact Info:
Brian Hazlehurst, PhD
Kaiser Permanente Center For Health Research
Brian.Hazlehurst@kpchr.org

Collaborators:
Richard Mularski, MD
MaryAnn McBurnie, PhD
Jon Puro, MPA-HA
Susan Chauvie, RN, MPA-HA

Funder:
Agency for Healthcare Research and Quality (AHRQ)
NLP to Measure Quality of Care in Diabetes: Lessons Learned

Alexander Turchin, MD, MS
Brigham and Women’s Hospital
Harvard Medical School
Project

Monitoring Intensification of Treatment for Hyperglycemia and Hyperlipidemia in Patients with Diabetes

Goal: to design process measures of quality of diabetes care that are tightly linked to patient outcomes

- Blood glucose
- Blood pressure
- Cholesterol

Process measures should be meaningful to providers:

- Medication intensification
- Lifestyle counseling
Project

Source: EMR
- Comprehensive
- Generalizable
- Efficient

Challenges:
- Large fraction of information needed is only in narrative documents (notes)
- No off-the-shelf NLP tools designed to identify concepts we needed

Solution: Design our own
Natural Language Processing

BEFORE YOU BEGIN
Start with a Business Case

Blood pressure extractor:

1. VS: BP **137/80**, HR 85, Weight 195 lbs

Blood pressures at home are **140-150/70-80**.

VNA called: patient has blood pressure **203/105**

Schools: patient's blood pressures daily: they range from **120/75** to **135/80**

Research: INCLUDE

Pay-for-Performance: DISCARD
Involve Domain Experts

- INSULIN
  - DIABETES MELLITUS
  - DIABETES
  - GESTATIONAL DIABETES
Involve Domain Experts

INSULIN

MEN

DIABETES MELLITUS

WOMEN

GESTATIONAL DIABETES
Involve Domain Experts

INSULIN: glargine, glulisine, detemir

INSULIN: regular, NPH, lispro, aspart

DIABETES MELLITUS

GESTATIONAL DIABETES
DESIGN
Hierarchical Processing

Full analysis

Note-level scan

Document Collection

Mandatory note-level concepts found

Mandatory sentence-level concepts found

Sentence-level scan
Custom Concept Classes

- Literal meaning: Start, increase
- Medical meaning: Prescribe
- Situational meaning: Call in
- Misspellings, abbreviations: Increase, incr.
Non-adherence to medications
- Significantly elevated BP (≥ 150/100)
- No intensification of anti-hypertensive medications

Blood pressures measured at home
- Notes with blood pressure ranges (e.g., 120-130/70-80)
Natural Language Processing

VALIDATION
Review by Health Professionals

Meds:

... 
Avapro 150 mg daily 
...

*Increase Avandia to 300 mg daily*
Meds:

... 
Avapro 150 mg daily 
...

*Increase Avapro to 300 mg daily*
Unbiased Validation

CORRECT

INCORRECT

Correction of reviewer rating errors **concordant** with software (e.g., both missed)

Correction of reviewer rating errors **discordant** with software (e.g., reviewer missed, software didn’t)

Compared to manual review
Average face-to-face lifestyle counseling rate (unique patients/hyperglycemic periods)

Percent of patients with A1c ≥ 7.0%

- ≥ 1 per month (468/489)
- between 1 per month and 1 per 6 months (9,659/14,089)
- < 1 per 6 months (9,999/12,406)

Time from first elevated A1c, days

Blood Pressure from Text for P4P

- Identified BP documented by physicians
- Frequently lower than that measured by clinic staff, thereby affecting quality measurement
- Must distinguish home from office BP measurements (home not acceptable for P4P)
Contact Information

Alexander Turchin, MD, MS
Brigham and Women’s Hospital
Harvard Medical School
aturchin@partners.org
Please submit your questions by using the chat box to the lower right of the screen.
Subgroup Discussion

Issue 1:
Copy-Paste in EMR
Copy-Paste: the Problem

Which doctor will achieve better diabetes control?
Text fragments are commonly copied between notes in EMR.

It is not known whether copied text reliably reflects care delivered to the patient.

**Question:** is copied lifestyle (diet, exercise, weight loss) counseling associated with lower blood glucose in patients with diabetes?
Study Population

- **5,914** patients with diabetes treated at primary care practices affiliated with BWH and MGH between 2000 and 2005
- **62,934** notes analyzed to identify lifestyle counseling
Study Design

- **Copied** counseling: sentence documenting counseling identical to that in the previous note by the same provider

- **Distinct** counseling: sentence not identical to previous note or no counseling in the previous note

- Primary **outcome**: time to A1c target (< 7.0%)
Was It Copied?

- The “Copy” button can only copy text within the same patient, not across patients.
- Templates created by provider can be used on any patient.
- Therefore, if identical text was the result of the use of templates, it would be evenly spread across all patients of the same provider.

\[
\frac{\text{Intra-patient prevalence}}{\text{Inter-patient prevalence}} = 31.1 \ (p < 0.0001)
\]
Multivariable analysis (Cox proportional hazards) adjusted for patient demographics, initial A1c, medication intensification, visit frequency, A1c measurement frequency and treatment with insulin:

<table>
<thead>
<tr>
<th>Counseling type</th>
<th>Hazard ratio for A1c normalization</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diet</td>
<td>4.98</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Exercise</td>
<td>3.50</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Weight loss</td>
<td>2.21</td>
<td>0.0011</td>
</tr>
<tr>
<td>Any counseling</td>
<td>4.35</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>
Distinct vs. Copied

- No significant relationship between duplicate (copied) lifestyle counseling documentation and time to A1c target

- No significant difference between effect of duplicate counseling and lack of any counseling documentation on time to A1c target
Subgroup Discussion

Issue 2: Scalability
Scalability

- **SPEED**
  - Speed vs. Accuracy
  - Real-time vs. Retrospective
  - Production System vs. External

- **COST**
  - Generalizable vs. Custom Designed
  - Probabilistic vs. Deterministic
Future

- Begin with basic functions (e.g., extraction of ejection fraction from echo reports) available in commercial EMRs
- Gradually develop more sophisticated / generalizable language models; EMRs will compete on better NLP capabilities
- Self-learning centrally (cloud?) available systems supporting multiple EMRs
Subgroup Discussion

Issue 3:
NLP and Quality Measurement
NLP and Quality Measurement

- **Structured data**
  - More precise / accurate
  - Easier / cheaper to process

- **Unstructured data**
  - Faster / easier to enter
  - Nonredundant
  - Better aligned with clinical workflow
CME/CNE Credits

To obtain CME or CNE credits:

Participants will earn 1.5 contact credit hours for their participation if they attended the entire Web conference.

Participants must complete an online evaluation to obtain a CE certificate.

A link to the online evaluation system will be sent to participants who attend the Web Conference within 48 hours after the event.