Speaker: *Zubin Jelveh*
Presentation title: The Lessons and Limits of Predicting Shooting Victimization
Co-authors: *Max Kapustin* (Cornell University, USA), *Sara B. Heller* (University of Michigan, USA), *Kimberley Smith* (University of Chicago, USA) and *Ben Jakubowski* (New York University, USA)

Almost 90,000 people become shooting victims each year in the U.S., costing society on the order of $100 billion. Algorithms to predict ex ante risk have the potential to reduce gun violence by improving how prevention efforts are allocated. But if their output "bakes in" biases present in the input data, then using them to guide police or other criminal justice efforts could exacerbate institutional biases and incur high costs from intervening with false positives. This paper considers an alternative use of algorithms to prevent gun violence: re-purposing police data to predict the risk of shooting victimization rather than arrest, and using the results to direct social services rather than law enforcement. We develop a proof-of-concept machine learning algorithm using Chicago Police Department data, and report on its sensitivity to important modeling choices and its out-of-sample performance both overall and across demographic groups. Our results demonstrate the significant potential of police data to identify people at high risk of becoming victims: Among the 500 people with the highest predicted risk, 56 are shot in the next 18 months---over 100 times the rate of the average Chicagoan. Providing this group with an intervention that reduces risk by 25% justifies spending at least $47K/person, given the estimated social cost of a gunshot victimization. But our results also highlight the limits of police data: The model fails to identify the vast majority of eventual victims, including almost all female victims. This high false positive rate underscores the danger of using such predictions to target law enforcement. Lastly, we document differences in predictive power across gender, race, and age groups, noting that differing causes of violence and representation in police data may help explain them.