COLUMBIA
COLUMBIA UNIVERSITY
DEPARTMENT OF
BIOMEDICAL INFORMATICS

COLUMBIA UNIVERSITY
School of Nursing

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

CONCERN

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
Haomiao Jia PhD
David Albers PhD
Chris Knapland Mphil
Jessica Schwartz BSN, RN

**Harvard / Brigham and Women's Hospital / Newton Wellesley Hospital**

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

***CO**mmunicating **N**arrative **C**oncerns **E**ntered by **RN**s

COLUMBIA
Columbia University Department of Biomedical Informatics
Columbia University School of Nursing
NewYork-Presbyterian Department of Nursing
HARVARD MEDICAL SCHOOL
BRIGHAM AND WOMEN'S HOSPITAL
PARTNERS HEALTHCARE

- 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.

**CO**mmunicating **N**arrative **C**oncerns **E**ntered by **RN**s

# CONCERN Predictive Model Purpose

Patients may be entering a risky state

~~Patients *already* in a risky state~~

## CONCERN Levels

= High: "Showing signs of deterioration"

= Medium: "Increased risk for deterioration"

= Low: "Low risk for deterioration"

COLUMBIA
COLUMBIA UNIVERSITY
DEPARTMENT OF
BIOMEDICAL INFORMATICS

COLUMBIA UNIVERSITY
School of Nursing

# CONCERN Flowsheet Ontology

**15** templates
**277** groups
**85** LDAs

**groups**

- Complex Assessment (IP COMPLEX ASSESSMENT)
  - Airway LDA (R LDA AIRWAY TRIGGER)
  - Arteriovenous Dialysis Access (LDA PHS AV DIALYSIS ACCESS)
  - Asia Impairment Scale (G PHS IP ASIA SCALE)

**Rows & Values**

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

*…..94,183,084 Flowsheet Row Data Points!*

- Complex Assessment (IP COMPLEX ASSESSMENT)
- Daily Cares/Safety (IP DAILY CARES/SAFETY)
- Intake/Output (IP INTAKE/OUTPUT)
- Intra Procedure Sedation (INTRA OP SEDATION DOCU
- IV Line Assessment (IP IV ASSESSMENT)
- Occupational Performance (T PHS IP OT OCCUPATION NAV)
- Psycho/Social/Spiritual (PHS PSYCHOSOCIAL INA NAV
- PT Ranking (PHS IP PT RANKING)
- QIDS-C (QUICK INVENTORY OF DEPRESSIVE SYMPT
- Screenings (SCREENINGS)
- Vital Signs (IP VITALS SIMPLE)
- Vital Signs Complex (IP VITALS ICU)

- Lumbar Drain (LDA PHS LUMBAR DRAIN)
- Mechanical Ventilation (G IP VENT SETTINGS NURSING)
- Modified Cranial Nerve Assessment (G PHS IP REFLEXES NEURO COMPLEX)
- Motor Function (G PHS IP MOTOR FUNCTION NEURO COMPLEX)
- Nasopharyngeal\Oropharyngeal Adjunct Airway (LDA PHS IP NASOPHARYNGEAL/OROPHARYNGEAL)
- Neuro External Ventricular Drain (LDA PHS NEURO EXTERNAL VENTRICULAR DRAIN)
- Neuro Monitoring Device (Non-EVD) (LDA PHS NEURO MONITORING DEVICE (NON-EVD))
- Neuro Symptoms (G PHS NEURO SYMPTOMS)
- Neurological (G PHS IP NEURO COMPLEX)
- Neurovascular Assessments (G PHS IP ICU NEUROVASCULAR ASSESSMENTS)
- NIH Stroke Scale (G CPN - NIH STROKE SCALE GROUP)
- POSS (G PHS IP POSS PASERO OPIOID SEDATION SCALE)
- RASS (G PHS IP RASS)
- Respiratory (G RESP ASSESSMENT)
- Respiratory Additional Assessments (PHS G IP ICU RESP ADDITIONAL ASSESSMENTS)

- Dressing Status (R DRESSING STATUS)
- Level cmH20 (R LDA VENTRICULOSTOMY LEVEL)
- Neuro EVD Properties (G NEURO EXTERNAL VENTRICULAR DRAIN PROPERTIES)
- Output (mL) (R DRAIN OUTPUT)
- Site Assessment (R NEURO LDA SITE ASSESSMENT)

# 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

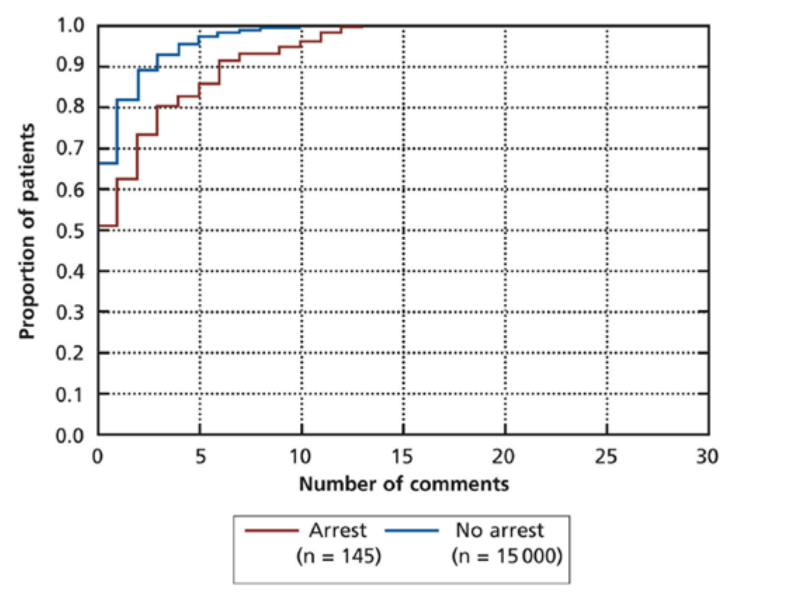
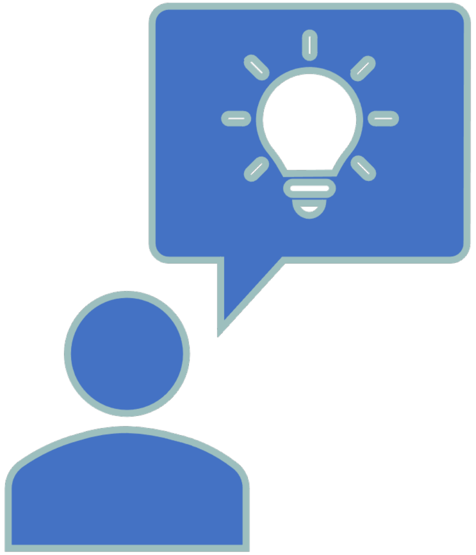



**Figure.**
Kolmogorov-Smirnov test for equality of distributional functions, $P < .001$.

*Collins S, Cato K, Albers D, Scott K, Stetson P, Bakken S, Vawdrey D. Relationship Between Nursing Documentation and Patients' Mortality. American Journal of Critical Care. 2013 Jul;22(4):306-13.

# Data Science & Applied Clinical Informatics Opportunity: Healthcare Process Modeling

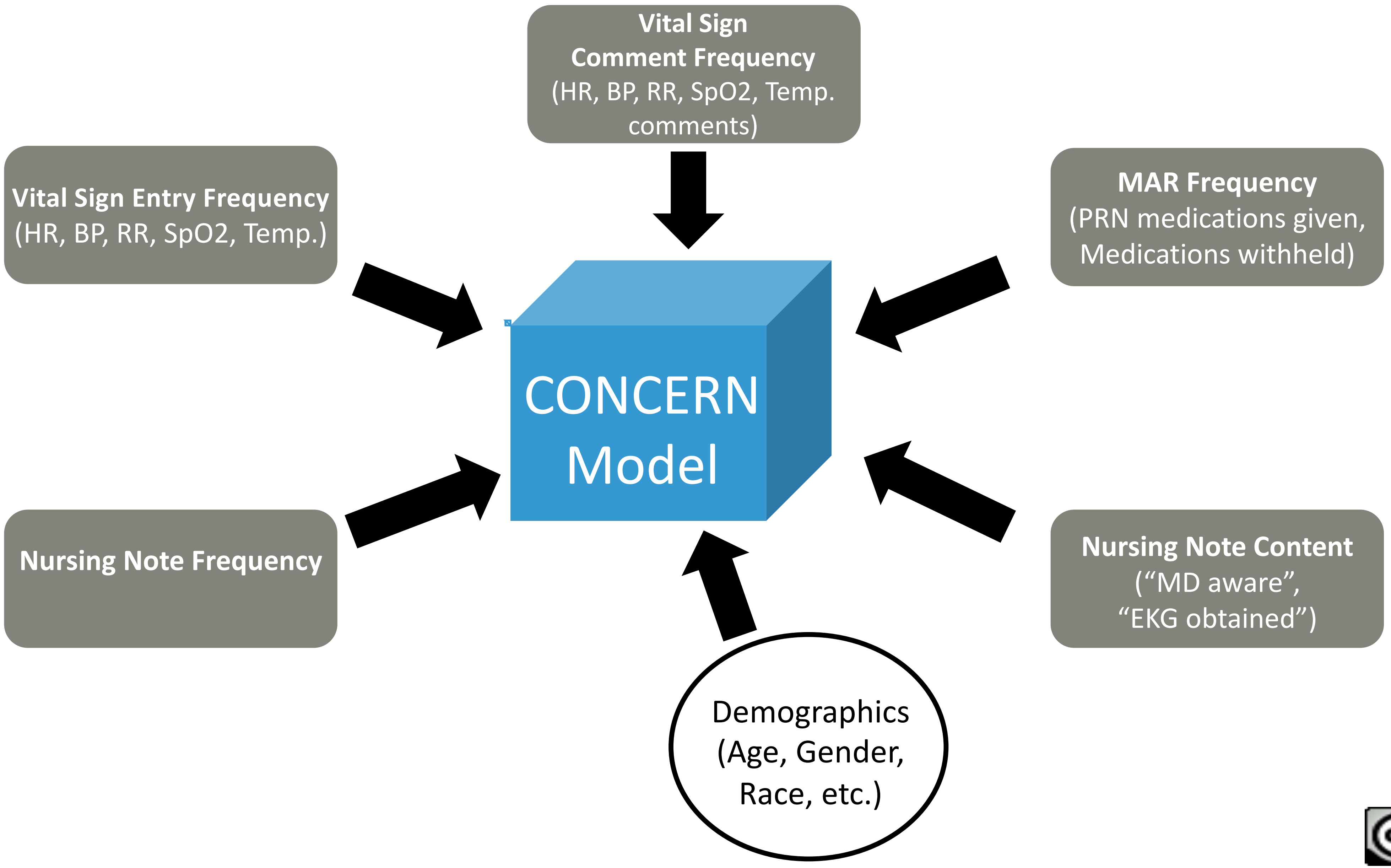
- **<u>Healthcare Process Models</u>** –
  - Identify features from user interaction with clinical systems which are patterns of clinical behaviors
  - Patterns interpreted as proxies of an individual's decisions, knowledge, and expertise
  - Use patterns in predictive models for associations with outcomes



- ***Clinical domain expertise is essential for accurate and comprehensive interpretations.***

# The CONCERN Predictive Model





COLUMBIA
COLUMBIA UNIVERSITY
DEPARTMENT OF BIOMEDICAL INFORMATICS

# The CONCERN Engine

COLUMBIA UNIVERSITY
School of Nursing

**Numerical Features**
found from classification model

**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”)

**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)

**Categorical Features**
Known confounders and variations on clinical workflow

**Demographics**
(Age, Gender, Race, Ethnicity)

**Hospital Location**
(ICU, ACU, Medicine, Surgical, etc.)

**Time Features**
(Hour of day, Day of week)

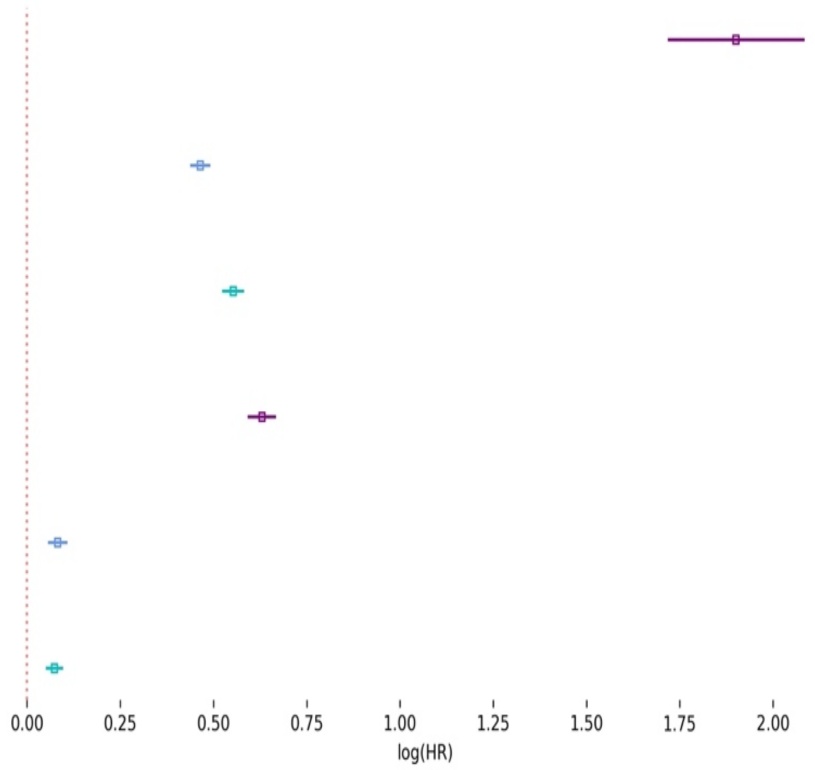
**Previous outcomes**
(Time since last ICU visit)

**Numerical and Categorical features combined into a time independent measure of clinical deterioration**
Assigned color (green, yellow, red) and score (1 10)

# Time Varying Survival Regression

Forest plot of the covariates used in Cox's time varying proportional hazard model and associated statistics

| | HR | log(HR) | Lower 95% | Upper 95% | z | p |
|---|---|---|---|---|---|---|
| CONCERN High Risk (Red score) | 6.6944 | 1.9013 | 1.7212 | 2.0813 | 20.6991 | < .0001 |
| NEWS High Risk | 1.5916 | 0.4647 | 0.4386 | 0.4909 | 34.8049 | < .0001 |
| MEWS High Risk | 1.7381 | 0.5528 | 0.5243 | 0.5814 | 37.9552 | < .0001 |
| CONCERN Moderate Risk (Yellow score) | 1.8772 | 0.6298 | 0.5925 | 0.6671 | 33.0793 | < .0001 |
| NEWS Moderate Risk | 1.086 | 0.0825 | 0.0573 | 0.1077 | 6.4223 | < .0001 |
| MEWS Moderate Risk | 1.0764 | 0.0736 | 0.0507 | 0.0965 | 6.3096 | < .0001 |




- 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

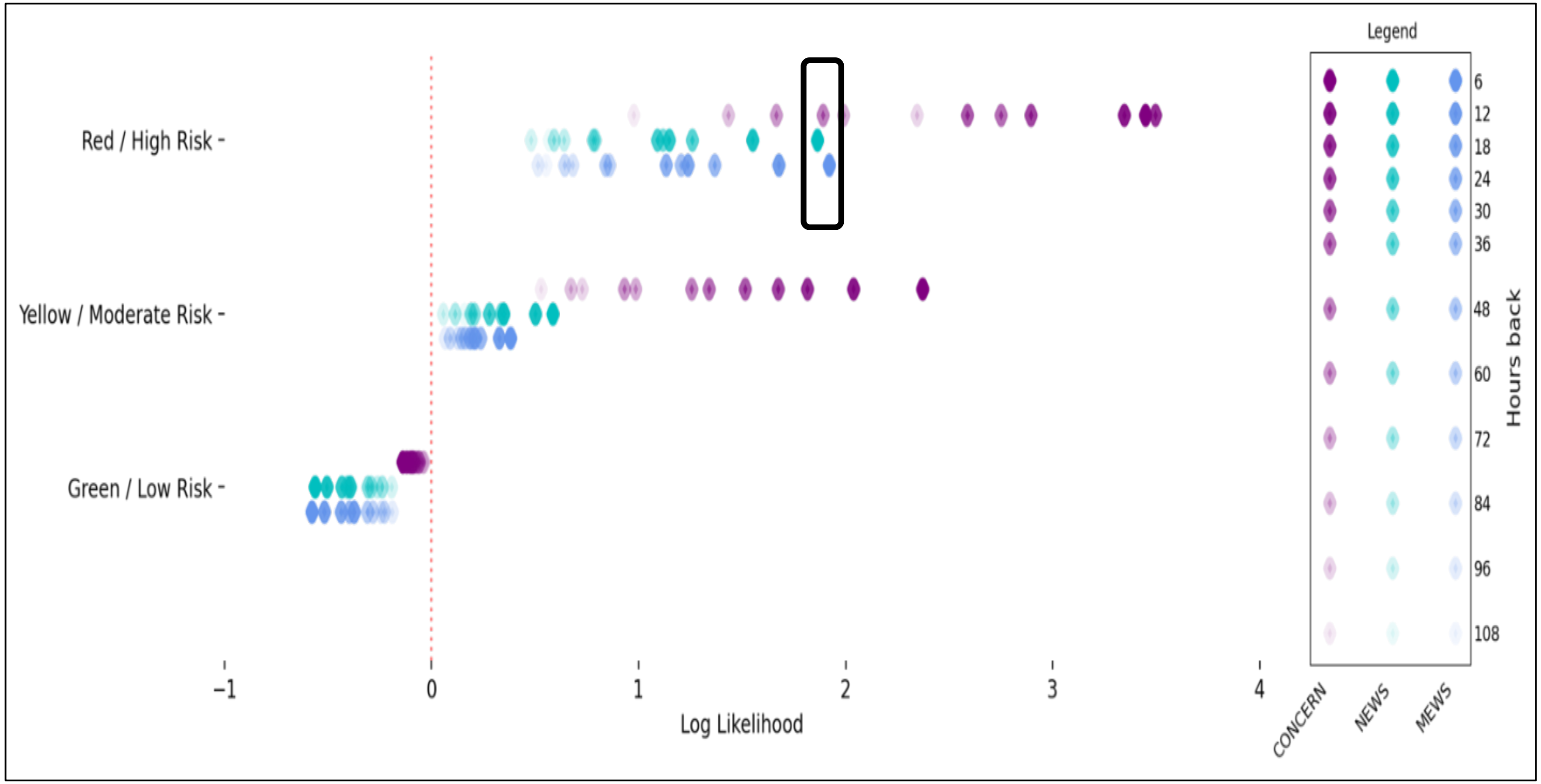
**Rossetti SC**, Knaplund C, Albers D, Dykes PD, Kang MJ, Korach TZ, Zhou L, Schnock K, Garcia J, Schwartz J, Fu LH, Klann JG, Cato K. Healthcare Process Modeling to Phenotype Clinician Behaviors for Exploiting the Signal Gain of Clinical Expertise (HPM-ExpertSignals): Development and Evaluation of a Conceptual Framework. Journal of the American Medical Informatics Association. Accepted Jan 2021.

COLUMBIA
Columbia University
Department of
Biomedical Informatics

# Comparison of Log Likelihood Ratios at Various Hours Before Event

Columbia University
School of Nursing

*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*) = P(*x* | patient has an event *h* hours in the future) / P(*x* | patient does not have an event *h* hours in the future). For example, L('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.

**Rossetti SC**, Knaplund C, Albers D, Dykes PD, Kang MJ, Korach TZ, Zhou L, Schnock K, Garcia J, Schwartz J, Fu LH, Klann JG, Cato K. Healthcare Process Modeling to Phenotype Clinician Behaviors for Exploiting the Signal Gain of Clinical Expertise (HPM-ExpertSignals): Development and Evaluation of a Conceptual Framework. Journal of the American Medical Informatics Association. Accepted Jan 2021.

# Structure-Process-Outcome Framework



The CONCERN Study is designed to leverage bias

**EHR Data Structures**

- Diversity in data set
- Intentional structures enable detecting bias

**Data Modeling and CDS* Processes**

- Take time to measure and check sources of bias

**Clinical Outcomes**

- Modify our algorithm based on findings

*CDS = clinical decision support

Columbia University Department of Biomedical Informatics

Columbia University School of Nursing

## Cumulative (Actual)

| Racial Categories | Ethnic Categories | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| | Not Hispanic or Latino | | | Hispanic or Latino | | | Unknown/Not Reported Ethnicity | | | |
| | F[illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible]known/ Not [illegible]eported | |
| American Indian/ Alaska Native | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0 | 749 |
| Asian | 10781 | 5258 | 0 | 962 | 862 | 0 | 366 | 440 | 0 | 18669 |
| Native Hawaiian or Other Pacific Islander | 1[illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0 | 1108 |
| Black or African American | 24[illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0 | 52761 |
| White | 139852 | 114833 | 2 | 15288 | 10856 | 0 | 7041 | 7773 | 0 | 295645 |
| More than One Race | 69 | 75 | 0 | 165 | 194 | 0 | 22 | 24 | 0 | 549 |
| Unknown or Not Reported | 2940 | 2860 | 0 | 10608 | 8985 | 0 | 14252 | 14114 | 20 | 53779 |
| **Total** | 178816 | 138362 | 2 | 31993 | 24466 | 0 | 24642 | 24959 | 20 | 423260 |

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

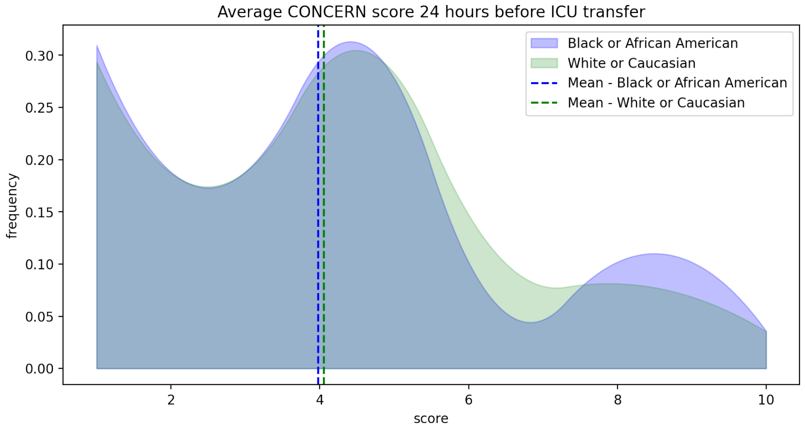
*Need to monitor model features in context of diverse data sets*

COLUMBIA
COLUMBIA UNIVERSITY
DEPARTMENT OF
BIOMEDICAL INFORMATICS

# Comparison of three Early Warning Scores

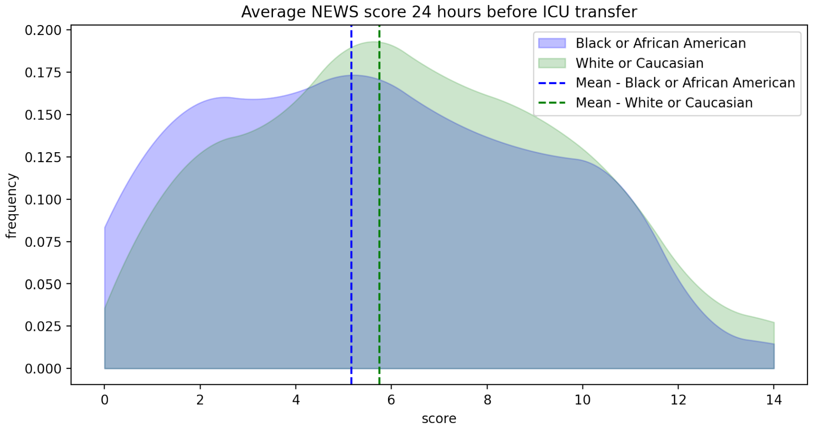
COLUMBIA UNIVERSITY
School of Nursing

## CONCERN



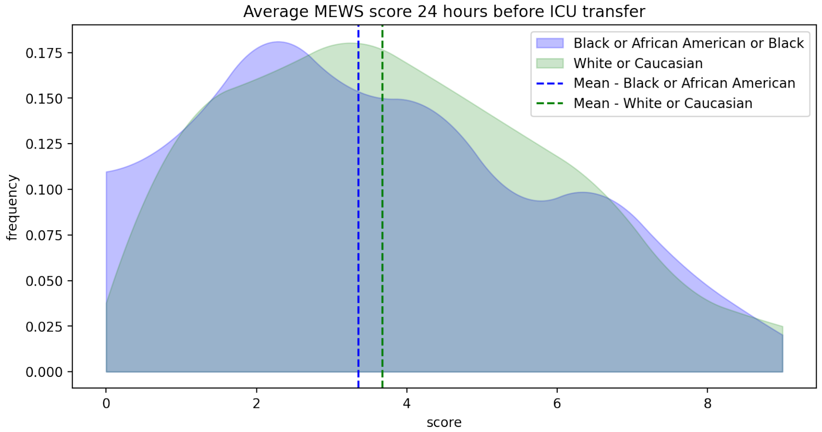

| Mean Black or African American | Mean White or Caucasian | P value |
|---|---|---|
| 3.976190 | 4.053905 | 0.210805 |

## National Early Warning Score (NEWS)




| Mean Black or African American | Mean White or Caucasian | P value |
|---|---|---|
| 5.161055 | 5.752857 | 0.009656 |

## Modified Early Warning Score (MEWS)




| Mean Black or African American | Mean White or Caucasian | P value |
|---|---|---|
| 3.355838 | 3.673857 | 0.046306 |

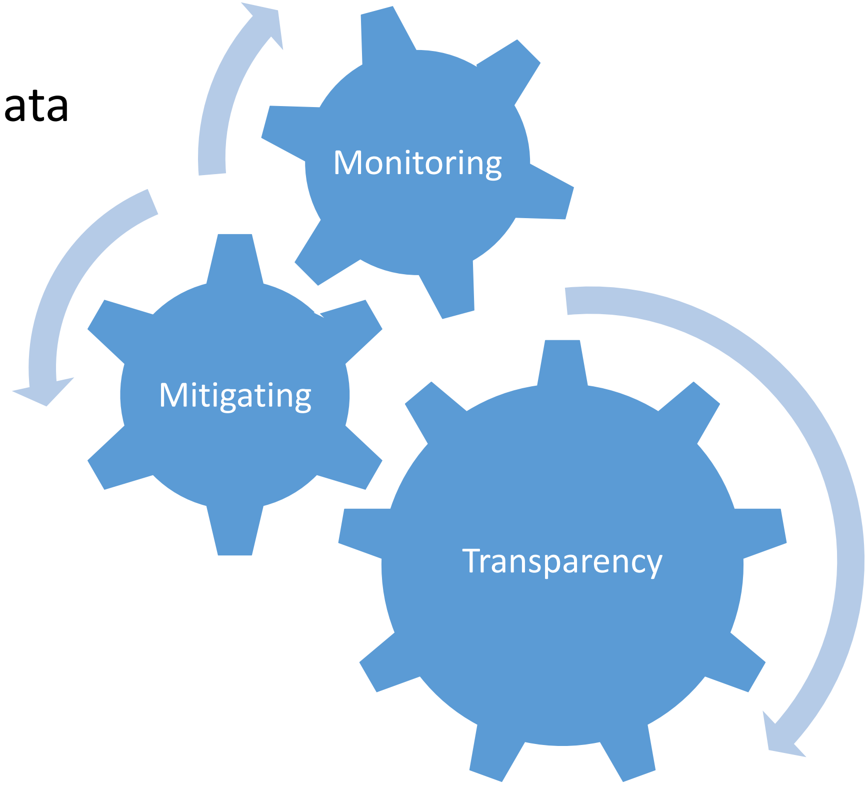
- Anticipated that race (and other patient demographics) would play a role in an EWS based on documentation patterns (CONCERN). Demographic information was included in the model building and postprocessing steps to reduce racial bias in the score.
- NEWS and MEWS based on a patient's physiological state and do not account for potential racial biases. White or Caucasian patients who are transferred to the ICU receive a statistically higher average scores then 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.


•Subbe, C. P., Kruger, M., Rutherford, P. & Gemmel, L. Validation of a modified Early Warning Score in medical admissions. *QJM* **94**, 521–6 (2001).
•Pimentel, M. A. F. *et al.* A comparison of the ability of the National Early Warning Score and the National Early Warning Score 2 to identify patients at risk of in-hospital mortality: A multi-centre database study. *Resuscitation* **134**, 147–156 (2019).

# 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?

COLUMBIA
Columbia University
Department of Biomedical Informatics

Columbia University
School of Nursing

# Discussion and Thank you


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

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

(NINR): 1R01NR016941-01: Communicating Narrative Concerns Entered by RNs (CONCERN): Clinical Decision Support Communication for Risky Patient States.