Towards Real-World Data for Dementia Studies:

Use of machine-learning to improve metrics

Sudeshna Das, PhD & Deborah Blacker, MD, ScD

Oct 15, 2020

MASSACHUSETTS Alzheimer's Disease Research Center

OVERVIEW

- Background
- Utilities and limitations of real-world data
- Process model
- Gold standard labels and baseline model
- Preliminary data from NLP approach to improve dementia diagnosis in EHR
- Future directions
- Questions

Work in progress, update from earlier talk December 2019

BACKGROUND: Harvard Initiative on Aging planning grant

Working group formed Spring 2019 growing out of experience on two projects focused on diagnosis of dementia in EHR, in many ways complementary

- MD review of local EHR to validate claims in affiliated ACO
- Use of structured data (dx, rx) in EHR to identify signals for drug repurposing

Recent doctoral thesis using NLP to diagnose depression in PCP notes

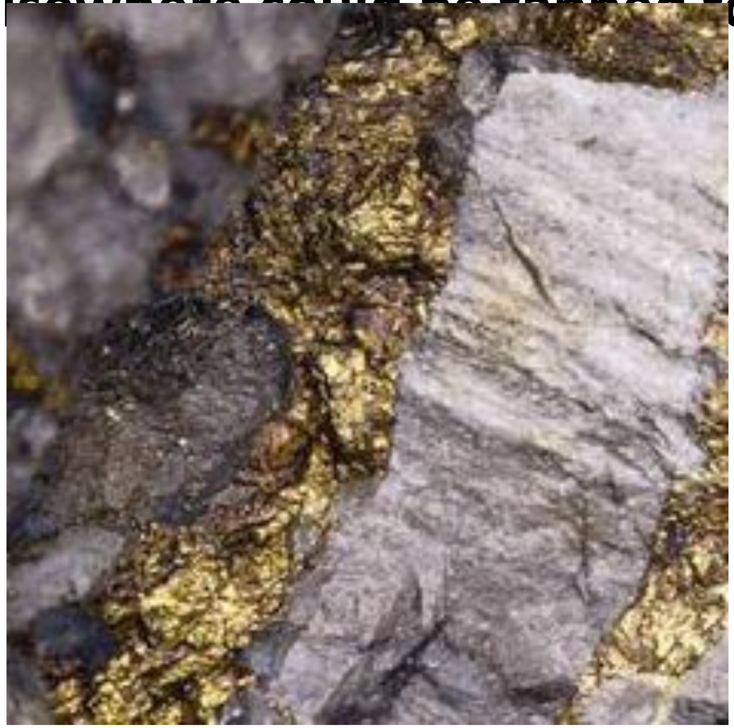
Utilities and Limitations of Real-world Data

Potential Uses of EHR and Claims-based diagnosis

Critical and common public health problem with enormous burden, cost, loss

Extensive clinical data in EHR, claims, and elsewhere could be tapped for many purposes, e.g.:

- Early recognition
- Targeted screening
- Follow up in cohorts
- Risk factors and genetics
- Drug repurposing signals
- Recruiting
- Health policy and finance



Limitations of EHR and Claims-based diagnosis

Extensive data and experience suggest under-recognition and delayed recognition of dementia—by patients, families, and physicians

Even when recognized, may not be formally indicated in claims or structured EHR variables

More limited data and common sense suggest that such under- and delayed recognition/under-coding varies by age, gender, education, etc.

Effort to quantify these issues to enable interpretation of these findings, to select best approach, to inform sensitivity analyses, etc.

Conceptual Model

CONCEPTUAL MODEL: Symptoms to Claims

Goal: understand process, taking particular note of potential sources of bias, to inform design and sensitivity analyses

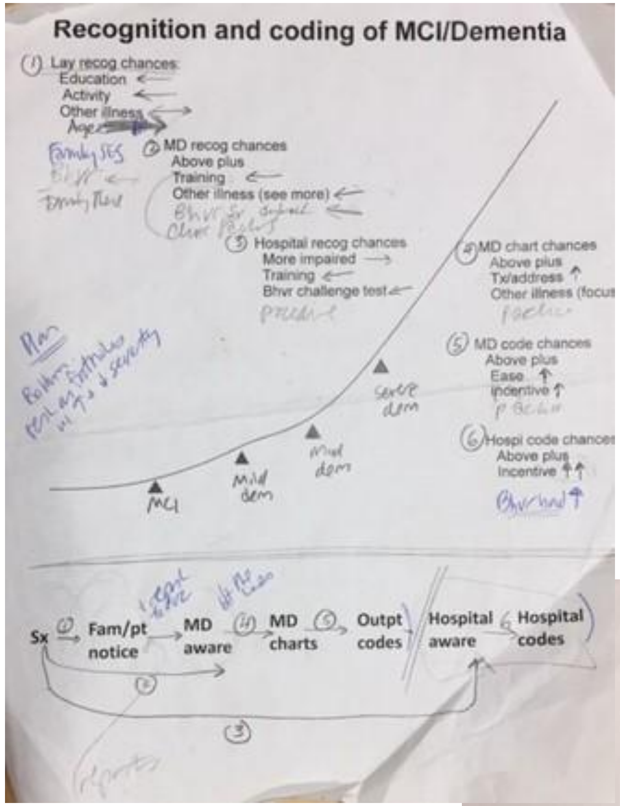
Initial model very complex, accounted for multiple pathways to diagnosis, skipping some steps, branches for hospitalization, etc.

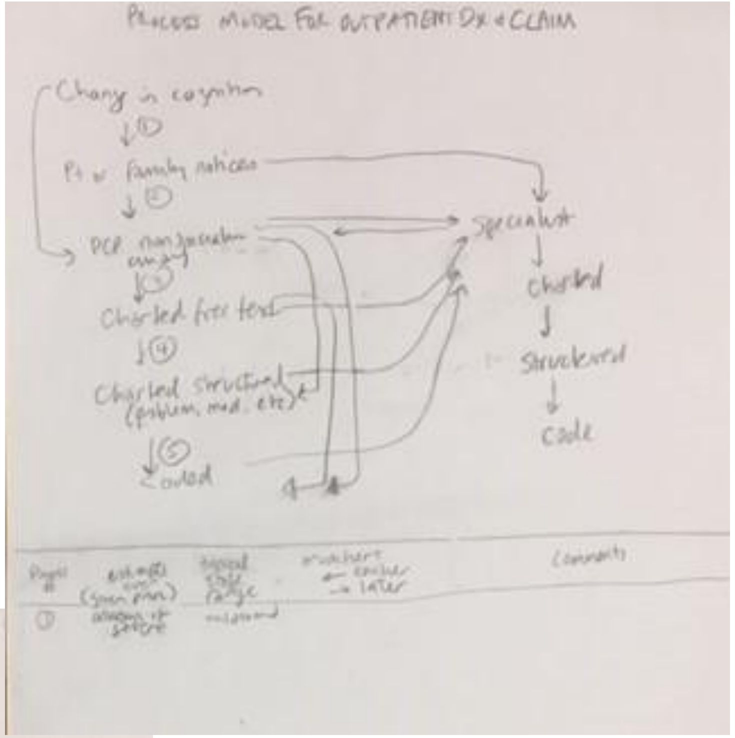
Current model simplified for clarity and ease of use; limited accounting for a very complex process

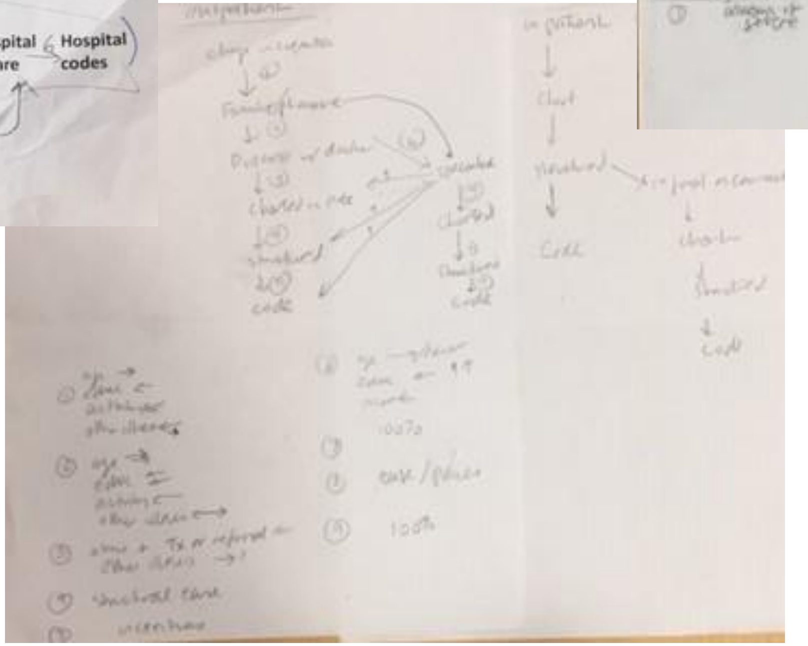
Estimates may be system specific: structured data depends on specific EHR system, options, ease of use; coding depends on ease of use and incentives

Estimates may depend on historical epoch and location based on above factors, plus time and location variation in practice patterns and screening

Original Process Models







Current (over)-simplified, more tractable, Process Model

Change in cognition

⇩ 1

Patient/family notices/concerned

⇩ 2

Report to MD* 6 ⇒ Specialist referral/consult**

⇩ 3

Charts free text in EHR

⇩ 4

Charts structured variable in EHR
(problem list, meds, etc.)

⇩ 5

Claims coded

⇩⇩⇩

****PCP focused, but applies more broadly**

****Specialist referral/consult generates more structured and unstructured data and codes**

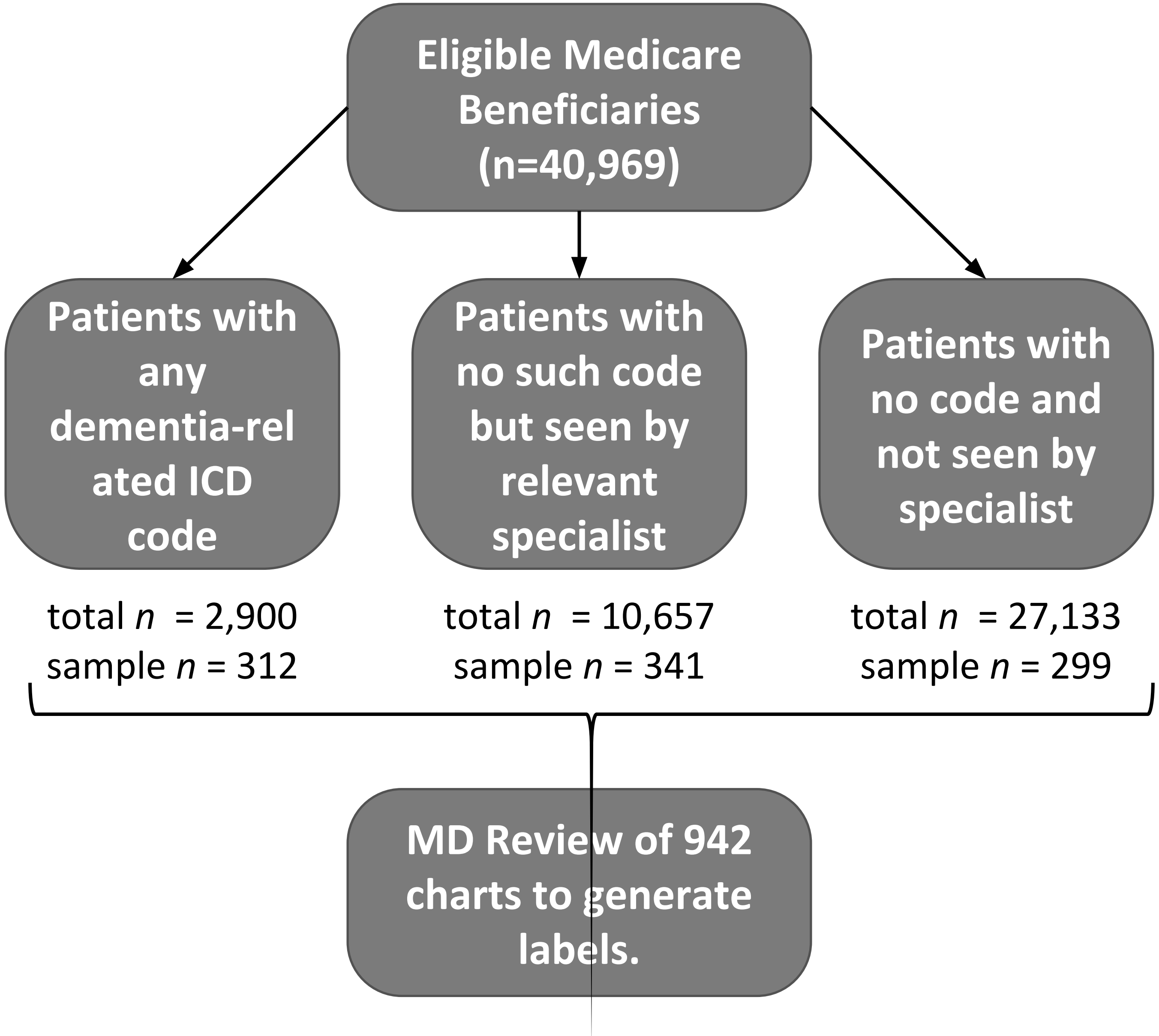
Notes for each step of Process Model

Step	Estimate overall (given prior step)	Modifiers: ⇓⇑ probability ⇐ shift earlier ⇒ shift later	Notes
1. *Change to Notice*	Early low, late higher	⇐ more education, younger, active, working, family involved ⇒ isolated, older	Other illness cuts both ways—more chance to observe, but may distract
2. *Notice to Report to MD*	Early low, late higher	Above plus: ⇑ better access (SES) ⇑ trust in MDs (URM, SES)	PCP may also notice directly (typically later in course unless there is screening or care presents challenge test [e.g., sliding scale insulin])
3. *Report to MD* to *Chart Free Text*	?50% estimated recognized overall	Above plus: ⇐ Training and screening ⇐ Hospitalization (vulnerability and detailed observations) ⇑ care or referral	Can skip steps between 3 and 5, will have some data on this

Step	Estimate overall (given prior step)	Modifiers ⇓⇑ probability ⇐ shift earlier ⇒ shift later	Notes
4. ***Chart Free Text* to *Structured Variable***	?	⇑ care or referral ⇑ Hospitalization ⇑ options available and ease of use	
5. ***Structured Variable* to *Billing Code***	?	⇑ care or referral ⇑ ease of coding options ⇑ incentives (especially if hospitalized where there are "coders" who scan chart)	
6. ***MD Notice* to *Specialist Referral***	Small	⇑ behavior issues (especially In hospital setting) ⇑ unusual presentation or early onset ⇑ PCP low comfort w/ dementia care ⇑ high SES patient	Assume if see a specialist or consult will chart and code (but presence of structured EHR elements likely depends on EHR system)

Generation of gold-standard dataset and comparison to claims

Procedure to Generate Labels



MD review of **942** charts for **3-year period (1/1/2016-12-31/2018)** as part of a colleague's effort to validate claims-based dx

MD blind to sampling scheme rated presence of a **cognitive concern** of any kind, **syndromic diagnosis**, and **dementia severity** (with goal of being able to note delayed as well as missed recognition).

For each rating, a **confidence level** of 1-4 was also included.

MD also noted whether there was evidence of **PD**, **FTD**, or other relatively **rare dementia subtypes**, any episodes of **delirium**, and the presence of behavior symptoms.

Gold-Standard Dataset Demographics

Characteristics	N (%)	
Age on 12/31/2018		
< 75 years	191	(20.3%)
75-79 years	243	(25.8%)
80-84 years	202	(21.4%)
>= 85 years	306	(32.5%)
Gender		
Female	559	(59.3%)
Male	383	(40.7%)
Duration of care	17.9 y	± 8.6
PCP within system	710	(75.4%)

Mean Age 74-81, 92-94% white

Gold-Standard Dataset Comparison to Claims

Cognitive concern as our “gold standard diagnosis”:

Cognitive concern akin to first criterion of MCI, that patient, a family member/ friend, or the clinician is concerned about cognition—casts a wide net for potential cognitive diagnosis, and might indicate those who would benefit from further evaluation

Compared expert labels of “cognitive concern” rated by expert raters with high to moderate confidence with EHR-based label

A visit diagnosis or problem list code of MCI or dementia (290.X, 294.X, 331.X, 780.93, G30.X and G31.X))

OR

An anticholinesterase inhibitor or memantine on medication list in EHR

Aware that cognitive concern is broader and MCI/dementia, but as a first step focused on how standard claims-style variables would map onto broadest categories

Gold-Standard Dataset Comparison to Claims

We compared the Expert-Adjudicated Labels to records of dementia-related ICD codes or medication in the patient's electronic health records (EHR)

Comparison of Dx-Rx with Gold-Standard Labels			
Clinician Adjudication	ICD code or medication	No ICD code or medication	Total
Cognitive Concern present*	273 (70.9%)	112 (29.1%)	385
Cognitive Concern absent*	43 (7.7%)	514 (92.3%)	557
***With a medium-to-high certainty score**			

Preliminary data from NLP project

Why use NLP?

- Signals in clinician notes often provide clue to cognitive dysfunction when there are no formal diagnosis codes or medications
- Clinicians may chart symptoms of cognitive issues in notes but may not make a formal diagnosis, refer to a specialist, or prescribe a medication for multiple reasons
    - Lack of time or expertise
    - No billing incentive
    - Stigma

Data Types	EHR	Claims
Visit Diagnosis	X	X
Prescriptions	X	X
Medication History	X	
Problem List	X	
Clinician Notes	X	
Out-of-system care		X

Using Machine-Learning to Detect Patients with Cognitive Concerns

Binary Classification of patients with or without concerns

Train and Test Data: Subset of Gold-standard with progress notes (N=767)
Split: 0.9 Train, 0.1 Test set for all models

Demographics:

Dataset	Age (Mean and SD)	% Female	Cognitive Concern (N)
Train (N = 690)	81.2 (7.4)	58.2%	308 (44.64%)
Test (N = 77)	80.19 (6.79)	54.5%	34 (44.16%)

NLP Models for Cognitive Concern Classification Using EHR

We developed 4 models with increasing complexity and deeper representation. Model 1 uses structured variables only and the other three use clinician notes.

- Model 1. Baseline Model with structured variables (diagnosis codes & medications)
- Model 2: Logistic Regression with counts of dementia-related concepts in notes
- Model 3. Logistic Regression with word vectorization (TF-IDF) on note
- Model 4. Transformer based language model (Longformer) on notes

Model #1: Baseline Model

Structured Data Only (Diagnosis Codes and Medication)

COVARIATES

COUNT OF

MCI or dementia ICD Code

+

COUNT OF

Meds (AChEI or memantine)

OUTCOME

Cognitive Concern (CC)

We built a logistic regression model:

CC ~ COUNT(ICD CODES) + COUNT(MEDS)

Model #2: NLP on Clinician Charts

Logistic Regression with Counts of Dementia-Related Concepts

COVARIATES

COUNT OF

Cognitive

+

COUNT OF

Behavioral

OUTCOME

Cognitive Concern

LOGISTIC REGRESSION MODEL

CC ~ SUM(Concept_1) …+ SUM(Concept_k)

<table>
<tr><th colspan="6">Cognitive Dysfunction / Impairment Domain</th></tr>
<tr><th>Memory</th><th>Executive Functioning</th><th>Confusion</th><th>Cognition</th><th>Orientation</th><th>Conherence</th></tr>
<tr><td>[*NOT*] remember</td><td>[*NOT*] follow instructions</td><td>confused</td><td>cognition</td><td>disoriented</td><td>disorganized</td></tr>
<tr><td>[*NOT*] recall</td><td>[*NOT*] making decisions |decsion making</td><td>sundowning</td><td>[*POOR*] mental status</td><td>[*NOT*] oriented</td><td>incoherent</td></tr>
<tr><td>forget | forgetful | forgot |f</td><td>[*NOT*] identify</td><td>confused</td><td>mental status</td><td>orientated to person</td><td>nonsensical |non-sensical |nor</td></tr>
<tr><td>repetitive questions</td><td>[*NOT*] plan | planning</td><td>Confusion</td><td>[*POOR*] cognition</td><td>[A+O |A&O] [X 1 |X 1-2| X2]</td><td>[*NOT*] to express coherent thoughts</td></tr>
<tr><td>0 memory | recall</td><td>[*NOT*] organizing | organize</td><td>sundowns</td><td>[*POOR*] cognitve skills</td><td>oriented to self</td><td>Illogical</td></tr>
<tr><td>[*POOR*] historian | history</td><td>[*NOT*] follow commands</td><td>[*NOT*] understand</td><td>cognition</td><td>oriented to name</td><td>thought process</td></tr>
<tr><td></td><td>[*NOT*] follow directions</td><td></td><td></td><td>[*NOT*] orient to time</td><td>talks nonsense</td></tr>
<tr><td></td><td>[*NOT*] follow steps</td><td></td><td></td><td>[*NOT*] know [where|who|wh</td><td>[*NOT*] make sense</td></tr>
<tr><td></td><td>[*POOR*] insight</td><td></td><td></td><td></td><td>[*NOT*] not rational</td></tr>
<tr><td></td><td>[*POOR*] judgement</td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>[*POOR*] problem solving</td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>[*NOT*] wordfinding | finding words</td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>[*NOT*] recoginize</td><td></td><td></td><td></td><td></td></tr>
</table>

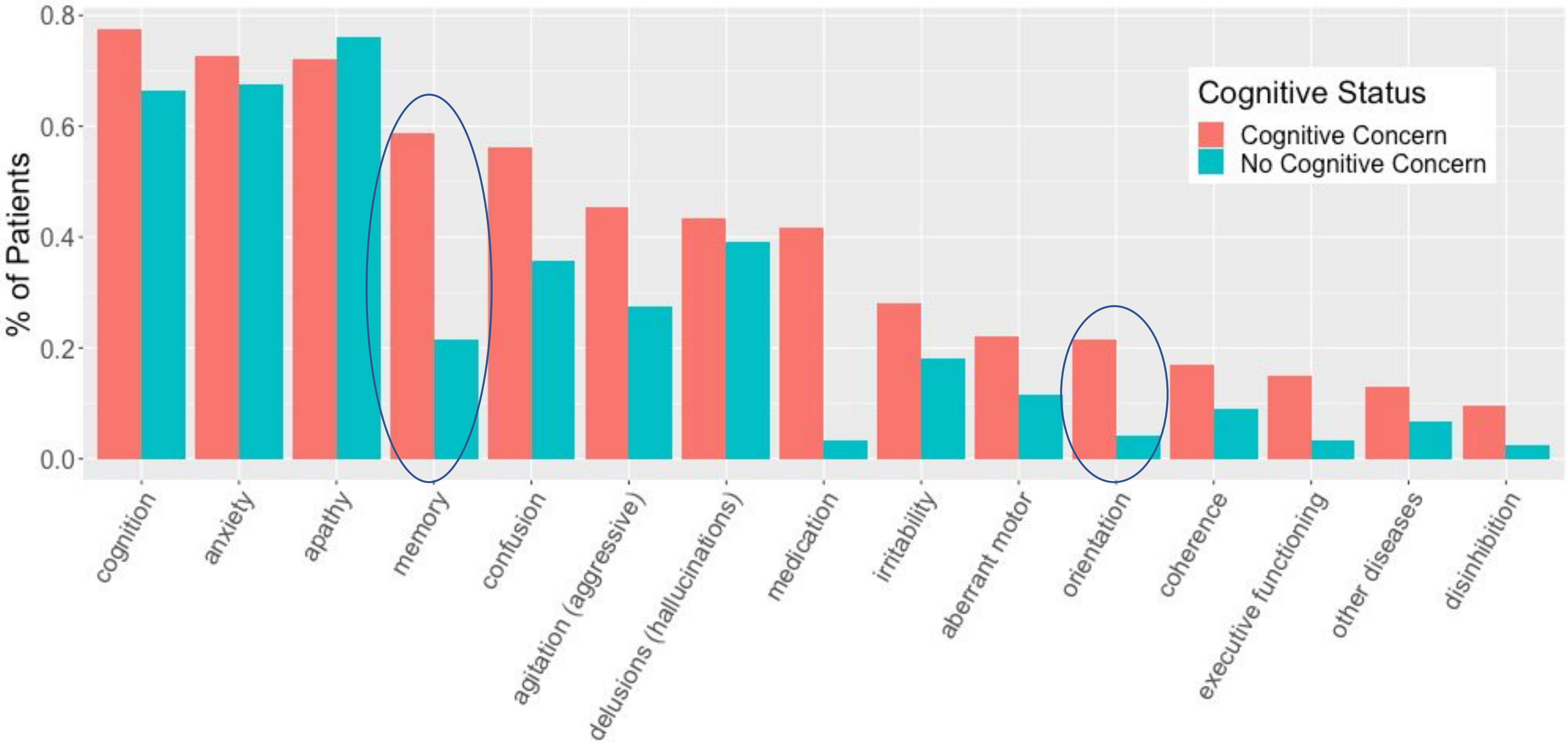
6 concepts in Cognitive Impairment Domain

7 concepts in Behavioral Impairment Domain

<table>
<tr><th colspan="7">Behavioral Dysfunction/ Impariment Domain</th></tr>
<tr><th>Delusions (Hallucinations)</th><th>Aberrant Motor</th><th>Agitation (Aggressive)</th><th>Apathy</th><th>Disinhibition</th><th>Anxiety</th><th>Irritability</th></tr>
<tr><td>delusional</td><td>restless</td><td>agitat</td><td>apathy</td><td>disinhibited</td><td>anxiety</td><td>irritable</td></tr>
<tr><td>paranoid</td><td>fidget</td><td>aggressive</td><td>apathetic</td><td>impulsive</td><td>anxious</td><td>defensive</td></tr>
<tr><td>delusions</td><td>pulling at</td><td>combative | combativenes</td><td>[*POOR*] motivation</td><td></td><td>scared</td><td>easily annoyed</td></tr>
<tr><td>hallucinat</td><td>rummaging</td><td>uncooperative</td><td>[*POOR*] interest</td><td></td><td>afraid</td><td>easily angry</td></tr>
<tr><td>suspicious</td><td>unpacking</td><td>[*NOT*] compliant</td><td>[*POOR*] affect</td><td></td><td>nervous</td><td>easily upset</td></tr>
<tr><td>Delusions</td><td>repetitive behavior</td><td>non-compliant</td><td>affect</td><td></td><td>fear | fearful</td><td>Labile |lability</td></tr>
<tr><td>paranoia</td><td>obsessive behavior</td><td>care resistance</td><td>unmotivated</td><td></td><td></td><td></td></tr>
<tr><td>delirum</td><td>Impulsive</td><td>verbally aggressive</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Hallucinating</td><td>pulling on</td><td>physically aggressive</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Hallucinations</td><td>fidgeting</td><td>Agitated</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Hallucinate</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Psychosis</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>wandering</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Gilmore-Bykovskyi AL et al. **Unstructured clinical documentation reflecting cognitive and behavioral dysfunction: toward an EHR-based phenotype for cognitive impairment**. J Am Med Inform Assoc. 2018;25(9):1206-12.

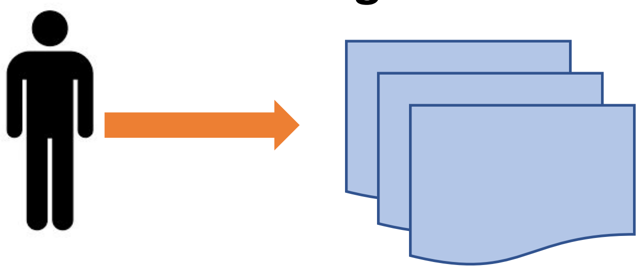
Model 2: Logistic Regression with counts of dementia-related concepts



Model #3: NLP on Vectorized Charts

Logistic Regression with Word Vectorization

Progress Notes



$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

TF-IDF
Term x within document y

$tf_{x,y}$ = frequency of x in y
df_x = number of documents containing x
N = total number of documents

~38,000 words

	W_1	W_2	W_3	...					...	W_K
P_1	0.5		2.1			0.2				
P_2				1.2						2.6
P_3	9.3									
...										
...										
P_N										

Each Patient's notes are converted into a word vector and the TF-IDF weight is computed. The correlation of the word vector with the outcome is computed.

Model #3: NLP on Vectorized Charts

We built a logistic regression model with the TF-IDF weights of top words

memory	0.33274388	behavioral	0.18553319
aricept	0.29681549	memantine	0.18413032
alzheimer	0.26634859	recall	0.18408992
dementia	0.25055677	accompanied	0.18073228
daughter	0.22997573	molst	0.17936733
donepezil	0.21860699	conversation	0.17851918
confused	0.21025071	remember	0.17835091
mental	0.20420438	executive	0.17742126
care	0.18963723	cognitive	0.17677409
decline	0.18701276	unable	0.17490906

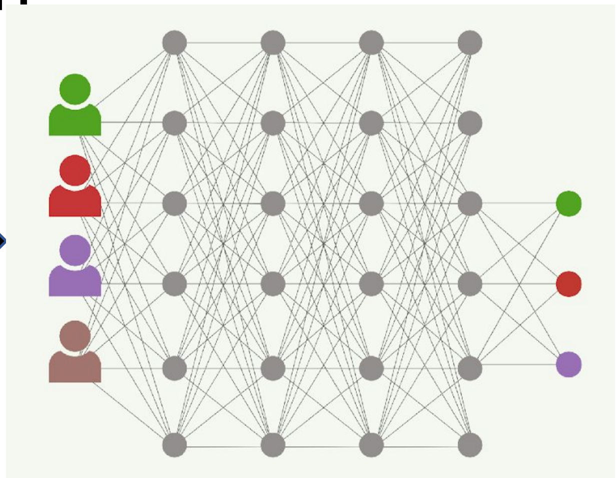
words most correlated with outcome

Model #4: Deep Learning NLP on Charts

Transformer Based Language Model

Sliding Window: 4,096 tokens, 20% overlap with previous

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.



DEEP NEURAL NETWORK

We built a deep learning model which computes the score of each window and aggregates those at a patient level. These models capture context of sentence in a bi-directional manner. We use a pre-trained model (BERT) and fine-tune to our classification task.

Model Performance

Model	AUC	Accuracy	FP	FN	Sensitivity	Specificity	PPV	NPV
Model 1	0.79	0.82	0	14	0.59	1.00	1.00	0.75
Model 2	0.88	0.84	4	8	0.76	0.91	0.87	0.83
Model 3	0.90	0.84	3	9	0.74	0.93	0.89	0.82
Model 4	0.93	0.87	1	9	0.74	0.98	0.96	0.82

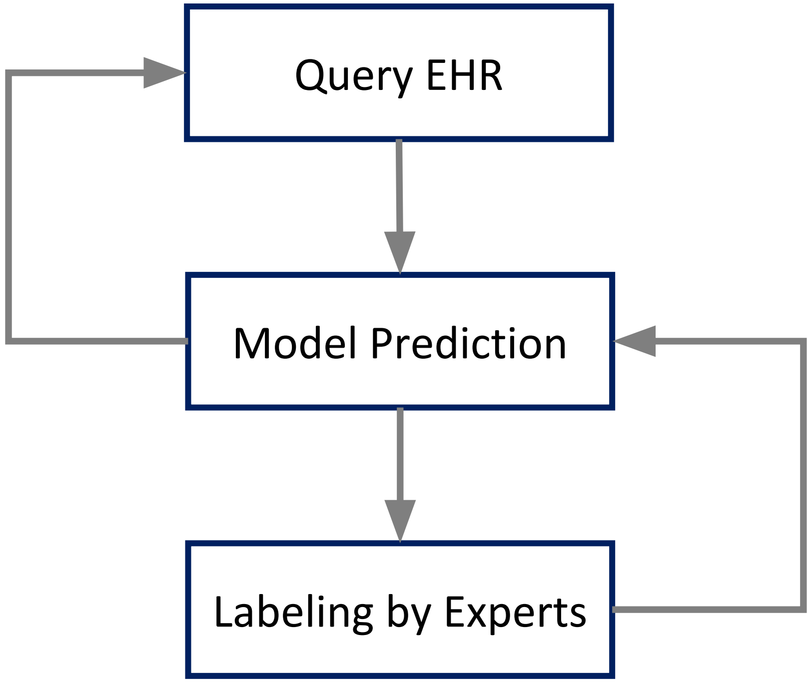
Sensitivity and specificity were computed at the threshold that provided highest accuracy.

Improving the model with active learning

- Refine the model with further labeled data.
- Oversample the cases with poor prediction and get expert labeling on the edge cases (i.e. model provides a poor confidence on the classification) .
- Developed an annotation tool for labeling that can be reused for other settings.

Next step: Active Learning Loop

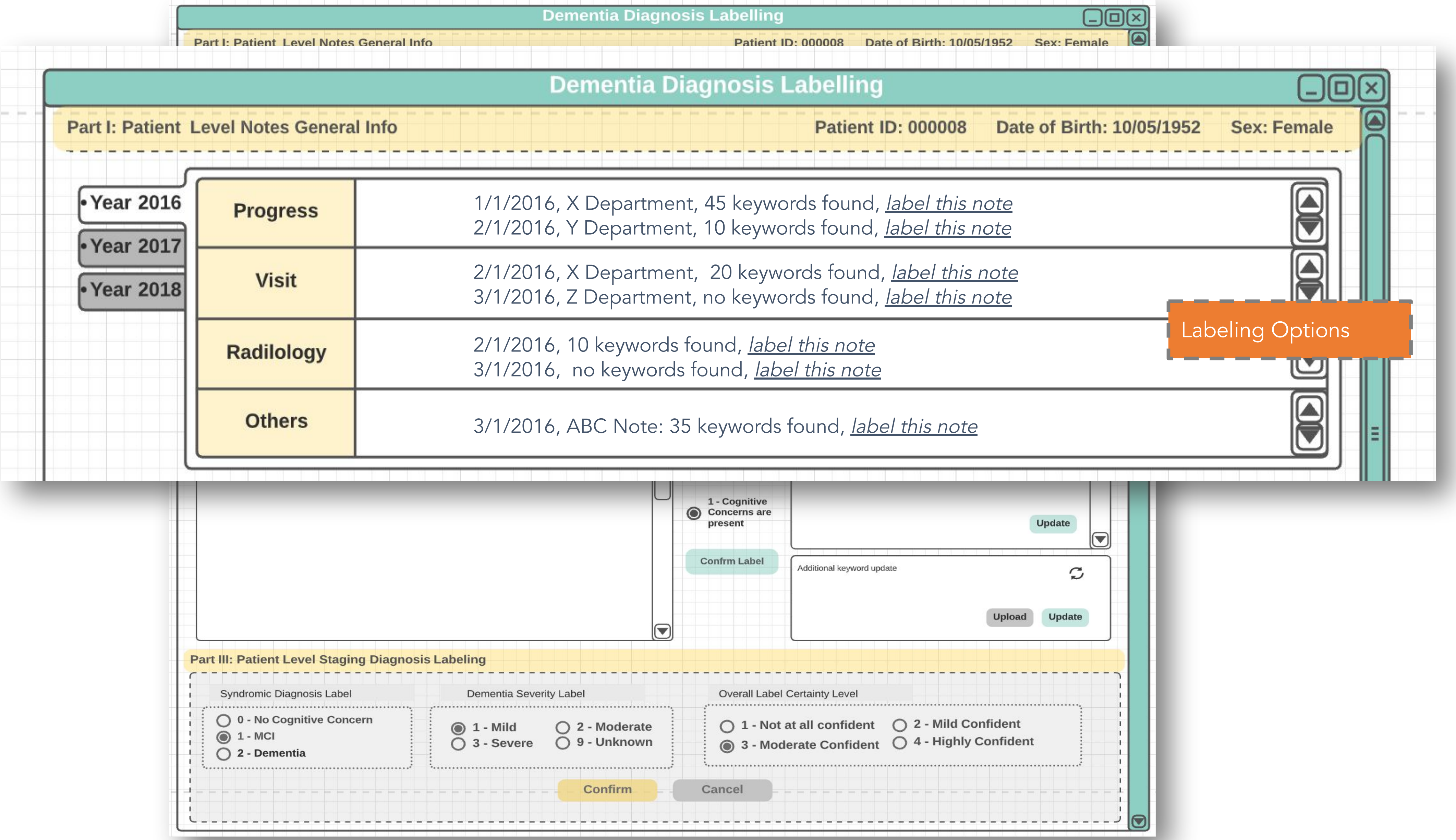
Use deep-learning model to predict labels. Model performance will be investigated to identify characteristics of poor performance

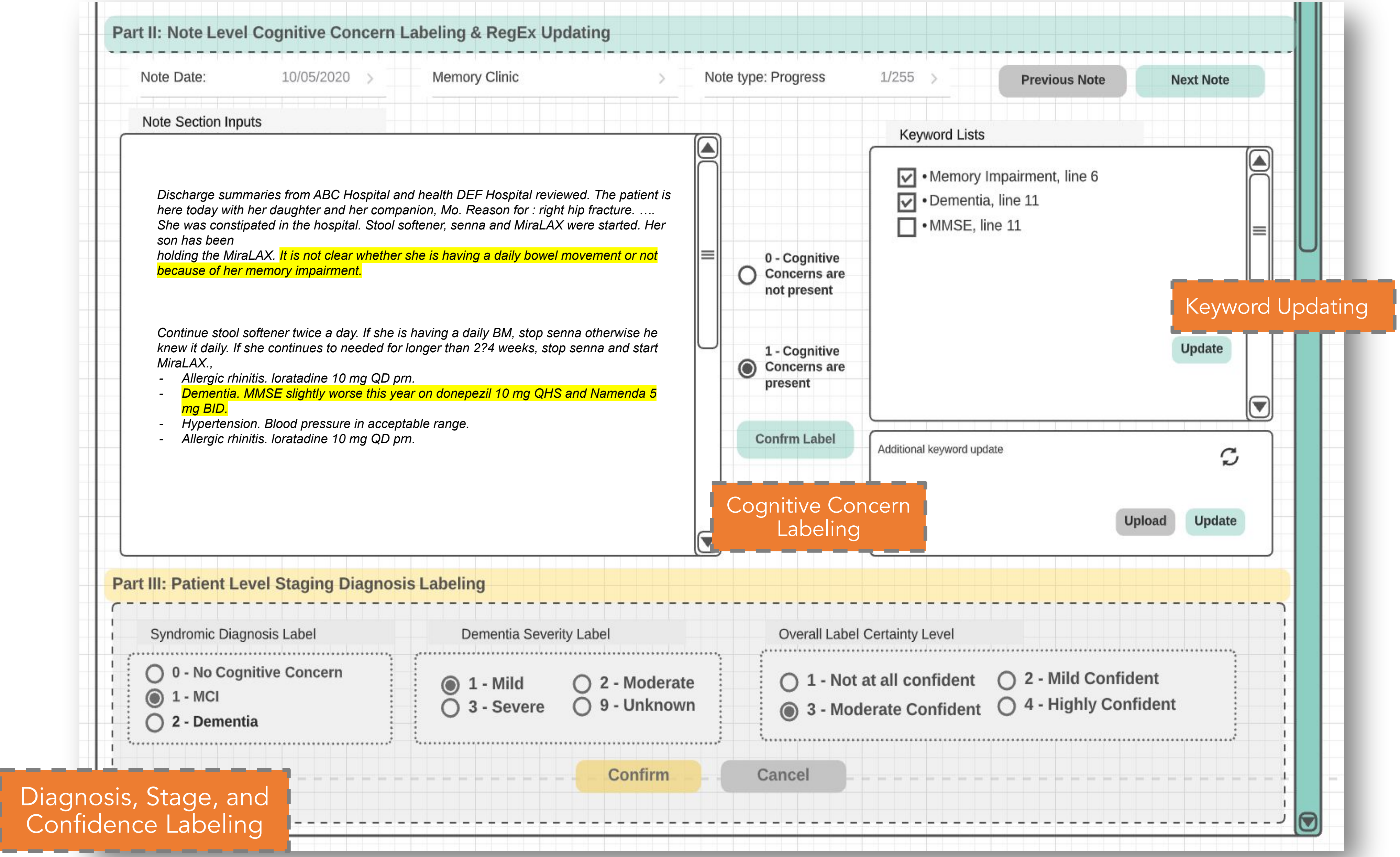


Query EHR database to generate another sample of patients

Patients predicted with poor confidence will be labeled by experts to generate expert-adjudicated labels

Annotation Tool – Assist Labeling Process





Highlights the keywords and provides context, such as the clinic and the provider, to facilitate clinical judgment.

Annotation tool has applications beyond this project. It can be used for efficient labeling and chart reviews.

Summary: How NLP can Help?

- Improve codes/meds-based case identification with linked EHR
- Better understanding of the predictive value of claims/meds-based case identification, and potential to understand nature and magnitude of biases in claims-based approach
- Improve efficiency of expert adjudication using new computer assisted annotation tool
- In clinical settings, NLP-based case identification could be used to identify patients (e.g., those with cognitive concerns who might benefit from cognitive screening or additional work-up)

Limitations of Study and EHR data

- Small training sample and very small test sample
- Sample is almost all white and beneficiaries of the Partners Accountable Care Organization (ACO), so with economic advantage
- EHR suffers from sporadic and missing data—missingness and non-missingness are both informative
- No data on those who don't seek care or whose decline is not noted at all
- Lots of patients will fall through the cracks…
- Limits to generalizability to other healthcare settings
- Not a population-based sample or a longitudinal study!

Future Directions

- Active learning to increase sample of gold-standard labels
- Combine structured data with NLP on clinician notes
- Classify stage of cognitive impairment
- Understand and quantify biases in EHR and Claims
- Identify predictors of biases

Acknowledgements

Zhuoqiao Mia Hong, **Colin G. Magdamo**, Yi-han Sheu, Prathamesh Mohite, Ayush Noori, Elissa M. Ye, Wendong Ge, Haoqi Sun, Laura Brenner, Greg Robbins, Shibani Mukerji, Sahar Zahar, Nicole Benson, Lidia Moura, **John Hsu**, Bradley T. Hyman, Michael B. Westover

Funding:

- Harvard Initiative on Aging Planning Grant
- National Institute on Aging
- Mass Life Sciences Internship Award