**Title**: Reis et al. (2009)

**Background**: Domestic abuse among emergency room patients is prevalent, but difficult to diagnose. Reis et al. apply machine learning techniques to mine electronic health records and identify factors that can predict future diagnoses of domestic abuse. In this case, an algorithm was trained on the longitudinal records (4+ years) for over 561,216 patients in a state-wide claims database covering hospital and ER encounters.

**Findings**: Reis et al. developed a naïve Bayes classifier to sort patients into low-risk and high-risk of eventually receiving a diagnosis of abuse. Different models were constructed for men and women, and the results were evaluated by timeliness of detection, sensitivity, specificity, and positive predictive value. Under one set of parameters, the classifier detects cases more than two years in advance (on average), correctly labels 51% of all positive cases as high-risk, and correctly labels 85% of negative cases as low risk. 10% of all high-risk patients would later be diagnosed with some form of abuse, assault, or intentional injury. This is almost three times higher than the overall prevalence of 3.4% These outcomes trade off under different parameters and case definitions of abuse, but the outcomes are generally similar. The most predictive model parameters were identified as alcohol/substance-related disorders, injuries from external causes, poisonings, frequency of clinical visits, and various mental disorders. These predictors were quantified using partial risk scores and shown to vary nontrivially between men and women.

**Conclusions**: The naïve Bayesian classifier was able to make sensitive, specific, and early prediction of risk of future domestic abuse, and its parameters identified various risk factors, stratified between men and women. This approach may be useful as part of an early warning system to identify high-risk patients for screening.