Welcome to the Sentinel Innovation and Methods Seminar Series

The webinar will begin momentarily

- Please visit www.sentinelinitiative.org for recordings of past sessions and details on upcoming webinars.
- Note: closed-captioning for today's webinar will be available on the recording posted at the link above.

Sentinel

Electronic Health Records (EHR) + Natural Language Processing + Machine Learning = Improved Sentinel Outcome Detection Algorithms

David Carrell
Susan Gruber

January 27, 2020

Sentinel Work Group Members*

FDA

Adebola Ajao
Robert Ball
Steven Bird
Sara Karami
Yong Ma
Michael Nguyen
Danijela Stojanovic
Mingfeng Zhang
Yueqin Zhao

Harvard Pilgrim

Adee Kennedy
Judy Maro
Mayura Shinde

Kaiser Washington

Maralyssa Bann
David Carrell
David Cronkite
James Floyd
Monica Fujii
Vina Graham
Kara Haugen
Ron Johnson
Jennifer Nelson
Mary Shea
Jing Zhou

Kaiser Northwest

Andrew Felcher
Brian Hazlehurst
Denis Nyongesa
Daniel Sapp
Matthew Slaughter

Putnam Data Science

Susan Gruber

Vanderbilt University

Cosmin (Adi) Bejan

HealthCore

Kevin Haynes

** alphabetical by last name*

Outline

1. Motivation and project goals -- anaphylaxis
2. Study design & EHR data
3. Natural language processing of clinician notes
4. Machine learned-models for outcome identification
5. Towards a General Framework

Outline

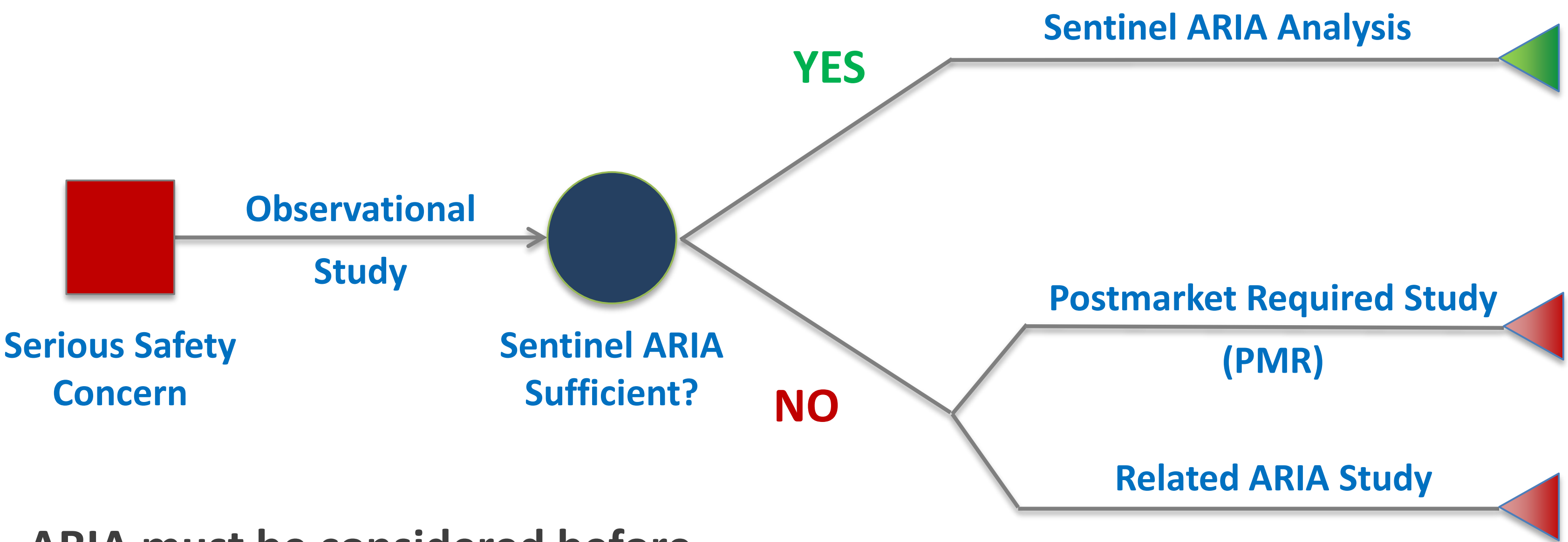
1. Motivation and project goals -- anaphylaxis
2. Study design & EHR data
3. Natural language processing of clinician notes
4. Machine learned-models for outcome identification
5. Towards a General Framework

What is ARIA?
(Active Risk Identification and Analysis)

- Electronic claims data, without manual medical record review
- Pre-defined, parameterized, and re-usable to enable faster safety surveillance (vs. protocol based assessments with fully customized programming)

slide courtesy of Michael Nguyen

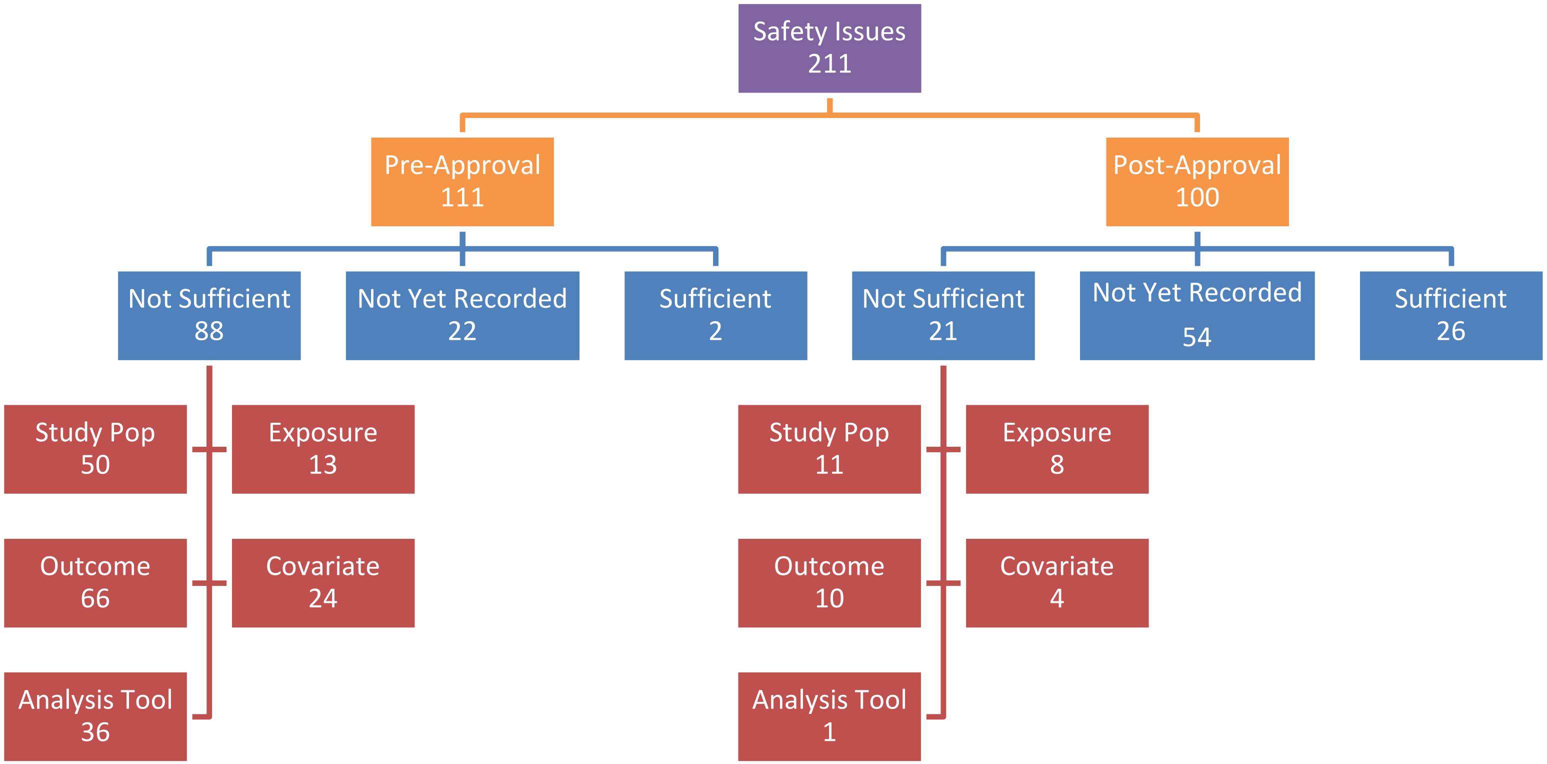
When is the ARIA Process Needed?



ARIA must be considered before a sponsor PMR can be issued

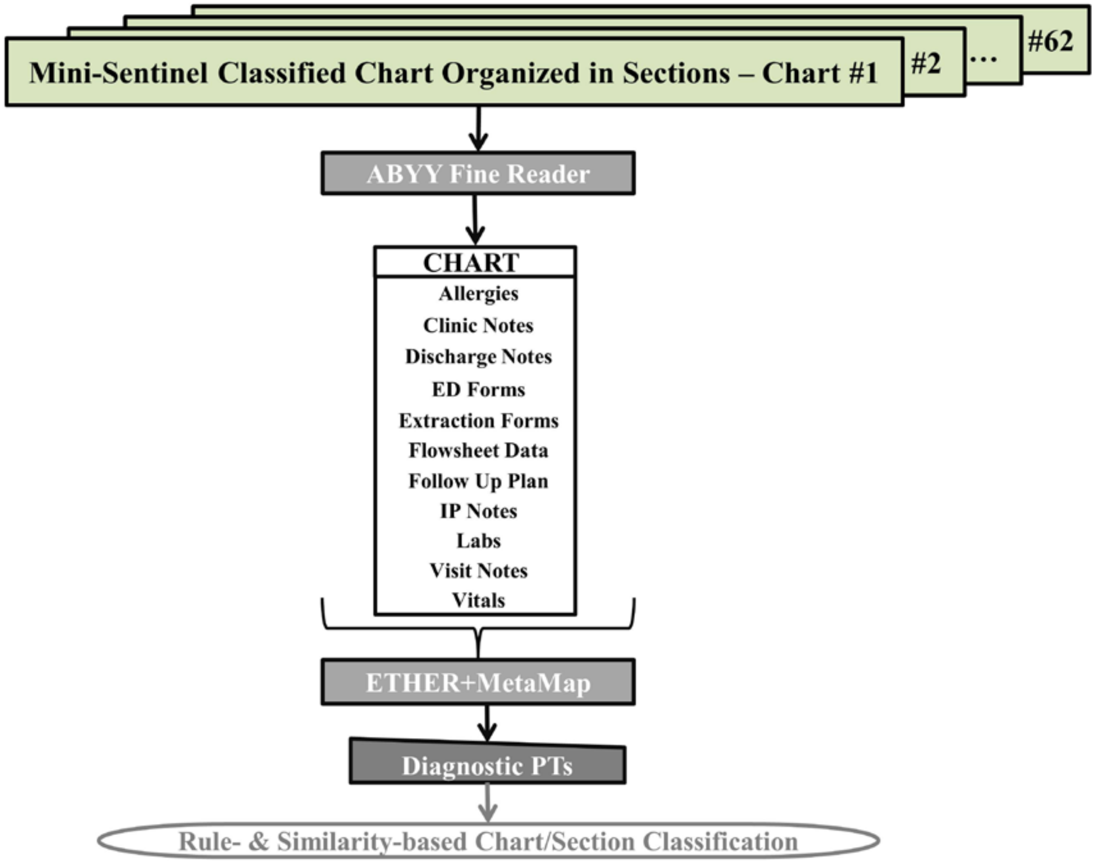
slide courtesy of Michael Nguyen

Reasons for ARIA Insufficiency



slide courtesy of Michael Nguyen

Health Outcome of Interest: Anaphylaxis



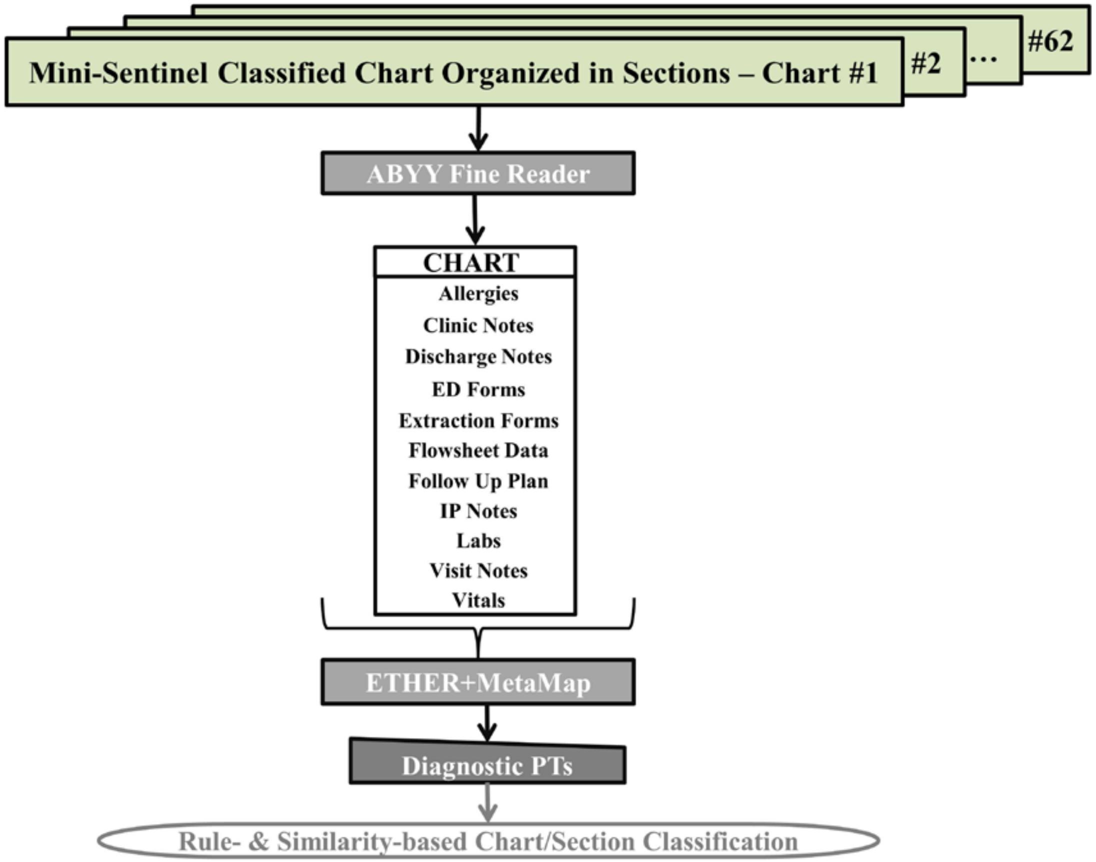
ETHER: Event-based Text-mining of Health Electronic Records; **ED**: Emergency Department; **IP**: Inpatient

KEY POINTS

- The previously developed natural language processing, rule- and similarity-based classification approaches demonstrated almost equal performance (F-measure: 0.753 vs. 0.729, recall 100% vs 100%, **precision 60.3% vs 57.4%**).*
- These algorithms might improve recall but had **similar precision (PPV 63.1% (95% CI: 53.9-71.7%)) to claims only algorithms from Mini-Sentinel**.**

*Ball R, Toh S, Nolan J, Haynes K, Forshee R, Botsis T. Evaluating automated approaches to anaphylaxis case classification using unstructured data from the FDA Sentinel System. Pharmacoepidemiology and Drug Safety, 27:1077–1084, 2018.
**Walsh KE et al. Validation of anaphylaxis in the Food and Drug Administration's Mini-Sentinel. Pharmacoepidemiology and drug safety 2013; 22: 1205–1213.

Health Outcome of Interest: Anaphylaxis



ETHER: Event-based Text-mining of Health Electronic Records; **ED**: Emergency Department; **IP**: Inpatient

KEY POINTS

- Reasons for misclassification included: **inability of the algorithms** the to make the same clinical judgments as human experts about the **timing, severity, or presence of alternative explanations; the identification of terms consistent with anaphylaxis but present in conditions other than anaphylaxis.**

Ball R, Toh S, Nolan J, Haynes K, Forshee R, Botsis T. Evaluating automated approaches to anaphylaxis case classification using unstructured data from the FDA Sentinel System. Pharmacoepidemiology and Drug Safety, 27:1077–1084, 2018.

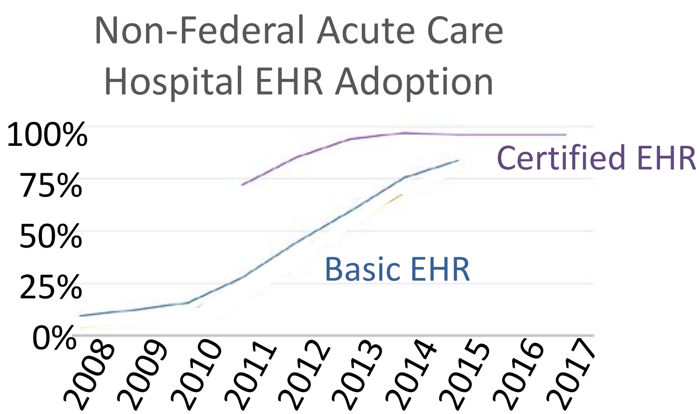
Project Goals

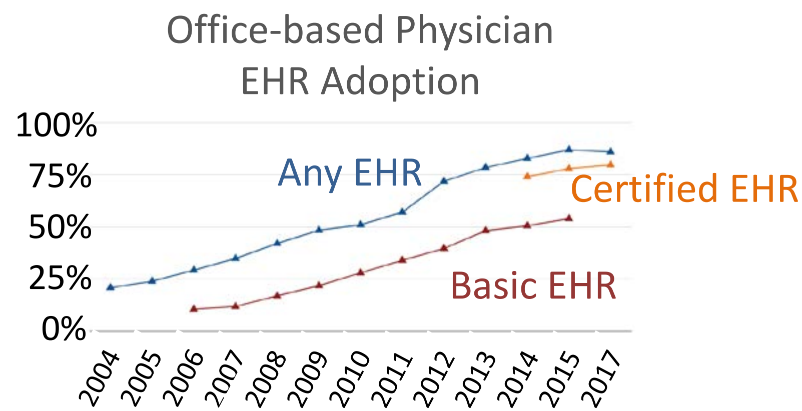
To improve classification accuracy for health outcomes of interest (HOIs) in Sentinel

- Create an outcome identification algorithm for anaphylaxis
  - extremely rare (~36 per 100,000 person-years)
  - complex diagnosis (clinical recognition of pattern of many symptoms
  - resembles similar conditions (severe allergic reaction)
  - accuracy of anaphylaxis diagnosis codes is poor (<2/3 are true cases)
- Develop a general framework
  - guiding principles for scaling up this methodology in Sentinel
  - tools for implementation

Key Areas of Innovation

- Leverage **EHR data**, including rich clinician notes to go beyond what is captured by structured data elements
  - Electronic health records (EHR) adoption approaching 100%*





- Extract relevant information with sophisticated natural language processing (**NLP**) methods
- Use advanced **machine learning** techniques for flexible modeling

*Office of the National Coordinator for Health Information Technology.
https://dashboard.healthit.gov/quickstats/pages/FIG-Hospital-EHR-Adoption.php Health IT Quick-Stat #47.
https://dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php Health IT Quick-Stat #50.

Outline

1. Motivation and project goals -- anaphylaxis
2. Study design & EHR data
3. Natural language processing of clinician notes
4. Machine learned-models for outcome identification
5. Towards a General Framework

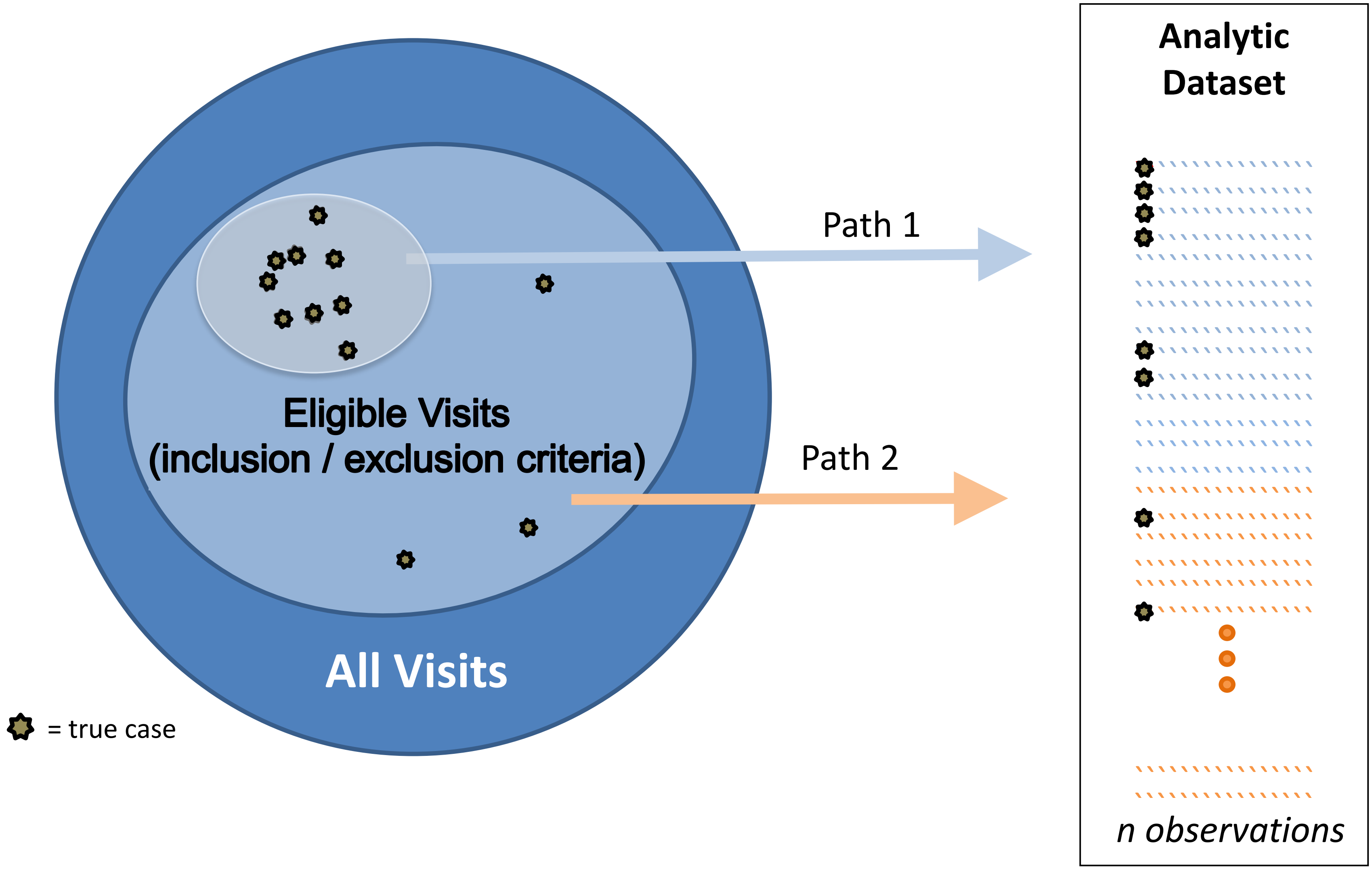
Anaphylaxis study design

- Study period: October 2015 – December 2018
- Population: Adults & children at KPWA or KPNW* w/either:
    1. Inpatient or ED anaphylaxis diagnosis code
    2. Outpatient anaphylaxis diagnosis code
    3. Angioedema, urticaria, or adverse effect of medication code (inpatient or ED)
- Gold standard outcome labels (via manual chart review)
- Structured covariates (features) defined by clinical experts
    - demographics, prescriptions, other diagnoses, procedures, etc.
- NLP-derived covariates from clinical notes corpora

*KPWA = Kaiser Permanente Washington
KPNW = Kaiser Permanente Northwest
Eligible patient had >1 year of continuous enrollment & no anaphylaxis code in 12 months prior

Stratified Random Sampling

Goal is to sample enough cases, while ensuring the analytic dataset faithfully represents the source population



Outline

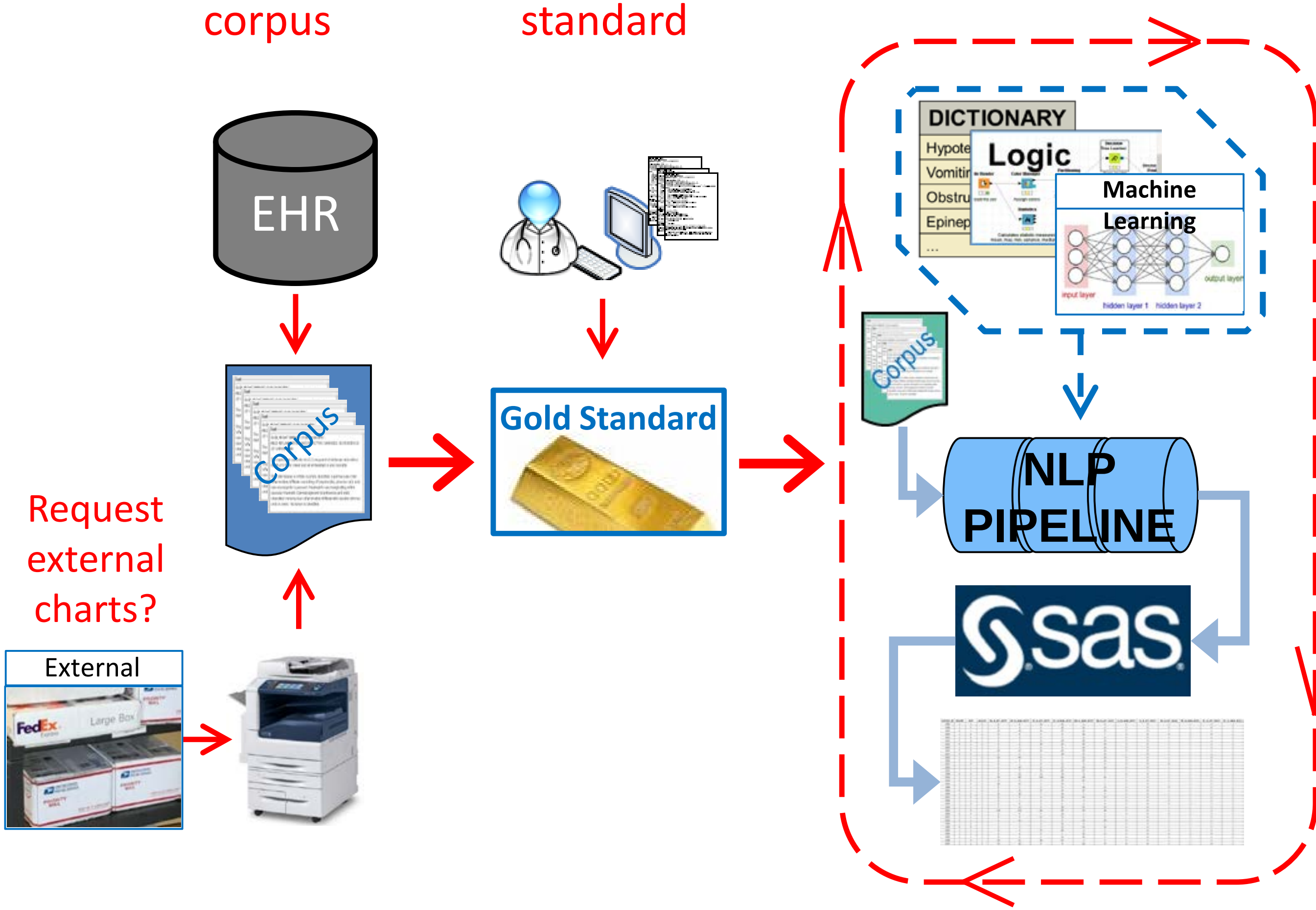
1. Motivation and project goals -- anaphylaxis
2. Study design & EHR data
3. Natural language processing of clinician notes
4. Machine learned-models for outcome identification
5. Towards a General Framework

Natural language processing (NLP): *Objective*

- Guided by clinical domain knowledge and …
- Combining methods from computer science, artificial intelligence and computational linguistics …
- Derive information from unstructured clinical text and represent it as structured "features" for use in …
- Developing automated algorithms to …
- Identify exposures, covariates, or outcomes of interest.

NLP: General approach

1) Assemble corpus
2) Create gold standard
3) Engineer NLP features



Manual approach:

- Gold standard
- Manual chart review
- Manual dictionary creation
- Manually-engineered features
- Manual or machine-learned algorithms

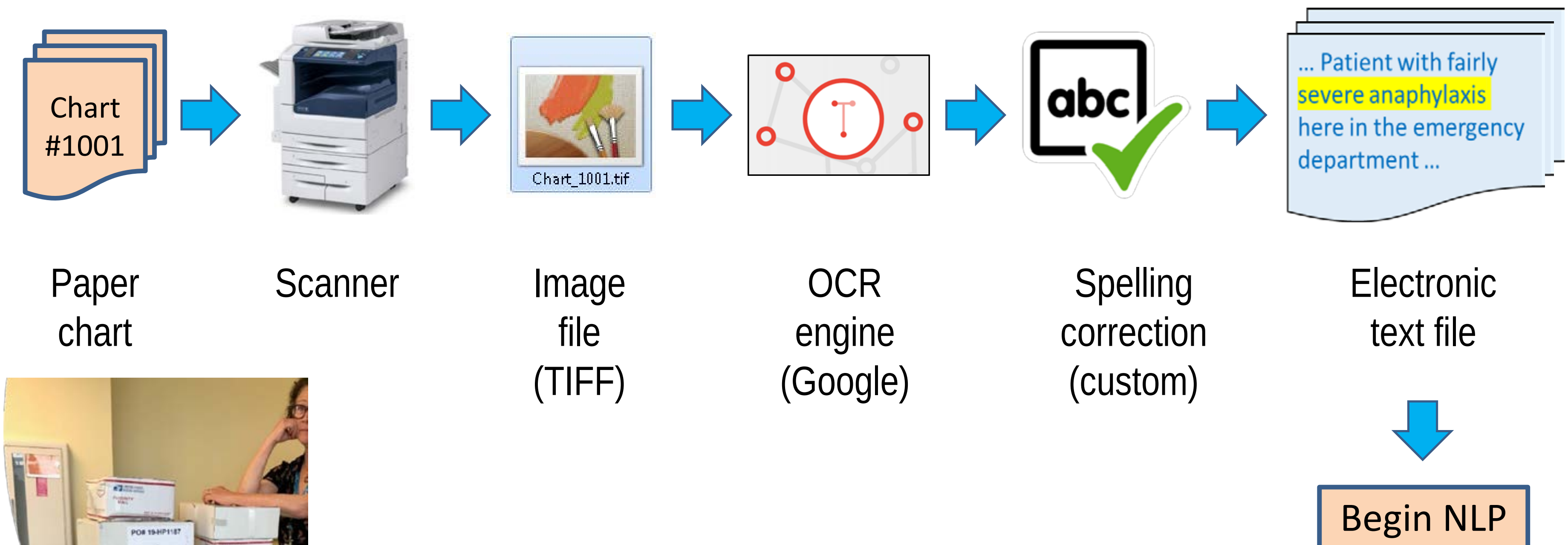
Automated approach:

- Silver standard
- Published knowledge sources
- Automated dictionary creation
- Automated feature engineering
- Automated algorithm development

Automated approach:
Yu et al. Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources JAMIA 2015;22:993–1000.
Yu et al. Surrogate-assisted feature extraction for high-throughput phenotyping. JAMIA 2017; e143–e149.
Yu et al. Enabling phenotypic big data with PheNorm. JAMIA 2018; 25(1), 54–60.

NLP: Electronic text from paper charts (OCR)

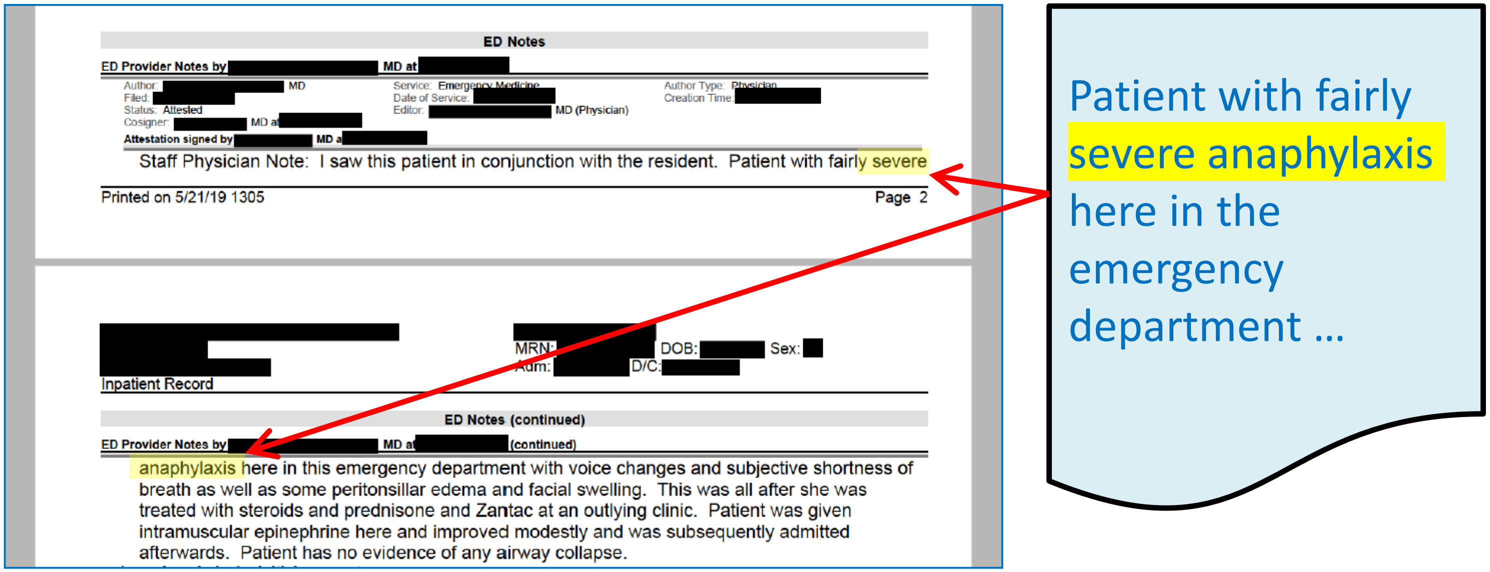
OCR: Optical Character Recognition



NLP: Electronic text from paper charts (OCR)

OCR-induced noise

- Character errors (“`bowel`” → “`bowe!`”)
- Page breaks



NLP: Manual gold standard creation

- KPWA (site #1)
  - Dual blind manual clinician review
  - Decisions recorded on spreadsheet
- KPNW (site #2)
  - Dual blind manual non-clinician abstractors follow protocol to populate REDCap form
  - Abstractors decide “easy” cases (10% MD QC review)
  - Clinician adjudicates difficult cases
  - REDCap useful for QA, data management, data preservation

NLP: Dictionary creation

1. Manual review of charts by clinician & informaticist
2. Exploratory query of clinical notes
3. Synonyms from dictionaries
4. Clinical knowledge sources
   - Automated Feature Extraction for Phenotyping (AFEP)

Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources

RECEIVED 24 October 2014
REVISED 25 February 2015
ACCEPTED 24 March 2015
PUBLISHED ONLINE FIRST 30 April 2015

Sheng Yu[1,2,3,*], Katherine P Liao[2,3], Stanley Y Shaw[4], Vivian S Gainer[5], Susanne E Churchill[5], Peter Szolovits[6], Shawn N Murphy[4,5], Isaac S Kohane[3,7], Tianxi Cai[8]

AMIA
OXFORD UNIVERSITY PRESS

NLP dictionary: 1. Manual review of charts

- Clinician & informaticist review, discuss, mark-up charts
  - "Think aloud" protocol
  - ~50 charts

Nose: No rhinorrhea.
Mouth: Mild swelling
Neck: Nontender, supple, no lymphadenopathy
Lymphatic: No lymphadenopathy noted.
Cardiovascular: Normal heart rate, normal rhythm, no murmurs, no rubs, no gallops. Intact distal pulses, no tenderness, no cyanosis, no clubbing.
Respiratory: Normal breath sounds, no respiratory distress, no wheezing, no chest tenderness. No severe stridor, severe wheezing
Abdomen: Bowel sounds are present. Abdomen is soft, no tenderness, no masses, no rebound or guarding. No organomegaly. No hernia.
GU: No CVA tenderness. Bladder is nontender and not distended.
Skin: Erythema noted about the face and minimally to the hands
Back: No tenderness
Musculoskeletal: No tenderness to palpation or major deformities noted. No back or cervical spine tenderness. Nc edema.

Pt after her CTA ABdomen she develop allergic /anaphylactic reaction in ED with nausea/vomting and tachycardia and hypotensive and she became hypoxic ,even so she had many ct with contrast without any reactions

She received multiple rounds of epinephrine , benadryl ,decadron ,pepcid

SHE FEEL MUCH BETTER NOW except some dizziness when she walk

- Dictionary terms are:
  - Clinically important (distinguish cases from non-cases)
  - Feasible for NLP

NLP dictionary: 2. Exploratory query

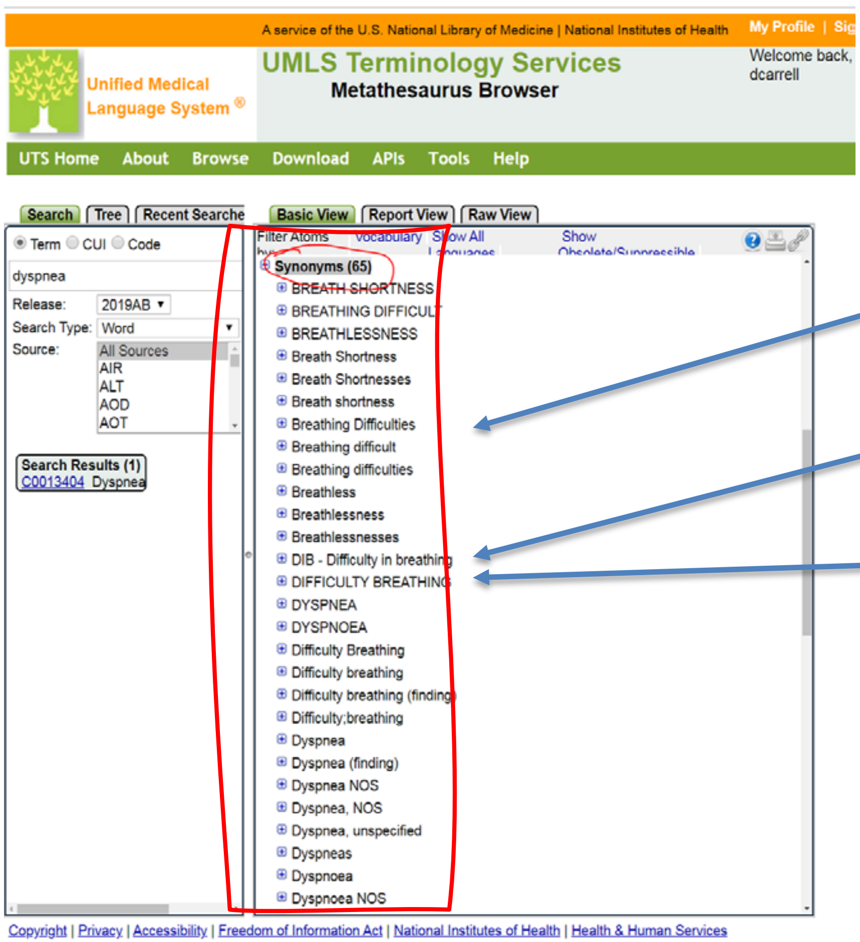
- Use relational database full-text indexing
- Find Synonyms of “dyspnea”
  - Known: “shortness of breath” and “trouble breathing”
  - Review notes with breath
    - 208 strings yield 5 new terms

Before_Term	Term	After_Term
was closing and wheezing and difficulty	breath	ing. She has some mild reactive airway d
and throat swelling. Having difficulty	breath	ing and a hard time swallowing saliva. W
rhythm. RESP: Clear to auscultation.	breath	ing comfortably. Jerico endorses feel
like this before. Feels like she cannot	breath	. Cannot swallow. Has not taken anything
omplaint: Allergic Reaction; Edema; and	breath	ing Problems HISTORY AND PHYSICAL E
tightening and it was a little hard to	breath	e so comes here for evaluation where she
ing Swelling around eyes, tears, no	breath	ing problems • Lovastatin • Sulfa (
en he began to cry and said he couldn't	breath	. He sent Mom a picture of his face- she
the first time. Pt apparently stopped	breath	ing briefly, was given epinephrine and a

NLP dictionary: 3. Synonyms

UMLS: Unified Medical Language System – Metathesaurus

“Dyspnea”



“breathing difficulties”

“DIB”

“difficulty in breathing”

…

NLP dictionary: Clinical knowledge sources

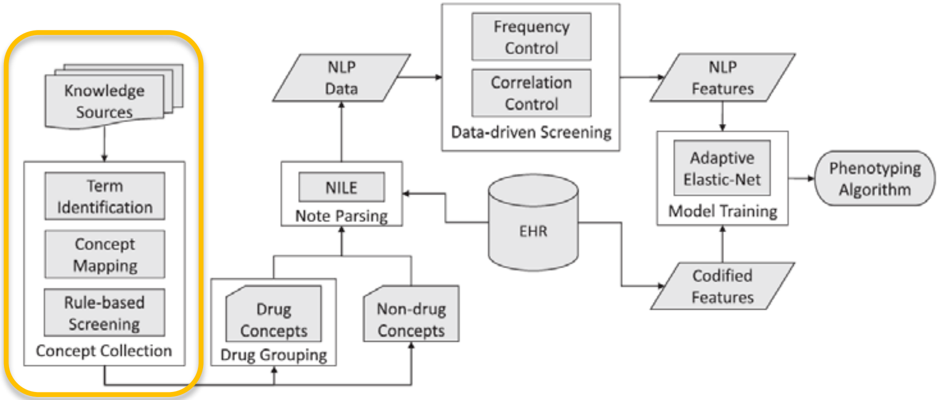
- 1st step in Yu and colleagues 2015 JAMIA paper “AFEP”

Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources

RECEIVED 24 October 2014
REVISED 25 February 2015
ACCEPTED 24 March 2015
PUBLISHED ONLINE FIRST 30 April 2015

AMIA OXFORD UNIVERSITY PRESS

Sheng Yu[1,2,3,*], Katherine P Liao[2,3], Stanley Y Shaw[4], Vivian S Gainer[5], Susanne E Churchill[5], Peter Szolovits[6], Shawn N Murphy[4,5], Isaac S Kohane[3,7], Tianxi Cai[8]



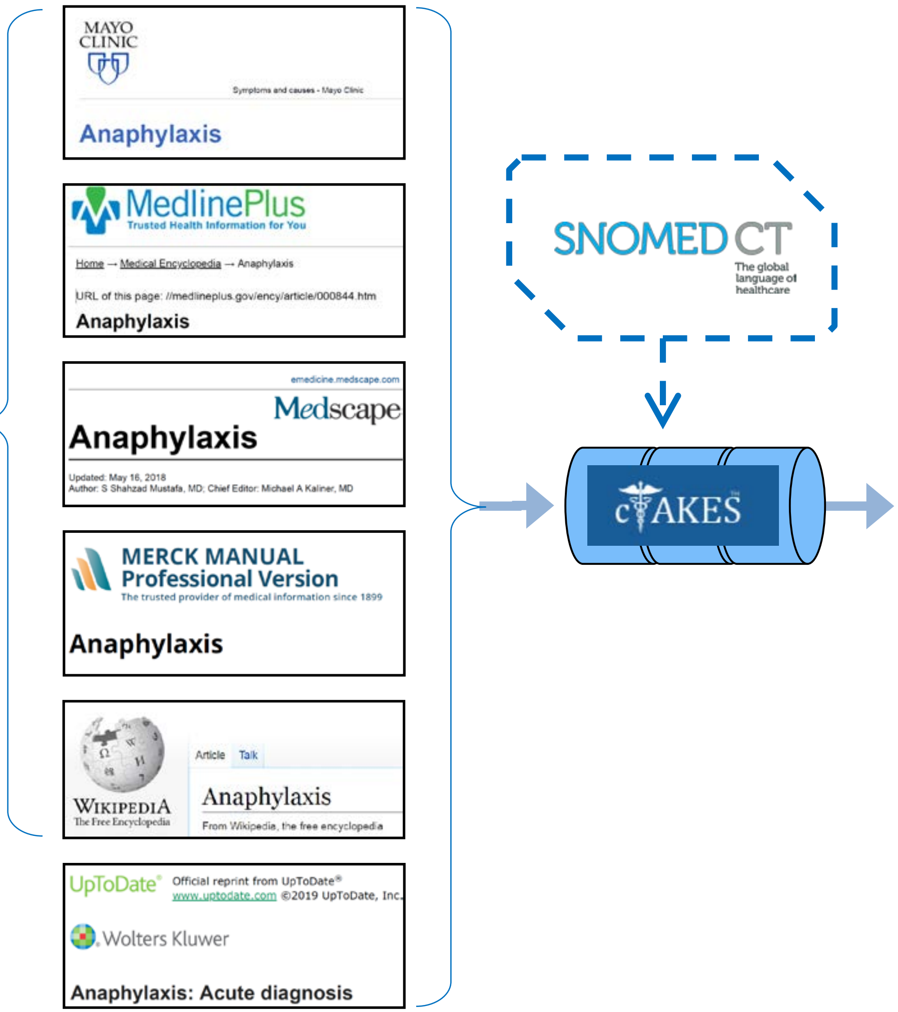
- Important terms will appear in ≥3 clinical knowledge base articles

Yu et al. Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources JAMIA 2015;22:993–1000.

NLP dictionary: Clinical knowledge sources

5 clinical knowledge base articles on the topic anaphylaxis

(+ UpToDate)



	Source	CUI_Code	Term
1	SNOMEDCT_US	C0663655	abacavir
2	SNOMEDCT_US	C0000726	Abdomen
3	SNOMEDCT_US	C1122087	adalimumab
4	SNOMEDCT_US	C0001443	Adenosine
5	SNOMEDCT_US	C3536832	Air
6	SNOMEDCT_US	C0001927	Albuterol
7	SNOMEDCT_US	C0002055	Alkalies
8	SNOMEDCT_US	C0002092	Allergens
9	SNOMEDCT_US	C0002508	Amines
10	SNOMEDCT_US	C0002575	Aminophylline
11	SNOMEDCT_US	C0002667	Amphetamines
12	SNOMEDCT_US	C0002771	Analgesics
13	SNOMEDCT_US	C0002792	anaphylaxis
14	SNOMEDCT_US	C0002932	Anesthetics
15	SNOMEDCT_US	C0002994	Angioedema
16	SNOMEDCT_US	C0003018	Angiotensins
17	SNOMEDCT_US	C0003232	Antibiotics
18	SNOMEDCT_US	C0003241	Antibodies
19	SNOMEDCT_US	C0003320	Antigens
20	SNOMEDCT_US	C0003360	Antihistamines
21	SNOMEDCT_US	C0003445	Antitoxins
22	SNOMEDCT_US	C0003450	Antivenin
23	SNOMEDCT_US	C0003467	Anxiety
24	SNOMEDCT_US	C0003483	Aorta
25	SNOMEDCT_US	C0003564	Aphonia
26	SNOMEDCT_US	C0233485	apprehension
27	SNOMEDCT_US	C0003842	Arteries
28	SNOMEDCT_US	C0004044	Asphyxia
29	SNOMEDCT_US	C0004057	Aspirin
30	SNOMEDCT_US	C1510438	Assay
31	SNOMEDCT_US	C0004096	Asthma
32	SNOMEDCT_US	C0231221	Asymptomatic
33	SNOMEDCT_US	C0392707	Atopy
34	SNOMEDCT_US	C0004259	Atropine
35	SNOMEDCT_US	C0004268	Attention
36	SNOMEDCT_US	C0004271	Attitude
37	SNOMEDCT_US	C0004398	Autopsy
38	SNOMEDCT_US	C0004521	Aztreonam
39	SNOMEDCT_US	C0004827	Basophils
40	SNOMEDCT_US	C0005558	Biopsy
41	SNOMEDCT_US	[illegible]	

367 unique SNOMED terms

90 terms appear in ≥3 sources

NLP dictionary: Clinical knowledge sources

90 terms in the Standard Nomenclature of Medicine, Clinical Terms (SNOMED CT) appeared in at least 3 anaphylaxis knowledge base articles on anaphylaxis.

Appearing in 5-6 articles		Appearing in 4 articles	Appearing in 3 articles	
Allergens	Blood	Angioedema	Air	Lung
Anaphylaxis	~~Cells~~[1]	Anxiety	Albuterol	Muscle
~~Diagnosis~~[1]	Dizziness	Atopy	Antigens	omalizumab
Diarrhea	Dyspnea	Basophils	Arteries	Ovum
~~Disease~~[1]	Exercise	Coughing	Asphyxia	Oxygen
Epinephrine	Heart	Edema	Autopsy	Panic
Hypersensitivity	Histamine	Esthesia	Chest	Proteins
Shock	Hypotension	Flushing	~~Complication~~[1]	receptor
Skin	Injection	Glucagon	Confusion	Redness
Urticaria	Latex	Hoarseness	Congestion	Seizures
Venoms	Nausea	Mastocytosis	Extravasation	~~Services~~[1]
Vomiting	Obstruction	Nose	Eye	~~Source~~[1]
Wheezing	Pain	Opioids	~~Gold~~[2]	Uterus
Abdomen	Palpitations	Rhinorrhea	Headache	Vaccines
Antibiotics	Pruritus	Stridor	Immunoglobulins	Vancomycin
Antibodies	Swelling	Tachycardia	Immunotherapy	Vasodilation
Antihistamines	Syncope	Tryptase	Lactams	Veins
Aspirin	Tongue		Larynx	
Asthma			Lightheadedness	
37 terms (13 in 6 and 24 in 5)		17 terms	36 terms	

[1] Terms unlikely to be useful for distinguishing anaphylaxis cases from non-cases.

[2] "Gold" is an author name appearing in 3 bibliographies (N Engl J Med 2008; 358:28).

NLP: Dictionary for anaphylaxis

- 843 unique terms
  - ~Half are for skin/mucosal involvement
- Median of 128 concepts extracted per chart (range: 9-2,092)

ID	CUI	TEXT	SOURCE	SOURCETYPE
3001	GI001	abd pain	GI	ABDOPAIN
6001	SM001	abdomen with erythema	GI	ABDOPAIN
3002	GI002	abdominal pain and shock	GI	ABDOPAIN
2001	BP001	acute hypotensive	BPREDUCED	HYPOTENSION
5001	RC001	acute hypoxic	RESPCOMP	HYPOXIA
5002	RC002	acute respiratory failure	RESPCOMP	RESPFAIL
5003	RC003	acute upper airway obstruction	RESPCOMP	AIRWAY
4001	OT001	admission diagnosis	OTHER	DIAGNOSIS
4002	OT002	admitting diagnosis	OTHER	DIAGNOSIS
5004	RC004	airway narrowing	RESPCOMP	AIRWAY CONSTRICTION
5005	RC005	airway obstruction	RESPCOMP	AIRWAY CONSTRICTION
6002	SM002	airway itch	SKINMUC	AIRWAY
6003	SM003	airway remains swolen	SKINMUC	ORALSWELL
6004	SM004	airway remains swollen	SKINMUC	AIRWAY
4003	OT003	alergic reacton	OTHER	ALLERGREACT
6005	SM005	all skin appears red	SKINMUC	RASH
4004	OT004	allergic reaction	OTHER	ALLERGREACT
4005	OT005	allergic reacton	OTHER	ALLERGREACT
4006	OT006	allergic to	OTHER	HYPO
4007	OT007	allergies	OTHER	HYPO
4008	OT008	allergy comment	OTHER	HYPO
2002	BP002	almost passed out	BPREDUCED	SYNCOPE
5006	RC006	altered mentation	RESPCOMP	ALTERED MENTATION
1001	AN001	anaphalytic shock	ANAPH	ANAPH SHOCK
1002	AN002	anaphylactic shock	ANAPH	ANAPH SHOCK
1003	AN003	anaphylaxis allergic shock	ANAPH	ANAPH SHOCK
4009	OT009	anaphylaxis	OTHER	ANAPH
2003	BP003	and hypotensive	BPREDUCED	HYPOTENSION
2004	BP004	and passed out	BPREDUCED	SYNCOPE
2005	BP005	and shock	BPREDUCED	SHOCK
6006	SM006	angioedema	SKINMUC	ANGIOEDEMA
1004	AN004	aphylactic shock	ANAPH	ANAPH SHOCK
6007	SM007	areas of erythema	SKINMUC	ERYTHEMA
6008	SM008	arms with erythema	SKINMUC	ERYTHEMA
2006	BP006	arrhythmia	BPREDUCED	CARDIACARRHYTH

NLP: Feature engineering (manual)

- Strategy: operationalize features expected to help distinguish true cases from non-cases
  - Sampson NIAID diagnostic criteria for anaphylaxis
  - 72 clinical concepts (from “abdominal pain” to “wheeze”)
  - Special features (e.g., admitted for observation, explicit dx)
  - Treatments (e.g., *epinephrine,* also via structured data)
  - Exposures associated with anaphylaxis (*structured data only*)
  - Competing diagnoses (*structured data only*)
- Rules of engagement
  - Do <u>not</u> use gold standard case status to improve engineering of features (reserve gold standard data for modeling)

NLP: Feature engineering (manual)

Anaphylaxis NLP features for Sampson/NIAID diagnostic criteria.		
Sampson Criterion	**Clinical criteria**	**NLP Features**
#1	Skin/mucosal involvement (SM), *plus either:* Respiratory compromise (RC) *or* Reduced blood pressure (BP)	SM+RC SM+BP
#2	Exposure to a likely allergen *for that patient*[1] *plus any 2:* Skin/mucosal involvement (SM) *or* Respiratory compromise (RC) *or* Reduced blood pressure (BP) *or* Gastrointestinal symptoms (GI)	~~SM+RC~~[2] ~~SM+BP~~[2] SM+GI RC+BP RC+GI BP+GI
#3	Exposure to a known allergen *for that patient*[1] *plus:* Reduced blood pressure (BP)	None[3]

1. Allergen exposure not operationalized because too difficult to do accurately via NLP.
2. This combination not included in criterion #2 because already in criterion #1.
3. Not operationalized because w/o allergen exposure reduced BP is non-specific.

Sampson HA, Muñoz-Furlong A, Campbell RL, et al. Second symposium on the definition and management of anaphylaxis: summary report – second national Institute of allergy and infectious disease/food allergy and anaphylaxis network symposium. J Allergy Clin Immunol. 2006;117:391–397

NLP: Feature engineering (manual)

- Illustrative rules …
  - "Rule of 2" (e.g., ≥2 mentions of "anaphylaxis")
  - Counts of key terms (e.g., N mentions of "airway restrictions")
    … count of *all* mentions
    … count of *affirmative* mentions
    … *normalize* counts …
  - Binary flags for mentions of any individual concepts
  - Binary flags for mentions in anaphylaxis symptom groups:
    1) Reduced BP, 2) GI, 3) Respiratory Comp., 4) Skin/Mucosal [5) Other]
  - Combinations that satisfy Sampson NIAID diagnostic criteria
  - Require multiple concepts in a short span of text
- Example: Any combination of terms satisfying NIAID criteria *plus* an explicit anaphylaxis diagnosis

NLP: Feature engineering (manual)

Subgroups of anaphylaxis concepts in the NLP dictionary (N unique terms).

- BRADYCARDIA (13)
- CARDIACARRHYTH (8)
- CARDIOCOLLAPSE (2)
- COLLAPSE (2)
- END ORGAN (2)
- HYPOTENSION (77)
- PALPITATIONS (3)
- SHOCK (3)
- SYNCOPE (30)
- TACHYCARDIA (9)
- ABDOPAIN (3)
- VOMIT (1)
- AIRWAY (4)
- AIRWAY CONSTRICTION (4)
- ALTERED MENTATION (1)
- APHONIA (3)
- BREATH (6)
- BRONCHOSPASM (1)
- CHEST DISCOMFORT (2)
- CHEST TIGHTNESS (9)
- COARSE BREATH SOUND (4)
- DYSPHONIA (1)
- DYSPNEA (55)
- HOARSENESS (7)
- HYPOXEMIA (6)
- HYPOXIA (3)
- IMPENDING DOOM (2)
- INTUBATION (6)
- LARYNGEAL OEDEMA (1)
- RESP COMPROMISE (3)
- RESP DISTRESS (2)
- RESPFAIL (1)
- RONCHI (2)
- STRIDOR (3)
- TACHYPNEA (5)
- THROAT CLOSURE (14)
- THROAT TIGHTNESS (34)
- TIGHTNESS BREATHING (1)
- VOICE QUALITY (1)
- WHEEZE (8)
- ANGIOEDEMA (102)
- DIFFICULTY SWALLOWING (14)
- DYSPHAGIA (1)
- EDEMA (4)
- ERYTHEMA (42)
- EYE SWELLING (33)
- FACIAL SWELLING (20)
- FLUSH (38)
- HIVES (68)
- ITCHING (14)
- ITCHY SOFT TISSUE (15)
- METALLIC TASTE (1)
- MOUTH (1)
- MOUTHSWELL (4)
- ORALSWELL (4)
- PRURITUS (15)
- RASH (7)
- REACTION (1)
- SOFT TISSUE SWELLING (4)
- SWELLING (31)
- THROAT (4)
- TINGLING (1)
- TINGLY SOFT TISSUE (14)
- URTICARIA (24)
- ALLERGREACT (5)
- ANAPH (5)
- COMPLAINT (12)
- DIAGNOSIS (8)
- DIFFERENTIAL (1)
- HYPO (6)
- IMPRESSION (1)

Groups: ● **REDUCED BLOOD PRESSURE** ● **GASTROINTESTINAL** ● **RESPIRATORY COMPROMISE** ● **SKIN/MUCOSAL** ● **OTHER**

NLP: Feature engineering (manual)

- "Special features" expected to be helpful
  - Explicit diagnoses ("`PRIMARY DX: Anaphylaxis`")
  - Rapid decline ("`collapsing`" or "`getting worse`")
  - Suddenness term near symptom term ("`rapid swelling`")
  - Need for observation ("`admit for observation`")
  - Having an epinephrine prescription

NLP: Feature engineering (manual)

- Summary of operationalized NLP features
  - 471 total features (many expected not to be of value)
  - Top 100 selected by informaticist as “best NLP features”
  - 25 selected by clinicians (includes 16 *not* in top 100)
  - 116 NLP features in analytic data set (top 100 + clinicians’)

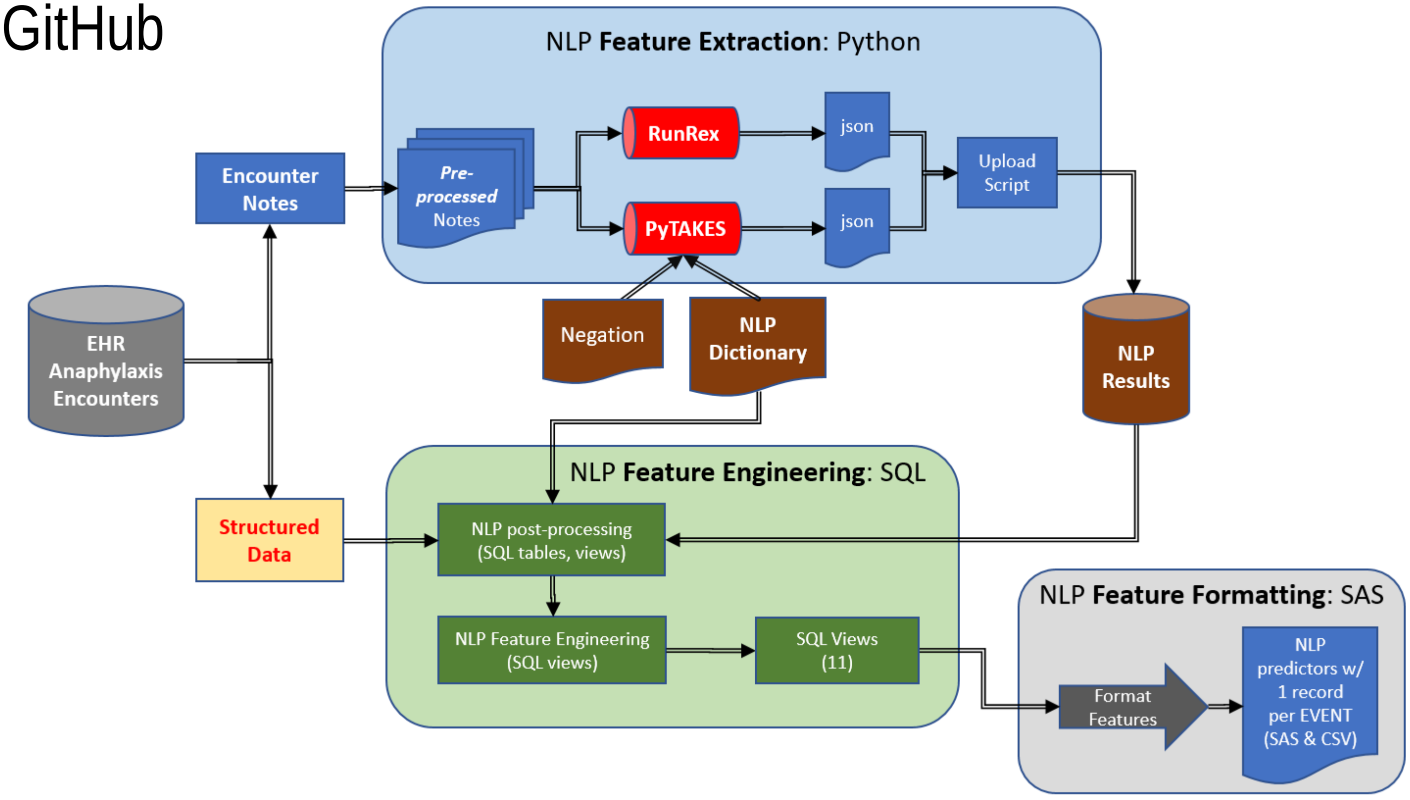
Counts of NLP features by feature engineering strategy.	
Anaphylaxis-related organ systems	10
66 anaphylaxis-related concepts	66
NIAID diagnostic criteria (combinations of organ systems)	30
Anaphylaxis terms	5
Special features (e.g., admit…)	5
TOTAL:	116

Structured data features

- 47 variables curated by clinicians + informaticists
- Structured feature categories
  - High-risk exposures (e.g., imaging dye, immunotherapy)
  - Competing diagnoses (asthma, COPD, serious infection)
  - Treatments (medications, procedures such as CPR)
  - Follow-up with immunology/allergist 45 days post index date
  - Type of anaphylaxis dx (e.g., food-related, venom, medicine)
  - History of anaphylaxis, allergic reaction
  - ED vs IP vs OP presentation setting
  - Demographics

Porting the NLP & structured data code to KPNW

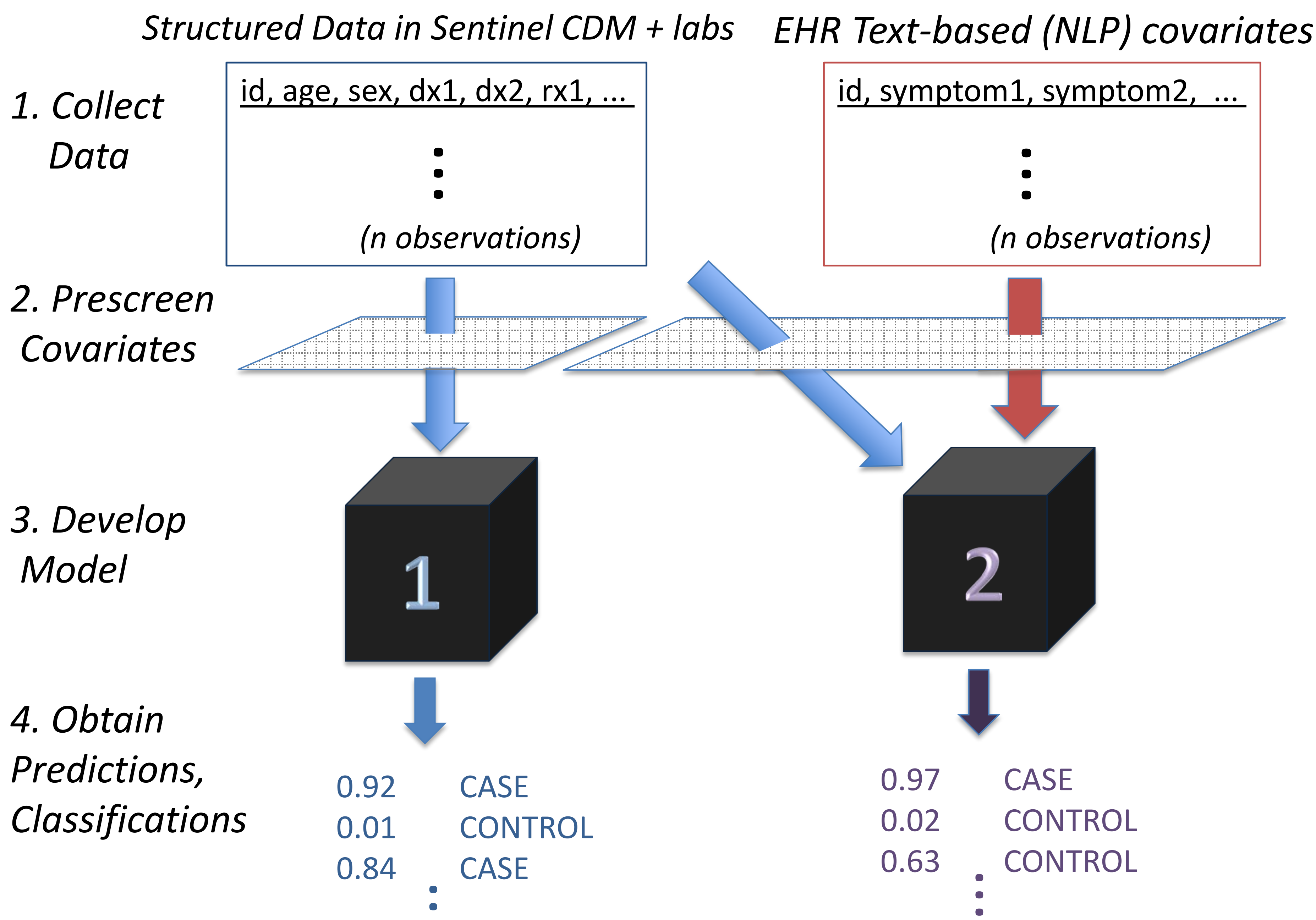
- Code package: NLP system (Python), SQL queries, SAS
- Documentation
- GitHub



Outline

1. Motivation and project goals -- anaphylaxis
2. Study design & EHR data
3. Natural language processing of clinician notes
4. Machine learned-models for outcome identification
5. Towards a General Framework

Model Development

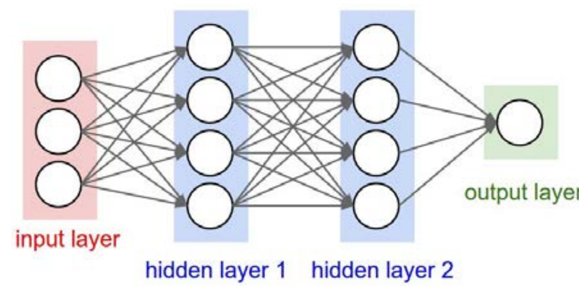


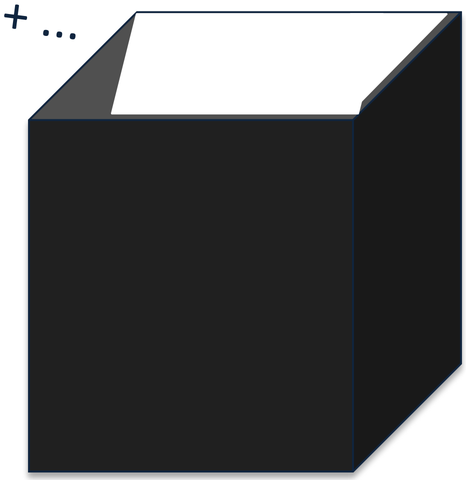
What’s in the box?

- Logistic regression
- Elastic net
- Bayesian Additive Regression Trees
- Neural network
- Boosted Trees

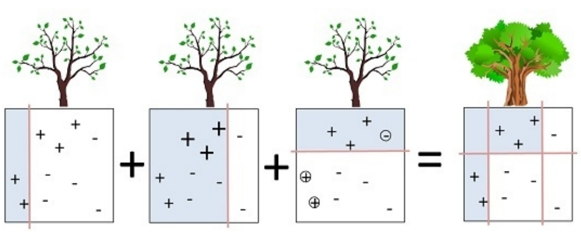
Super Learner
(a weighted combination)

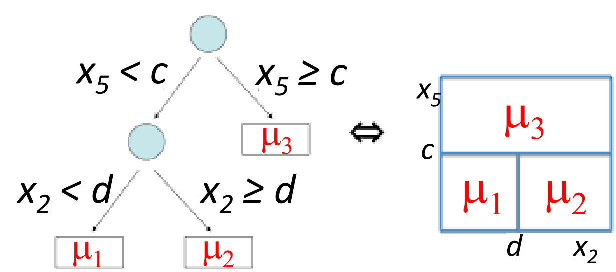
$\beta_0 + \beta_1 * age + \beta_2 * ICD10 + \ldots$





Boosted Regression Tree is a hierarchical and supervised machine learning method that combines weak learners (binary splits) to strong prediction rules that allow a flexible partition of the feature space.





75 Models

Algorithm	R package name	Notes on tuning parameters
1. Logistic regression	(base)	
2. Elastic net	glmnet	10-fold cross validation to select optimal alpha and lambda
3. Gradient boosting	xgboost	Variant 1: maximum tree depth = 2 Variant 2: maximum tree depth = 4
4. Bayesian Additive Regression Trees	dbarts	Variant 1: k = 2 (default), Variant 2: k=1 (reduced regularization prior)
5. Neural network (feed forward)	neuralnet	Variant 1: 1 hidden layer containing 1 node Variant 2: 1 hidden layer containing 3 nodes
6. Super Learner	SuperLearner	AUC-based calculation of the optimal weighted combination of predictions from the other algorithms under consideration

3 x (3 x 8 + 1) = 75

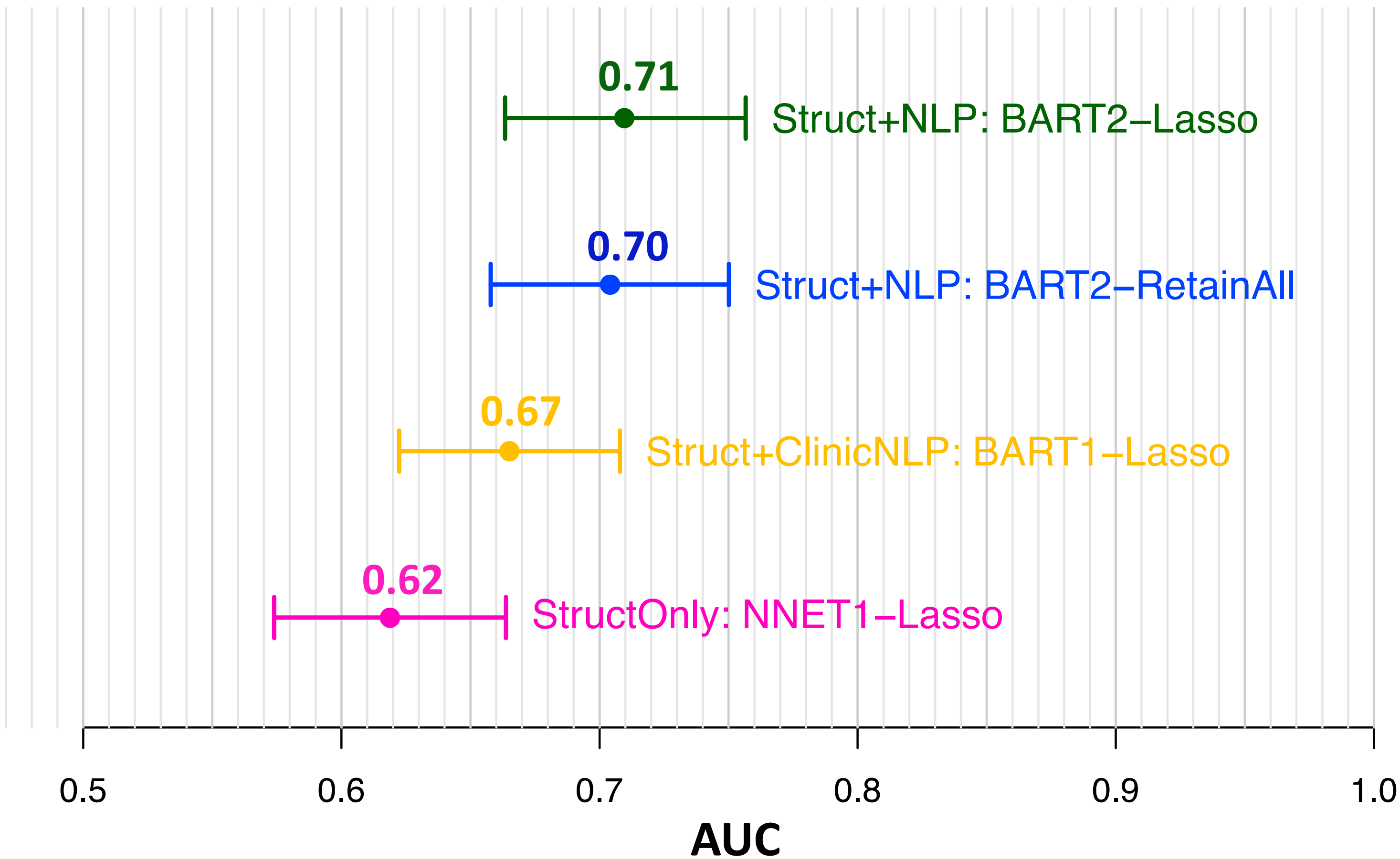
3	(3	8	1)
Datasets	Covariate Selection	Variants of six prediction algorithms	SL weighted combination
structured data structured+NLP struct+clinicianNLP	none lasso clustering		

Results

Path	KPWA (n=239)		KPNW (n=277)	
	Cases	Controls	Cases	Controls
1	106 (65.8%)	55 (34.2%)	115 (70.6%)	48 (29.4%)
2	48 (61.5%)	30 (38.5%)	65 (57.0%)	49 (43.0%)
all	154 (64.4%)	85 (35.6%)	180 (65.0%)	97 (35.0%)

Results

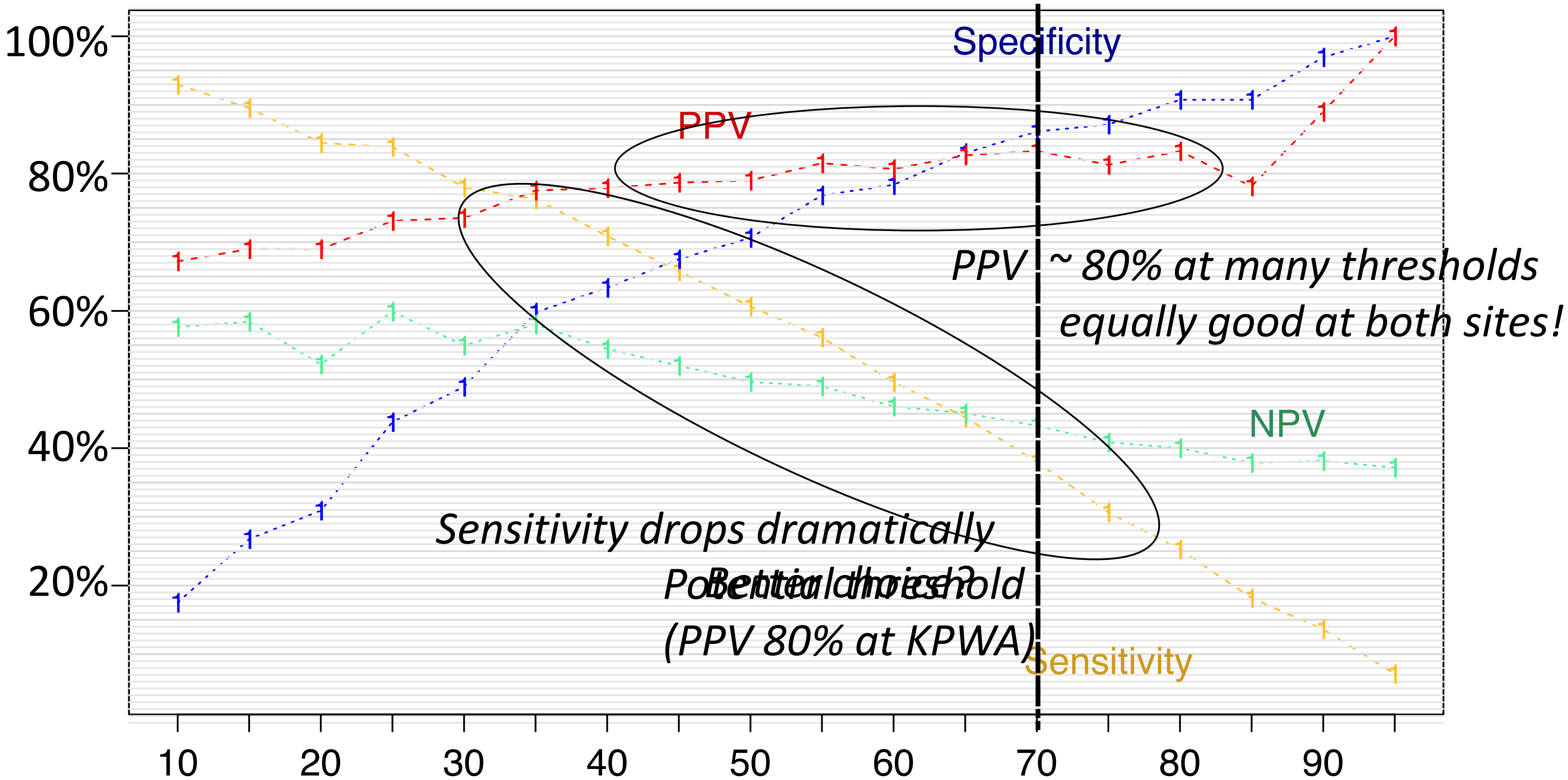
Cross-validated AUCs for best models for each KPWA data set



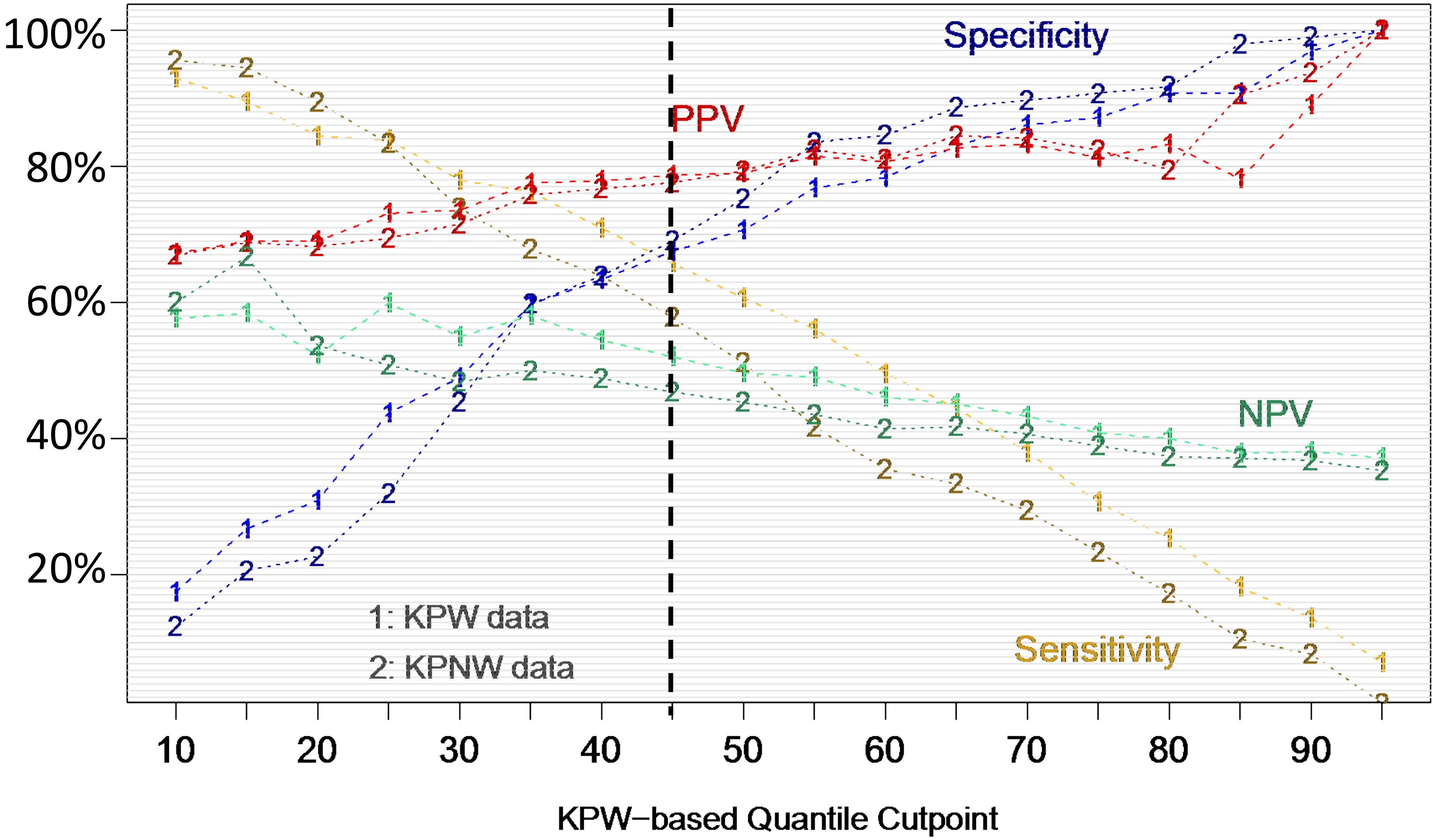
Results

- Two versions of Bayesian Additive Regression Trees combining structured data with NLP-derived covariates were nearly identical
- BART2-RetainAll generalized best to KP Northwest external validation set
  - cvAUC at KPWA = 0.70, cvAUC at KPNW = 0.67
  - Next step: Choose a prediction risk threshold for classification
    - if risk >= *threshold*, classify as a case, otherwise a control
    - most interested in high positive predictive value (PPV), high sensitivity (% cases identified)

Results: Performance Metrics



Results: Performance Metrics

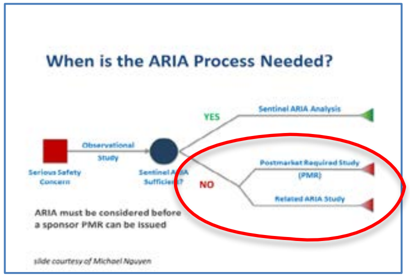


Outline

1. Motivation and project goals -- anaphylaxis
2. Study design & EHR data
3. Natural language processing of clinician notes
4. Machine learned-models for outcome identification
5. Towards a General Framework

Towards a General Framework

- … to improve electronic phenotyping via ML+NLP
  - Guiding principals
    - **Improve ARIA sufficiency**
    - Transportability → Sentinel Data Partners
    - Reusability (tools, resources)



Towards a General Framework

- DRAFT General Framework:

  Step 0: Systematically assess fitness-for-purpose

  Step 1: Create reference standard (Gold labels)

  Step 2: Feature engineering (NLP & structured data)

  Step 3: Model development

  Step 4: Evaluate the model (AUC, PPV, sensitivity, …)

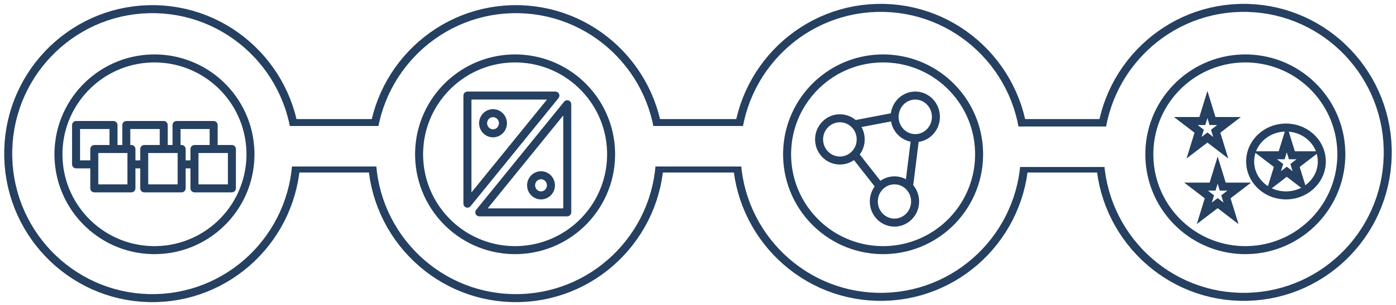
Questions & Discussion

Susan Gruber – sgruber@putnamds.com
David Carrell – david.s.carrell@kp.org

Additional Slides

Priorities	Goals	Initiatives	Outputs
Establishing data infrastructur e	Establishing a Sentinel electronic health record (EHR) network requires determining where to source and how to structure the data, as well as implementation of robust governance, harmonization, and quality assurance (QA) processes.	• Horizon scan of EHR databases • Adding unstructured data to the Sentinel common data model • Assessment and validation of source data mappings to improve the reliability and reproducibility of real-world data sources • Harmonizing EHRs from heterogenous systems • Developing and integrating approaches to identifying date and cause of death • FHIR implementation preparedness	• EHR data partners • Set of necessary EHR data elements • EHR common data model • Data governance process • Data harmonization and QA strategy • Data quality metrics • Sentinel death index • FHIR strategy
Expanding feature engineering	Frameworks and tools are needed for extracting critical information from EHR data to enable and enhance EHR-based computable phenotyping and to support EHR-based descriptive, inferential, and detection queries in Sentinel.	• Extending machine learning methods development in Sentinel: follow-up analyses for anaphylaxis algorithm and formalization of a general phenotyping algorithm • Scalable automated natural language processing- (NLP-) assisted chart abstraction • Advancing scalable NLP approaches for unstructured EHR data • Improving probabilistic phenotyping of incident outcomes through enhanced ascertainment with NLP	• Computable phenotyping framework • NLP tools for cohort identification, exposure assessment, covariate ascertainment, and outcome identification • Chart review automation approaches • Automated feature extraction tool to improve confounding control in EHR data • NLP-assisted chart abstraction tool
Enhancing causal inference	Developing, evaluating, and implementing advanced epidemiologic and statistical methods will enable Sentinel to make best use of EHR data to increase Active Risk Identification and Analysis (ARIA) sufficiency and expand the acceptance and use of real-world data for regulatory decision-making.	• Empirical evaluation of the causal inference effects of utilizing best practices for pharmacoepidemiologic studies • Enhancing causal inference in the Sentinel system: an evaluation of targeted learning and propensity scores • Approaches for handling missing laboratory data • Subset calibration for detecting and correcting for bias • Development of performance metrics and reporting standards • Advancing distributed regression in Sentinel	• Causal inference design and analysis framework • Super learner, target maximum likelihood estimation, complex treatment strategy analysis, missing data, subset calibration, and distributed regression tools • Inferential query performance metrics and reporting standards
Advancing detection analytics	Building safety signal detection approaches for specific use cases and in EHR data, in general, will substantially enhance Sentinel's capabilities for ensuring medical product safety but requires special design and analytic methods.	• Evaluation of existing approaches to EHR-based signal detection • Empirical comparison of EHR-based approaches to signal detection in Sentinel • Developing and advancing EHR-based signal detection methods • Advancing methods for safety signal detection for pregnancy and birth outcomes • Developing and evaluating a cancer signal detection tool	• Methodological framework for EHR-based signal detection • General safety signal detection tool for EHR data • Enhanced methods for signal detection for pregnancy and birth outcomes • Tool for cancer safety signal detection

slide courtesy of Joshua Gagne



Data infrastructure

- Data partners
- Data elements
- Governance
- Harmonization
- Data quality assurance

Feature engineering

- Natural language processing
- Automated feature extraction
- Computable phenotyping

Causal inference

- Target trial design
- Advanced, semi-automated analytics
- Subset calibration
- Distributed methods

Detection analytics

- Methodological framework
- Statistical methods
- Cancer outcomes
- Pregnancy and birth outcomes

slide courtesy of Joshua Gagne

Variable Importance (struct. + all NLP)

Top 5 structured:

1. Number of prior years with allergic reaction diagnoses (-)
2. Allergic reaction diagnosis in the prior year (-)
3. Same-day exposure to any imaging procedure (-)
4. Prescription for antihistamines @discharge (-)
5. Prescription for corticosteroids @discharge (-)

Top 5 NLP-derived:

1. ≥2 affirmative mentions of hypotension
2. Any description of respiratory compromise and reduced BP near a mention of either anaphylaxis as a diagnosis, epinephrine administration, suddenness of onset, or admission for observation
3. ≥2 affirmative mentions of skin/mucosal involvement and either respiratory compromise or reduced blood pressure near anaphylaxis as a diagnosis
4. ≥2 affirmative mentions of wheezing
5. any description of skin/mucosal involvement and reduced blood pressure near a mention of either anaphylaxis as a dx, epinephrine administration, suddenness of onset, or admission for observation

General Framework for Developing an Adverse Event Identifying Model (a.k.a. electronic phenotyping)

Step 0. Systematically assess *fitness-for-purpose*. (What purpose? Which HOIs? What population? What data (notes, labs, images)? Which data partners and with adequate N? What NLP/ML method?)

Step 1. Create *reference standard* (gold labels)

- Design/select the sample
- Determine events criteria
- Develop adjudication & abstraction protocol
- Obtain records
- Train adjudicators
- Review records & conduct deep annotation
- Conduct ongoing QC

Step 2. *Feature engineering* (NLP & structured)

- Identify signs, symptoms, dx codes, procedures…
- NLP development (prepare/QC note corpus, create dictionary, iteratively enrich with related terms & winnow down the feature set)
- Select structured features
- Operationalize features

Step 3. *Model development*

Step 4. *Evaluate the Model* (AUC, sens, spec,...)

- gold-standard labeled validation data