Machine learning analysis of maternal pregnancy clinical notes to predict newborns at risk for neonatal abstinence syndrome

R / Medicine Conference
New Haven, CT
September 7, 2018

Joseph Chou, MD PhD
jchou2@mgh.harvard.edu
Neonatal Intensive Care, MGHfC

HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL

MassGeneral Hospital *for* Children℠

Who am I?

MassGeneral Hospital for Children

- Clinical neonatologist at MGH
  - no formal training as a programmer or data scientist
  - started learning R in 2014

- Clinician perspective
  - data-driven improvement of clinical care delivery
  - analytic reports, QI dashboards, clinical decision support tools

- I have no conflicts of interest to disclose
- This study was approved by the Partners Institutional Review Board

Opioid Use Disorder and Pregnancy

MassGeneral Hospital for Children™

- Opioid use disorder is a public health emergency
  - heroin, oxycodone, fentanyl, morphine
- 2.3% of women who delivered in MA with opioid use disorder
- Opioids cross the placenta to the fetus
  - withdrawal symptoms in the newborns after birth
  - ***Neonatal Abstinence Syndrome*** (NAS)
- Rate of NAS in MA is almost three times the national average
  - 16 newborns per 1,000 live births

Neonatal abstinence syndrome (NAS)

MassGeneral Hospital for Children™

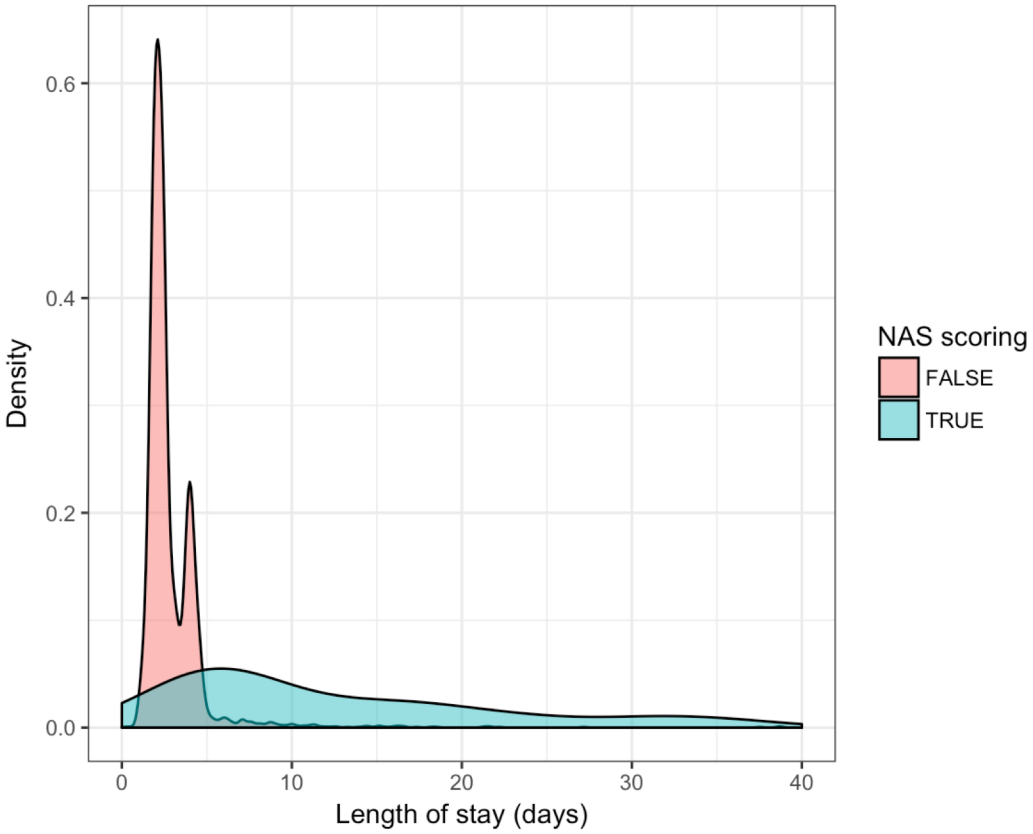
- Neurologic
  - Increased muscle tone, tremors, hyperactive reflexes, irritability and restlessness, difficulty sleeping, **seizures**, skin excoriation
- Autonomic dysfunction
  - Yawning, nasal stuffiness, sweating, sneezing, fever, skin mottling
- Gastrointestinal abnormalities
  - Diarrhea, regurgitation, vomiting, dysmature swallowing, **poor feeding, failure to thrive**
- Respiratory
  - increased respiratory rate, cessation of breathing
- Symptom monitoring; e.g., ***Finnegan score***
  - ***ALL*** newborns believed at risk for NAS are scored

NEONATAL ABSTINENCE SCORING SYSTEM

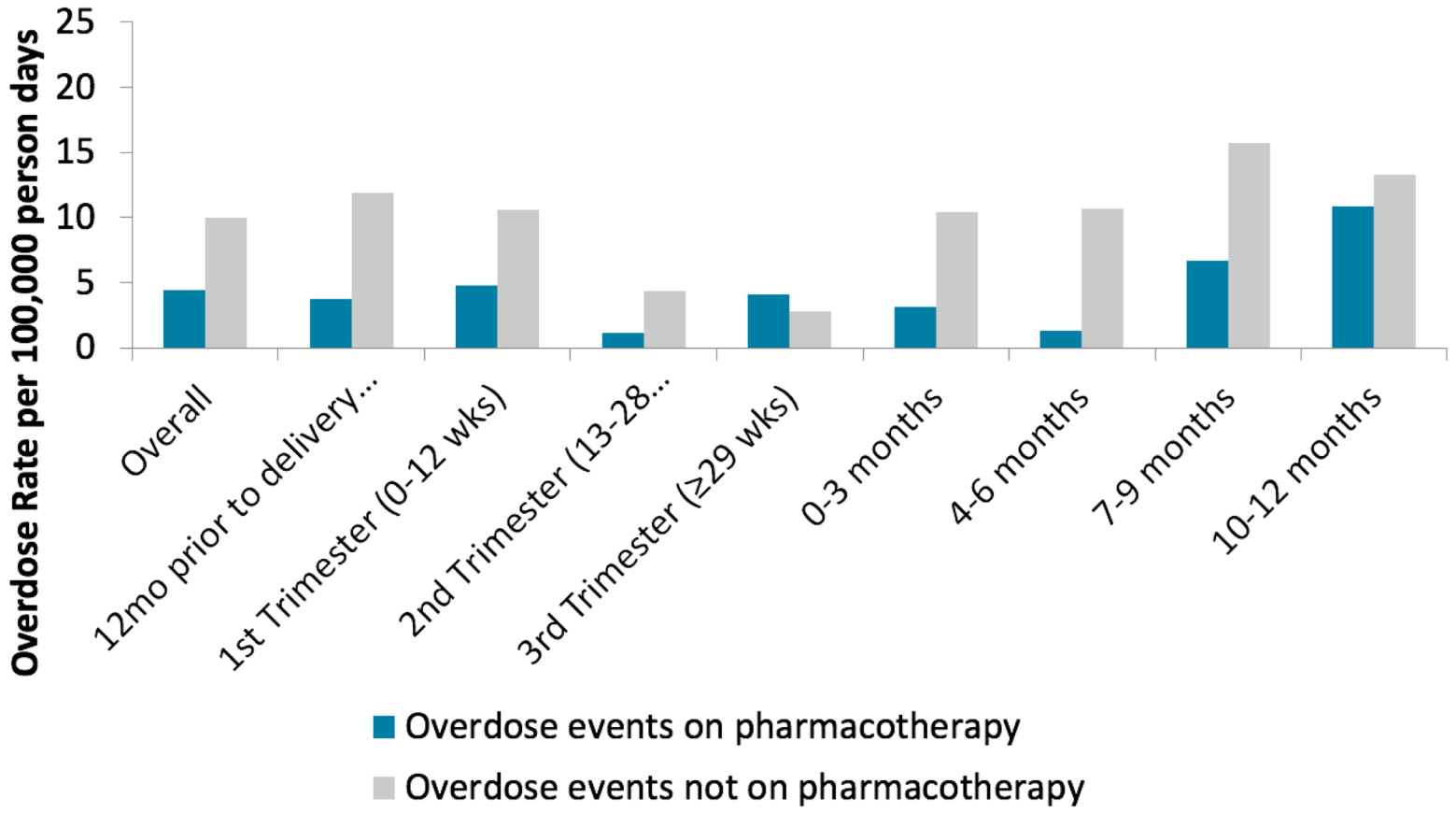
SYSTEM	SIGNS AND SYMPTOMS	SCORE		AM
CENTRAL NERVOUS SYSTEM DISTURBANCES	Continuous High Pitched (or other) Cry	2		
	Continuous High Pitched (or other) Cry	3		
	Sleeps <1 Hour After Feeding	3		
	Sleeps <2 Hours After Feeding	2		
	Sleeps <3 Hours After Feeding	1		
	Hyperactive Moro Reflex	2		
	Markedly Hyperactive Moro Reflex	3		
	Mild Tremors Disturbed	1		
	Moderate-Severe Tremors Disturbed	2		
	Mild Tremors Undisturbed	3		
	Moderate-Severe Tremors Undisturbed	4		
	Increased Muscle Tone	2		
	Excoriation (Specific Area)	1		
	Myoclonic Jerks	3		
	Generalized Convulsions	5		
METABOLIC/VASOMOTOR/RESPIRATORY DISTURBANCES	Sweating	1		
	Fever 100.4°-101°F (38°-38.3°C)	1		
	Fever > 101°F (38.3°C)	2		
	Frequent Yawning (>3-4 times/interval)	1		
	Mottling	1		
	Nasal Stuffiness	1		
	Sneezing (>3-4 times/interval)	1		
	Nasal Flaring	2		
	Respiratory Rate >60/min	1		
	Respiratory Rate > 60/min with Retractions	2		
GASTRO-INTESTIONAL DISTURBANCES	Excessive Sucking	1		
	Poor Feeding	2		
	Regurgitation	2		
	Projectile Vomiting	3		
	Loose Stools	2		
	Watery Stools	3		

Resource costs of NAS

- NAS prolongs hospital length of stay
- Average cost per NAS baby is $66,700



Reduction in overdoses when pregnant women are identified and treated



- treatment (*e.g.*, methadone, suboxone) significantly reduces overdose risk
- new MGH clinic launched in 2018 for pregnant women with opioid use disorder
- ***how can we identify mothers to enroll?***

Goal

Use maternal clinical free text notes to predict newborns at risk for neonatal abstinence syndrome

Methods (all done in R)

MassGeneral Hospital for Children™

- Retrospectively identified all mothers who delivered at MGH between April 2016 and June 2018
  - 80% training set and a held-out 20% test set
- Predictors
  - Maternal clinical notes during the first two trimesters of pregnancy
  - Generated a *'bag-of-words'* for each mother
- Prediction target: newborns suspected to be at risk for NAS
  - Defined by > 5 Finnegan NAS scores during the birth encounter

Dataset and demographics

MassGeneral Hospital for Children™

	Training Set	Test Set
total number	***5,554***	1,387
number at risk for NAS (%)	68 (***1.22%***)	16 (1.15%)
total notes	103,217	25,775
number notes per pregnancy (IQR)	17 (11 – 24)	16 (11 – 24)
unigram features	***57,105***	
gestational age at birth (weeks)	39.4	39.4
birth weight (grams)	3350	3340
birth encounter length of stay (days)	2.4	2.4

- significant class imbalance
- each mother's notes converted to a vector of TF-IDF values (length 57,105)

Term Frequency – Inverse Document Frequency

MassGeneral Hospital for Children™

For each mother:

- generate bag-of-words from clinical notes from the first two trimesters
- each word a feature (unigram, case insensitive, no numbers, not stemmed)

For each unigram word feature:

- **term frequency (TF)**, *within each mother's bag-of-words*
  = # occurrences in document / total # of words in document
- **inverse document frequency (IDF)**, *within the entire training corpus*
  = ln(# documents / # documents with the term)
- **TF-IDF** = TF * IDF

Each mother's notes was converted to a vector of **TF-IDF** values (length 57,105), roughly:

- the frequency of a term in a particular mother's notes (TF),
- normalized by the frequency of that term appearing in any mother's notes (IDF)

Predictive model fitting

MassGeneral Hospital for Children™

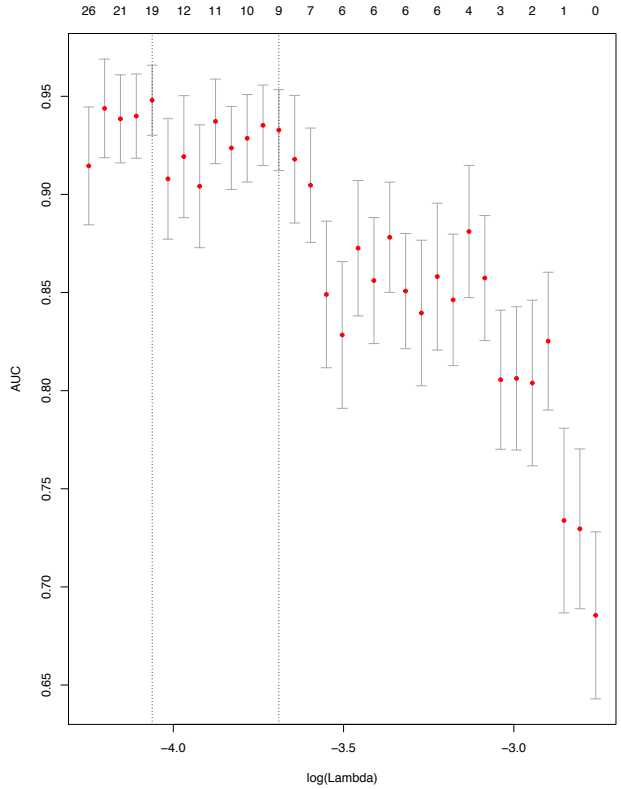
- Binary outcome (at risk for NAS) modeling via penalized logistic regression
  - elastic-net (*glmnet*)
  - predictor matrix of 5,554 patients x 57,105 TF-IDF word features
  - took advantage of *glmnet* support for sparse input matrices

- $p >> n$, high dimensionality predictor for fairly small dataset
  - requires regularization / feature selection
  - preference for sparse and interpretable model
  - lasso (L1 regularization) performs implicit feature selection
  - elastic-net penalty heavily weighted to lasso (alpha = 0.99)

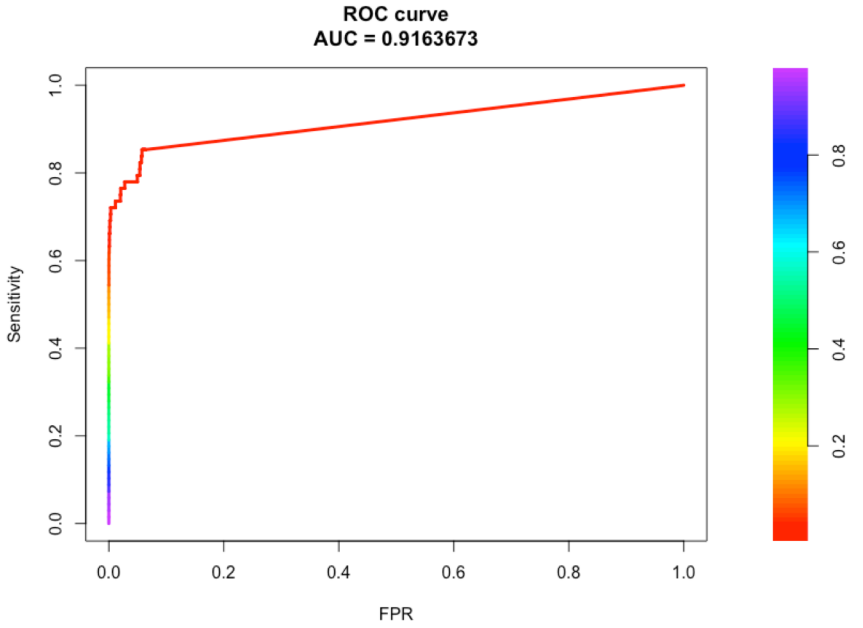
- twenty-fold cross validation to select overall penalty strength (lambda)

Training set modeling

CV for lambda selection



AU-ROC curve = 0.916



- used penalty weight *lambda = lambda.1se*

Selected features predictive of newborns at risk for NAS were ***sparse*** and ***interpretable***

MassGeneral Hospital for Children™

Word	Beta	IDF
nas	244	4.65
heroin	106	4.60
clonidine	102	5.85
opioid	84.9	5.01
subutex	53.6	5.16
methadone	47.2	4.88
suboxone	35.4	5.22
abuse	25.5	2.71
detox	9.46	5.79
(intercept)	-4.74	*NA*

Unigram features from maternal clinical notes during the first two trimesters

- Beta: logistic regression coefficients to be used with TF-IDF predictors
- IDF: inverse document frequency from the training corpus (need for new predictions)

Predictive Model Performance

MassGeneral Hospital for Children™

	Training Set	Test Set
total number, n	5,554	1,387
number at risk for NAS (%)	68 (1.22%)	16 (1.15%)
AU-ROC	0.916	0.930
sensitivity	72.1%	68.8%
specificity	99.7%	99.8%
positive predictive value	73.1%	78.6%
negative predictive value	99.7%	99.6%
prediction accuracy (accuracy > no information rate)	99.3% ($p < 0.0001$)	99.4% ($p = 0.021$)

- threshold chosen to limit the false positive rate on the training set to <1%
- test set performance similar
- adequate performance for clinical utility

Another use case – identifying newborns

MassGeneral Hospital for Children™

- Previously, intentional limitation of information available for modeling
  - notes limited to first two trimesters of pregnancy
  - goal was to identify mothers while still pregnant

- ***Another use case: at point-of-care on newborns after delivery***
  - use ***ALL*** prenatal notes to identify newborns who need to be monitored

Identifying newborns at-risk for NAS using ***all*** maternal notes up to delivery

MassGeneral Hospital for Children™

	Training Set	Test Set	Full, no date limit
total number, n	5,554	1,387	6,941
number at risk for NAS (%)	68 (1.22%)	16 (1.15%)	84 (1.21%)
AU-ROC	0.916	0.930	***0.952***
sensitivity	72.1%	68.8%	***81.0%***
specificity	99.7%	99.8%	99.7%
positive predictive value	73.1%	78.6%	77.3%
negative predictive value	99.7%	99.6%	99.8%
prediction accuracy (accuracy > no information rate)	99.3% ($p < 0.0001$)	99.4% ($p = 0.021$)	99.5% ($p < 0.0001$)

- Sensitivity improved to 81% by including more recent maternal notes
- Can further improve performance by incorporating other predictors, *e.g.* ***toxicology screens, medications, billing problem codes*** (data not shown)

Generalizable approach

- Use same TF-IDF predictor matrix to identify features associated with ***other outcomes***

- **Preterm** delivery before 35 weeks
  - features generally related to **multiple gestation**
  - triplet, mfm, monochorionic, mono, twins, pprom, ttts, twin
    (maternal-fetal medicine, preterm premature rupture of membranes, twin-to-twin transfusion syndrome)

- **Large for gestation** newborns (>90th percentile)
  - features suggestive of **diabetes**
  - macrosomia, diabetes, insulin, retinopathy, joslin

- Could this be used in the future for ***hypothesis generation / new knowledge***?

Conclusions

- Simple natural language processing of unstructured clinical notes can yield well-performing predictive models that are of clinical utility
  - newborns at risk for NAS predicted with pregnancy clinical notes
  - ***sparse and interpretable*** word features
  - ***minimal prior knowledge*** and ***minimal subject matter expertise***
  - ***automated*** data extraction from the EHR without manual chart review
  - ***minimal computational resources***
  - ***minimal cost*** – R / RStudio / publicly available packages

- ***Applications*** include identification of:
  - ***mothers for enrollment*** in substance use disorder clinics
  - at risk ***newborns at the point of care*** who require symptom screening (Shiny app!)
  - patient population cohorts for ***retrospective epidemiological studies***

Acknowledgments

- Dr. Paul Lerou
  - Division Chief of MGH*f*C Newborn Medicine

- Dr. Davida Schiff
  - Director of the MGH ***HOPE*** Clinic
  - ***H***arnessing support for ***O***pioid and substance use disorder in ***P***regnancy and ***E***arly childhood

- All the contributors to the R community who make this possible

Thank you!

References

- **caret**: Kuhn M et al. (2018). caret: Classification and Regression Training. R package version 6.0-80.
- **glmnet**: Jerome Friedman, Trevor Hastie, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22.
- **pROC**: Robin X et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, p. 77.
- **PRROC**: Grau J et al. (2015). PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. Bioinformatics (31) 15, pp. 2595-2597.R package version 1.3.1.
- **RODBC**: Brian Ripley and Michael Lapsley (2017). RODBC: ODBC Database Access. R package version 1.3-15
- **stringr**: Hadley Wickham (2018). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.3.1.
- **tidyverse**: Hadley Wickham (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1.
- **tidytext**: Silge J, Robinson D (2016). “tidytext: Text Mining and Analysis Using Tidy Data Principles in R.”

- Marafino BJ, Boscardin WJ, and Dudley RA (2015). “Efficient and sparse feature selection for biomedical text classification via the elastic net: Application to ICU risk stratification from nursing notes.” Journal of Biomedical Informatics, 54, 114-120.