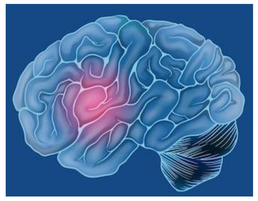
RWD51

Identifying Migraine in Primary Care Workflows: Prevalence, Risk Signals, and Predictive Tools

Taylor P. Ikner, MS and Richard W. Hass, PhD

College of Population Health, Thomas Jefferson University, Philadelphia, PA 19107

Background and Significance



- Migraine affects >1 billion people globally and is a leading contributor to disability and reduced quality of life
- Despite its burden, migraine remains under-recognized and underdiagnosed in primary care settings

•Primary care providers often encounter nonspecific symptoms (e.g., fatigue, generalized pain), which may obscure migraine diagnosis

•Delayed or missed diagnosis may lead to increased emergency department utilization and progression to chronic migraine

•The widespread use of electronic medical records (EMRs) enables the use of routinely collected clinical and social data for earlier identification

•Predictive modeling approaches may help identify high-risk patients who could benefit from preventive care and timely intervention

Aims

- Identify demographic, clinical, and social risk factors associated with migraine
- Evaluate the performance of machine learning and statistical models for detecting migraine risk using EMR-derived data

Methods

- A cross-sectional analysis was conducted using de-identified EMR data from outpatient primary care encounters across Jefferson Health clinics.

Study Population

- N = **59,088** adult outpatient visits
- Adult patients receiving routine primary care services
- Multi-site clinic representation including urban and suburban settings

Predictor Variables

- **Demographics Factors:** Age, Sex, Clinic Location
- **Clinical factors:** Pain-related conditions, psychiatric comorbidities, vital signs
- **Social risk indicators:** Health-related social needs (HRSN) documented in the EMR

Model Development & Evaluation

- Two models evaluated:
  - L1-regularized logistic regression
  - Random Forest
- Data split into **70% training and 30% testing subsets**
- Model performance assessed using:
  - ROC-AUC and PR-AUC
  - Calibration and Brier score
  - Threshold-based diagnostic metrics

Key Findings

HRSN Status and Migraine Prevalence

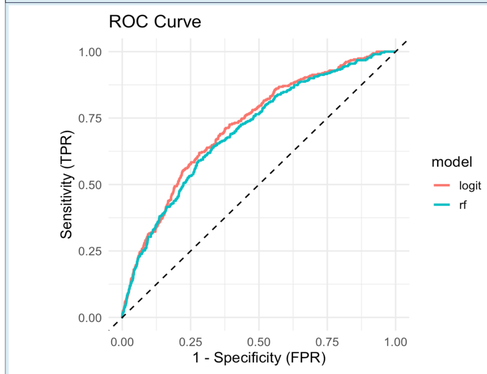
Any HRSN	Migraine Flag	N	%
No	No	53009	**97.53**
No	Yes	1345	**2.47**
Yes	No	4584	**96.83**
Yes	Yes	150	**3.16**

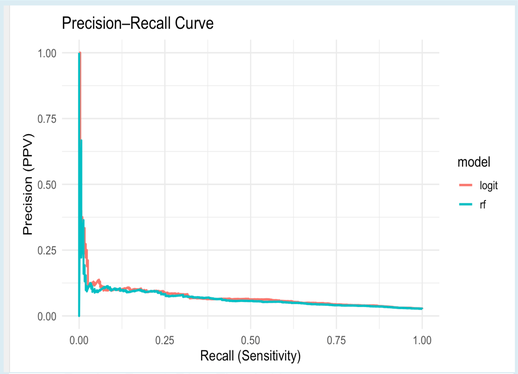
Variable Importance Based on Impurity Index

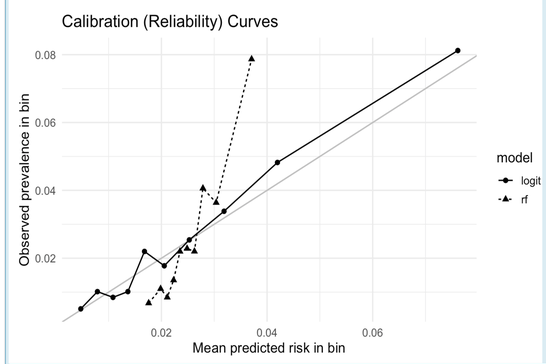
Variable	Importance
Age	**41.15**
Sex - Male	**27.09**
Sex - Female	**26.75**
BMI	**26.22**
Asthma Dx Yes	**15.19**
Fibro Dx No	**14.94**
Depression Dx Yes	**14.62**
Anxiety Dx No	**14.55**

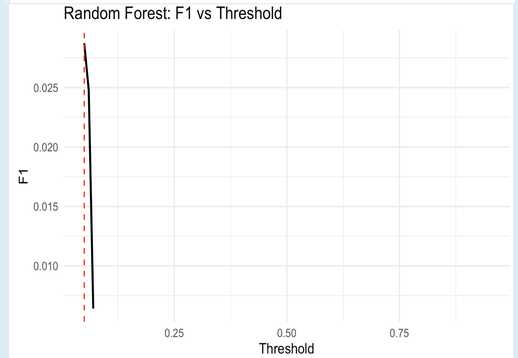
Sensitivity Analyses: Model Performance on Test Set

Scenario	Model	ROC-AUC	PR-AUC	Brier
Same-Day Dx Only	Random Forest	**0.736**	**0.083**	**0.024**
Same-Day Dx Only	L1- Logistic	**0.710**	**0.055**	**0.024**
Exclude HRSN	L1- Logistic	**0.720**	**0.073**	**0.025**
Exclude HRSN	Random Forest	**0.705**	**0.069**	**0.025**
WHO BMI + ACC/AHA BP	Random Forest	**0.745**	**0.101**	**0.024**
WHO BMI + ACC/AHA BP	L1- Logistic	**0.731**	**0.071**	**0.024**









Threshold	ACC	Sensitivity	Specificity	PPV	NPV	F1	Youden
0.06	**0.925**	**0.212**	**0.9439**	**0.0928**	**0.97803**	**0.1293**	**0.1568**
0.05	**0.891**	**0.303**	**0.9071**	**0.0808**	**0.97972**	**0.1276**	**0.2103**

Discussion

- Migraine prevalence was **2.53% (n = 1,495)**, suggesting underdiagnosis in primary care
- Higher prevalence observed among younger adults and female patients
- Strong associations with migraine were observed for psychiatric comorbidities, chronic pain conditions, and demographic factors such as age and sex
- Models showed moderate ROC-AUC **(0.705–0.745)** but low PR-AUC **(0.055–0.101)** due to rare outcomes
- Logistic regression showed better calibration and more reliable probabilities than Random Forest
- Clinical and social factors modestly improved predictive signal
- These findings support the growing recognition that migraine burden is influenced by complex biopsychosocial interactions
- EMR-based risk stratification tools may support conservative screening and guide referral to specialty care

Limitations

- Cross-sectional design limits assessment of causality and temporal patterns
- ICD-10 coding may underestimate prevalence due to undiagnosed or undocumented cases
- Missing data (e.g., BP categories, HRSN) may have impacted model performance
- Limited clinical detail (e.g., headache symptoms, medications, referrals)
- Low prevalence (2.53%) led to class imbalance and reduced sensitivity
- Findings may have limited generalizability due to a single health system dataset

Conclusion

- EMR-based predictive models can identify meaningful demographic, clinical, and social risk signals associated with migraine in primary care populations
- The low observed prevalence (2.53%) suggests substantial underdiagnosis in routine care
- Logistic regression provided more reliable and well-calibrated probability estimates for clinical use
- Models are best applied as **risk-flagging tools** to identify patients who may benefit from further evaluation and follow-up
- Future work should focus on improving sensitivity through enhanced data inputs and modeling approaches

References shared upon request
txi036@jefferson.edu

@ResearchAtJeff