Diagnosis Prediction Case Study:

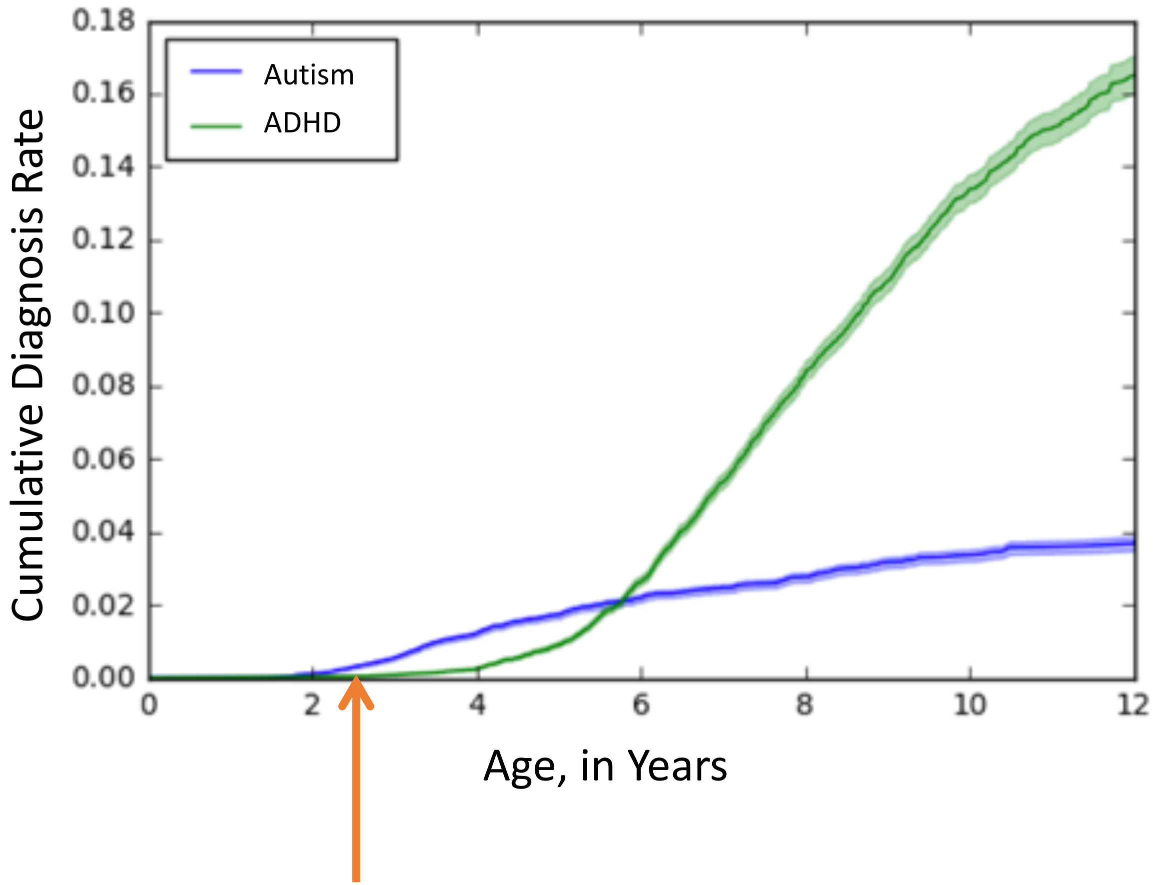
Predicting ASD & ADHD Risk from the EHR

Matthew Engelhard

Early Autism & ADHD prediction from the EHR

- Long-term benefits of early treatment depend on effective early diagnosis

- Profound racial, ethnic, and socioeconomic disparities in:
  - Rates of early diagnosis
  - Time between problem recognition and diagnosis
  - Length of follow-up



Goal: Diagnose autism before age 3

Duke Forge, NICHD 1P50HD093074, NIMH/NINDS 1 R01 MH121329

Perinatal and Neonatal Factors (# studies)	Results Across Studies	Summary Effect Estimate (95% CI)
Presentation		
Abnormal presentation (15)	10–, 5↑	1.44 (1.07–1.94)
Breech (4)		1.81 (1.21–2.71)
Other perinatal factors		
Cord complications (14)	13–, 1↑	1.50 (1.00–2.24)
Fetal distress (4)	3–, 1↑	1.52 (1.09–2.12)
Birth injury or trauma (6)	6–	4.90 (1.41–16.94)
Twins or multiple birth (10)	7–, 3↑	1.77 (1.23–2.55)
Maternal hemorrhage (4)	3–, 1↑	2.39 (1.35–4.21)
Birth weight and size		
Total birth weight (decreased) (15)	12–, 2↑, 1↓	
Low birth weight (<2500 g) (15)	8–, 7↑	1.63 (1.19–2.33)
Small for gestational age (10)	7–, 3↑	1.35 (1.14–1.61)
Clinical impression		
Congenital malformation (11)	4–, 7↑	1.80 (1.42–2.82)
Apgar score		
Low 5-minute Apgar score (8)	6–, 2↑	1.67 (1.24–2.26)
Neonatal Status		

PEDIATRICS®
OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

Review Article

Perinatal and Neonatal Risk Factors for Autism: A Comprehensive Meta-analysis

Hannah Gardener, Donna Spiegelman and Stephen L. Buka

Pediatrics August 2011, 128 (2) 344-355; DOI: https://doi.org/10.1542/peds.2010-1036

Research Letter

FREE

Nov 2012

Prenatal and Perinatal Risk Factors for Attention-Deficit/Hyperactivity Disorder

Jochen Schmitt, MD, MPH; Marcel Romanos, MD

› Author Affiliations | Article Information

Arch Pediatr Adolesc Med. 2012;166(11):1074-1075. doi:10.1001/archpediatrics.2012.1078

Table. Sample Characteristics and Risk Factors of ADHD in Children and Adolescents

Characteristic/Exposure (Reference for Regression Analyses)	Sample Characteristics, No. (%)[a]		Logistic Regression Analysis, OR (95% CI)	
	Children With ADHD (n = 660)	Children Without ADHD (n = 12 828)	Bivariable (Unadjusted) Analysis	Multivariable (Adjusted) Analysis[b]
Sex (reference: female)	133 (20.2)	6604 (51.5)	4.20 (3.47-5.10)	4.42 (3.56-5.49)
Age, y, mean (SD)	9.8 (4.3)	11.3 (3.4)	1.08 (1.06-1.11)	1.09 (1.07-1.11)
Socioeconomic position[c]				
Upper class (reference)	114 (17.4)	3486 (27.3)	1 [Reference]	1 [Reference]
Middle class	325 (49.5)	6087 (47.7)	1.63 (1.31-2.03)	1.57 (1.23-2.00)
Lower class	218 (33.2)	3202 (25.1)	2.08 (1.65-2.62)	2.04 (1.56-2.68)
Maternal gestational diabetes mellitus (reference: absent)	24 (4.1)	256 (2.2)	1.93 (1.26-2.95)	1.91 (1.21-3.01)
Maternal smoking during pregnancy (reference: never)	158 (24.6)	2081 (16.4)	1.66 (1.38-2.00)	1.48 (1.19-1.84)
Maternal alcohol consumption during pregnancy (reference: never)	96 (14.8)	1775 (14.0)	1.07 (0.86-1.34)	1.02 (0.79-1.33)
Perinatal health problems (reference: absent)[d]	235 (36.2)	2955 (23.2)	1.88 (1.60-2.22)	1.69 (1.40-2.03)
Breastfeeding (ever vs never fully breastfeeding)	345 (56.7)	7943 (67.5)	0.63 (0.54-0.74)	0.83 (0.69-0.996)
Atopic eczema (ever vs never)	132 (20.2)	1820 (14.4)	1.51 (1.24-1.84)	1.62 (1.30-2.02)

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; OR, odds ratio.
[a] Numbers represent number (proportion) of children per exposed for discrete variables and means (SD) for continuous variables.
[b] Adjusted for all exposures listed in the Table; analysis based on 11 222 observations without any missing data.
[c] Classified based on parental education, professional qualification, professional status, and family net income according to Winkler and Stolzenberg.[4]
[d] Breathing problems, maladaptation, infections, icterus, low birth weight/premature delivery, and/or inpatient treatment.

Our Predictive Modeling Journey…

1. Getting the data & making sense of it (or trying to)

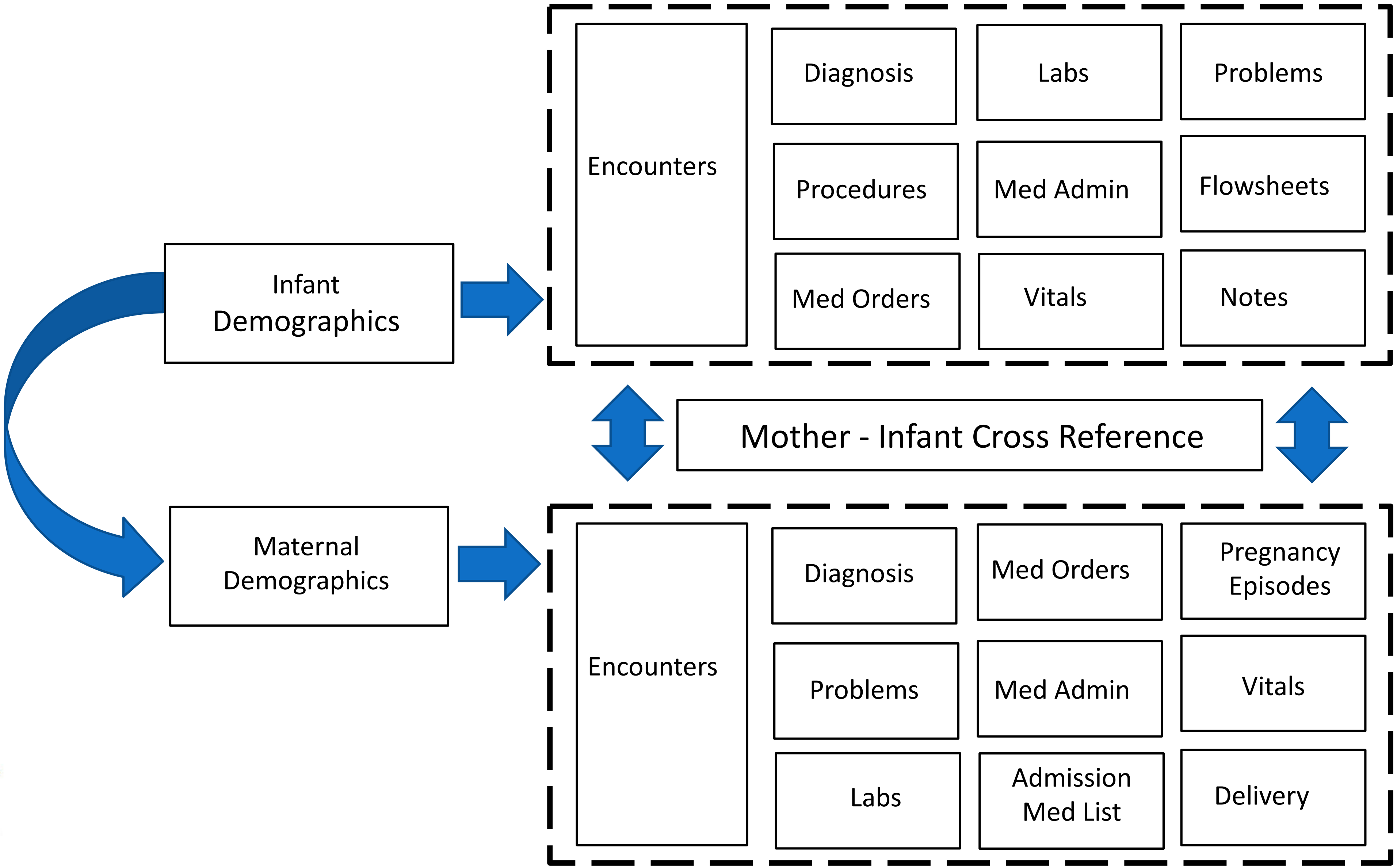
2. Defining the outcome (i.e. what are we even predicting?)

3. Structuring the prediction task (i.e. but what are we *really* predicting?)

4. Incorporating clinical notes (and all the ways it can go wrong)

1. Get the data and (try to) make sense of it

ML of EHR for ASD and ADHD Risk Assessment



FORGE

Where is the data we need?

DukeHealth

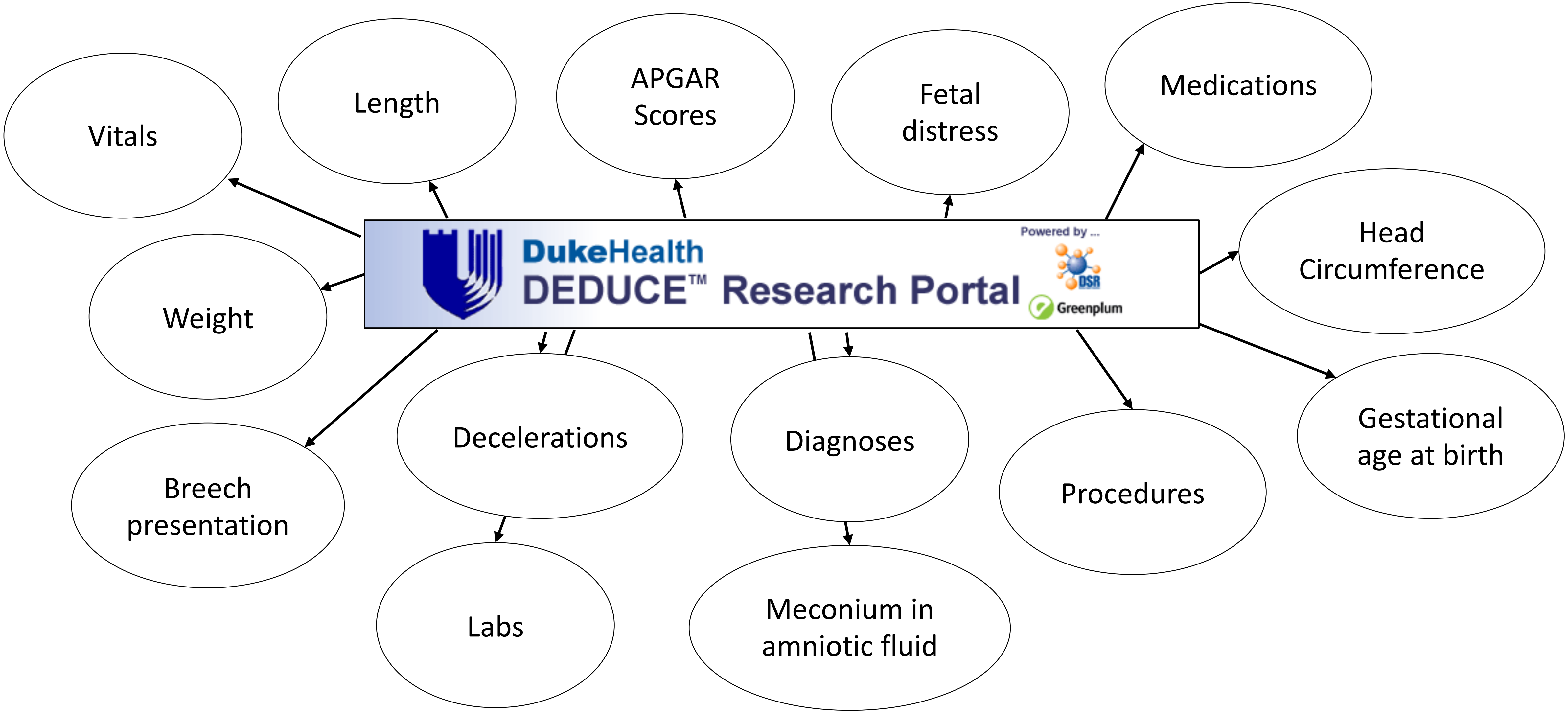
DEDUCE™ Research Portal

Powered by ...

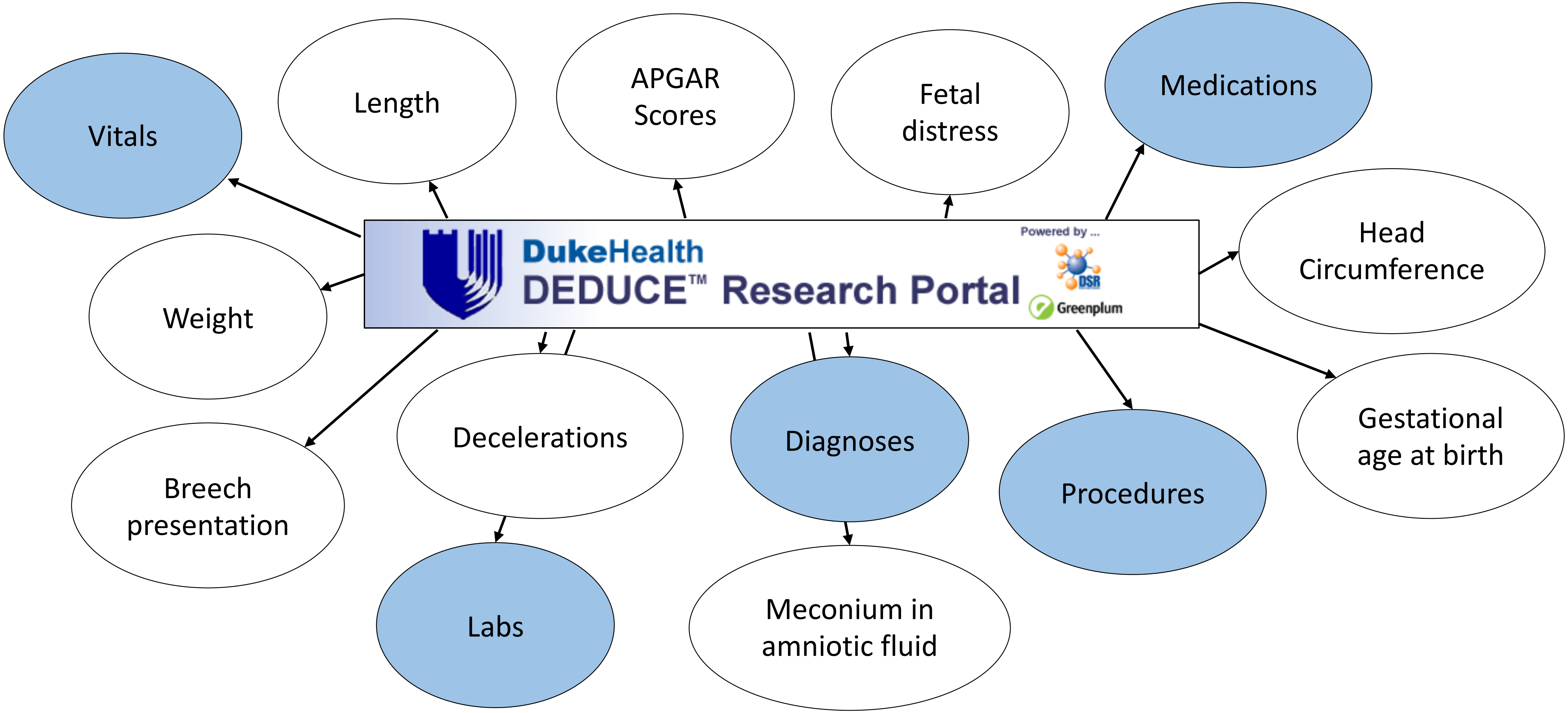
DSR

Greenplum

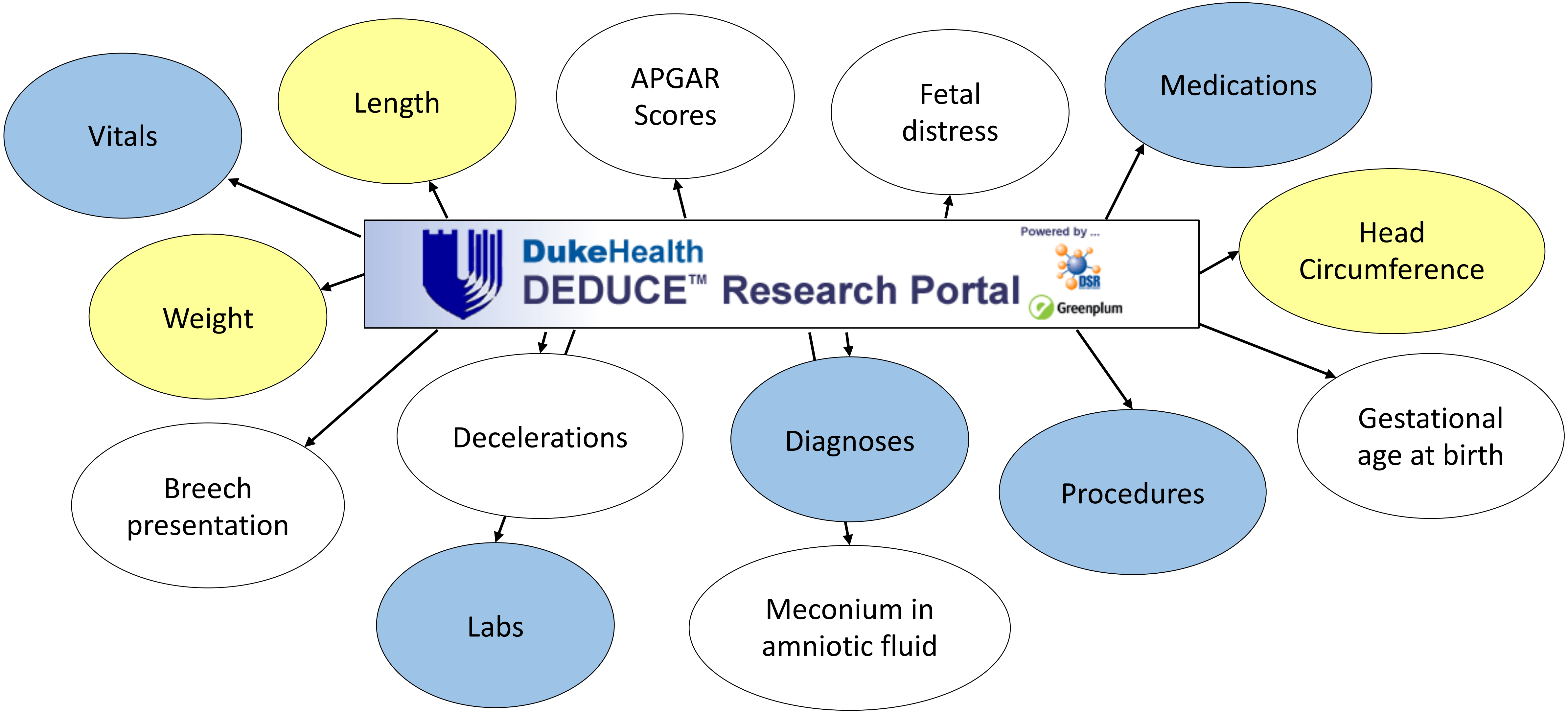
Where is the data we need?



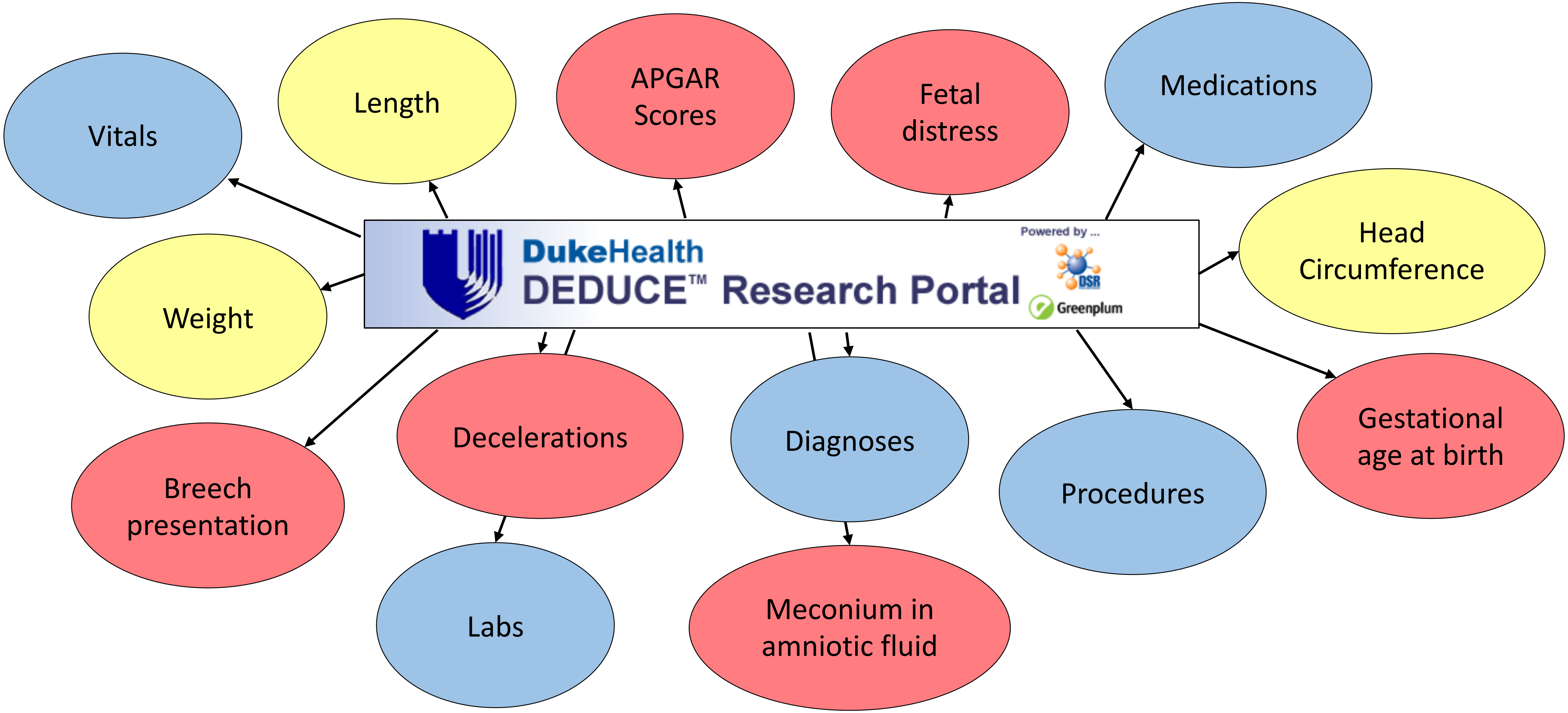
Where is the data we need?



Where is the data we need?



Where is the data we need?



Which values can we trust?

```
In [18]: df['Clinic Service or Specialty'].value_counts()
Out[18]:
Pediatrics                                  405265
PED. BEHAVIORAL DEVELOPMENT & GEN           315404
Urgent Care                                  44272
Physical and Occupational Therapy            27803
Speech Pathology                             27114
CHART RESPONSIBLE MD                         23164
Ophthalmology                                15562
Primary Care                                 12959
COMMUNITY AND FAMILY MEDICINE                11053
Radiology                                    10307
Pediatric Psychiatry                         10027
General Surgery                               9005
Lab                                           8417
Missing or invalid                            7793
```

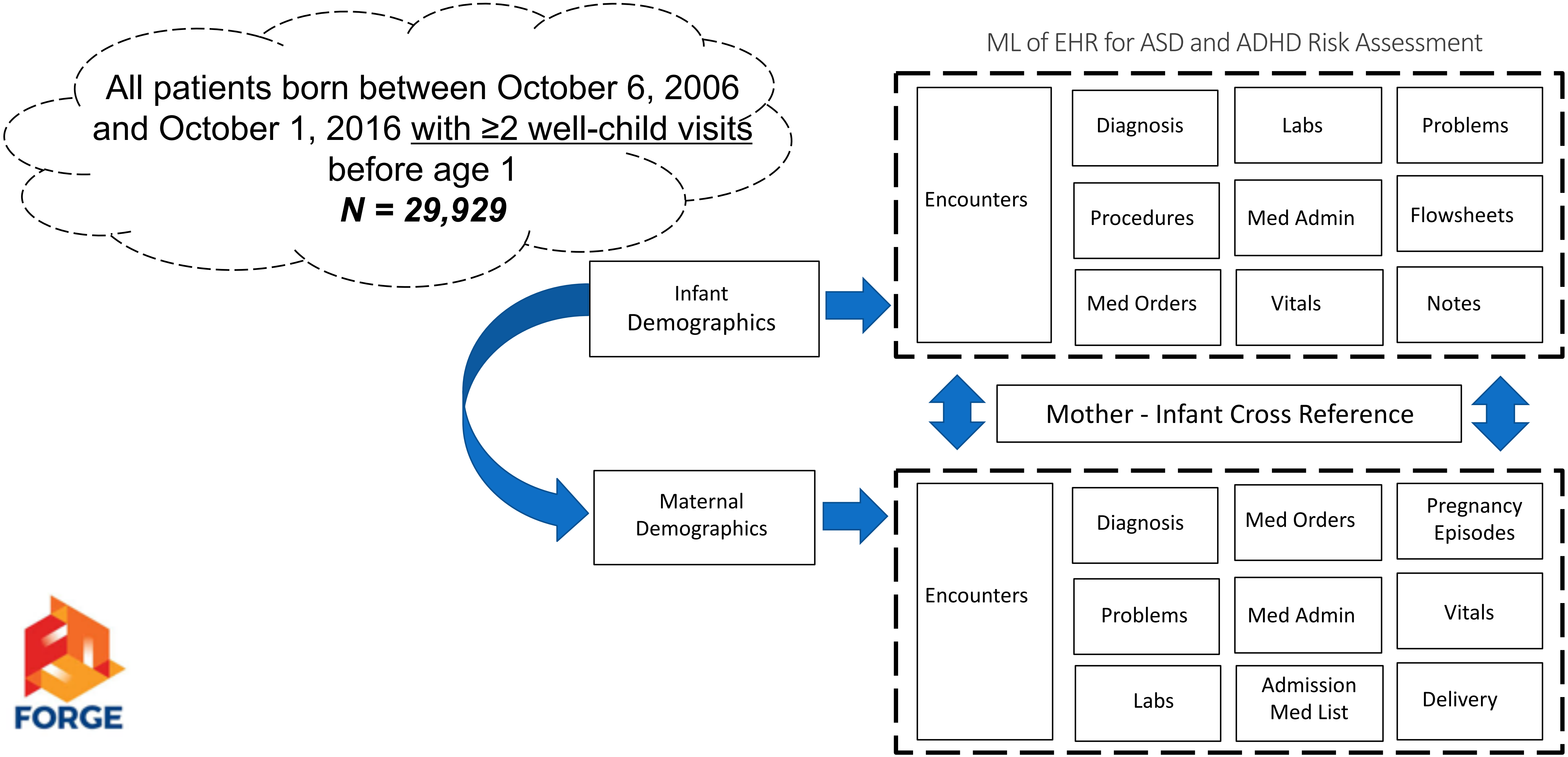
Which values can we trust?

```
In [18]: df['Clinic Service or Specialty'].value_counts()
Out[18]:
Pediatrics                                405265
PED. BEHAVIORAL DEVELOPMENT & GEN         315404
Urgent Care                                44272
Physical and Occupational Therapy          27803
Speech Pathology                           27114
CHART RESPONSIBLE MD                       23164
Ophthalmology                              15562
Primary Care                               12959
COMMUNITY AND FAMILY MEDICINE              11053
Radiology                                  10307
Pediatric Psychiatry                       10027
General Surgery                             9005
Lab                                         8417
Missing or invalid                          7793
```

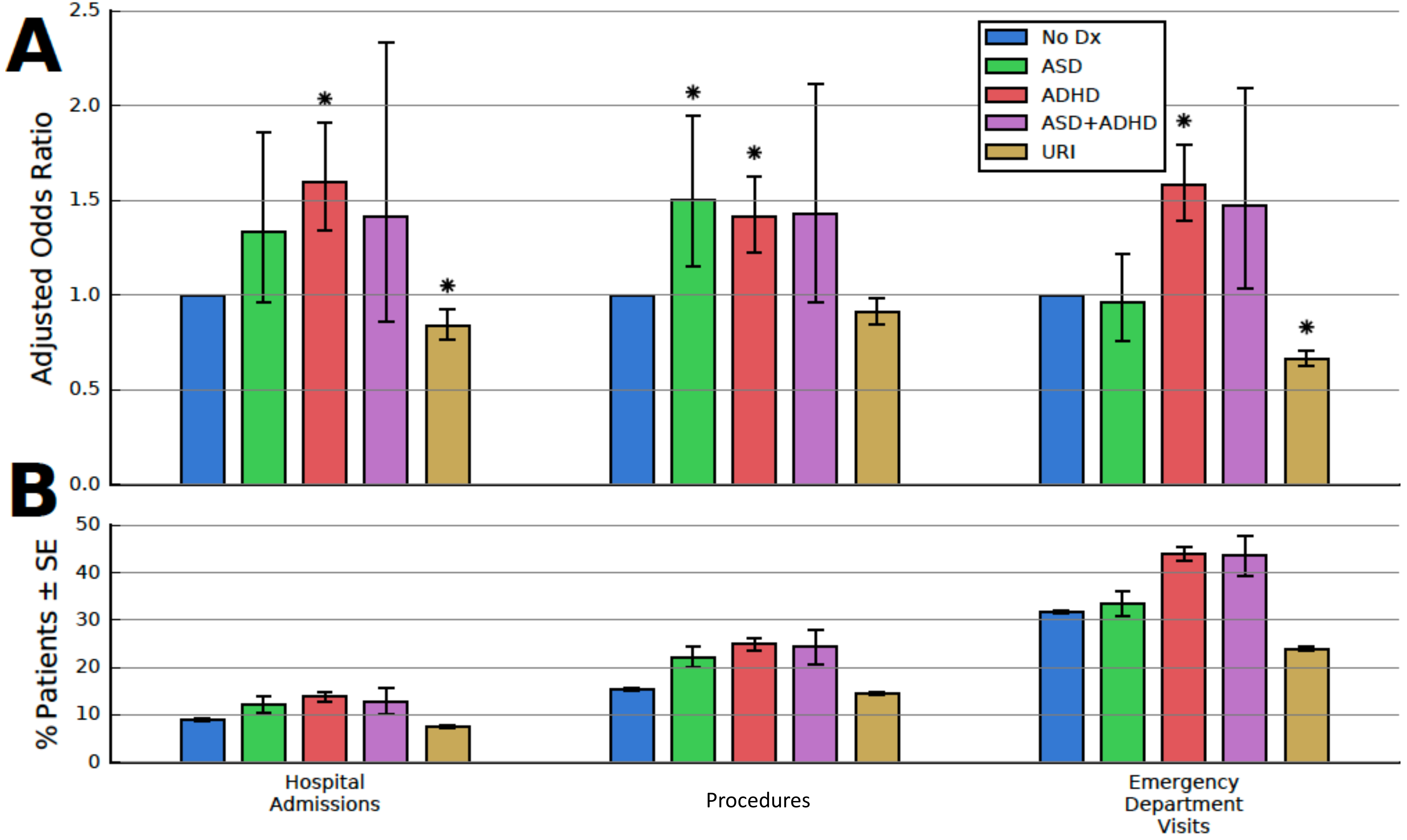
Which values can we trust?

```
In [18]: df['Clinic Service or Specialty'].value_counts()
Out[18]:
Pediatrics                              405265
PED. BEHAVIORAL DEVELOPMENT & GEN       315404
Urgent Care                              44272
Physical and Occupational Therapy        27803
Speech Pathology                         27114
CHART RESPONSIBLE MD                     23164
Ophthalmology                            15562
Primary Care                             12959
COMMUNITY AND FAMILY MEDICINE            11053
Radiology                                10307
Pediatric Psychiatry                     10027
General Surgery                           9005
Lab                                       8417
Missing or invalid                        7793
```

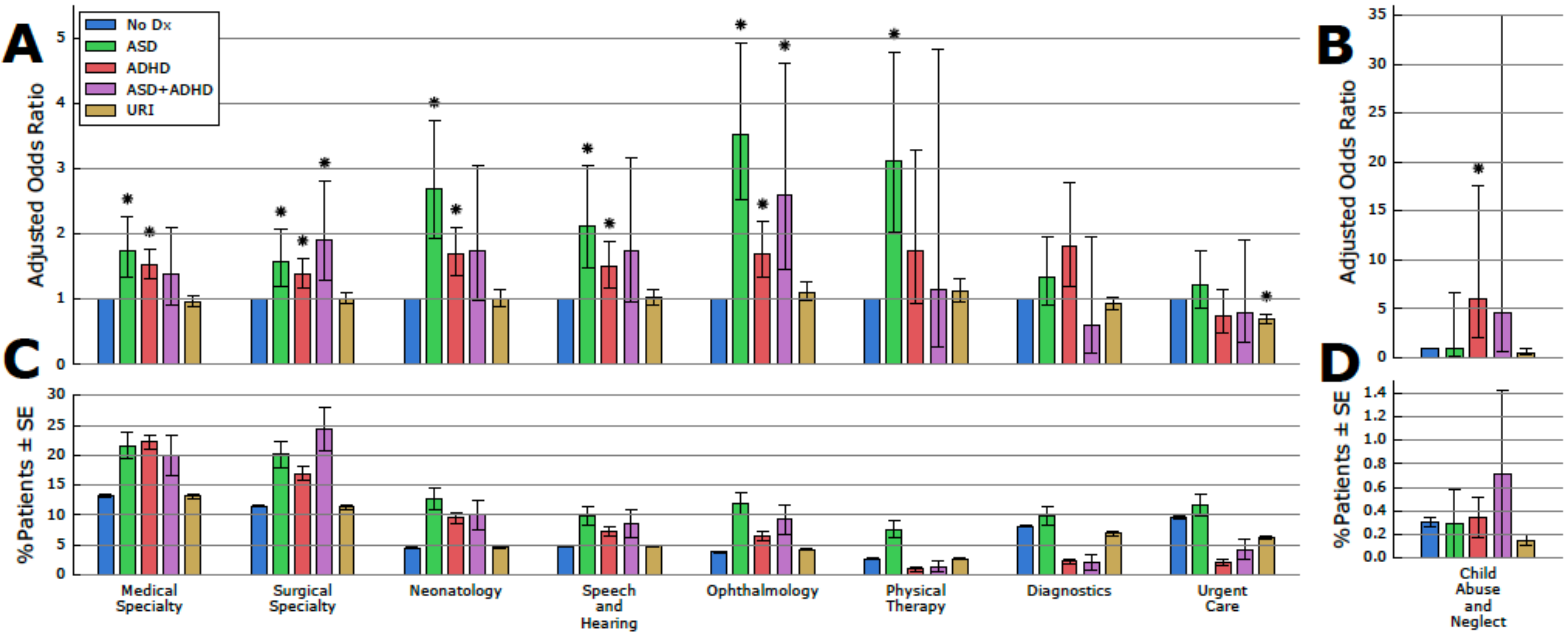
Defining the Cohort



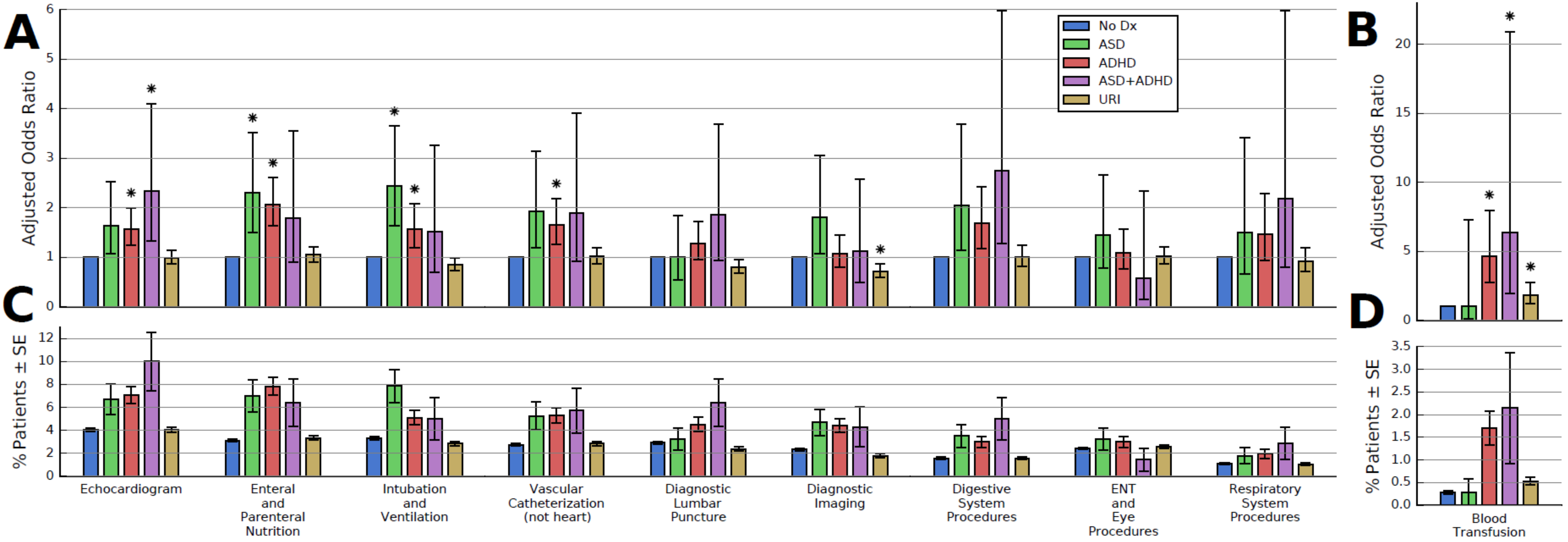
Increased Hospital Admissions, Procedures, and ED Visits before age 1



Distinctive Patterns of Outpatient Specialty Care before age 1



Distinctive Patterns of Procedures before age 1



Prediction Model -> Clinically Meaningful Performance Measures

	Model-based referral	No model-based referral
Child has autism	Earlier Diagnosis (true positive)	Referral via Current Mechanisms (false negative)
Child does not have autism	Unnecessary Specialist Visit (false positive)	No Action Taken or Needed (true negative)

<u>Decision</u>: initiate referral

model is NOT used to rule out diagnoses

<u>Measures of Success:</u>

- How much earlier, on average, can we diagnose and intervene?
- How many children are unnecessarily referred?

2. Figure out what we're predicting

ADHD Computable Phenotype:

- 2 codes, or 1 + med

Validation of the Use of Electronic Health Records for Classification of ADHD Status

Siobhan M Gruschow [1], Benjamin E Yerys [1] [2], Thomas J Power [1] [2], Dennis R Durbin [1] [2], Allison E Curry [1]

Affiliations + expand

PMID: 28112025 PMCID: PMC5843549 DOI: 10.1177/1087054716672337

Free PMC article

Abstract

Objective: To validate an electronic health record (EHR)-based algorithm to classify ADHD status of pediatric patients.

Method: As part of an applied study, we identified all primary care patients of The Children's Hospital of Philadelphia [CHOP] health care network who were born 1987-1995 and residents of New Jersey. Patients were classified with ADHD if their EHR indicated an International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis code of "314.x" at a clinical visit or on a list of known conditions. We manually reviewed EHRs for ADHD patients (n = 2,030) and a random weighted sample of non-ADHD patients (n = 807 of 13,579) to confirm the presence or absence of ADHD.

Results: Depending on assumptions for inconclusive cases, sensitivity ranged from 0.96 to 0.97 (95% confidence interval [CI] = [0.95, 0.97]), specificity from 0.98 to 0.99 [0.97, 0.99], and positive predictive value from 0.83 to 0.98 [0.81, 0.99].

Conclusion: EHR-based diagnostic codes can accurately classify ADHD status among pediatric patients and can be used by large-scale epidemiologic and clinical studies with high sensitivity and specificity.

Keywords: accuracy; adolescents; attention deficit disorder; medical records; sensitivity.

ASD Computable Phenotype:

- 2 codes, or 1 from pediatric psych

PEDIATRICS®
OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

Article

Accuracy of Autism Screening in a Large Pediatric Network

Whitney Guthrie, Kate Wallis, Amanda Bennett, Elizabeth Brooks, Jesse Dudley, Marsha Gerdes, Juhi Par

Pediatrics October 2019, 144 (4) e20183963; DOI: https://doi.org/10.1542/peds.2018-3963

Article | Figures & Data | Supplemental | Info & Metrics | Comments

Download PDF

Abstract

Two Competing Approaches

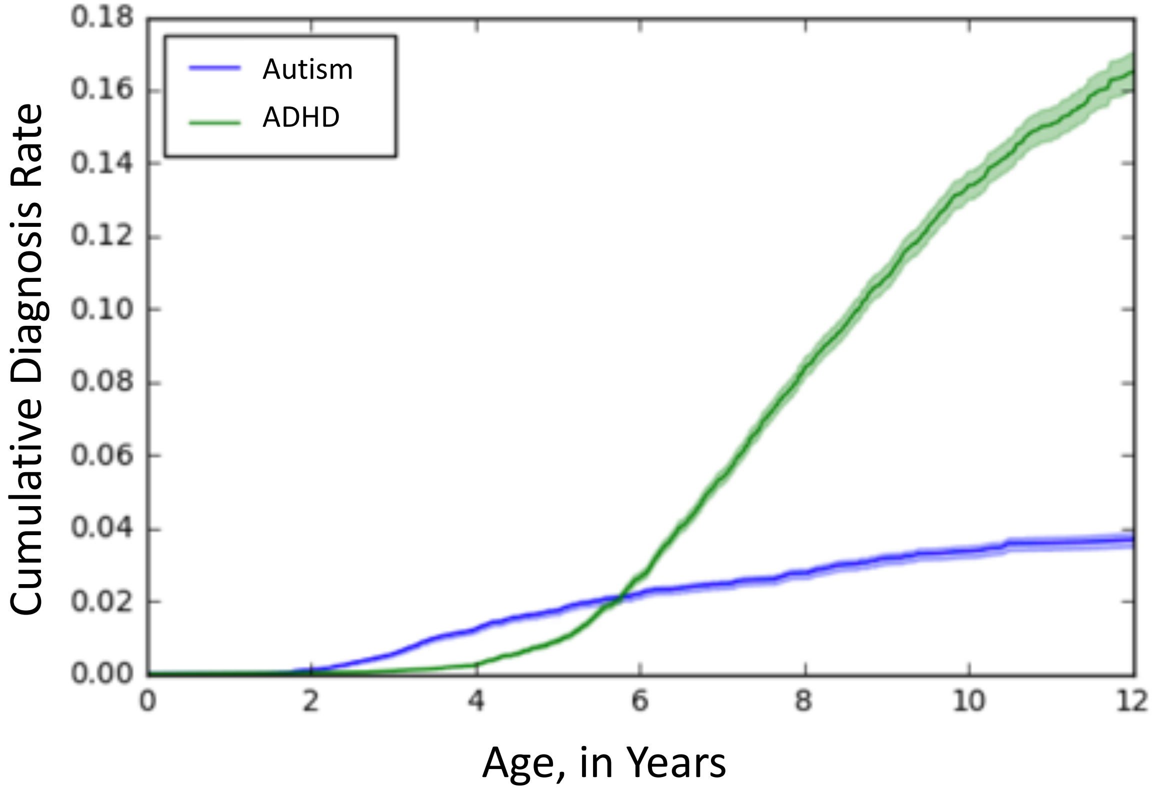
1. Predict risk of *diagnosis*
   - Need gold standard diagnoses…
   - Evaluate computable phenotype against gold standard
   - *maybe* you can evaluate TPR. Good luck with the rest.

2. Predict risk of an ASD- or ADHD-related diagnosis code
   - If we're building a screening tool, this is probably good enough
   - So far, appears to give us similar results

3. Figure out what we’re *really* predicting

Dx Timing vs Probability vs Risk

- Diagnosis Timing
  - known disparities
  - we don't care to replicate
- Diagnosis Probability
  - probably closer to what we really care about
  - figure out who will be diagnosed and diagnose them earlier
- Diagnosis Risk
  - e.g. Cox PH models
  - blends timing and probability



Standard time to event model: the right setup?

- Suppose we have data $D = \{\boldsymbol{x}_i, t_i, s_i\}_{i=1}^{N}$, where the $\boldsymbol{x}_i \in \mathbb{R}^d$ are features for individual i, the $t_i \in (0, \infty)$ are associated times, and the $s_i \in \{0,1\}$ denote whether the t_i are event times ($s_i = 1$) or right-censoring times ($s_i = 0$).

- We suppose event times are drawn independently from $f_\theta(t|\boldsymbol{x}_i)$, which has associated survivor function $F_\theta(t|\boldsymbol{x}_i) = 1 - \int_0^t f_\theta(\tau|\boldsymbol{x}_i)d\tau$.

- In the standard time to event framework, and assuming non-informative censoring, the parameters θ may be chosen to maximize the likelihood as follows:

$$\theta_{ML} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{N} \{s_i \log f_\theta(t|\boldsymbol{x}_i) + (1 - s_i) \log F_\theta(t|\boldsymbol{x}_i)\}$$

Neural Mixture Cure: Separate Probability from Timing

- We're interested hidden variables $c_i^j \in \{0,1\}$ representing finite event occurrence – in other words, whether the event $j \in \{1, \dots, M\}$ will ever occur in individual i.

- We use a binary stochastic neural network layer to model the occurrence of many potentially interrelated events and their effects on:
  a) the timing of other events
  b) (potentially) censoring times (e.g. length of follow-up)

- Supposing $c_i^j | \boldsymbol{x}_i \sim \mathrm{Bern}(\sigma\left(h_\phi(\boldsymbol{x}_i)\right))$, where $\sigma(\cdot)$ is the logistic function and $h_\phi(\cdot)\colon \mathbb{R}^d \to \mathbb{R}$ has parameters ϕ, we can maximize a lower bound on the expected log-likelihood over $\boldsymbol{c}_i$:

$$\log p_{\theta,\phi}\,(D) \geq \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbb{E}_{\boldsymbol{c}_i \sim p_\phi(\boldsymbol{c}_i|\boldsymbol{x}_i)} \log p_\theta(t_i^j, s_i^j | \boldsymbol{c}_i, \boldsymbol{x}_i)$$

- In the above, $p_\theta(t_i, s_i | c_i, \boldsymbol{x}_i) \mathrel{\underset{\sim}{\propto}} \epsilon^{s_i(1-c_i)} f_\theta(t_i|\boldsymbol{x}_i)^{s_i} F_\theta(t_i|\boldsymbol{x}_i)^{(1-s_i)c_i}$inable parameter penalizing incorrect prediction of $c_i = 0$ when $s_i = 1$.

Results: It works with fake data!

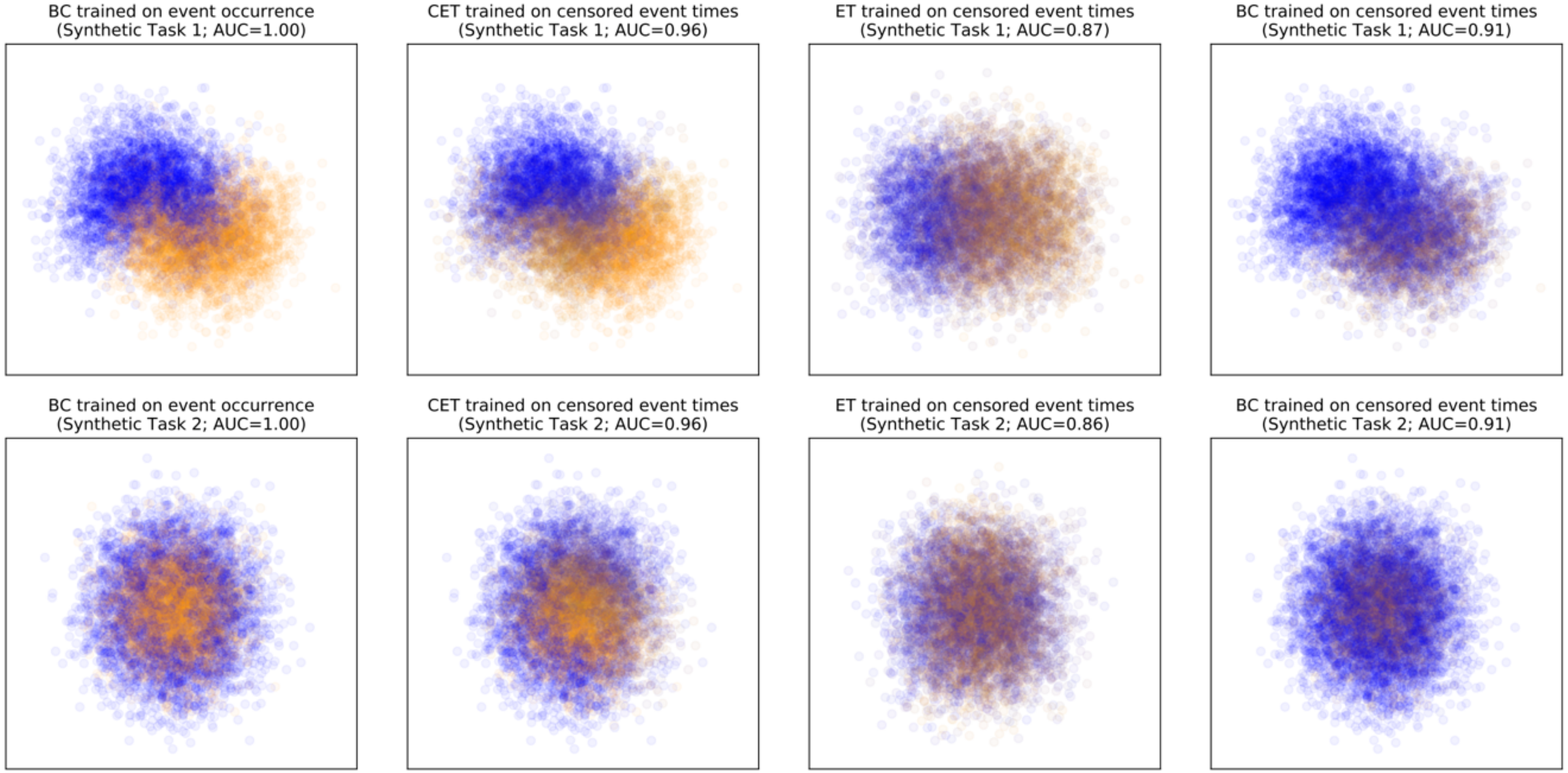


Figure: Prediction of event occurrence on two synthetic datasets. When no censoring is present, a binary classifier predicts event occurrence perfectly (left). When censoring is present, our model (middle left) predicts event occurrence more effectively than the standard event time model (middle right) and binary classifier (right).

Results: It works with real data!!!

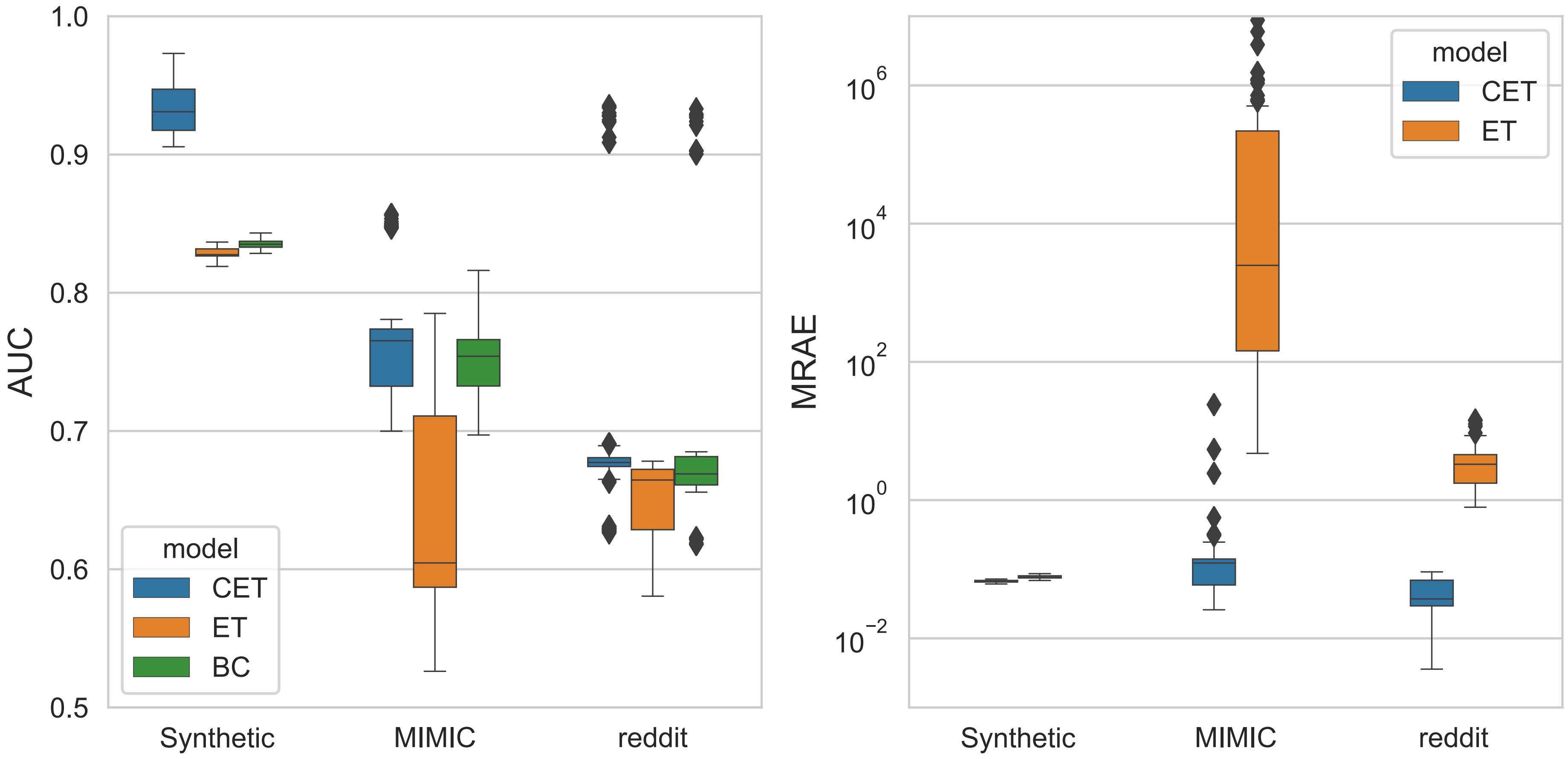
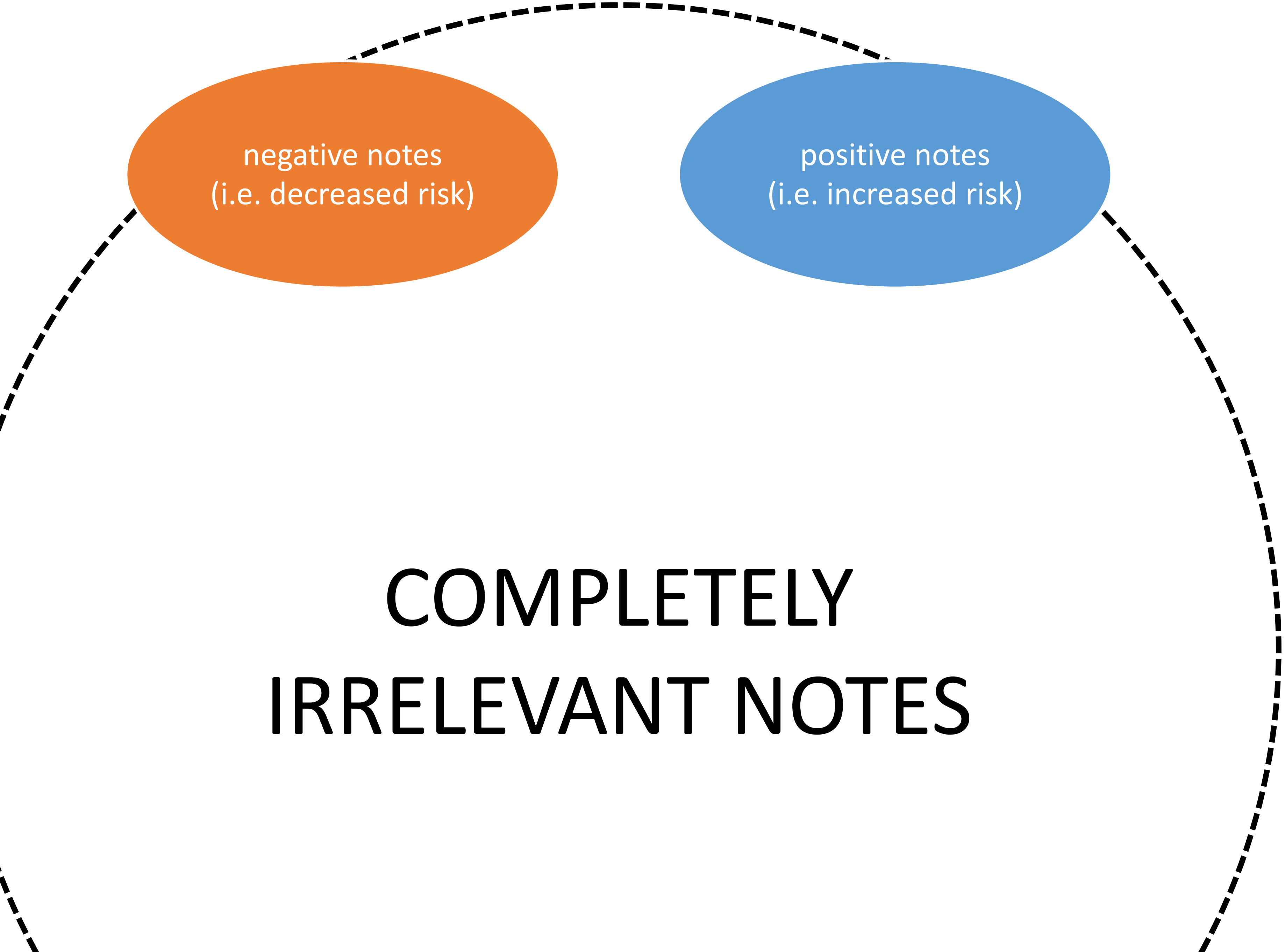


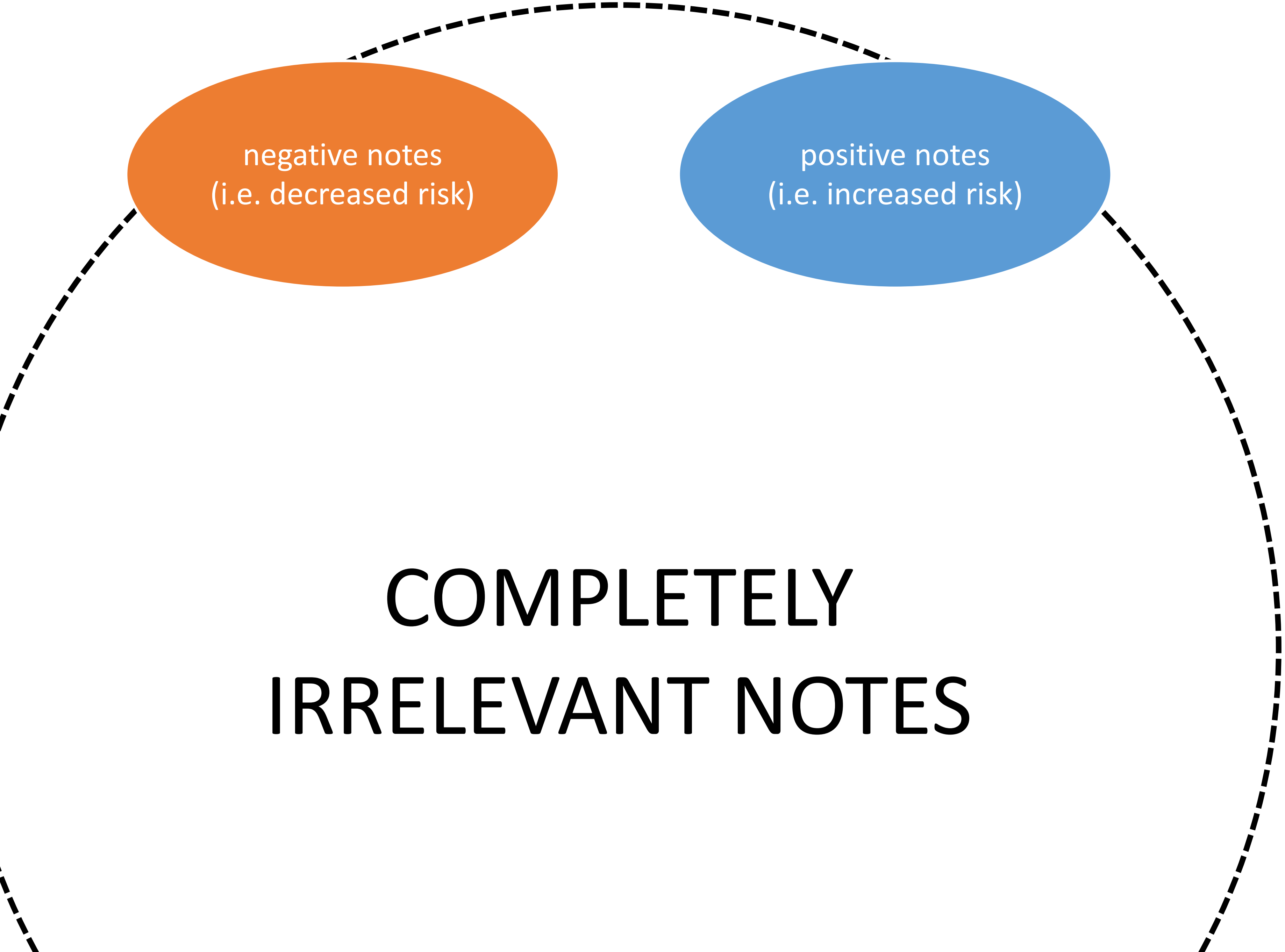
Figure: AUC (left) and mean relative absolute error (right) for all prediction tasks.

4. An impenetrable wall of text

negative notes
(i.e. decreased risk)

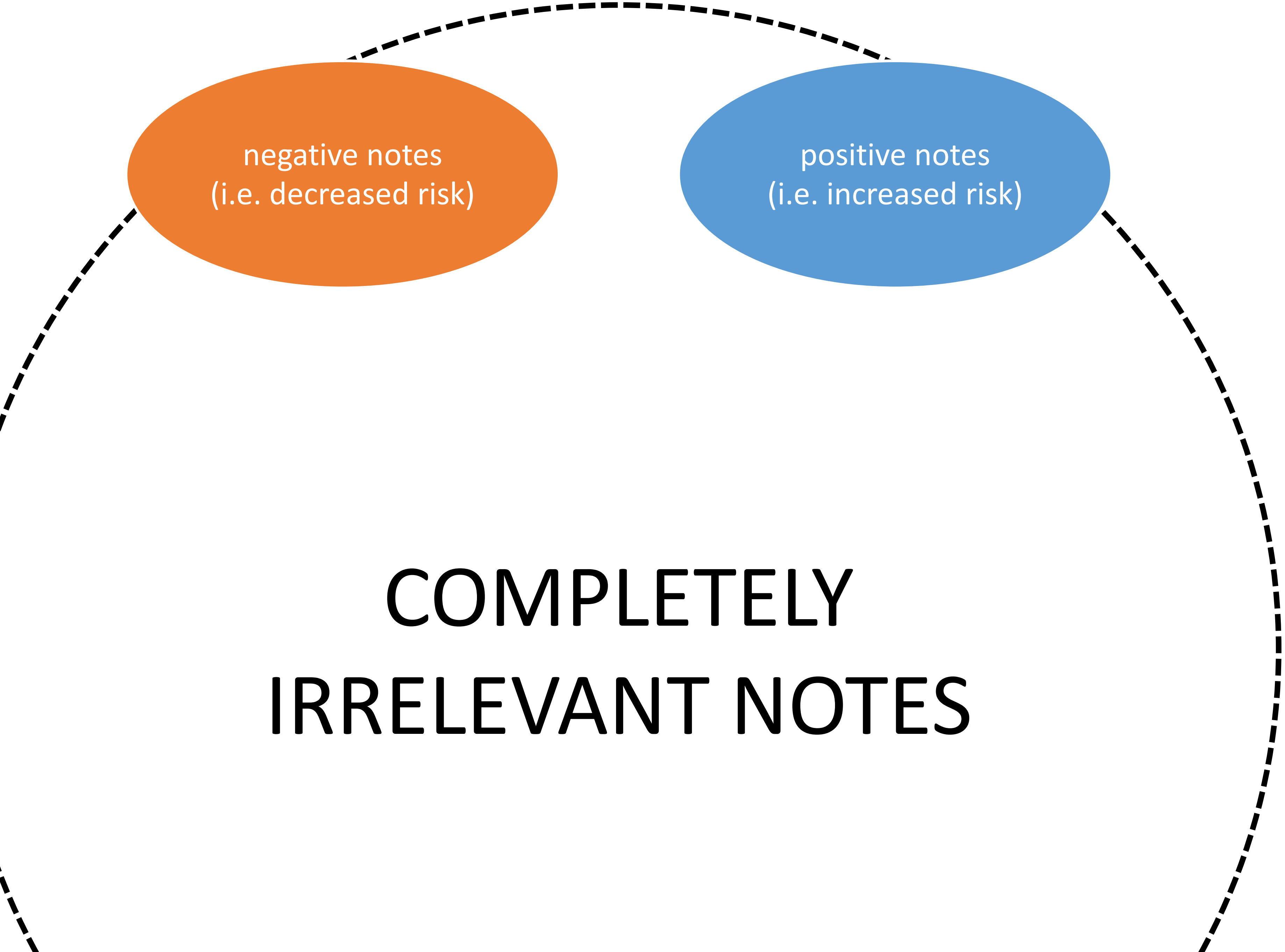
positive notes
(i.e. increased risk)


negative notes
(i.e. decreased risk)
positive notes
(i.e. increased risk)
COMPLETELY
IRRELEVANT NOTES



Predict on a note level or patient level?

- Conceptual differences
- Computational challenges
- Find a way to remove irrelevant notes



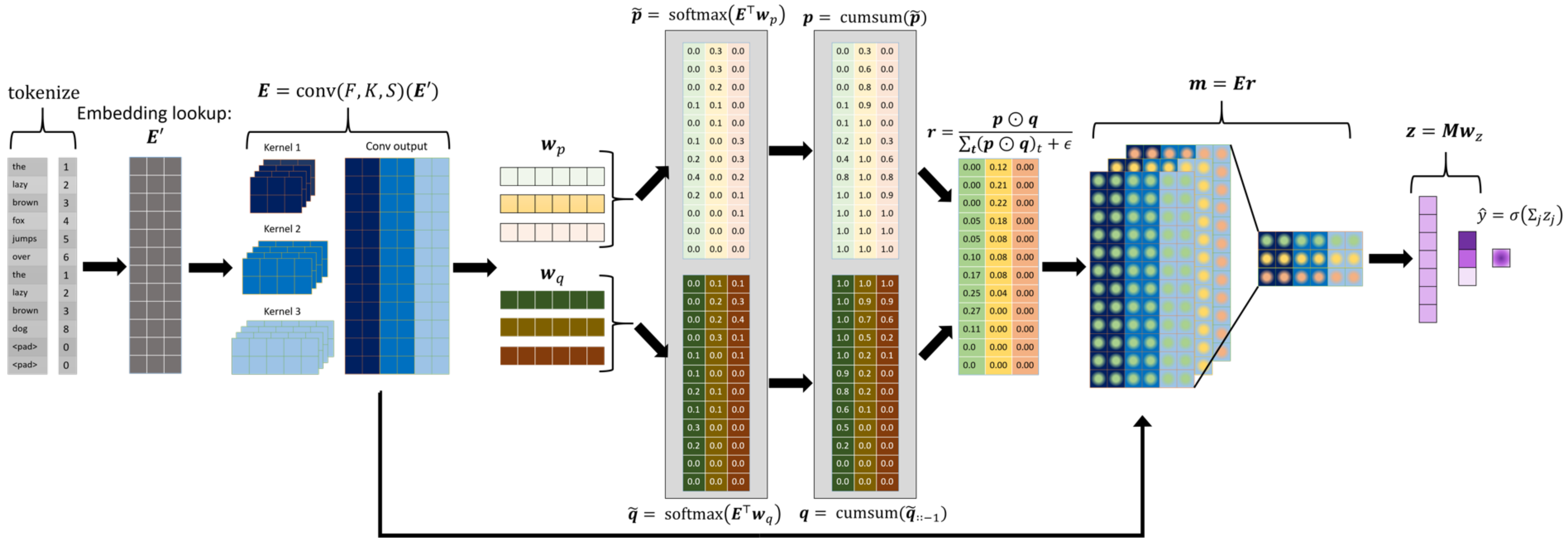
Bonus challenge:

Because we are looking at *early* risk factors, providers can't necessarily tell us which notes belong in which bucket

Predictive Extraction: Identifying predictive content within (clinical) text

Given a document X and its associated binary label y, select contiguous sequences of text called *spans* that, jointly, are sufficient to predict the label y effectively. Furthermore, we assign each span a score reflecting its contribution to the prediction y.

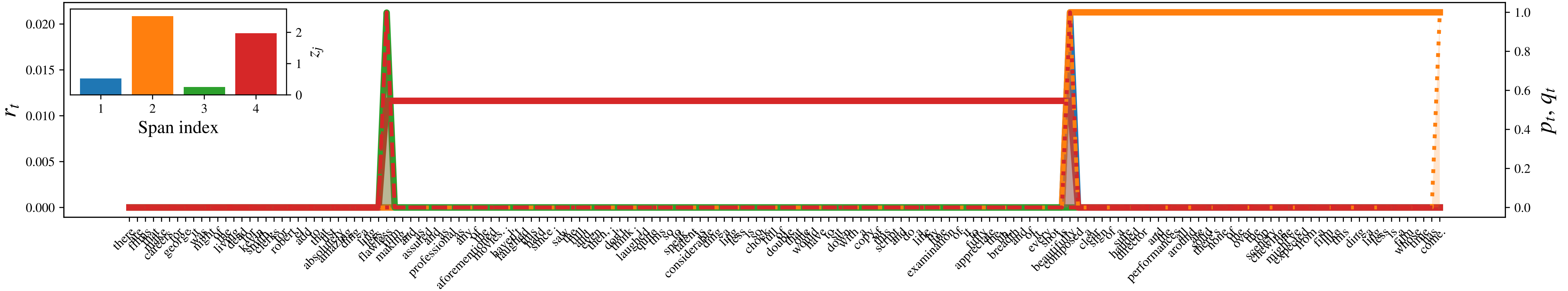
Our Predictive Extraction Architecture



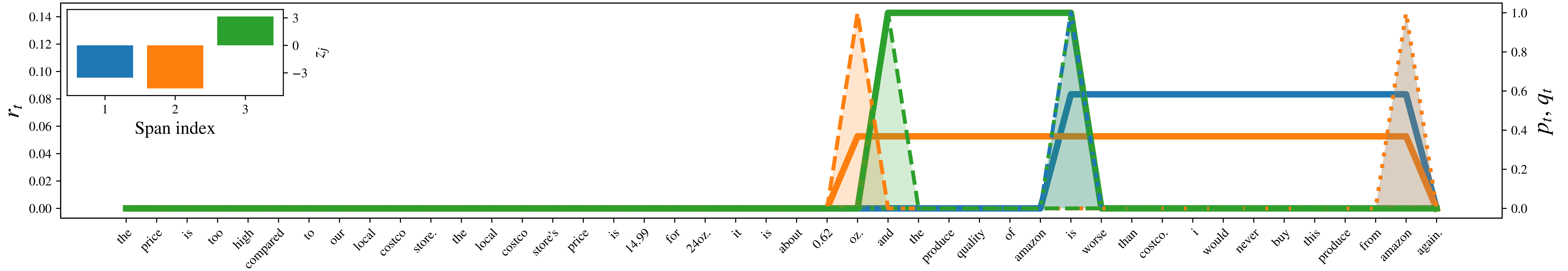
Subramanian V, Engelhard MM, Berchuck SI, Henao R, Chen L, Carin L.
SpanPredict: Extraction of Predictive Document Spans with Neural Attention.
Submitted to 16th conference of the European Chapter of the Association for Computational Linguistics (EACL).

Example Span Selection

IMDb. Label: 1. P(y=1): 0.78785

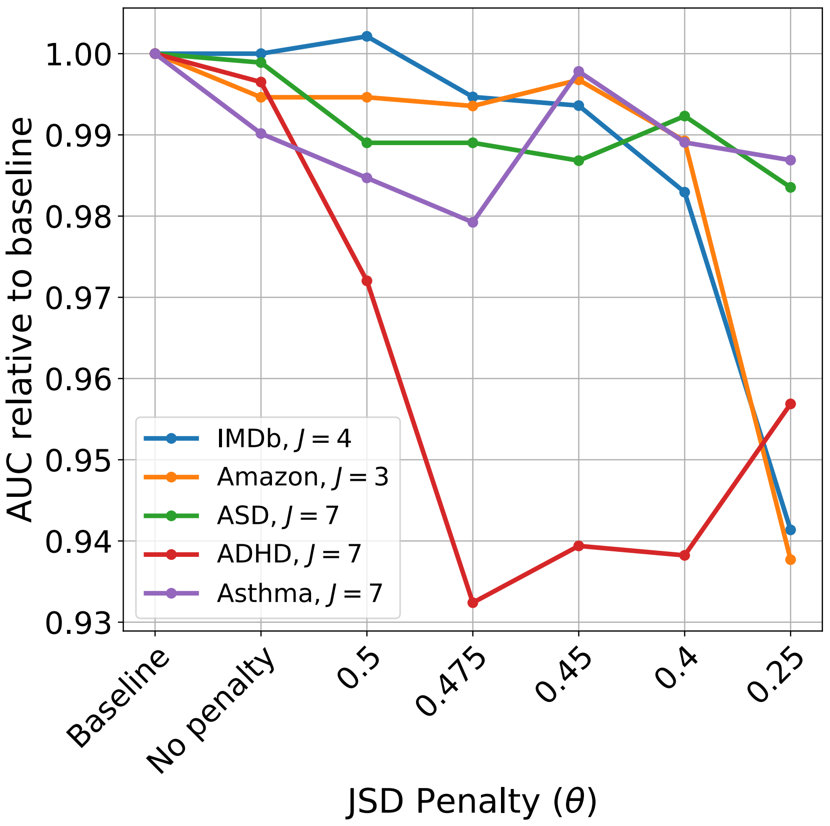


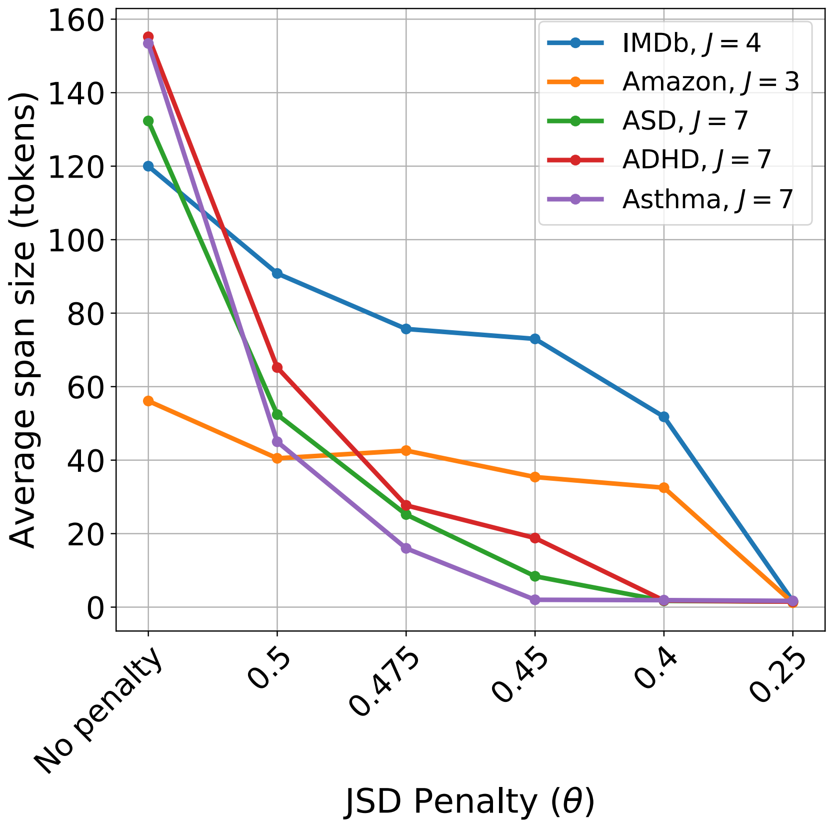
Amazon Food Reviews. Label: 0. P(y=1): 0.15540



Subramanian V, Engelhard MM, Berchuck SI, Henao R, Chen L, Carin L. SpanPredict: Extraction of Predictive Document Spans with Neural Attention. Submitted to 16th conference of the European Chapter of the Association for Computational Linguistics (EACL).

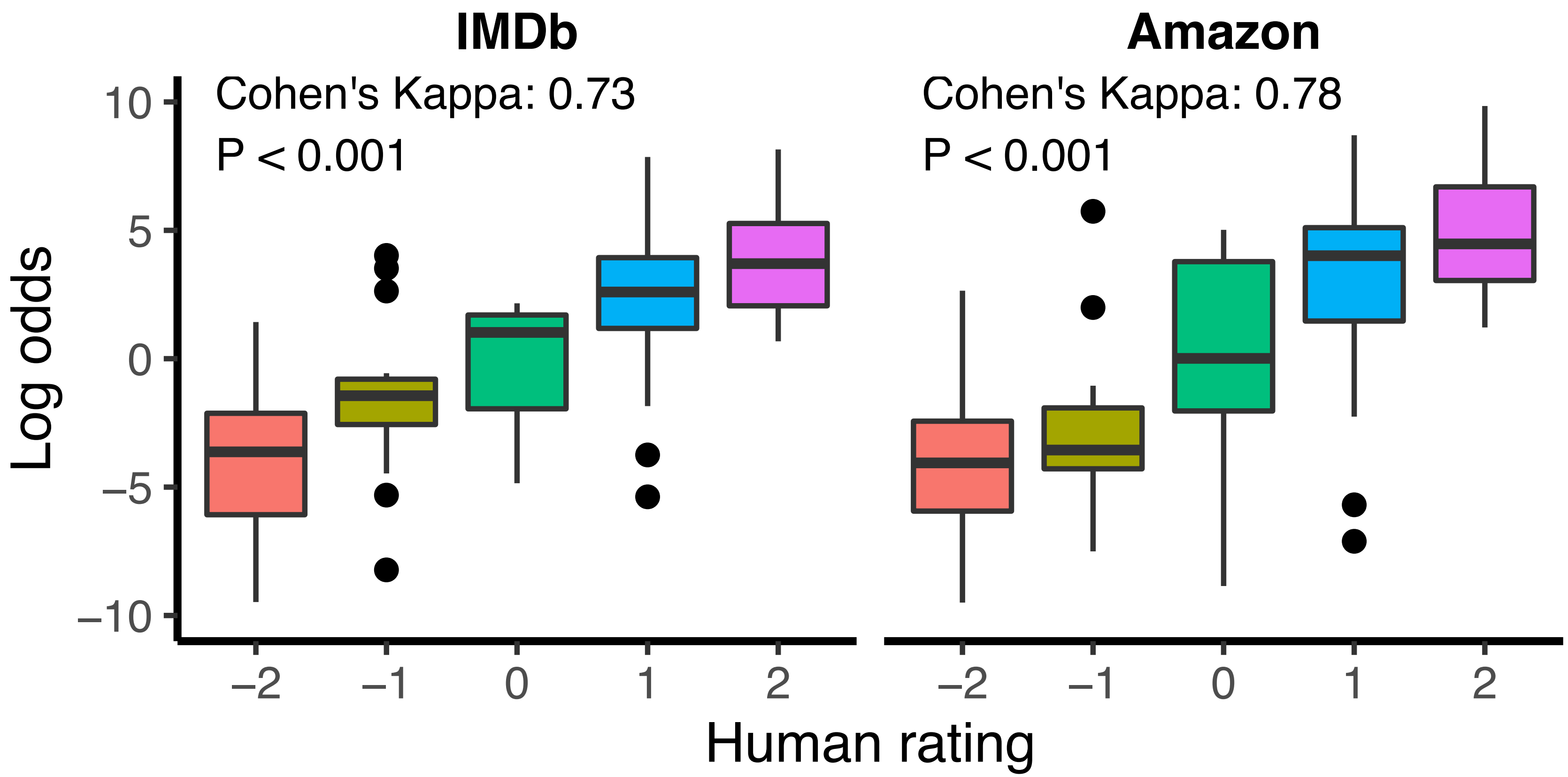
Control span size by penalizing overlap via JS divergence





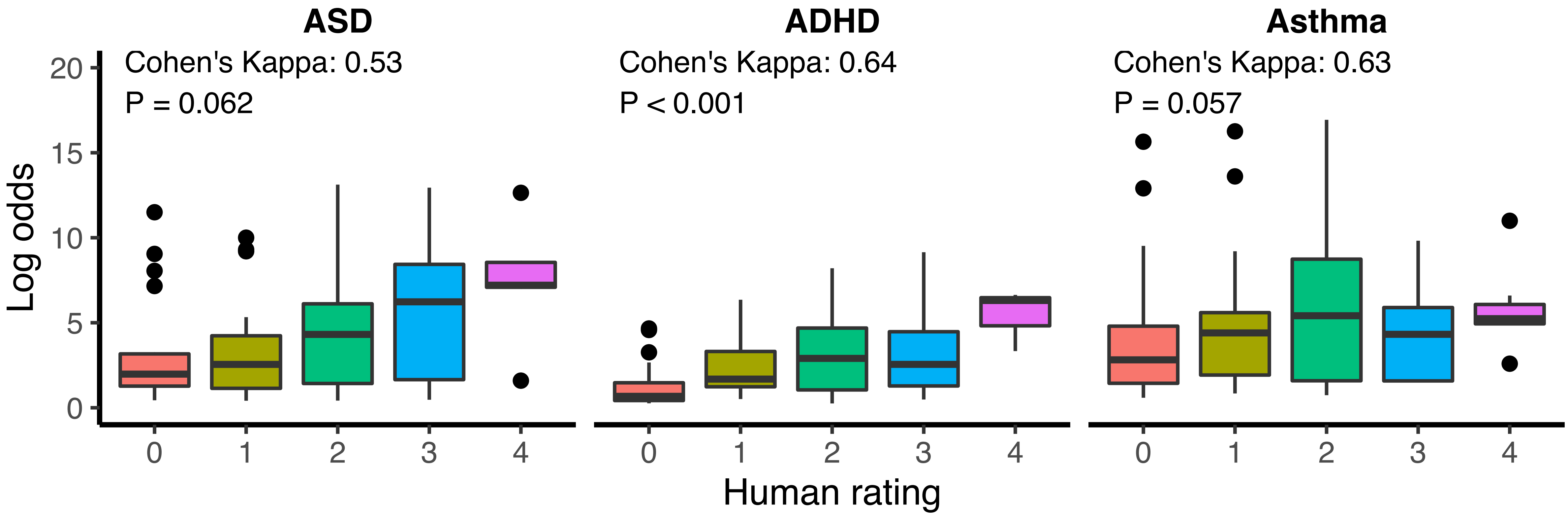
Subramanian V, Engelhard MM, Berchuck SI, Henao R, Chen L, Carin L. SpanPredict: Extraction of Predictive Document Spans with Neural Attention. Submitted to 16th conference of the European Chapter of the Association for Computational Linguistics (EACL).

Non-medical text: high agreement with humans



Subramanian V, Engelhard MM, Berchuck SI, Henao R, Chen L, Carin L. SpanPredict: Extraction of Predictive Document Spans with Neural Attention. Submitted to 16th conference of the European Chapter of the Association for Computational Linguistics (EACL).

Clinical notes: partial agreement with clinicians



Subramanian V, Engelhard MM, Berchuck SI, Henao R, Chen L, Carin L. SpanPredict: Extraction of Predictive Document Spans with Neural Attention. Submitted to 16th conference of the European Chapter of the Association for Computational Linguistics (EACL).

Passages most predictive of autism diagnosis

Passage (from note)	Change in predicted dx log-odds
subjective intake chief complaint problems with sleep, inattention, and behavioral concerns both in the home and school setting. DATE, recently more anger and recent tic like behavior	+6.95
psychologist presenting problem NAME is a 3 year, 4 month old female who was referred for a neurodevelopmental assessment due to concerns regarding her overall development, behavior, and social emotional functioning and to assess for autism spectrum disorder	+6.82
problem list diagnosis • disruptive behavior disorder • impaired speech articulation • daytime enuresis • other subjective visual disturbances • hypermetropia of both eyes • adhd attention deficit	+6.81
problem list diagnosis • anemia of prematurity • history of colitis • meconium tox for thc • extreme immaturity of newborn, 27 completed weeks • nasal congestion of newborn • presumed	+6.78
motor delay DATE • hypotonia DATE • clasped thumb DATE • polydactyly DATE • developmental	+6.74
therapy NAME was seen for developmental support during rop eye exam today. the	+6.65

Developmental and behavioral concerns are highly predictive

Premature birth and perinatal complications are also highly predictive

Subramanian V, Engelhard MM, Berchuck SI, Henao R, Chen L, Carin L. SpanPredict: Extraction of Predictive Document Spans with Neural Attention. Submitted to 16th conference of the European Chapter of the Association for Computational Linguistics (EACL).

Duke Forge, NICHD 1P50HD093074, NIMH/NINDS 1 R01 MH121329

Next steps: new methods, evidence for impact

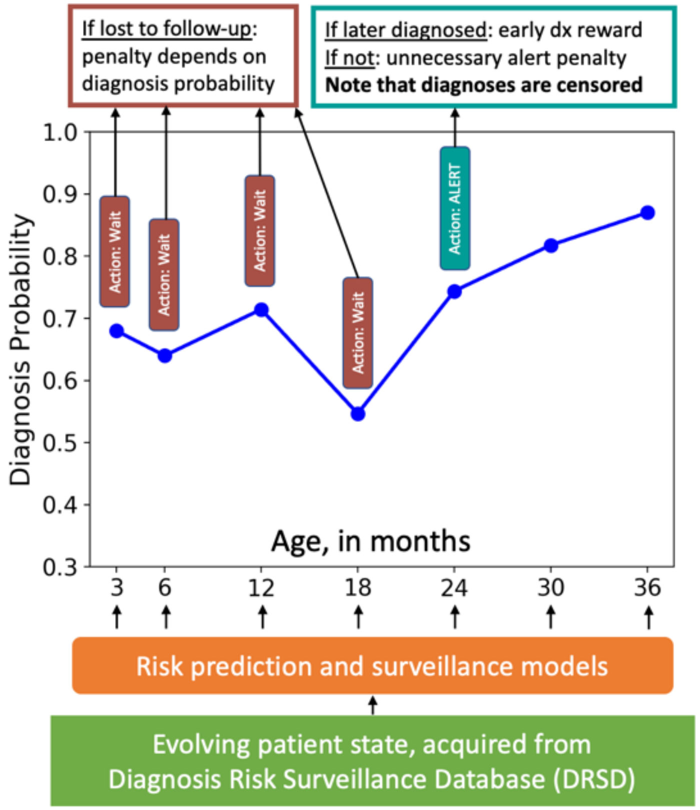


Figure 6: Conceptual Diagram of Risk Surveillance

- Silent deployment of autism and ADHD prediction in Duke EHR (K01 proposal)

Duke Clinical Research Institute
From Thought Leadership to Clinical Practice

Duke Surgery

- Prediction of retinopathy of prematurity in Pediatrix cohort
- Prediction of surgery risk

Thank you!

Contact: m.engelhard@duke.edu