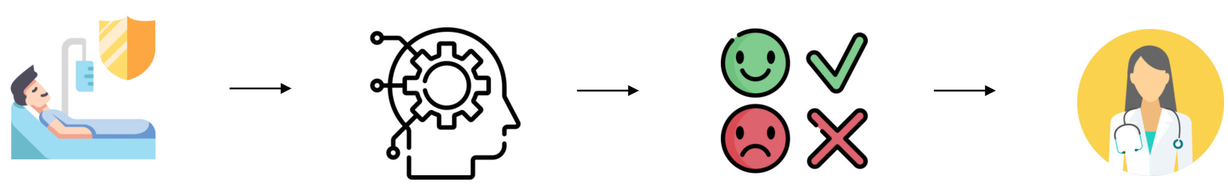
# Evaluating a machine learning model of inpatient clinical deterioration for bias on sociodemographic factors

Michael Colacci MD[1-3], Chloe Pou-Prom MSc[1], Amol Verma MD, MPhil[1-3], Muhammad Mamdani PharmD, MA, MPH[1-3]

[1] Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, Toronto, Canada, [2] Department of Medicine, University of Toronto, Toronto, Canada
[3] Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

## BACKGROUND

5-10% of hospitalized patients die or require ICU admission. Machine learning tools can effectively prevent patient deterioration. CHARTWatch is one such tool in use at St. Michael's Hospital (University of Toronto)



Machine learning tools do not always perform equally well in all patient populations. Few evaluations have been conducted among models in clinical use.

**Research Question:** Among hospitalized general medicine patients, is CHARTWatch model performance for predicting inpatient deterioration consistent across patient subgroups?

## METHODS

**Study Design**: Retrospective cohort study of patients hospitalized to general medicine at St. Michael's Hospital between Nov 2016 – Dec 2019

**Sociodemographic subgroups:**

- Age (< 60 years, 60-80 years, >80 years)
- Sex
- Homelessness
- Neighborhood material deprivation[a]
- Neighborhood ethnic concentration[b]

**Outcome**: Model performance: Balanced error rate (average of false positive and negative rate), precision, recall, F1-score, specificity, AUC, Brier score and overall accuracy for each subgroup

Subgroups compared using Balanced Error Rate Ratio (<0.8 or >1.25 considered to be a meaningful difference)

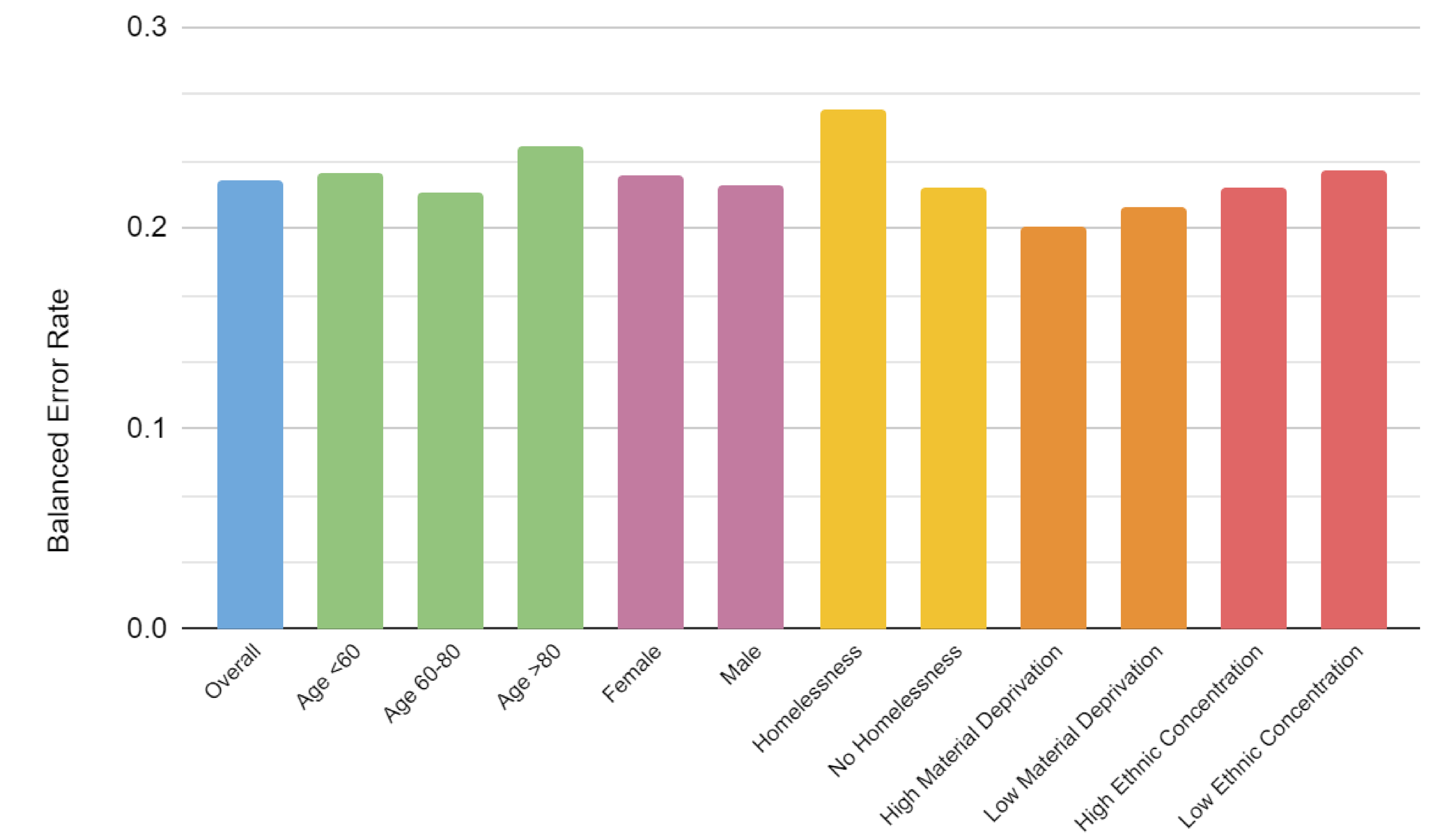
**Model outcome**: Inpatient death or ICU transfer

Legend: [a]Material Deprivation: Neighborhood level measure of inability for individuals and communities to access and attain basic material needs. The indicators included in this dimension measure income, quality of housing, educational attainment and family structure characteristics.
[b]Ethnic Concentration: Neighborhood level measure of people who are recent immigrants and/or people belonging to a 'visible minority' group (defined by Statistics Canada as "persons, other than aboriginal peoples, who are non-Caucasian in race or non-white in color")

## RESULTS

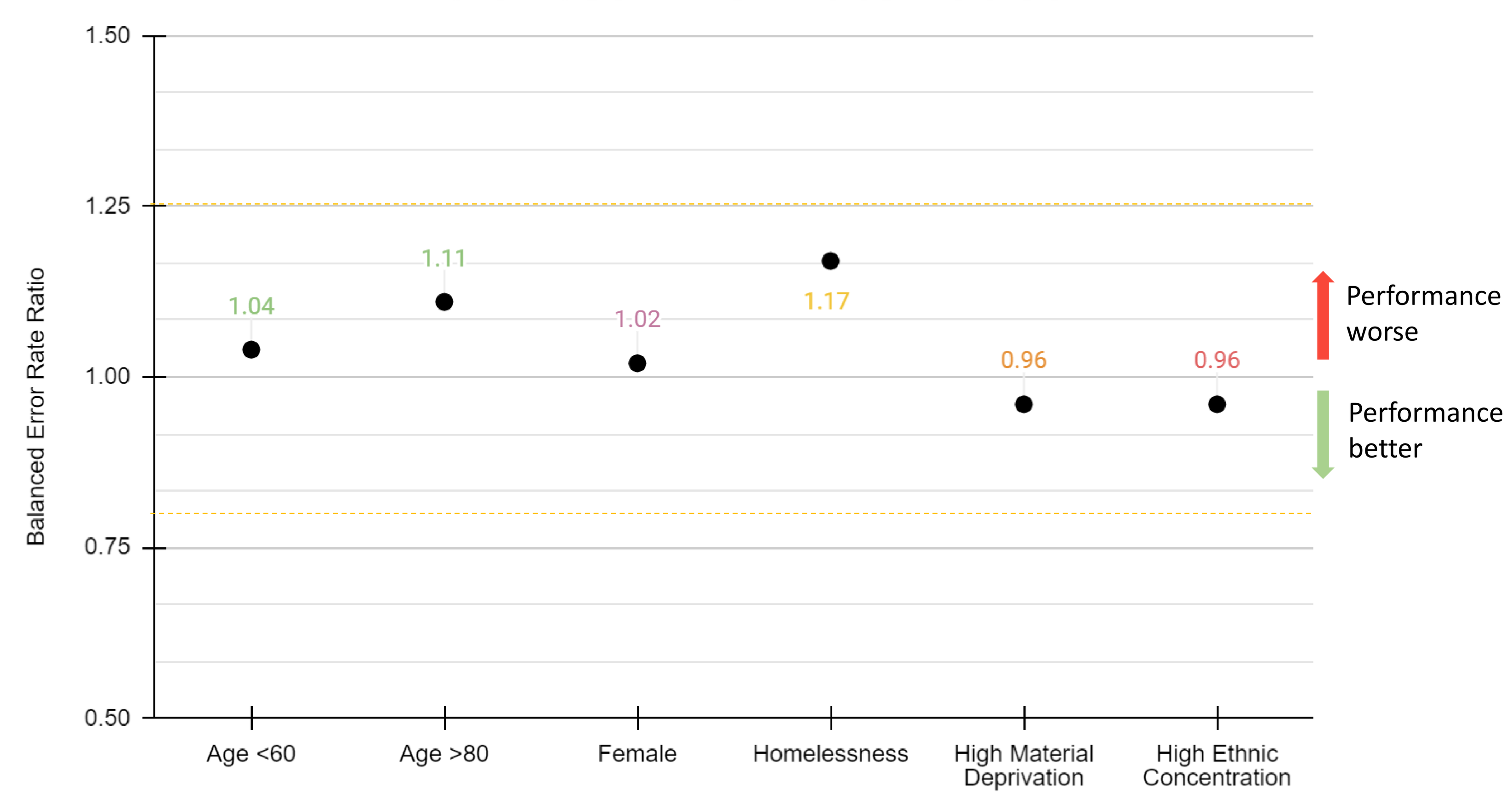
| Subgroup | N (%) | ICU Transfer or Death |
|---|---|---|
| Overall | 9259 | 720 (7.8%) |
| Age <60 years | 3446 (37.2%) | 183 (25.4%) |
| Age 60-80 years | 3618 (39.1%) | 300 (41.6%) |
| Age >80 years | 2195 (23.7%) | 237 (32.9%) |
| Female | 3810 (41.1%) | 295 (40.9%) |
| Homelessness | 1478 (16.0%) | 65 (9.0%) |
| High Ethnic Concentration | 3175 (34.3%) | 231 (32.1%) |
| High Material Deprivation | 2788 (30.1%) | 226 (31.4%) |

### Balanced Error Rates



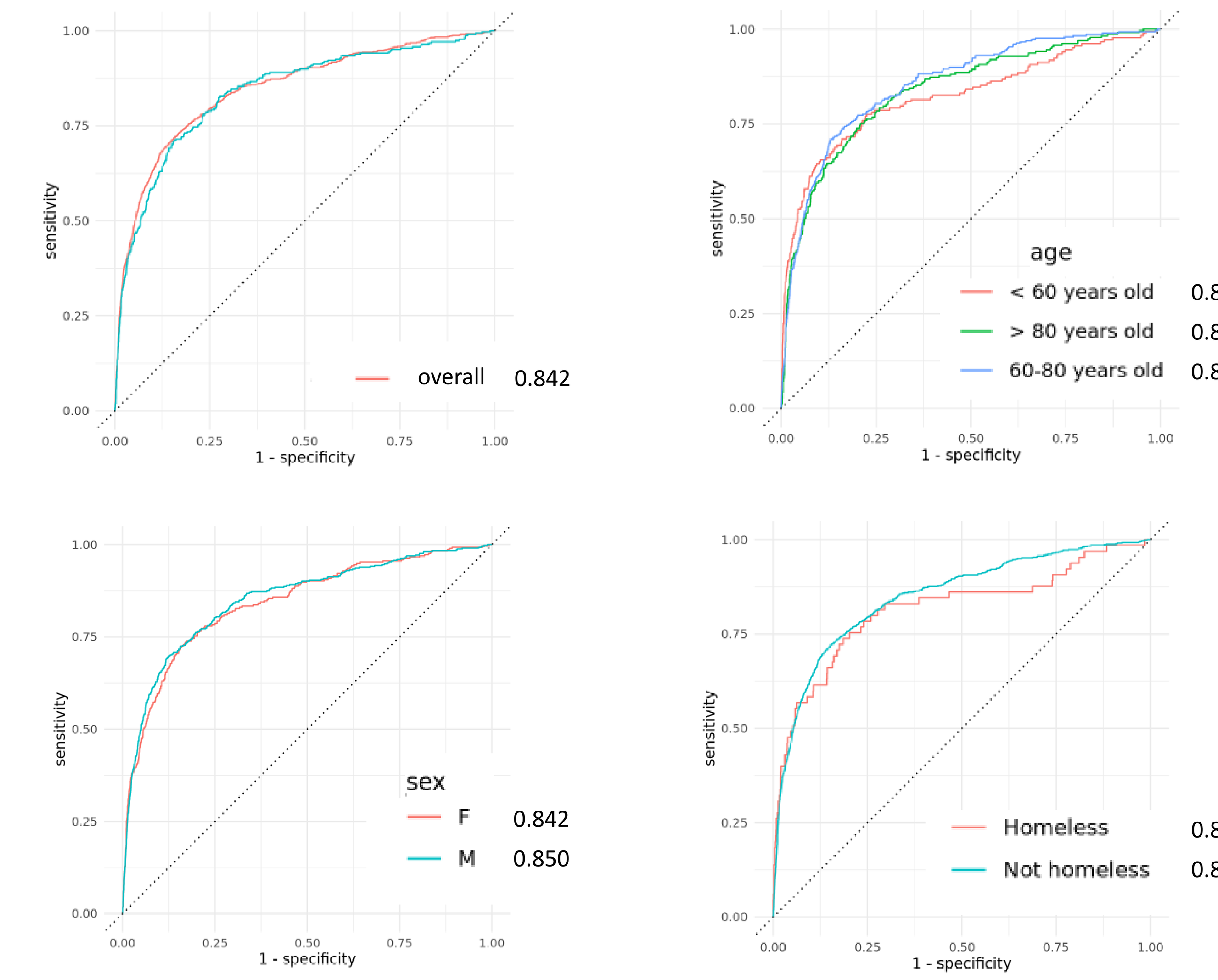

Legend: Balanced Error Rate (average of false positive rate and false negative rate) for each sociodemographic subgroup. Higher balanced error rates indicate poorer model subgroup performance.

### Balanced Error Rate Ratios




Legend: Balanced Error Rate Ratio for each sociodemographic subgroup (calculated by dividing subgroup balanced error rate by reference group balanced error rate). Reference groups were Age 60-80, Male, Not Experiencing Homelessness, Low Material Deprivation and Low Ethnic Concentration, respectively. All BERR were >0.8 and <1.25 (dotted lines).

### AUC




## DISCUSSION

We identified that a machine learning model for inpatient clinical deterioration showed strong performance among all patient subgroups.

None of the observed differences were greater than the pre-specified thresholds, but these are novel metrics, and the balanced error rate ratio was greatest among those >80 years old (1.11) and those experiencing homelessness (1.17)

It is critical to interrogate model performance for fairness across patient subgroups

**REFERENCES:**

1. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science (80). 2019. doi:10.1126/science.aax2342
2. Barocas S, Selbst A. Big Data's Disparate Impact. Calif Law Rev. 2016. doi:10.15779/Z38BG31
3. Matheson FI, Moloney G, van Ingen T; Ontario Agency for Health Protection and Promotion (Public Health Ontario). 2016 Ontario marginalization index: user guide. 1st revision. Toronto, ON: St. Michael'sHospital (Unity Health Toronto); 2022. Joint publication with Public Health Ontario
4. Juhn YJ, Ryu E, Wi C, King KS, Malik M, Romero-Brufau S, Weng C, Sohn S, Sharp RR, Halamka JD. Assessing socioeconomic bias in machine learning algorithms in health care: a case study of the HOUSES index, Journal of the American Medical Informatics Association, Volume 29, Issue 7, July 2022, Pages 1142–1151, https://doi.org/10.1093/jamia/ocac052
5. Burlina P, Joshi N, Paul W, Pacheco KD, Bressler NM. Addressing Artificial Intelligence Bias in Retinal Diagnostics. Transl Vis Sci Technol. 2021 Feb 5;10(2):13. doi: 10.1167/tvst.10.2.13. PMID: 34003898; PMCID: PMC7884292.
6. S. Khor, E.E. Hahn, E.C. Haupt, V. Shankaran, S. Clark, P. Rodriguez, Y. Chen, A. Bansal. The Impact of Including Race and Ethnicity in Risk Prediction Models on Racial Bias. Value in Health. June 2021. Volume 21(1). DOI:https://doi.org/10.1016/j.jval.2021.04.007

UNITY HEALTH TORONTO

Dalla Lana School of Public Health