Back to School webinar: Real World Risk Prediction in Learning Health Systems

September 8, 2020
2:00 – 3:00 p.m. ET
The audio and slide presentation will be delivered directly to your computer

• Speakers or headphones are required to hear the audio portion of the webinar.

• If you do not hear any audio now, check your computer’s speaker settings and volume.

• Please take the questionnaire after watching the webinar.
Disclaimer

This webinar is brought to you by AcademyHealth through the **HSRProj program**. HSRProj is a joint effort of AcademyHealth and the Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, funded by the National Library of Medicine, National Institutes of Health, U.S. Department of Health & Human Services.

The opinions and recommendations expressed in this presentation are those of the individual presenters, and do not necessarily reflect those of AcademyHealth or the National Library of Medicine.
Welcome!

Margo Edmunds, Ph.D.
Vice President, Evidence Generation and Translation
AcademyHealth

Patricia Gallagher, M.L.S, M.A., AHIP
Project Officer
U.S. National Library of Medicine (NLM)
NLM Resources for HSR

https://www.nlm.nih.gov/nichsr

Finding and Using Health Statistics

About Health Statistics

How many Americans are affected by cardiovascular disease each year? Did the number of Americans with health insurance increase significantly following the passage of the Affordable Care Act? How do the leading causes of death in the United States differ from those worldwide?

The answers to these important questions and many others lie in the data collected by government, private, and non-profit agencies and organizations. The health statistics reported by these groups are integral to monitoring trends in the health status of populations, planning the allocation of health care resources, and evaluating the effectiveness of public health interventions.

This course for librarians and students in health sciences describes different types of health statistics, how they are collected, and where they can be found.

Course Goals

1. Understand what health statistics measure and how to use them to improve general health
2. Gain a basic knowledge of the statistical terms commonly used when reading about health statistics
3. Learn different ways health information can be collected, and the pros and cons of each
4. Become familiar with a variety of online sources for health statistics
5. Create a set of strategies to find specific health statistics
Request for Information (RFI): Information and Data Resources Needed by the Health Services Research Community for Research and Practice.

Vision
A world in which evidence informs decisions for **optimal health for all.**

Mission
AcademyHealth improves health and health care for all by advancing evidence to **inform policy and practice.**
Vision

• **I. Impact**
  Accelerate and amplify efforts to ensure that high quality, trustworthy data, valid measures and evidence are used for decisions in policy and practice.

• **II. Workforce**
  Develop and sustain a diverse workforce to respond to the changing needs of stakeholders who need evidence to advance health and health care improvement.

• **III. Engagement**
  Enhance our engagement with the individuals and organizations who use evidence to drive health improvement and health equity in the future.

• **IV. Innovation**
  Embrace innovation, technology and other societal trends to advance and inform new and relevant evidence to achieve health improvement for all.
HSR is the multidisciplinary field of scientific investigation that studies how social factors, financing systems, organizational structures and processes, health technologies, and personal behaviors affect access to health care, the quality and cost of health care, and ultimately our health and well-being.

National Academy of Medicine, 2018
NLM and AcademyHealth collaborate to host professional development activities.

The Back to School Webinar has been an annual tradition since 2015.

The HSRProj Research Competition for Students has been an annual tradition since 2017.
Presenters and Objectives
Learning Objectives:

• At the conclusion of this webinar, participants will be able to:
  • Describe ways big data can be applied to enhance public health and health services research
  • Define predictive modeling
  • Identify practical considerations in the implementation of predictive modeling

Speaker: Erin Holve
Introduction to data science tools and methods
Ernest Moy
TOP MACHINE LEARNING ALGORITHMS YOU SHOULD KNOW

- Linear Regression
- Logistic Regression
- Linear Discriminant Analysis
- Classification and Regression Trees
- Naive Bayes
- K-Nearest Neighbors (KNN)
- Learning Vector Quantization (LVQ)
- Support Vector Machines (SVM)
- Random Forest
- Boosting
- AdaBoost

Real World Predictive Analytics in Learning Health Systems

- Predictive Analytics: Use of current & past data to predict future outcomes.
- Real World
  - Model has to work reliably.
  - Data has to be processed to give answers in time to act.
  - Tool has to be acceptable and understandable by users.
- Learning Health System
  - Model improves health.
  - Model evolves, e.g., to reduce inequities.
Veterans are more likely than Non-Veterans to fill opioid prescriptions in recent years.

Development of the predictive model and the randomized trial to test the application of the model

Jodie Trafton
Design of a predictive model

• Needs to align with its use case!!!
• Many design decisions will need to be made as you develop and optimize your model
• These should be informed by your intended use of the model
• Here, I will walk you through:
  • the genesis of our use case for the STORM predictive model,
  • how that informed design and development of the model,
  • and our implementation of the model into clinical practice.
In 2010: Clinical Practice Guidelines for Chronic Opioid Therapy and Substance Use Disorder

- Guidelines available
- Recommended effective treatments and risk mitigation strategies
- Not consistently implemented in practice
How could we facilitate guideline-based practice?

- Worked with guideline authors
- Operationalized guideline concepts into codable data elements based on VHA medical record data
- Incorporated into a set of key guideline adherence metrics
- Built into computerized decision support that presented key risk factors and tracked patient tailored recommendations for care
- Pilot tested in primary care practice
Why develop a predictive model?

• Decision support was highly rated by clinicians, but....
  • Was only consistently adopted by clinicians who were already following guidelines at high rates
  • Simplified following guidelines, but risk mitigation was still time consuming to do
  • Clinicians wanted to know when they really needed to prioritize opioid-related risk mitigation

• Clinical Question:
  • Given that I don’t have time to implement all risk mitigation with all patients, when should I prioritize these interventions over other clinical priorities?
    – Which patients are likely to suffer harm from not receiving full risk mitigation?
A second use case

- Meanwhile, we were starting a national Overdose Education and Naloxone Distribution (OEND) program
  - Set up protocols and clinician and patient trainings to facilitate overdose prevention, identification and rescue protocols
  - Providers wanted to know which patients needed to be prioritized to receive OEND

- Clinical Question:
  - Which patients with access to opioids are most likely to experience an accidental overdose or suicidal ideation/behavior?
Modeling Goal: Predicting what for whom?

- Develop a predictive model that estimates risk of overdose or suicide ideation/behavior among patients exposed to opioids
  - Modeled in two cohorts
    - Patients who received an opioid prescription
    - Patients with opioid use disorder

- Do we need separate models for suicide and overdose risk?
  - Modeled separately and together
    - Extremely similar. Combined for usability.
Modeling decision 1: Predictor Inclusion

• Initially considered all codable data elements defined with the guideline authors
• Found that nearly all risk mitigation strategies were positive predictors of risk
  • Selectively applied to higher risk patients in clinical practice
  • Removed these from model
    – Goal of model was to increase use of these interventions
    – Intervening would increase risk estimation
    – Confusing and discouraging to clinicians
• Included all patient factors
• Excluded clinical interventions except for prescription fills
  • Wanted to account for risk due to medications taken
  • Incorporated clinical interventions in decision support
Strong diagnostic and health care event risk factors for overdose or suicide-related events

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Odds Ratio</th>
<th>Model Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior overdose or suicide-related event</td>
<td>23.1</td>
<td>2.62</td>
</tr>
<tr>
<td>Detoxification treatment</td>
<td>18.5</td>
<td>.06</td>
</tr>
<tr>
<td>Inpatient mental health treatment</td>
<td>16.6</td>
<td>1.0</td>
</tr>
<tr>
<td>Sedative use disorder diagnosis</td>
<td>11.2</td>
<td>.23</td>
</tr>
<tr>
<td>Stimulant use disorder diagnosis</td>
<td>8.1</td>
<td>.73</td>
</tr>
<tr>
<td>Opioid use disorder diagnosis</td>
<td>8.0</td>
<td>.31</td>
</tr>
<tr>
<td>Mixed substance use disorder</td>
<td>8.0</td>
<td>.33</td>
</tr>
<tr>
<td>Cannabis use disorder</td>
<td>5.9</td>
<td>.27</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>5.8</td>
<td>.82</td>
</tr>
<tr>
<td>Alcohol use disorder</td>
<td>5.3</td>
<td>.36</td>
</tr>
<tr>
<td>Other mental health disorder</td>
<td>5.7</td>
<td>.73</td>
</tr>
<tr>
<td>Major Depression</td>
<td>4.8</td>
<td>.61</td>
</tr>
<tr>
<td>Emergency Department visit</td>
<td>3.4</td>
<td>.72</td>
</tr>
<tr>
<td>Fall or accident</td>
<td>2.9</td>
<td>.44</td>
</tr>
<tr>
<td>PTSD</td>
<td>2.6</td>
<td>.34</td>
</tr>
<tr>
<td>Tobacco use disorder</td>
<td>2.2</td>
<td>.18</td>
</tr>
<tr>
<td>AIDS</td>
<td>2.2</td>
<td>.20</td>
</tr>
<tr>
<td>Liver Disease</td>
<td>2.2</td>
<td>.15</td>
</tr>
<tr>
<td>Other neurological disorder</td>
<td>2.1</td>
<td>.18</td>
</tr>
<tr>
<td>Electrolyte disorders</td>
<td>2.0</td>
<td>.19</td>
</tr>
</tbody>
</table>
MH/SUD and Non-Opioid Related Factors Have Higher Odds Ratios than Opioid-Related Factors in VHA Predictive Model

Odds Ratios for Overdose/Suicide-Related Events

Each additional MG of opioid dose: Risk increased by 0.3%
(100 MG: 30% increase in risk)
<90 day and >90 day prescription: Risks same

Risk increased slightly with increasing MEDD
• e.g., 120 MEDD would increase modeled risk by about as much as a PTSD or AUD diagnosis

STORM Analysis: Oliva et. al. Psych. Services 2017
High Odds Ratios for Other Evidence-Based Sedating Pain Medications

Odds Ratios for Overdose/Suicide-Related Events

- Having TCAs, SNRIs and Anticonvulsants is associated with increased risk
  - Association could be related to unmanaged pain, cumulative sedation, depressive symptoms, etc.

Oliva et al., Psychol Serv 2017
Modeling decision 2: What modeling strategy?

- Experimented with multiple modeling strategies using the same core dataset
  - Cohort 1: All patients with an opioid prescription in FY2010
  - Cohort 2: All patients seen for an opioid use disorder in FY2010
  - Predicting overdose and suicide through FY2011

- Findings:
  - Similar model performance with multiple modeling approaches
  - Good prediction for opioid therapy cohort.
  - Poor model performance for Opioid Use Disorder cohort
    - Whole cohort at high risk with no obvious predictors that substantially distinguished risk between patients
  - Final model included: Random effects for region and health care system, interaction terms to handle commonly co-occurring predictors (e.g. inpatient treatment and detoxification procedures)
Why implement the logistic regression model?

• Simplicity
  • Clinicians were typically familiar and comfortable with regression models
    – facilitated trust
  • Relatively easy to calculate nightly in our SQL-based data warehouse

• Included all clinically expected predictors
  • Clinicians expressed discomfort with excluding predictors
    – Wanted all of the factors they felt were clinically important searched for and considered in the model and decision support

• Performed as well as other models
  • Similar AUC
  • Similar risk enrichment in top cohorts
### Model performance

*Mean Predicted Risk Scores, Actual Overdose- or Suicide-Related Event Rates, and Sensitivity and Specificity for Varying Risk Cohorts.*

*Note.* 23,790 total overdose or suicide-related events in FY2011 among 1,135,601 patients (2.1%).

<table>
<thead>
<tr>
<th>Risk Cohorts (N subjects with the highest risk scores)</th>
<th>Mean Predicted Risk Score (range 0-100)</th>
<th>Actual overdose/suicide-related event rate in FY2011</th>
<th>True Positive Sensitivity for this Cut Point</th>
<th>False Negative 1-Specificity for this Cut Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>57.9</td>
<td>53.7% (~1 in 2)</td>
<td>0.025</td>
<td>0.00041</td>
</tr>
<tr>
<td>10,000</td>
<td>38.1</td>
<td>36.2% (~1 in 3)</td>
<td>0.152</td>
<td>0.0057</td>
</tr>
<tr>
<td>100,000</td>
<td>11.8</td>
<td>11.9% (~1 in 10)</td>
<td>0.502</td>
<td>0.079</td>
</tr>
<tr>
<td>500,000</td>
<td>4.0</td>
<td>4.1% (~1 in 25)</td>
<td>0.850</td>
<td>0.432</td>
</tr>
<tr>
<td>1,000,000</td>
<td>2.3</td>
<td>2.3% (~1 in 50)</td>
<td>0.985</td>
<td>0.878</td>
</tr>
</tbody>
</table>
### Did the model meet the clinical need?
**Improved efficiency of identification of patients with adverse events over individual predictors or target populations**

<table>
<thead>
<tr>
<th>Risk approaches for VA patients prescribed opioids</th>
<th>Risk-model based (STORM top 20,000 patients)</th>
<th>MEDD &gt; 200 mg</th>
<th>Opioid Use Disorder diagnosis</th>
<th>Co-prescribed sedative medication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number in VA in FY2010</td>
<td>20,000</td>
<td>19,496</td>
<td>20,871</td>
<td>185,477</td>
</tr>
<tr>
<td>Total number of opioid or suicide-related events in FY2011</td>
<td>5780</td>
<td>882</td>
<td>2779</td>
<td>4951</td>
</tr>
<tr>
<td>% of all opioid or suicide-related events in FY2011 (N=23,790)</td>
<td>24.3%</td>
<td>3.7%</td>
<td>11.7%</td>
<td>20.8%</td>
</tr>
<tr>
<td>% of risk cohort with an opioid or suicide-related event in FY2011</td>
<td>28.9%</td>
<td>4.5%</td>
<td>13.3%</td>
<td>2.7%</td>
</tr>
<tr>
<td>Minimum risk score in cohort</td>
<td>17.3%</td>
<td>0.3%</td>
<td>0.4%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Median risk score in cohort</td>
<td>26.8%</td>
<td>2.5%</td>
<td>6.5%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Maximum risk score in cohort</td>
<td>79.8%</td>
<td>78.2%</td>
<td>79.8%</td>
<td>79.8%</td>
</tr>
</tbody>
</table>

**Comparison of Different Risk Approaches (Risk-Model Based [STORM] versus Individual Risk-Factor Based)**
How did we use it?

• Used VA Corporate Data Warehouse and Business Intelligence platform
  • SQL database with nightly extracts of all 130 VA medical records
  • Access controls for report permissions based on staff medical record access

• Built decision support as a MS SSRS report
  • Set up nightly extraction of risk predictors
  • Estimated risk for all patients with active opioid prescriptions
  • Posted on reports designed for:
    – Facility/Team/Provider summary
    – Population Management
    – Patient look-up
  • Focus on encouraging and tracking risk mitigation
  • Patient stratification by modeled risk
STORM—Family of decision support tools to support safe care of patients exposed to opioids

Includes: Predictive analytics for risk stratification, flexible population management, summary information on risk mitigation implementation for targeting QI and education, recommendation and tracking of risk mitigation, and patient level care review.
A randomized policy evaluation to test the application of the model

Jodie Trafton
Mandating Interdisciplinary Case Review for Patients estimated at “very high” risk

Turning predictive modeling-based decision support into a targeted prevention intervention.

Implemented at VHA’s 141 Health Care Systems

2 randomized components:
• Timing of expansion of “very high” risk population
• Centralized oversight and Action planning
**Randomization**

<table>
<thead>
<tr>
<th>Time Point</th>
<th>T₀ (Policy)</th>
<th>T₁ (9 mo.)</th>
<th>T₂ (15 mo.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facilitation</td>
<td>1%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>No Facilitation</td>
<td>1%</td>
<td>1%</td>
<td>5%</td>
</tr>
</tbody>
</table>

Risk Groups defining Very High Risk

*All VAMCs are randomized to facilitation or no facilitation arm when the policy is issued. VAMCs within each arm are also randomized to increase STORM risk threshold from 1% to 5% at one of two subsequent time points.*

**Risk Mitigation Strategies**

- Naloxone use
- Non-opioid pain treatments
- Psychosocial treatments

**Clinical Outcomes: Opioid-related serious adverse events (SAEs)**

- Overdose
- Overdose death
- Suicide-related events
- Accidents

**Implementation Factors**

- Aims 1a and b: Number & type of implementation strategies
- Aim 2: Barriers/facilitators to policy notice implementation

---

**Fig. 1** Design for clinical (PEPReC) and implementation evaluations (CHERP)

---

1Partnered Evidence-Based Policy Resource Center
2Center for Health Equity Research and Promotion

Chinman et al., 2019, Implement Sci
Outcomes Associated with Targeting Interdisciplinary Case Review to Patients Estimated at “Very High” Risk of Overdose or Suicide-Related Events

Stepped-wedge designed expansion of population characterized as “very high” risk

<table>
<thead>
<tr>
<th>Patients with risk scores between 1% and 5%</th>
<th>Step 1</th>
<th>Step 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinics 1-70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinics 71-140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timeline (month)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

Patient enters the 1% to 5% risk range in a white facility-month cell

Patient enters the 1% to 5% risk range in a green facility-month cell
Design

STORM trial was a 23-month, multi-center stepped wedge cluster randomized trial.

All 140 VA medical centers were included and received the intervention by the end of the study.

Each medical center entered the study on April 18, 2018, and randomly crossed over into the intervention conditions in two waves: study month 11 and month 17.

Participants

- Eligible participants were VA patients with an active prescription for opioids with a predictive risk of SAE in the top 1-5% of all patients, after the start of the trial.

- Patients with OUD and/or risk scores in the top 0-1% had previously received the intervention and were ineligible for the study analyses.
Comparing Control and Treatment Patients

- Predominantly white, male
- Average age is 58 at baseline
- Sample is evenly balanced between control and treatment conditions
- We also examined balance for 31 different comorbidities

<table>
<thead>
<tr>
<th></th>
<th>Control (n=41,816)</th>
<th>Treatment (n=22,967)</th>
<th>Standardized difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>85.6%</td>
<td>85.0%</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>69.9%</td>
<td>71.7%</td>
<td>0.04</td>
</tr>
<tr>
<td>Black</td>
<td>23.8%</td>
<td>21.9%</td>
<td>0.04</td>
</tr>
<tr>
<td>Other</td>
<td>6.32%</td>
<td>6.9%</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Marital Status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>41.1%</td>
<td>42.3%</td>
<td>0.03</td>
</tr>
<tr>
<td>Single/Never Married</td>
<td>14.8%</td>
<td>14.2%</td>
<td>0.02</td>
</tr>
<tr>
<td>Div/Sep/Widowed</td>
<td>43.7%</td>
<td>43.1%</td>
<td>0.01</td>
</tr>
<tr>
<td>N/A</td>
<td>0.4%</td>
<td>0.4%</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (avg.)</td>
<td>58.4</td>
<td>59.1</td>
<td>0.05</td>
</tr>
<tr>
<td>Homeless</td>
<td>12.4%</td>
<td>10.4%</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Outcomes of Interest

• The primary outcomes of interest were opioid-related serious adverse events (SAEs) and all-cause mortality within 127-days following the intervention.

• SAE's included:
  • Opioid overdose, sedative overdose, acetaminophen overdose, other drug overdose, motor vehicle accident, accidental falls, other accidents, and possible and confirmed suicide-related events.
  • A measure of ‘any’ SAE
Regression Models

- A patient-level logistic mixed model regression was used to estimate the impact of treatment on the likelihood of outcomes, controlling for time, facility and patient characteristics.

- A statistically significant estimate indicates the odds ratio of experiencing an opioid related SAE or all-cause mortality due to being included in the STORM “very high risk” cohort mandated for risk review.
Effect of being mandated to receive an interdisciplinary case review
(patients in top 1-5%)

Top 1-5% STORM dashboard patients had 23% lower odds of all-cause mortality in the next 4 months when labeled “very high” risk and subject to mandate for interdisciplinary case review.

*Effect of being included as “very high” risk due to mandate expansion, regardless of whether the mandated patient received a case review.
Implications

• The odds of all-cause mortality for STORM dashboard patients relative to control patients was 0.773 (95% CI: 0.639, 0.937).

• This translates to approximately 180 lives saved in the first 4 months after identification in the 1-5% risk group.

• Ascertainment bias is a potential concern in detecting SAE’s. Mortality is a more reliably captured outcome.

• Mechanism may be due to increased probability of receiving a case review.

• The odds of receiving a case review for STORM dashboard patients relative to control patients was 6.263 (95% CI: 3.946, 8.580).
Outcomes associated with getting an interdisciplinary case review among patients in top 1%

Among the top 1% risk patients, when patients' case review status was “completed” (from month they were case reviewed plus 12 months) their risk of Any SAE or death were reduced

*These findings are not yet peer-reviewed

**Effect of getting a review among those always mandated to receive one**

- Any SAE = 0.71
- Mortality = 0.72
Current work looking at algorithmic racial and gender bias in the model that may guide future refinement

Suzanne Tamang
Algorithmic Bias: What is it?

Should we be concerned?

Why should we be concerned?
The **ROC curve** is created by plotting the **true positive rate** (TPR) against the **false positive rate** (FPR) at various threshold settings.

The true-positive rate is also known as **sensitivity, recall or probability of detection** in machine learning. The false-positive rate is also known as **probability of false alarm** and can be calculated as \(1 - \text{specificity}\). The ROC curve is the sensitivity or recall as a function of fall-out.
ROC Curve and AU-ROC: Race x Ethnicity

2016-2017

In Sample ROC: Non-Hispanic

Out Sample ROC: Hispanic

Out Sample ROC: Non-Hispanic
Precision Recall Curve

The **Precision Recall curves** are created by plotting the **Precision**, also known as the **positive predictive value** and **Recall**, the **true-positive rate**. Recall is more commonly called **sensitivity** in medicine and is the probability the model will predict all positive cases for the outcome.

In contrast to the ROC curves and ROC-AUC statistics, the Precision-Recall Curve and the PR-AUC performance metric provide *more information on prediction scenarios that involve rare binary events.*
Race X Sex

**Male**

In Sample

Out Sample

**Female**

In Sample

Out Sample
False Negative Parity

The **false-negative rate** represents the percentage of true positives missed by the prediction model.

**False-negative parity** describes the closeness of the false positive rate (false positives/true positives) across different subgroups of interest. It is a commonly reported in algorithmic bias analyses.
PR Curve and AU-PRC: Race x Age

Under 50
50-65
Over 65

In Sample

Out Sample
Calibration is defined as the following property:

“If we assign some group a risk of $x$, the actual outcome incidence rate should also be $x$”

For example, if we assign a group of people a risk of 10%, the actual overdose/suicide-related incidence rate should also be 10%.
Calibration: Race

In Sample

Out Sample
SAE Trends X Race during modeling period

Sharp jump in drug poisoning rates between 2015 and 2018

Increase varied by race/ethnicity

Large relative increase in drug poisoning rates in Black population:

<table>
<thead>
<tr>
<th>Year</th>
<th>Black</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>12.2</td>
<td>21.1</td>
</tr>
<tr>
<td>2017</td>
<td>20.6</td>
<td>27.5</td>
</tr>
</tbody>
</table>

Emphasizes the need for on-going calibration of predictive models, particularly when population risk is evolving rapidly.

Discussion
Alyce Adams, Jodie Trafton, Suzanne Tamang, & Ernest Moy
References


Thank you

Please take a moment to fill out the brief evaluation available on the webinar page