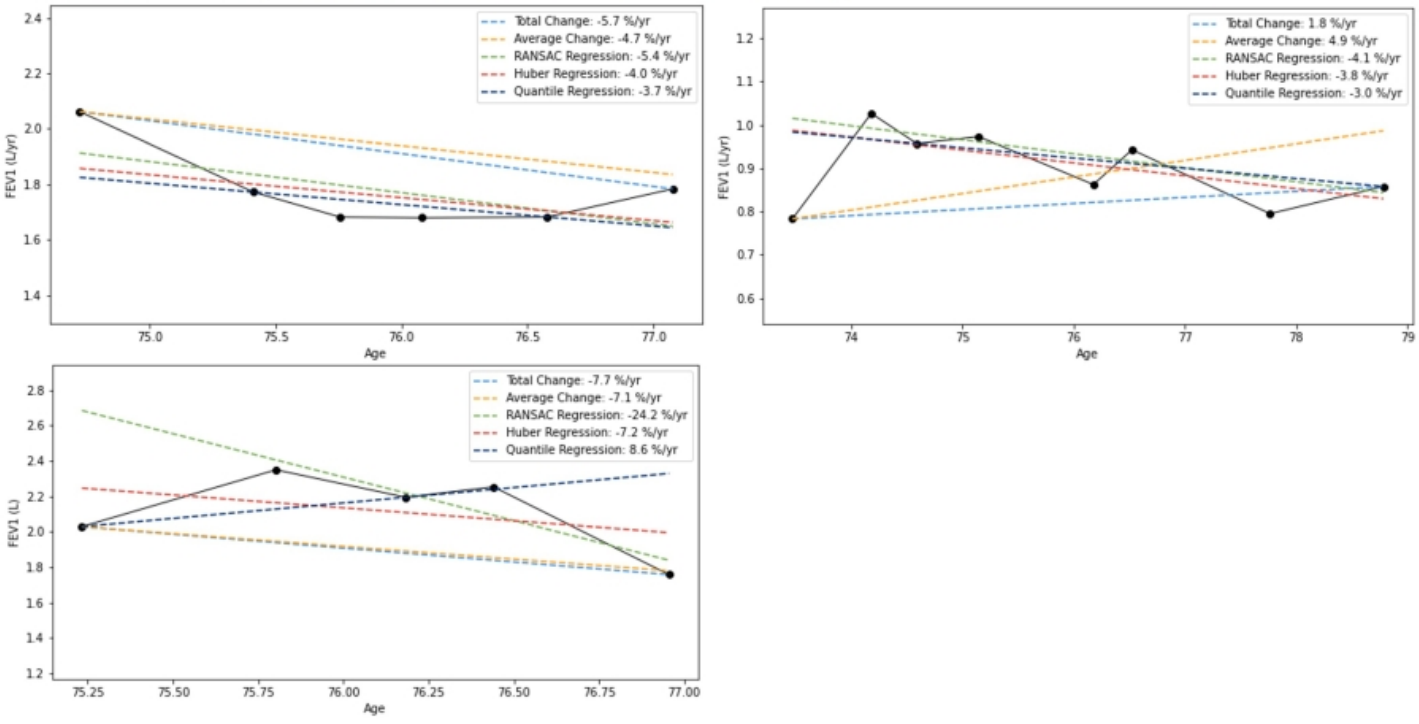
Estimating Rate of Lung Function Change Using Clinical Spirometry Data

A. Balasubramanian[1], C. Cervantes[2], A. Gearhart[2], N. Putcha[3], A. Fawzy[1], M. C. Mccormack[4], A. C. Singh[5], R. A. Wise[6], N. N. Hansel[7]; [1]Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, MD, United States, [2]Applied Physics Laboratory, Johns Hopkins University, Laurel, MD, United States, [3]Medicine, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, MD, United States, [4]John Hopkins Univ, Baltimore, MD, United States, [5]Pulmonary/Critical Care, Pittsburgh, PA, United States, [6]Pulmonary and Critical Care, Johns Hopkins Univ School of Medicine, Baltimore, MD, United States, [7]Johns Hopkins Univ, Baltimore, MD, United States.

Corresponding author's email: abalasu2@jhmi.edu

Rationale: In Chronic Obstructive Pulmonary Disease (COPD), accurately estimating lung function from electronic health record (EHR) data would be beneficial but requires addressing complexities in clinically-obtained lung function testing. This study compared analytic methods for estimating rate of change in forced expiratory volume in one second (FEV_1) obtained from EHR data. Methods: We estimated rate of FEV_1 change in COPD patients who had ≥3 outpatient tests spanning at least one year from the Johns Hopkins Healthcare System. Estimates were calculated as both an absolute mL/year and a relative %/year using non-regressive (Total Change, Average Change) and regressive (Quantile, Random Sample Consensus [RANSAC], Huber) methods. We compared distributions of the estimates across methods with particular attention to extreme values. Associations between estimates and frequency of all-cause or COPD hospitalizations were performed using univariate zero-inflated negative binomial regressions. Results were validated in an external cohort of similar individuals with COPD. Results: Among 1417 participants, estimated median rate of change across most methods was approximately -30 mL/yr or -2%/yr. Non-regressive methods frequently generated erroneous estimates due to outlier first measurements or short intervals between tests; Average Change yielded the most extreme estimates (minimum=-3761 mL/yr). Regressive methods, and Huber specifically, minimized extreme estimates and dispersion (Huber coefficient of variation -3.54). Huber regression included all FEV_1 values in estimation, weighting outliers rather than excluding them, as is explicit in RANSAC estimation or implicit in Quantile estimation (Figure 1). Huber, Total Change, and Quantile estimates were associated with all-cause hospitalizations (Huber incidence rate ratio 0.98, 95%CI 0.97-0.99, p<0.001). More rapid decline in FEV_1 as estimated by Huber regression was also associated with smoking status, increased comorbidities, and prior hospitalizations. Conclusions: Using EHR data to estimate FEV_1 rate of change is highly clinically applicable but sensitive to variable numbers of tests, intervals of testing, and outlier values resulting from factors including clinical indication for testing or quality of test. While no analytic method will fully overcome these complexities, we identified Huber regression as consistently minimizing extremes and avoiding erroneous estimates. Huber regression may be a useful analytic method in future analyses of lung function decline using EHR data.

Figure 1. Pulmonary Function Timelines and Estimated Slopes by Method



This abstract is funded by: Precision Medicine Center of Excellence, Highmark Health