



# 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, Presenters, and Disclosures

---

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



# Background

---

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....”*



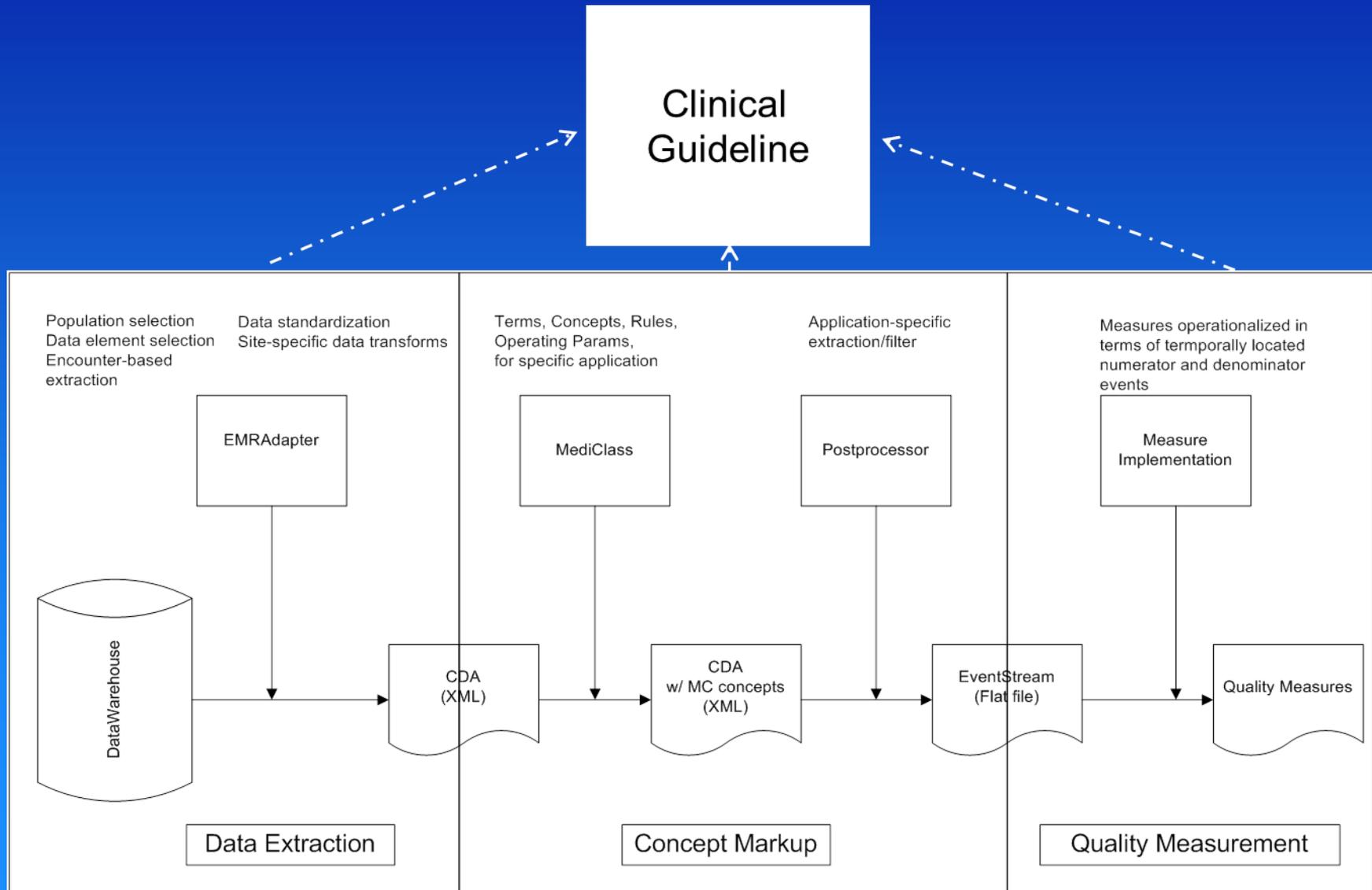
# McGlynn/RAND Conclusions (NEJM, June 2003)

---

- 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





# 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.

Hazlehurst, Frost, Sittig, Stevens. MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record. *JAMIA*. 2005 Sep-Oct;12(5):517-29.



# Asthma Care Quality Measure Set (partial)

Quality Measure	Denominator criteria [Index Date]	Numerator criteria [Measure Interval]	Operationalization Comments
Patients with the diagnosis of persistent asthma should have a historical evaluation of asthma precipitants.	Patients with persistent asthma [PA Qualification Date]	Patients with a subjective evaluation of precipitants or triggers [observation period]	Probably only found in the text progress notes.
Patients with the diagnosis of persistent asthma should have spirometry performed annually.	Patients with persistent asthma [PA Qualification Date]	Patients with orders for PFTs or documentation of office spirometry or of PFT results [subsequent 12 months]	Numerator satisfied with documentation of referral to pulmonary specialist if no PFT known available.
Patients with the diagnosis of persistent asthma should have available short acting beta-2 agonist inhaler for symptomatic relief of exacerbations.	Patients with persistent asthma [PA Qualification Date]	Prescription for a short acting beta-2 agonist to use PRN [subsequent 12 months]	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.
All patients seen for an acute asthma exacerbation should have current medications reviewed.	Patients with persistent asthma meeting criteria for outpatient exacerbation [Exac. Encounter]	Documentation that medications reviewed by provider [same visit]	Numerator satisfied if provider documents asthma specific medication history in notes or active management of current medication list.



# Clinical Events Dataset File (portion)

Patient segment

Provider segment

Event 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,

.....

.....

# Clinical Events Dataset File (cont.)



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,,,,,'Persi  
stent',,Main Note,

pat3,KPNW,19xx,M,'?','?','?',prov3,FP,'loc4',Outpatient,enc4,MedsReview,Note,20010111,,,,,Anci  
llary 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





# Asthma Care Quality (ACQ) Findings

- 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.



# Chart Review Validation

- 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.



# Chart Review Validation

- 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



# Outcomes Related to ACQ Measures

- 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



# Ongoing Work

- 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

CER Hub - Mozilla Firefox

File Edit View History Bookmarks Tools Help

cerhub.org https://www.cerhub.org/index.html

CER Hub



## www.cerhub.org

Welcome About Us Mission & Membership Research Projects Resource Center Contact Us



### 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.

[Learn More »](#) [View CER Projects »](#)

#### 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.



# Asthma Care Quality (ACQ) Study

---

## Contact Info:

Brian Hazlehurst, PhD

Kaiser Permanente Center For Health Research

[Brian.Hazlehurst@kpchr.org](mailto: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:

Research: **INCLUDE**

1. VS: BP **137/80**, HR 85, RR 18, SpO<sub>2</sub> 98% on 1L O<sub>2</sub>

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

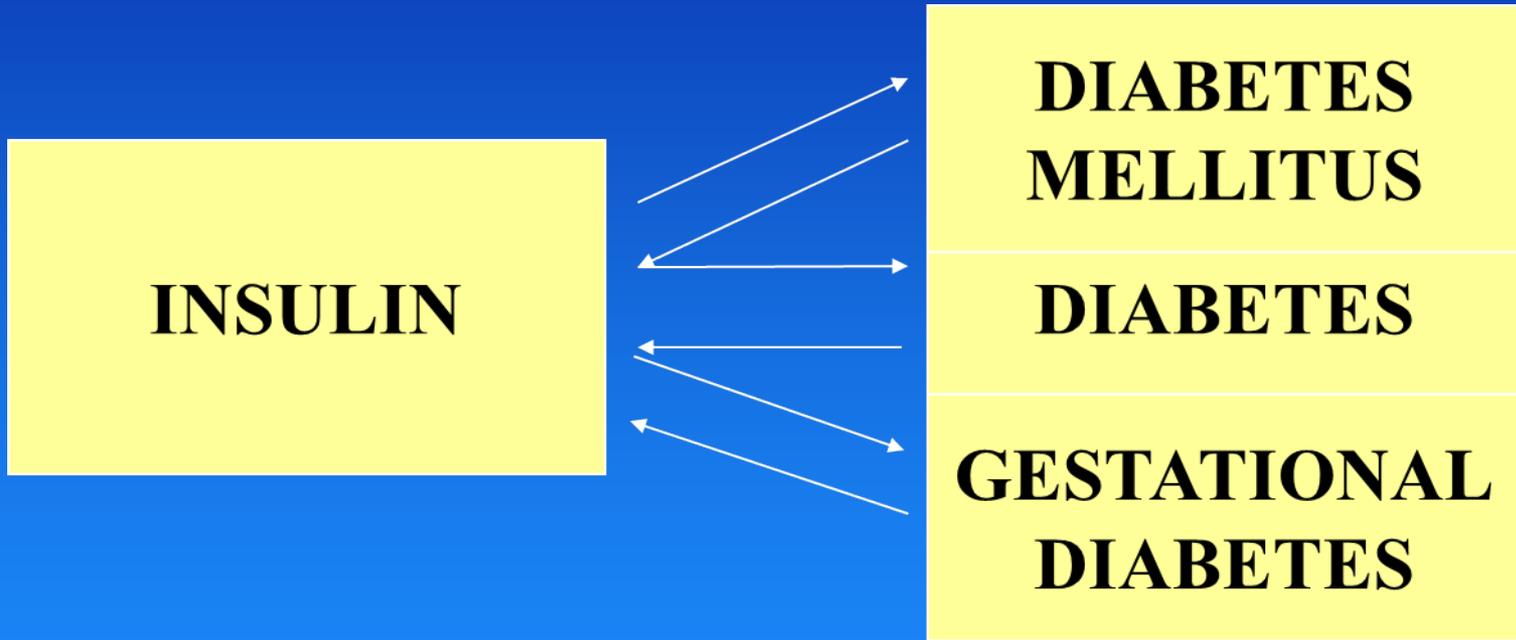
**VNA called:** patient has blood pressure **203/105**

**School:** patient has blood pressures

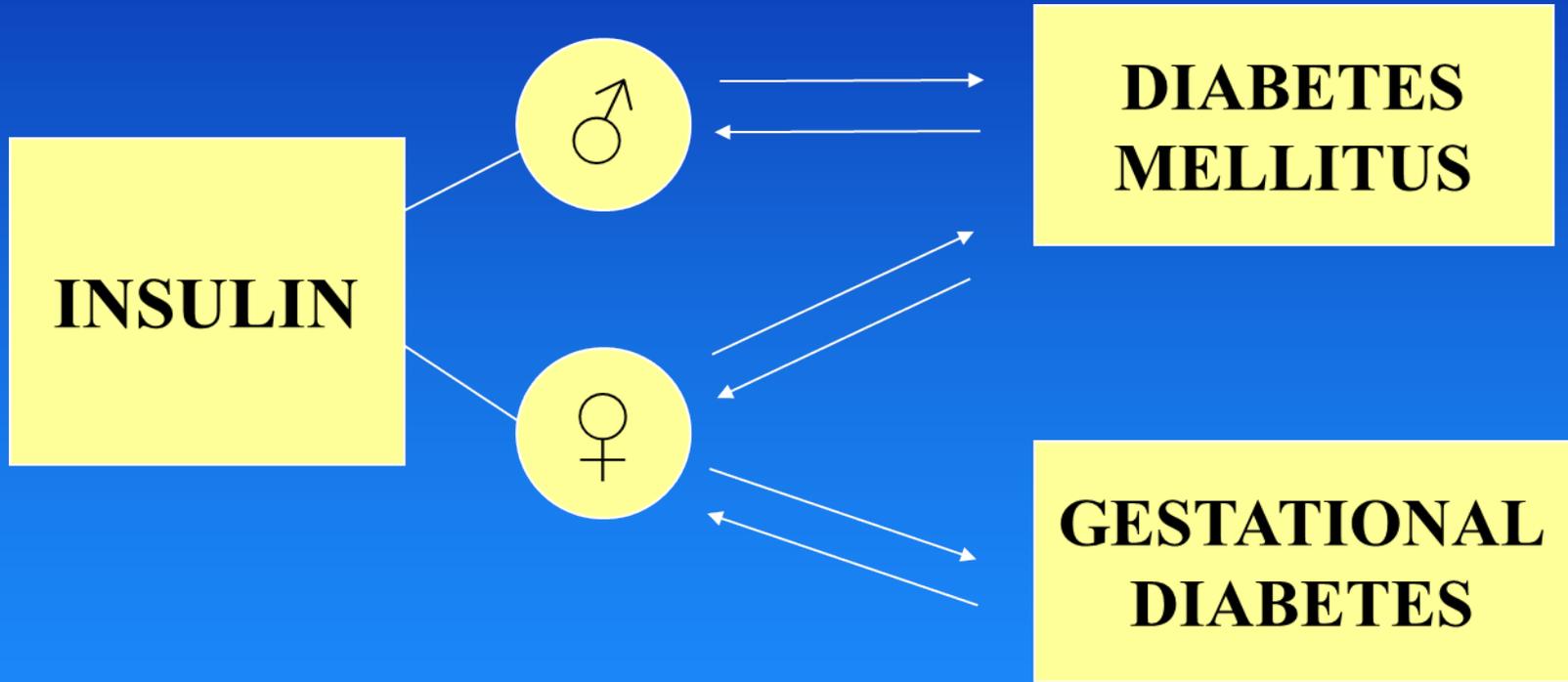
Pay-for-Performance: **DISCARD**

daily: they range from **120/75** to **135/80**

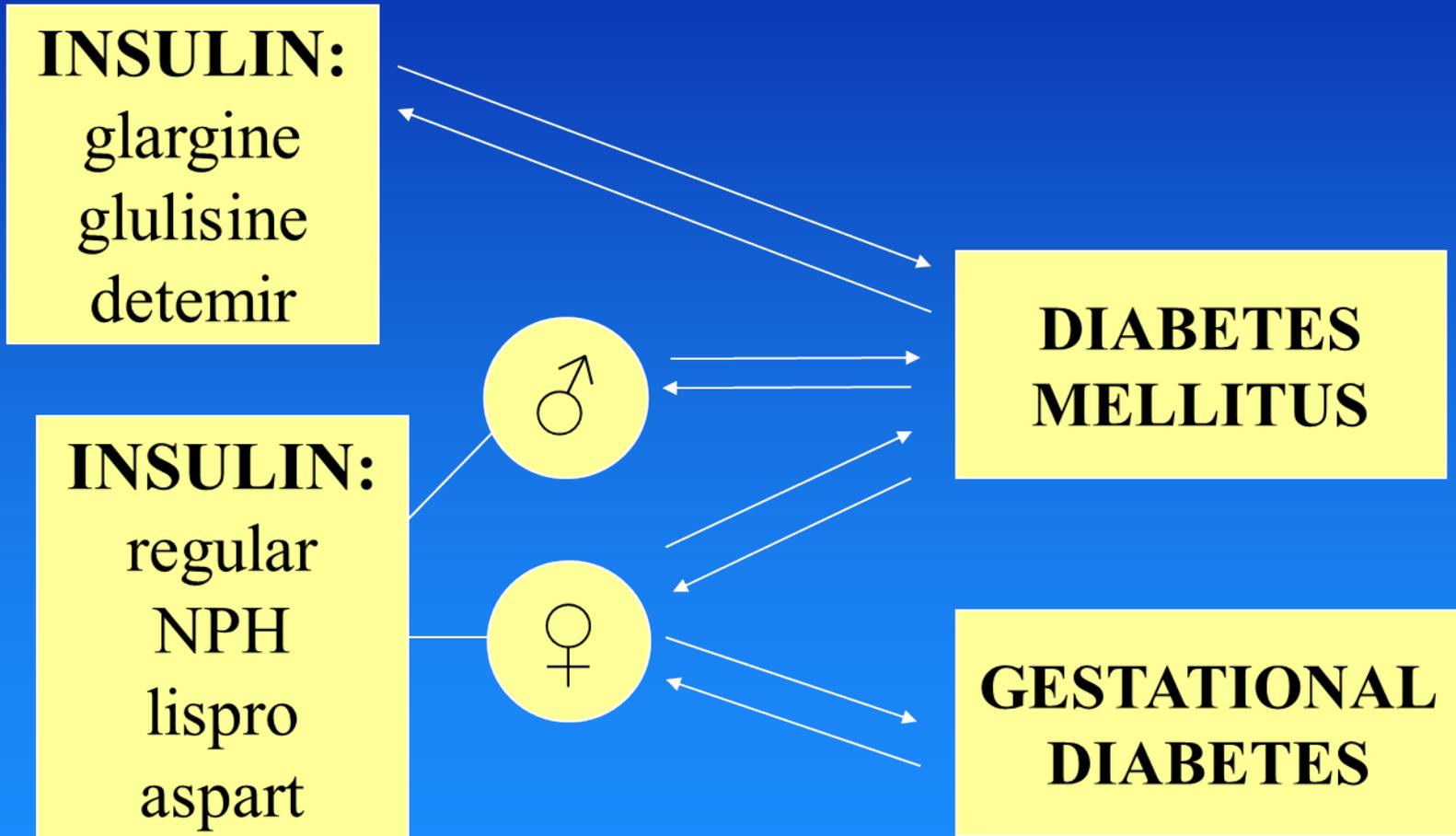
# Involve Domain Experts



# Involve Domain Experts



# Involve Domain Experts



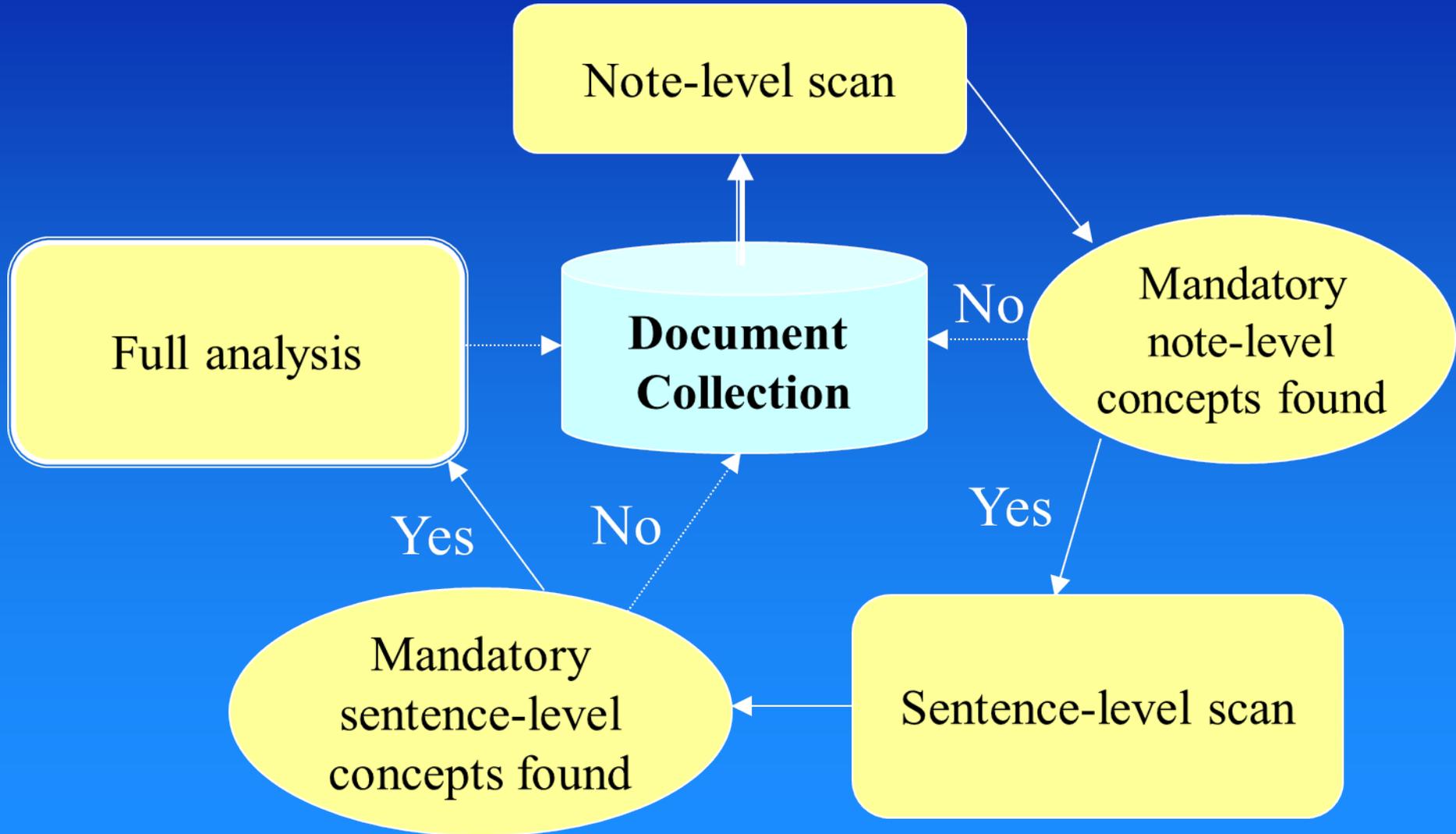


# Natural Language Processing

---

**DESIGN**

# Hierarchical Processing



# Custom Concept Classes

Literal meaning	<i>Start, increase</i>
Medical meaning	<i>Prescribe</i>
Situational meaning	<i>Call in</i>
Misspellings, abbreviations	<i>Increase, incr.</i>



# Enrichment of Data Sources

---

- Non-adherence to medications
  - Significantly elevated BP ( $\geq 150/100$ )
  - 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*



# Review by Health Professionals

Meds:

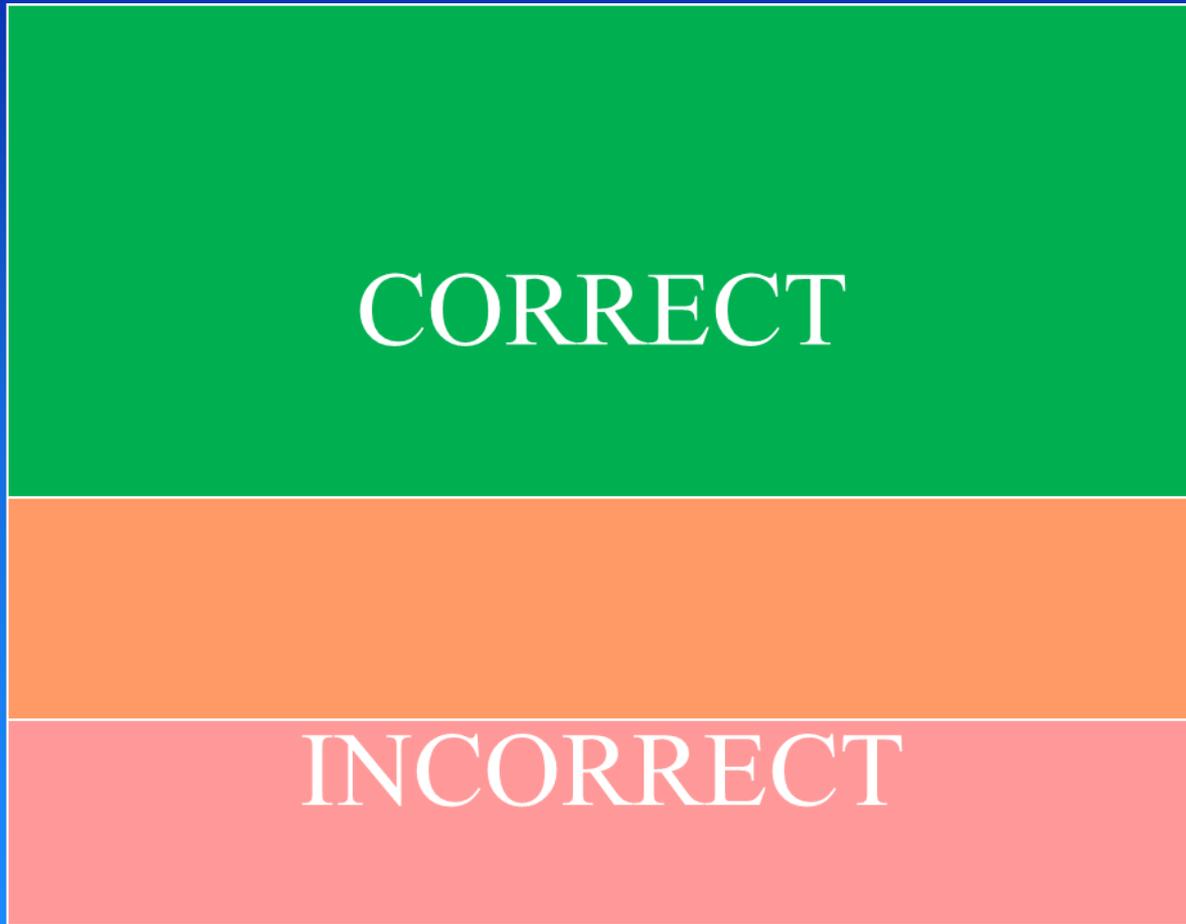
...

Avapro 150 mg daily

...

*Increase **Avapro** to 300 mg daily*

# Unbiased Validation

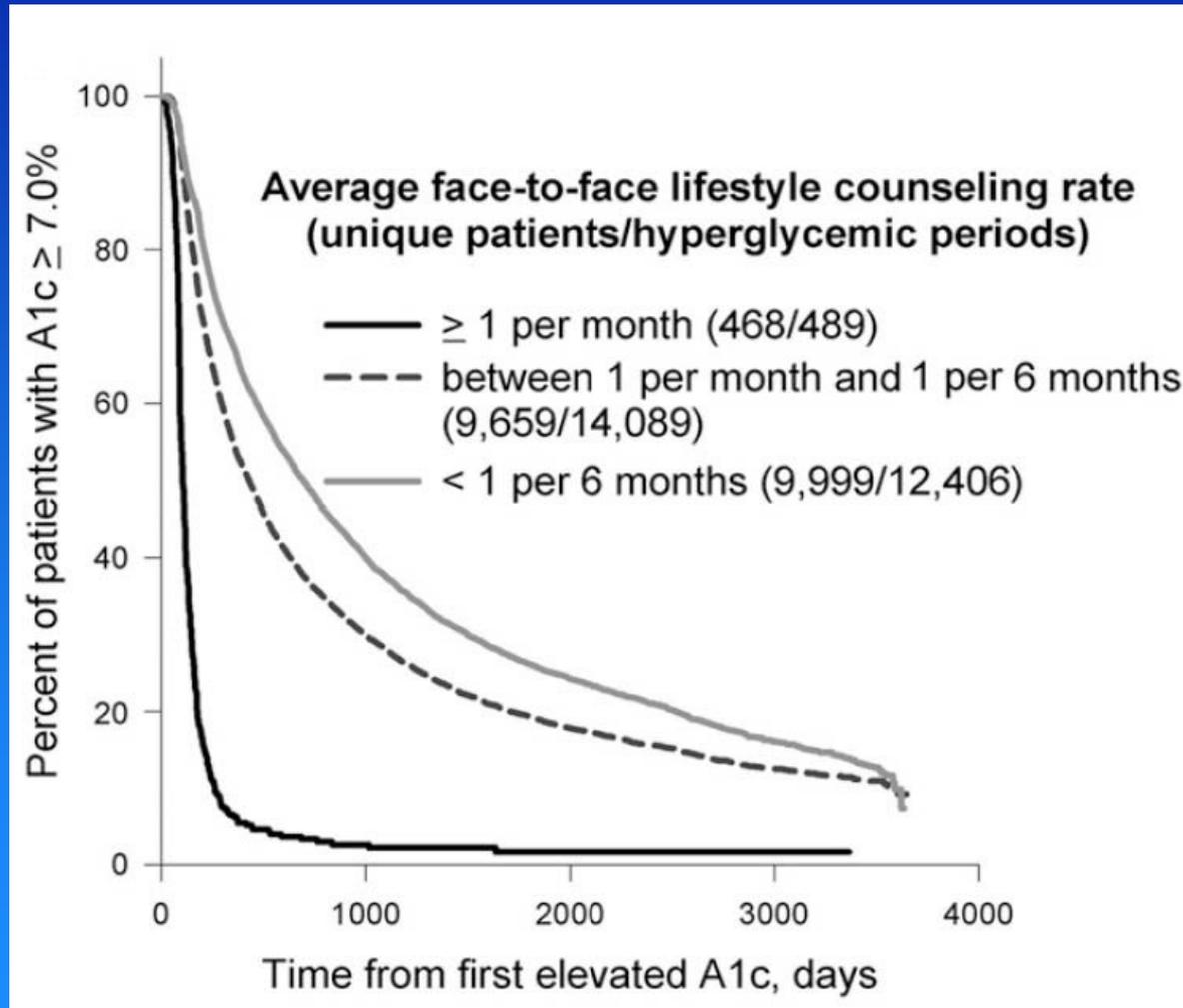


Correction of reviewer rating errors **concordant** with software (e.g., both missed)

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

Compared to manual review

# Project: Results





# Project: Implementation

---

- *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](mailto:aturchin@partners.org)



# Q & A

---

Please submit your questions by using the chat box to the lower right of the screen.

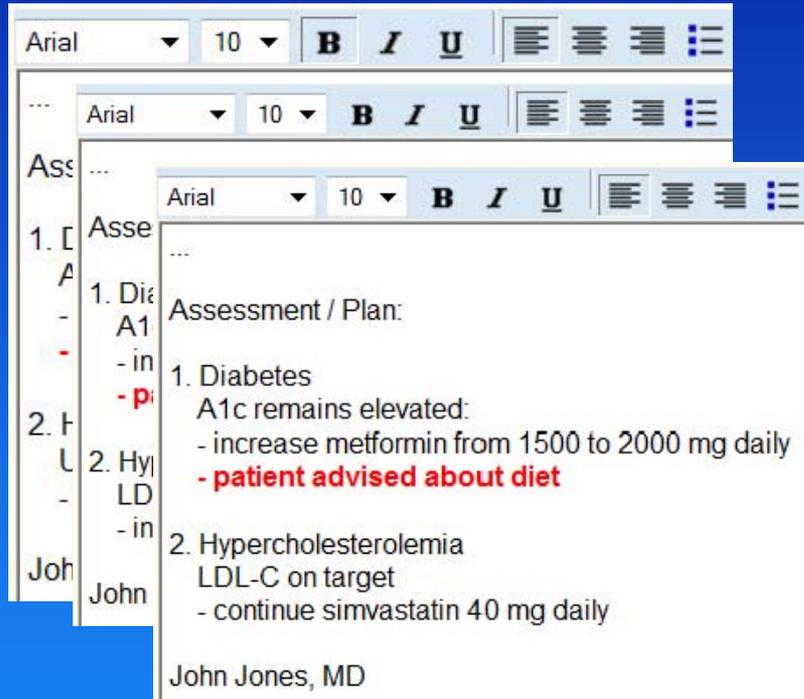


# Subgroup Discussion

---

Issue 1:  
Copy-Paste in EMR

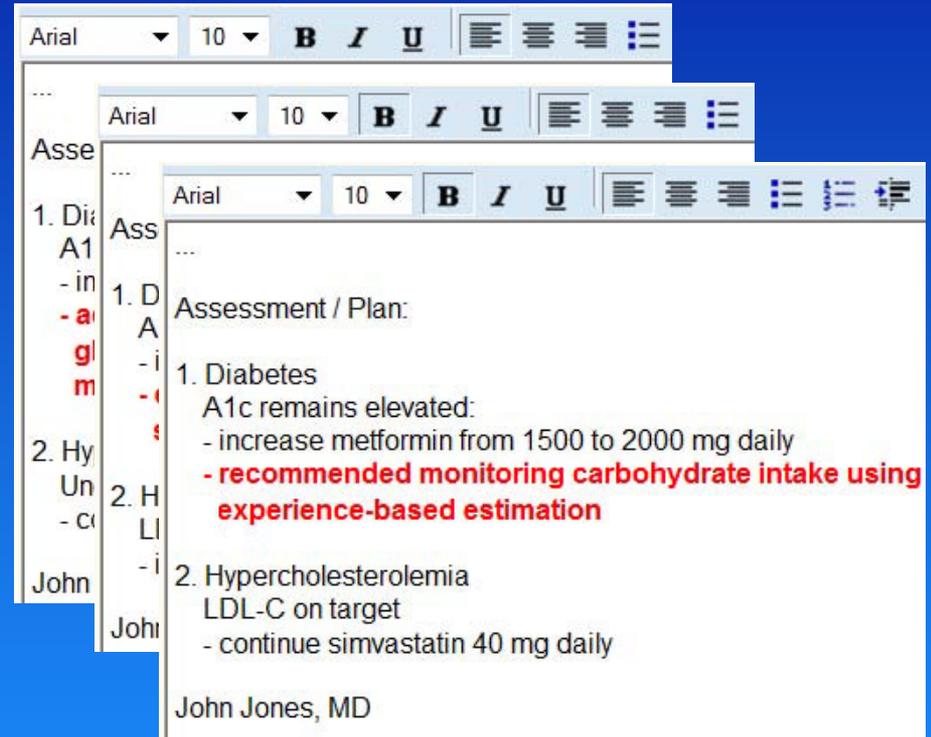
# Copy-Paste: the Problem



Assessment / Plan:

- Diabetes  
A1c remains elevated:  
- increase metformin from 1500 to 2000 mg daily  
**- patient advised about diet**
- Hypercholesterolemia  
LDL-C on target  
- continue simvastatin 40 mg daily

John Jones, MD



Assessment / Plan:

- Diabetes  
A1c remains elevated:  
- increase metformin from 1500 to 2000 mg daily  
**- recommended monitoring carbohydrate intake using experience-based estimation**
- Hypercholesterolemia  
LDL-C on target  
- continue simvastatin 40 mg daily

John Jones, MD

Which doctor will achieve better diabetes control?



# Copy-Paste in EMR

---

- 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

$$\frac{\text{Intra-patient prevalence}}{\text{Inter-patient prevalence}} = 31.1 \text{ (p < 0.0001)}$$



# Distinct Counseling & A1c

Multivariable analysis (Cox proportional hazards) adjusted for patient demographics, initial A1c, medication intensification, visit frequency, A1c measurement frequency and treatment with insulin:

Counseling type	Hazard ratio for A1c normalization	P-value
Diet	4.98	< 0.0001
Exercise	3.50	< 0.0001
Weight loss	2.21	0.0011
Any counseling	4.35	< 0.0001



# 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.