Leveraging Biases in EHR Data Patterns to Predict Risky Patient States: Opportunities to Monitor and Mitigate Racial and Ethnic Biases in Predictive Models

This study is funded by the National Institute of Nursing Research (NINR): 1R01NR016941-01

Communicating Narrative Concerns Entered by RNs (CONCERN): Clinical Decision Support Communication for Risky Patient States.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
The CONCERN Study

**Columbia University**
- Sarah Collins Rossetti PhD RN (MPI)
- Kenrick Cato PhD RN (MPI)
- Suzanne Bakken PhD RN

**Harvard / Brigham and Women’s Hospital / Newton Wellesley Hospital**
- Haomiao Jia PhD
- David Albers PhD
- Chris Knapland Mphil
- Jessica Schwartz BSN, RN

- Patricia Dykes PhD RN (site-PI)
- Jeff Klann PhD
- Kumiko Schnock PhD
- Li Zhou MD PhD

- Min Jeoung Kang, RN, PhD
- Tom (Zfania) Korach, MD, PhD
- Frank Chang, MSE

**Consultants**
- Yalini Senathirajah PhD
- Matthew Fred MD

**Advisory Board**
- David Bates, MD, MSc
- Bonnie Westra PhD, RN, FAAN, FACMI
- Charles Pozner MD

This project is supported by grant number 1R01NR016941-01 from the National Institute for Nursing Research (NINR)

The CONCERN* Study: Clinical Decision Support Communication for Risky Patient States

*Communicating Narrative Concerns Entered by RNs*
• I have no conflicts to disclose
What is CONCERN?

• Early warning system (EWS) for patient deterioration based on nursing documentation patterns or “signals”.

• Detects the nurses' expert clinical judgment when it perceives changes in a patient’s clinical state.

• Alerts earlier than other EWSs, because these subtle patient changes usually occur well before physiological alterations in the patient.

• Leverages existing documentation, preventing increases to documentation burden.

Communicating Narrative Concerns Entered by RNs
CONCERN Predictive Model Purpose

Patients may be entering a risky state

CONCERN Levels

- High: “Showing signs of deterioration”
- Medium: “Increased risk for deterioration”
- Low: “Low risk for deterioration”
### CONCERN Flowsheet Ontology

<table>
<thead>
<tr>
<th>Templates</th>
<th>Groups</th>
<th>LDAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>277</td>
<td>85</td>
</tr>
</tbody>
</table>

From only only 1 site, 1 year of data.....

.....94,183,084 Flowsheet Row Data Points!
We leverage signals of variability in nursing EHR behaviors as one type of bias

Experts are often unable to articulate the cues that guide them (Benner et al., 2009; Kahneman et al., 2009)

Great potential for predictive modeling if associated behaviors can be measured

1. Nurses increase their frequency of surveillance – and subsequently frequency of documentation – for patients that fit a concerning pattern

2. Nurses synthesize their clinical assessments - otherwise buried in structured flowsheet fields – in short comments associated with specific flowsheet values
   - e.g. highlight the relationship between Sp02 and supplemental oxygen to the physician as an indicator of deteriorating status
Bias in metadata patterns as signals of clinical concern

Focus only on values of EHR data will miss healthcare processes and nursing interventions activated far before a patient’s vital signs are abnormal.

Approach can shift how we understand and leverage clinical observational skills and clinician entered data within a patient’s chart.

The act of documenting a free-text comment or other optional data in a flowsheet row.

Information that the nurse likely determined an event or observation was clinically significant enough to record.

Data Science & Applied Clinical Informatics
Opportunity: Healthcare Process Modeling

- **Healthcare Process Models** –
  - Identify features from user interaction with clinical systems which are patterns of clinical behaviors
  - Patterns interpreted as proxies of an individual’s decisions, knowledge, and expertise
  - Use patterns in predictive models for associations with outcomes

- Clinical domain expertise is essential for accurate and comprehensive interpretations.
The CONCERN Predictive Model

Vital Sign
Comment Frequency
(HR, BP, RR, SpO2, Temp. comments)

Vital Sign Entry Frequency
(HR, BP, RR, SpO2, Temp.)

Nursing Note Frequency

Nursing Note Content
(“MD aware”, “EKG obtained”)

MAR Frequency
(PRN medications given, Medications withheld)

Demographics
(Age, Gender, Race, etc.)
The CONCERN Engine

### Numerical Features
- **Vital Sign Entry Frequency** (HR, BP, RR, SpO2, Temp.)
- **Vital Sign Comment Frequency** (HR, BP, RR, SpO2, Temp. comments)
- **MAR Frequency** (PRN medications given, Medications withheld)
- **Nursing Note Frequency**
- **Nursing Note Content** ("MD aware", "EKG obtained")

### Categorical Features
- **Demographics** (Age, Gender, Race, Ethnicity)
- **Hospital Location** (ICU, ACU, Medicine, Surgical, etc.)
- **Time Features** (Hour of day, Day of week)
- **Previous outcomes** (Time since last ICU visit)

### Continuous Time dependent scaling
- What time period in the day does the data come from? (e.g. 3:00 – 15:00)
- How long has the patient been in the hospital (e.g. 27 hours)

### Numerical and Categorical features combined into a time independent measure of clinical deterioration
- Assigned color (green, yellow, red) and score (1–10)
Forest plot of the covariates used in Cox’s time varying proportional hazard model and associated statistics

- CONCERN high risk score implied greater hazard than both the MEWS and NEWS high risk scores (6.69 versus 1.74 and 1.59)
- CONCERN moderate risk score also implied greater hazard than both the MEWS and the NEWS moderate risk scores

The likelihood of an event occurring 48 hours after observing a CONCERN high risk score is comparable to the likelihood of an event occurring 6 hours after observing a high risk MEWS or NEWS score — a difference of 42 hours.

Caption: The likelihood ratio, defined as $L(x,h) = \frac{P(x | \text{patient has an event} \ h \ \text{hours in the future})}{P(x | \text{patient does not have an event} \ h \ \text{hours in the future})}$. For example, $L(\text{CONCERN score = yellow}, 6)$ quantifies how well the CONCERN algorithm separate the probability measures induced by whether the patient has an event 6 hours in the future after observing a ‘yellow’ score. Larger values represent more weight given to the numerator versus the denominator, while smaller values represent more weight given to the denominator.

The CONCERN Study is designed to leverage bias. EHR Data Structures, with diversity in data set and intentional structures enabling detecting bias, lead to Data Modeling and CDS*. Processes involve taking time to measure and check sources of bias. Clinical Outcomes require modifying the algorithm based on findings.

*CDS = clinical decision support
<table>
<thead>
<tr>
<th>Racial Categories</th>
<th>Hispanic or Latino</th>
<th>Hispanic or Latino</th>
<th>Unknown/Not Reported Ethnicity</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Hispanic or Latino</td>
<td>Hispanic or Latino</td>
<td>Unknown/Not Reported Ethnicity</td>
<td>Total</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>616</td>
<td>237</td>
<td>21</td>
<td>749</td>
</tr>
<tr>
<td>Asian</td>
<td>10781</td>
<td>5258</td>
<td>0</td>
<td>18669</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>45</td>
<td>35</td>
<td>0</td>
<td>1108</td>
</tr>
<tr>
<td>Black or African American</td>
<td>24</td>
<td>27</td>
<td>0</td>
<td>52761</td>
</tr>
<tr>
<td>White</td>
<td>139852</td>
<td>114833</td>
<td>2</td>
<td>295645</td>
</tr>
<tr>
<td>More than One Race</td>
<td>69</td>
<td>75</td>
<td>0</td>
<td>549</td>
</tr>
<tr>
<td>Unknown or Not Reported</td>
<td>2940</td>
<td>2860</td>
<td>0</td>
<td>53779</td>
</tr>
<tr>
<td>Total</td>
<td>178816</td>
<td>138362</td>
<td>2</td>
<td>423260</td>
</tr>
</tbody>
</table>

Need robust data set to make sure model is not amplifying biases in the dataset

Need to monitor model features in context of diverse data sets.
Comparison of three Early Warning Scores

• Anticipated that race (and other patient demographics) would play a role in an EWS based on documentation patterns (CONCERN). Demographic information was included in the model building and postprocessing steps to reduce racial bias in the score.

• NEWS and MEWS based on a patient’s physiological state and do not account for potential racial biases. White or Caucasian patients who are transferred to the ICU receive a statistically higher average scores than Black or African American patients.

• CONCERN, NEWS and MEWS scores were generated every hour for the 24-hour period preceding an unanticipated transfer to ICU. The average score was calculated for each patient.

• The dataset was comprised of 157 Black or African American patients and 1600 White or Caucasian patients.

• Race identified by race field in the EHR.


<table>
<thead>
<tr>
<th>Score</th>
<th>Mean Black or African American</th>
<th>Mean White or Caucasian</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONCERN</td>
<td>3.976190</td>
<td>4.053905</td>
<td>0.210805</td>
</tr>
<tr>
<td>NEWS</td>
<td>5.161055</td>
<td>5.752857</td>
<td>0.009656</td>
</tr>
<tr>
<td>MEWS</td>
<td>3.355838</td>
<td>3.673857</td>
<td>0.046306</td>
</tr>
</tbody>
</table>
Bias in EHR Data: Opportunity & Responsibility to Identify and Intervene

- **Monitoring** predictive data models for bias
  - EHR data has biases, therefore predictive models derived from EHR data have high likelihood of bias

- **Mitigating** bias in predictive models
  - At beginning stages of this work
  - Building in model features to mitigate bias

- **Transparency** in data models are foundational for this work
  - Required to check model assumptions
  - Important for evaluating new implementations in different populations
Conclusion & Future Directions

- Structuring nursing science research databases to enable
  - Diversity in datasets
  - Monitoring for bias
  - Transparency of derived data models
  - Mitigating bias in models used in clinical decision support

- Criteria for diversity in data sets? Favor multi-site studies?
- What is the variability of EHR bias across different clinical sites?
- Approaches for bias checking in all CDS models used in nursing and new site implementations?
- Mitigation process once bias detected in an existing CDS/predictive model in broad clinical use?
- How do we measure impact of clinical staff diversity on EHR biases and patient outcomes?
Discussion and Thank you

Sarah Collins Rossetti, RN, PhD, FAMIA, FACMI
sac2125@cumc.Columbia.edu

Kenrick Cato, RN, PhD, FAAN
kdc2110@cumc.columbia.edu