

Can We Identify Implicit Bias in Clinical Notes, and Are Biases Stable Across Time and Geography?

STUDY DESIGN

- Clinical notes on critically ill adults admitted to the University of California, San Francisco (UCSF) (2012-2022) and Beth Israel Deaconess Medical Center (BIDMC) (2001-2012)
- Trained, unsupervised, word-embedding algorithms to measure the cosine similarity of context between racial or ethnic descriptors and stigmatizing target word(s)

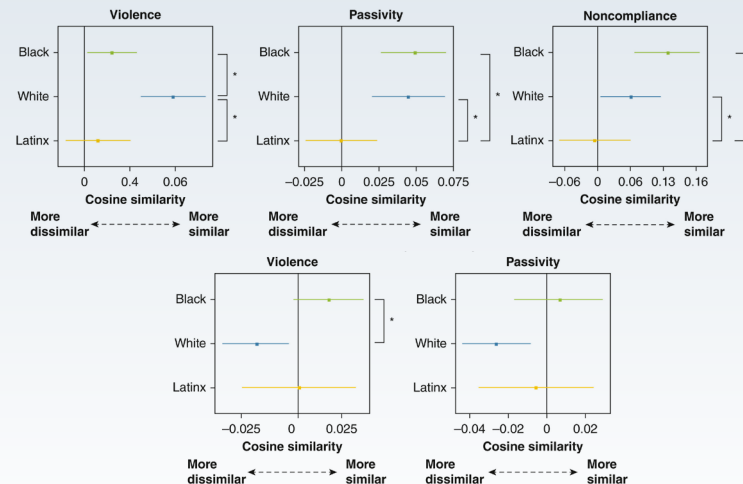
RESULTS

UCSF notes:

- Black descriptors less likely to be similar contextually to violent words compared with White descriptors
- Black descriptors more similar contextually to passivity and noncompliance words compared with Latinx descriptors

BIDMC notes:

- Black descriptors more likely to be similar contextually to violent words compared with White descriptors



Implicit race bias is identifiable in ICU notes, but relationships between race and ethnicity descriptors and stigmatizing words are not stable across time and geography, suggesting that natural language processing algorithms relying on clinical notes could reinforce disparities.