The Israeli Institute for Applied Research in Computational Health

EPA 2020
28TH EUROPEAN CONGRESS OF PSYCHIATRY
4 - 7 July 2020

EUROPEAN PSYCHIATRIC ORGANISATION

Enhancing the detection of postpartum depression from electronic health records using machine learning algorithms

Guy Amit, PhD.
July 2020

Background and goals

Motivation

- Postpartum depression is one of the most common complications of pregnancy and childbirth, with estimated prevalence of 10-15%
- PPD risk is associated with biological, phycological and sociodemographic factors
- There are no quantitative tools for risk estimation, and screening is typically based on symptom questionnaires (such as the Edinburgh postnatal depression scale)
- Early identification of PPD risk during or before pregnancy may enable effective early intervention

Suggested solution

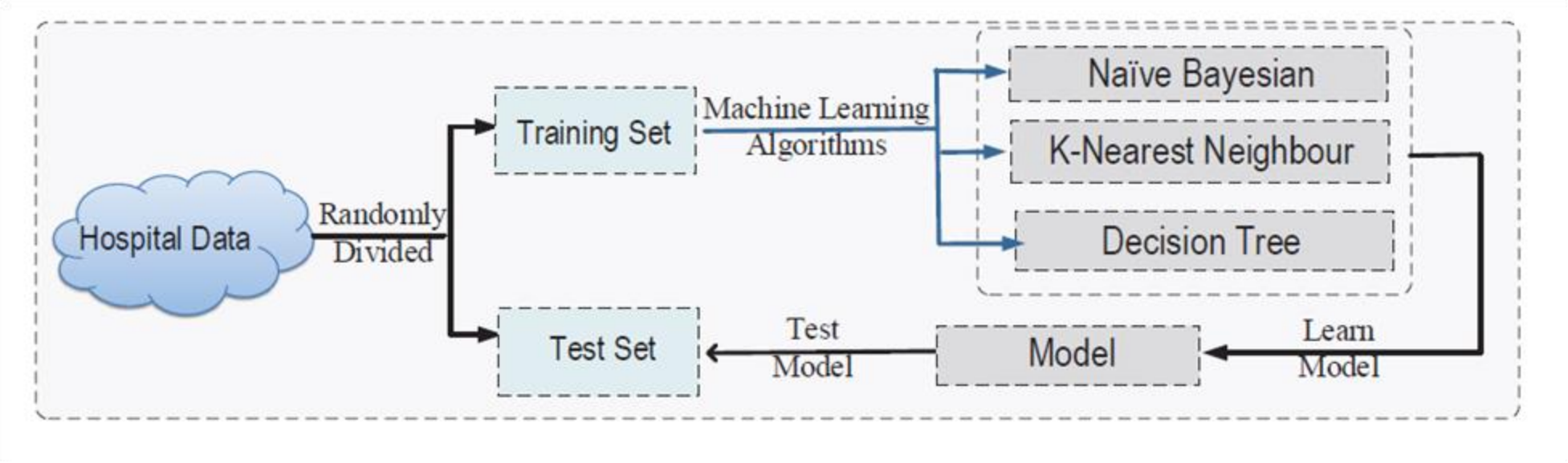
- A predictive model that uses electronic health records (EHR) for learning to identify patients at risk
- May enable early identification of patients at risk
- May be used to augment current screening tools

Expected impact

- Improved outcome for mother and child by early intervention

Machine learning in healthcare

Research approach: Use Machine Learning to identify signals within readily available EHR data to indicate individuals and sub-populations with higher risk for health-related outcomes



Data and patient cohort

- Dataset: UK primary care electronic health records (IQVIA-IMRD, ~18M patients)

1,084,394
First pregnancies
(1-Jan-2000 to 31-Dec-2018)

exclusions

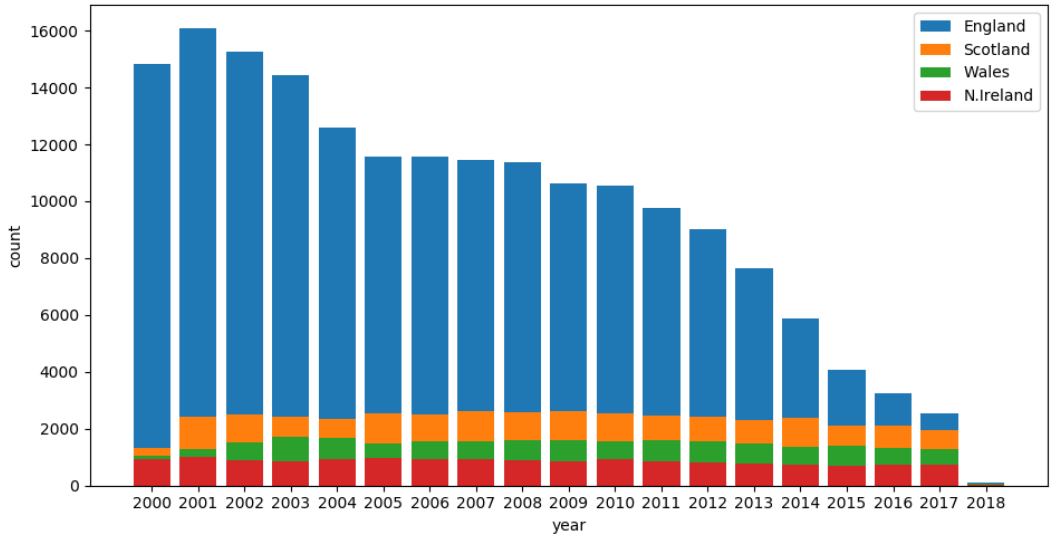
74,220
Non-valid age
(<18, >45)

855,204
Live births

508,438
Non-valid medical record

6,002
ambiguous outcome

266,544 Included
(age 18-45, valid & active medical record, non-ambiguous outcome)



Temporal validation
- 178,017 (68%) Train set (1/00 to 4/10)
- 82,568 (32%) Test set (5/10 to 12/18)

Geographical validation
- 177,833 (68%) Train set (England)
- 82,752(32%) Test set (S,W,NI)

Holdout set
- 5959 EPDS

PPD Outcome definition

At least one of the following indications <u>during the first year after birth</u>:

1. Women with depression related diagnoses (excluding depression symptoms and anxiety)
2. Women with new antidepressant prescriptions (excluding drugs with dual indications)
3. Women with depression related non-pharmacological treatment (NPT)

	Recorded Depression Tx	No recorded Depression Tx	Total (% pts)
Recorded Depression Dx	**16,284**	**5,869**	22,153 (8.3%)
No recorded Depression Dx	**13,555**	**230,836**	
Total (% pts)	29,839 (11.2%)		35,708 (13.4%)

Tx: treatment (antidepressants or NPT); Dx: diagnosis code

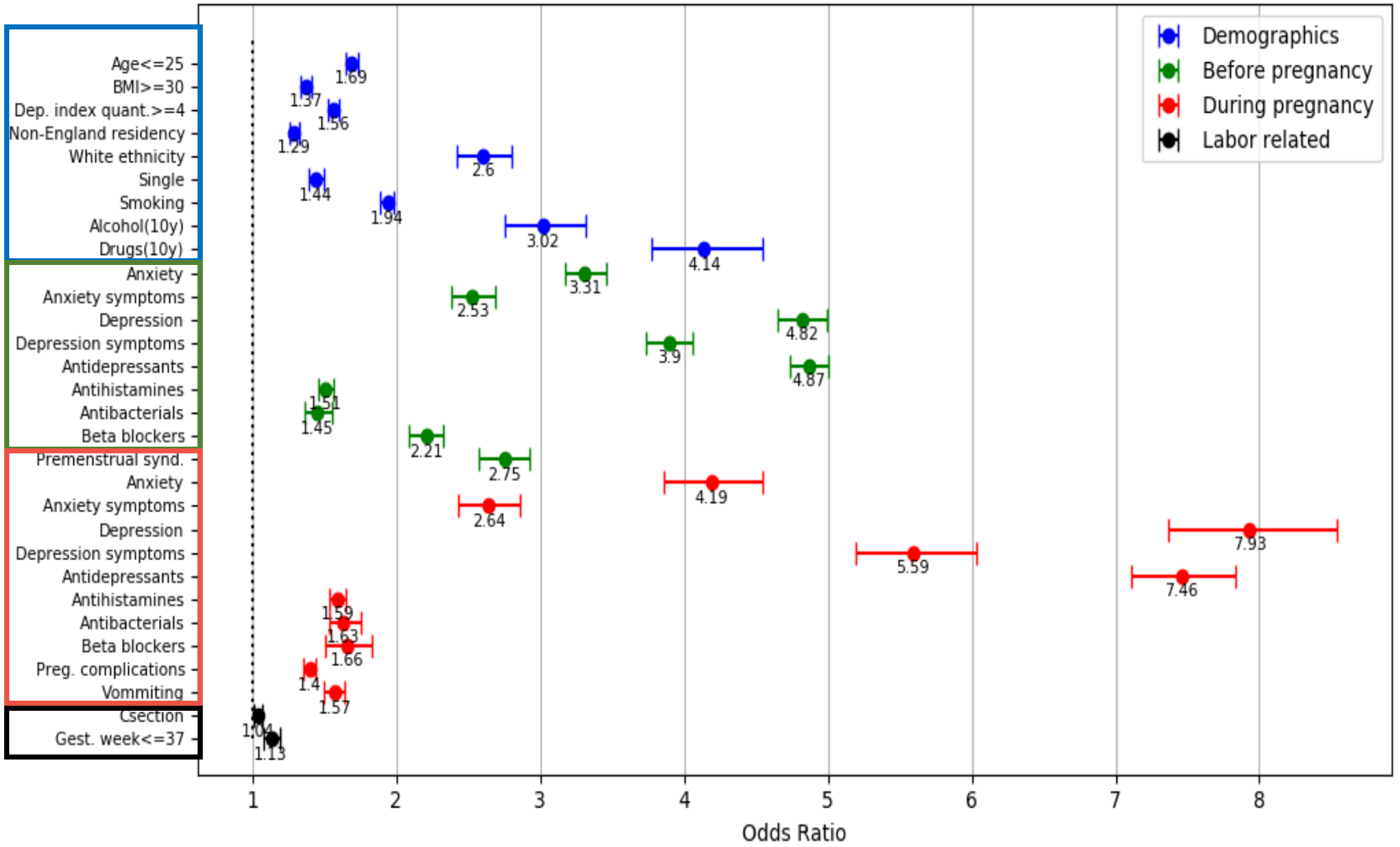
PPD prevalence	Train set	Test set	Holdout set
Geographical validation	12% (E)	15% (S,W,NI)	20% (18%,26%) E S,W,NI
Temporal validation	14% (00-10)	12% (10-18)	20% (18%,29%) 00-10 10-18

Predictor variables

- Demographic and socioeconomic
  - Age, BMI, marital status, ethnicity
  - Deprivation index, smoking, alcohol-use, drug-use
- Medical diagnoses during pregnancy
  - Mental disorders and symptoms: depression, anxiety, psychosi
  - Pregnancy complications: GDM, preeclampsia, vomiting
  - Miscellaneous health conditions: migraine, diarrhea
- Labor and infant-related
  - Labor complications: cesarean section, episiotomy
  - Infant-related: gestational week, birth weight, APGAR, feeding type
- Medical diagnoses before pregnancy (2y)
  - Mental disorders and symptoms, PMS
  - Visit count
- Drug prescriptions during / before pregnancy (2y)
  - Antidepressants, antibacterial, antihistamines, beta-blocking
  - Drug prescription count

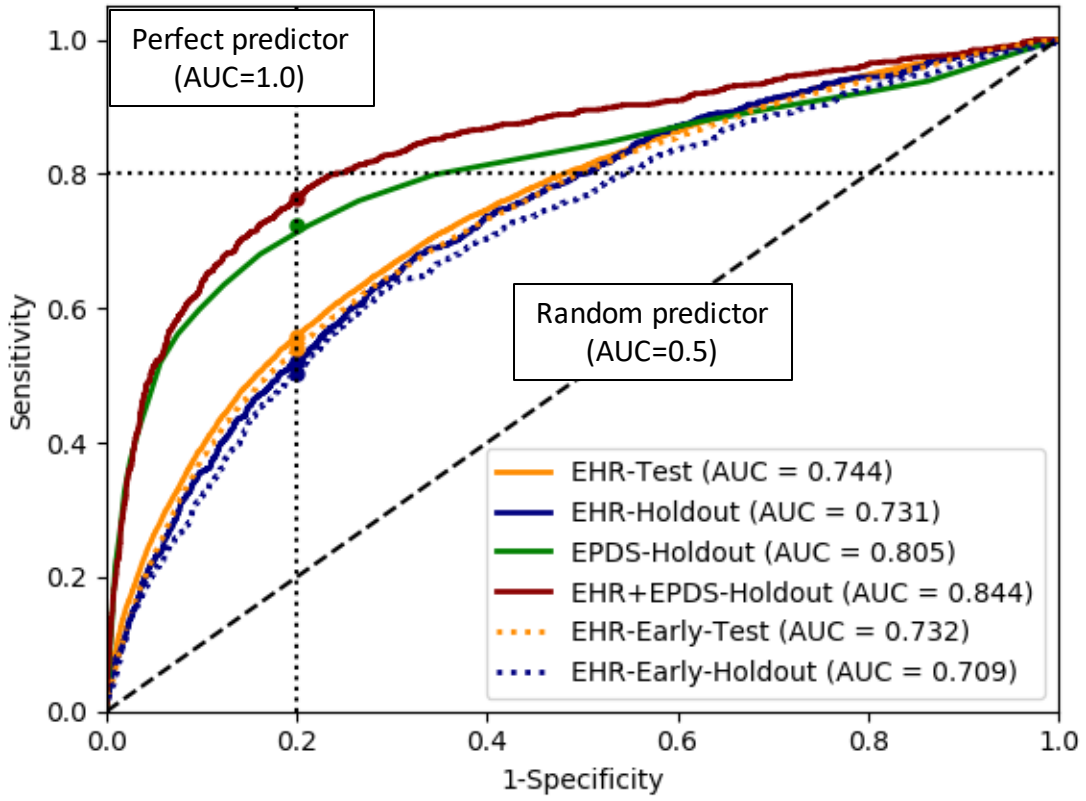
Patient characteristics and univariate analysis

Characteristic	Value
N	266544
Age (yrs)	30.0±5.8
Ethnicity	
White	99971 (37.5%)
Asian	7367 (2.8%)
Black	3167 (1.2%)
Other	2412 (0.9%)
Unknown	152573 (57.2%)
Marital status	
Single	34145 (12.8%)
Married	62929 (23.6%)
Unknown	169470 (63.6%)
Country	
England	182506 (68.5%)
Scotland	42113 (15.8%)
Wales	26565 (10.0%)
N. Ireland	15360 (5.8%)
Deprivation index (quantile)	3.03±1.3
Pre-pregnancy BMI	25.4±5.0
Cesarean section	51151 (19.2%)
Smoking	64778 (24.3%)
History of depression	17384 (6.5%)

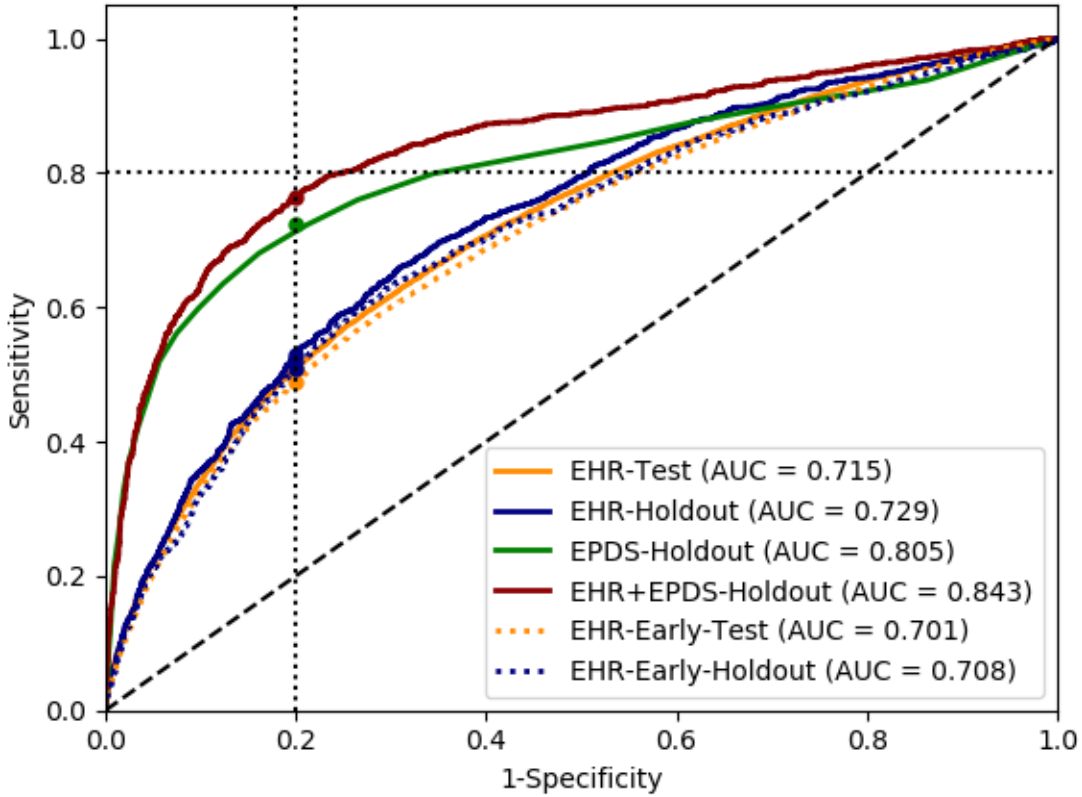


Prediction performance (gradient boosting trees models)

Temporal validation

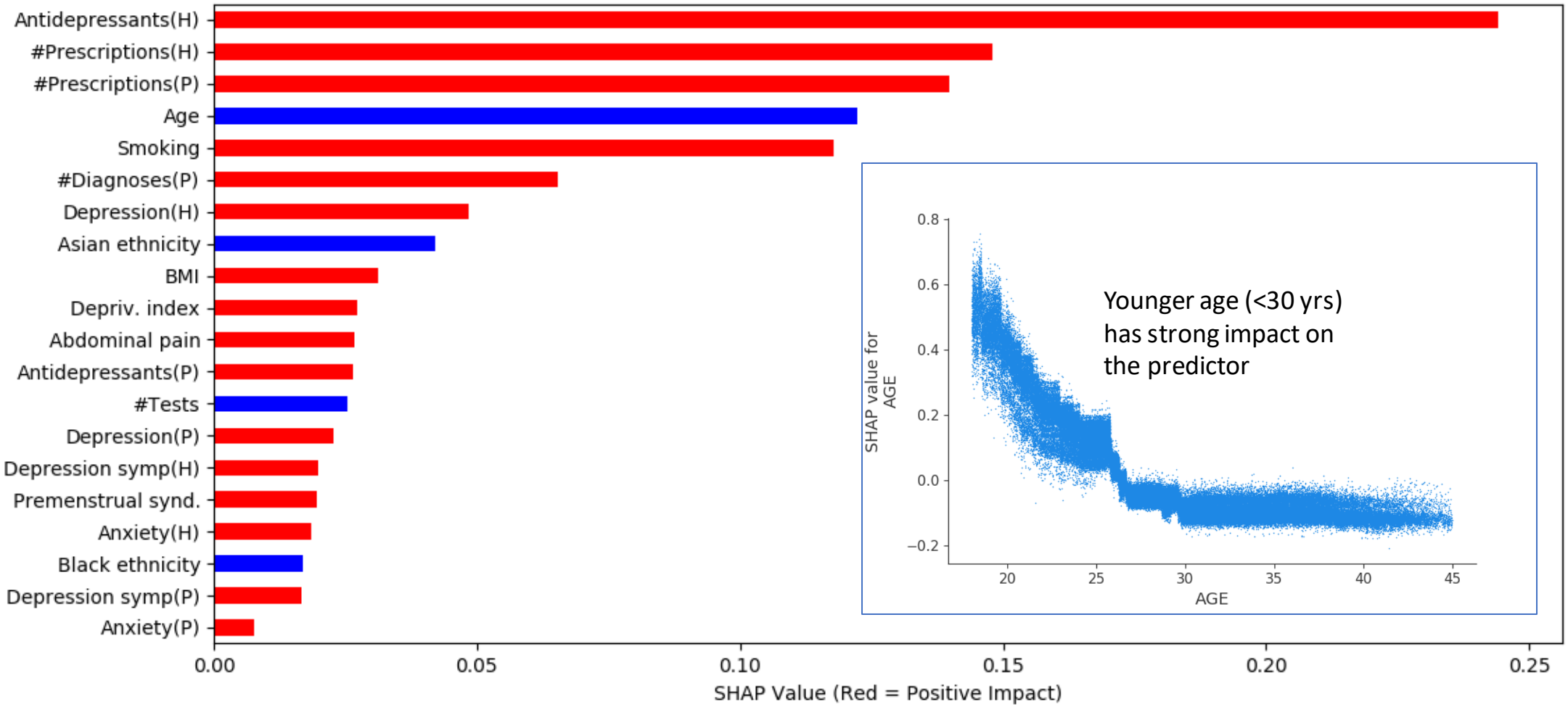


Geographical validation



- Combining EHR-based prediction with EPDS score improved of EPDS-alone (sensitivity 0.76 vs. 0.72, specificity 0.8)
- Early EHR-based prediction (before pregnancy) is nearly as accurate as late prediction (after labor)

Variable importance (SHAP analysis†)



RED = increase risk ; **BLUE = decreased risk**

P=during pregnancy, H=History (2y before pregnancy)

Higher SHAP value = stronger variable contribution

† Lundberg SM, Lee SI, A Unified Approach to Interpreting Model Predictions, NeurIPS 2017

Summary and conclusions

- PPD can be predicted from EHR data with fair accuracy, even before pregnancy
- EHR-based prediction can improve the accuracy of EPDS for PPD screening
- Although some of the risk factors are well known, integrating and quantifying them into an accessible risk score may have clinical value
- Future work:
  - External validation
  - Deployment in a clinical environment
  - A prospective study

Collaborators

- **Weil Cornell Medicine**
  - Jyotishman Pathak
  - Yiye Zhang
  - Alison Herman
  - Rochelle Joly
  - Meghan R. Turchioe
- **Sheba Medical Center**
  - Vered Bar
- **Our team at KI**
  - Irena Girshovitz
  - Pinchas Akiva

Contact: guy@kinstitute.org.il
http://kinstitute.org.il