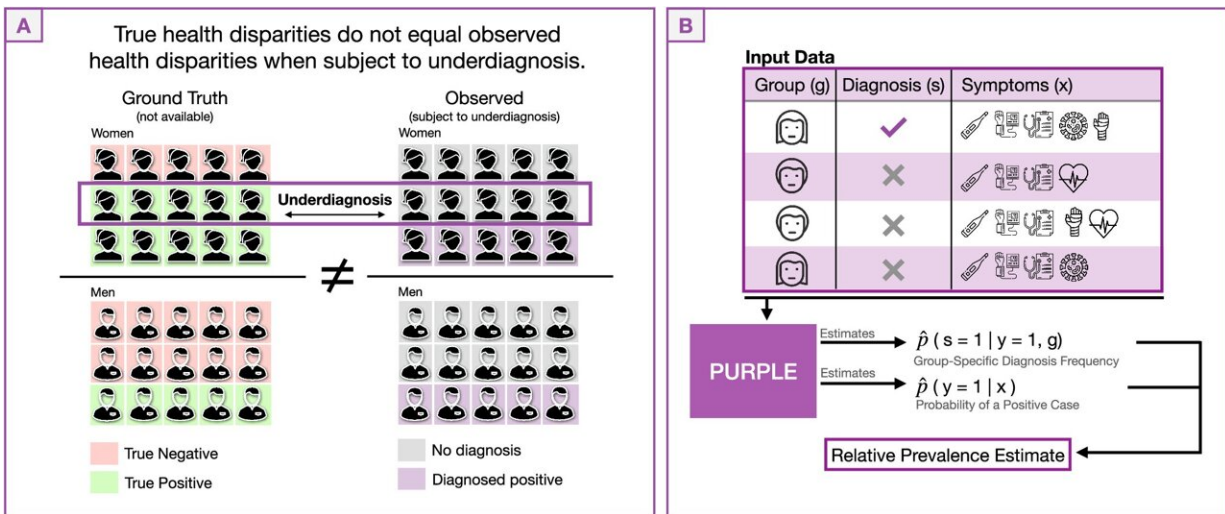# New model shows more realistic picture of intimate partner violence

June 24 2024, by Patricia Waldron

Underreporting can skew observed relative prevalences and conceal health disparities. PURPLE is designed to estimate the relative prevalence while correcting for underreporting. **A** Underreporting leads to inaccurate observed relative prevalences. Understanding the relative prevalence of a health condition between groups g—for example, men and women—is important to effective medical care. However, these estimates are often based on diagnoses s (i.e., positive diagnosis vs. no diagnosis) instead of the true patient state y (sick vs. not sick). Underreporting, which is known to vary by demographic groups, leads to inaccurate relative prevalence estimates that can hide the groups most affected by a condition. **B** PURPLE uses data on patient diagnoses s, symptoms x, and group membership g to accurately estimate the relative prevalence of a condition. PURPLE first estimates the group-specific diagnosis probability, $p(s = 1|y = 1, g)$, and disease likelihood, $p(y = 1|x)$, up to constant multiplicative factors and then combines these estimates to compute the relative prevalence.

We show this is possible under three widely-made assumptions: no false positives, random diagnosis within groups, and constant $p(y = 1|x)$ between groups. Credit: *npj Women's Health* (2024). DOI: 10.1038/s44294-024-00011-5

Intimate partner violence is notoriously underreported and correctly diagnosed at hospitals only around a quarter of the time, but a new method provides a more realistic picture of which groups of women are most affected, even when their cases go unrecorded.

PURPLE, an algorithm developed by researchers at Cornell and the Massachusetts Institute of Technology, estimates how often underreported health conditions occur in different demographic groups. Using hospital data, the researchers showed that PURPLE can better quantify which groups of women are most likely to experience intimate partner violence compared with methods that do not correct for underreporting.

The new method was developed by Divya Shanmugam, formerly a doctoral student at MIT who will join Cornell Tech as a postdoctoral researcher this fall, and Emma Pierson, the Andrew H. and Ann R. Tisch Assistant Professor of computer science at the Jacobs Technion-Cornell Institute at Cornell Tech and in the Cornell Ann S. Bowers College of Computing and Information Science.

The researchers describe [their approach](link) in "Quantifying Disparities in Intimate Partner Violence: a Machine Learning Method to Correct for Underreporting," published in *npj Women's Health*.

"Often we care about how commonly a disease occurs in one population versus another, because it can help us target resources to the groups who need it most," Pierson said. "The challenge is, many diseases are

underdiagnosed. Underreporting is intimately bound up with societal inequality, because often it tends to affect groups more if they have worse access to health services."

Shanmugam became interested in intimate partner violence after Pierson recommended the book "No Visible Bruises: What We Don't Know About Domestic Violence Can Kill Us" by Rachel Louise Snyder. She realized that the pervasive issue of underreporting was something statistical methods could help address. The result was PURPLE (Positive Unlabeled Relative PrevaLence Estimator), a machine learning technique that estimates the relative prevalence of a condition when the true numbers of affected people in different groups are unknown.

The researchers applied PURPLE to two real-life datasets, one that included 293,297 emergency department visits to a hospital in the Boston area, and a second with 33.1 million emergency department visits to hospitals nationwide. PURPLE used demographic data along with actual diagnoses of intimate partner violence and associated symptoms, like a broken wrist or bruising, which could indicate the condition even when the patient was not actually diagnosed.

"These broad datasets, describing millions of emergency department visits, can produce relative prevalences that are misleading using only the observed diagnoses," Shanmugam said. "PURPLE's adjustments can bring us closer to the truth."

PURPLE indicated that patients who are nonwhite, not legally married, on Medicaid, or who live in lower-income or metropolitan areas are all more likely to experience intimate partner violence. These results match up with previous findings in the literature, demonstrating the plausibility of PURPLE's results.

The results also show that correcting for underreporting is important to

produce accurate estimates. Without this correction, the [hospital](link) datasets do not show a straightforward relationship between [income level](link) and rates of victimization. But PURPLE clearly shows that rates of violence are higher for women in lower income brackets, a finding that agrees with the literature.

Next, the researchers hope to see PURPLE applied to other often-underreported women's health issues, such as endometriosis or polycystic ovarian syndrome.

"There's still a lot more work to be done to measure the extent to which these outcomes are underdiagnosed, and I think PURPLE could be one tool to help answer that question," Shanmugam said.

The new technique also has potential applications beyond health conditions. PURPLE could be used to reveal the relative prevalence of underreported police misconduct across precincts or the amounts of hate speech directed at different demographic groups.

Kaihua Hou, a doctoral student at the University of California, Berkeley, contributed to the study. Pierson also has an appointment with Weill Cornell Medicine.

Provided by Cornell University