

Application of Machine Learning Methodologies to Understand Violence against Women Using National Survey Data

Anita Raj, Nabamallika Dehingia, Lotus McDougal, Abhishek Singh, Kaushalendra Kumar, Julian McAuley
Center on Gender Equity and Health, University of California San Diego



INTRODUCTION

- Sexual violence (SV) is a prevalent and pervasive violation of human rights
- In India, 6% of ever-married women have experienced marital SV, while 1% have experienced non-marital SV
- Machine learning (ML) methods can explore low prevalence data to offer insights
- Study aimed to generate hypotheses regarding correlates of SV

METHODS

- Data from Demographic Health Survey (2015-16) in India- we included all variables from the dataset in the analysis to identify potential correlates
- Used ML models (*lasso* and *ridge* regression) combined with iterative thematic analysis (ITA)- an approach developed by our team that involved qualitative coding of ML generated predictive variables from iterative rounds, into relevant themes
- ITA output provides a collection of themes/concepts that are likely correlated to our outcome, SV

CONCLUSION

- ML-ITA is a valid analytic approach that can elucidate variables associated with a given outcome that may not otherwise have been hypothesized, particularly for low prevalence outcomes where traditional statistical models fail to perform
- SV is most likely to occur in marriage and linked to a history of family and partner violence, lesser freedom of movement and economic agency, and lesser access to media and health care
- Findings are consistent with prior work demonstrating associations between SV and other violence experiences, but also showed novel correlates; our analysis highlights several themes that may be promising avenues to identify and support women experiencing violence



RESULTS

- Eleven themes were identified to be associated with experience of marital SV, while eight themes were related to non-marital SV

MARITAL SV

- Exposure to violence
- Sexual behavior
- Autonomy
- Socio-demographics
- Access to media
- Health knowledge
- Health system access
- Partner control
- Economic agency
- Reproductive and maternal health
- Health status

NON-MARITAL SV

- Exposure to violence
- Geography
- Sexual behavior
- Health system access
- Autonomy
- Reproductive and sexual health
- Economic circumstances
- Maternal health

Figure: Themes correlated to the two forms of sexual violence (SV), identified using ML-ITA

- Past exposure to violence was most predictive of both types of SV
- Even though the two forms of SV had similar predictors, SV reported by unmarried adolescents correlated with higher wealth and mobility, suggesting divergent risks for SV based on age and relationship to the perpetrator