BMJ Open Respiratory Research

Estimating rate of lung function change using clinical spirometry data

Aparna Balasubramanian ,[1] Christopher Cervantes,[2] Andrew S Gearhart,[3] Nirupama Putcha,[1] Ashraf Fawzy,[1] Meredith C McCormack,[1] Anil Singh,[4] Robert A Wise,[1] Nadia N Hansel[1]

To cite: Balasubramanian A, Cervantes C, Gearhart AS, *et al*. Estimating rate of lung function change using clinical spirometry data. *BMJ Open Respir Res* 2024;**11**:e001896. doi:10.1136/bmjresp-2023-001896

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjresp-2023-001896).

AB and CC are joint first authors.

Received 19 June 2023
Accepted 10 September 2024

[1]Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, Maryland, USA
[2]Asymmetric Operations Sector, Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, USA
[3]Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, USA
[4]Pulmonary and Critical Care Medicine, Alleghany Health Network, Pittsburgh, Pennsylvania, USA

Correspondence to
Dr Aparna Balasubramanian; abalasu2@jhmi.edu

ABSTRACT

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 testing. This study compared analytic methods for estimating rate of forced expiratory volume in one second (FEV_1) change from EHR data.

Methods We estimated rate of FEV_1 change in patients with COPD from a single centre who had ≥3 outpatient tests spanning at least 1 year. Estimates were calculated as both an absolute mL/year and a relative %/year using non-regressive (Total Change, Average Change) and regressive (Quantile, RANSAC, Huber) methods. We compared distributions of the estimates across methods focusing on extreme values. Univariate zero-inflated negative binomial regressions tested associations between estimates and all-cause or COPD hospitalisations. Results were validated in an external cohort.

Results Among 1417 participants, median rate of change was approximately −30 mL/year or −2%/year. 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/year), while regressive methods, and Huber specifically, minimised extreme estimates. Huber, Total Change and Quantile FEV_1 slope estimates were associated with all-cause hospitalisations (Huber incidence rate ratio 0.98, 95% CI 0.97 to 0.99, p<0.001). Huber estimates were also associated with smoking status, comorbidities and prior hospitalisations. Similar results were identified in an external validation cohort.

Conclusions Using EHR data to estimate FEV_1 rate of change is clinically applicable but sensitive to challenges intrinsic to clinically obtained data. While no analytic method will fully overcome these complexities, we identified Huber regression as useful in defining an individual's lung function change using EHR data.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Spirometry is a cornerstone of assessment in many lung diseases including chronic obstructive pulmonary disease, with prior studies demonstrating that rapid declines in lung function are associated with morbidity and mortality. Being able to understand and appropriately analyse electronic health record pulmonary function testing data, which is readily available but expectedly noisy, is a growing need.

WHAT THIS STUDY ADDS

⇒ This study compares how different analytic methodologies handle key challenges associated with clinical spirometry data and identifies Huber regression as a useful method to accurately estimate an individual's rate of lung function change over time.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Accurately estimating lung function change for an individual from clinical data will allow clinicians to better assess progression of disease and offer opportunities to use electronic health record spirometry data in research studies on lung disease.

INTRODUCTION

Spirometry is a cornerstone of assessment in chronic obstructive pulmonary disease (COPD), where repeated measures are used to evaluate lung function over time. In 1977, Fletcher *et al* cemented the importance of spirometry by documenting lung function decline in smokers.[1] Since then, numerous studies have observed a decline in forced expiratory volume in one second (FEV_1) among patients with COPD, and rate of lung function decline has been used in pharmacologic therapy trials as an outcome of interest.[2–12] Prognostically, rapid FEV_1 decline has been associated with increased hospitalisations and mortality.[13 14] Despite the importance of quantifying lung function change over time, it remains difficult to accurately estimate an individual's lung function trajectory in clinical practice.

Much of the related work has relied on longitudinal studies where spirometry is conducted at regular intervals under controlled settings over a prespecified time frame.[12 15–17] Unfortunately, this is not how pulmonary function testing (PFT) or interpretation occurs in clinical practice, raising the concern that methods to define lung function trajectories, in particular simple slope methods, developed from

research studies may be inadequate in the clinical setting. More sophisticated lung function modelling often relies on methods, which compare individuals to population-level averages, acknowledging that even among healthy individuals, there is a large degree of individual variability around those averages. Defining an individual's rate of lung function change, without necessarily estimating population-level averages, however, is particularly relevant clinically as population averages in clinical cohorts can be biased by the indications for testing and heterogeneity in the population across healthy and disease states. Furthermore, common clinical practice when estimating lung function decline is to examine all available FEV_1 measurements, often without context on the indication, setting or quality of those tests. By aligning methods with this heuristic, lung function decline assessment may be linked more closely with clinical decision-making and clinical outcomes. Additionally, real-world clinical data offer the opportunity to use the rich characterisation available in electronic health records (EHR) to explore novel factors that contribute to both FEV_1 decline and improvement.

With these numerous considerations in mind, methods for quantifying rate of change must be examined as a first step in defining an individual's lung function trajectory using clinical data. EHR pulmonary function data pose unique analytic challenges due to high variability that is both intrinsic to the test and related to clinical fluctuations, differing numbers of tests, indications for testing and inconsistent time intervals between tests. This study compares methods of quantifying lung function change using clinical data with the goal of identifying a reliable and accurate method to characterise a patient's longitudinal trajectory.

METHODS

Study population and design

This retrospective observational study was conducted using EHR data from the Johns Hopkins Healthcare System (JHHS) on individuals with COPD who had at least three outpatient PFT conducted between 1 December 2015 and 14 July 2022 (online supplemental methods). We restricted our analyses to outpatient testing to minimise values captured during acute events. Individuals were excluded for (1) <1 year between first and last PFT, (2) lung transplantation or lung resection surgery with an unknown date. If surgical dates were known, only PFTs obtained prior to that date were included.

We obtained an external validation cohort from the Highmark Allegheny Health Network (HAHN) of 124 individuals with ≥3 PFTs spanning at least 1 year. Location of PFTs and surgical details were unavailable from this data, therefore we did not restrict to outpatient PFTs nor apply surgical exclusions. The aim of validation was to demonstrate that the method that performed best in the JHHS population, performed as well in a distinct study population with potentially different practice patterns, leading to obtaining testing and less information regarding the context of testing. The institutional review board approved this study at both sites with waived consent (IRB00251249).

PFT data and clinical characterisation

All PFTs were conducted and interpreted using ATS/ERS guidelines.[18] We abstracted test date, location, absolute FEV_1, FVC, FEV_1/FVC, diffusing capacity of the lung for carbon monoxide (DLCO), age at each PFT, biologic sex at birth, self-reported race, ethnicity, height and weight to calculate body mass index, smoking status at the first study-eligible PFT and number of all-cause and COPD exacerbation (ICD-10 J44.1) hospitalisations. Comorbidities were defined by ICD-10 diagnosis codes for coronary artery disease, congestive heart failure, diabetes mellitus, obstructive sleep apnoea, chronic kidney disease and pulmonary hypertension.

FEV_1 rate of change

We selected methods for estimating an individual's rate of lung function change, λ, based on commonly used methods in the literature. We hypothesised these methods would generate different λ estimations for a given individual, demonstrate varying differences between estimated and observed values and yield varying associations between λ estimations and clinical outcomes. Population-level estimates were intentionally not analysed, focusing instead on estimating λ based solely on an individual's own PFT data. We compared five methods: Total Change, Average Change, Random Sample Consensus (RANSAC) regression, Quantile regression and Huber regression. Total Change was defined as the slope between the first and last PFT in an individual's timeline, while Average Change is the average of the slopes between consecutive PFTs. RANSAC is a regression method used to estimate model parameters based on iteratively sampling a random subset of PFTs, testing whether the remaining PFTs fit the parameters within a prespecified margin, and optimising the model parameters to maximise for inliers or values that meet those prespecified margins.[19] RANSAC, therefore, will select the regression model parameters that include the most PFT values, and as a result will also detect and exclude outliers. Quantile regression is a parameter estimation method that minimises the absolute loss across the data, which in our case finds the median.[20] Huber regression[21] minimises the squared loss of inlier samples and the absolute loss of outlier samples. Analyses were performed using an absolute change metric, millilitres per year (mL/year), and a relative change metric, percent change from index FEV_1 per year (%/year). These analyses focus on FEV_1, given its pre-eminence in COPD assessment; FVC and DLCO change, although both still clinically relevant in COPD, were not analysed, as within our study population, there was limited variability in FVC values over time and limited sample size for DLCO. The time scale used was calendar

time between first and last PFT, centred at the age of each individual's first PFT. Further details on all methods are provided in online supplemental file.

Statistical analysis

To first assess how well each method's estimates reflected an individual's PFT timeline, we compared distributions of λ estimates generated by each method using summary statistics with a particular focus on extreme values (<−15 or ≥15%/year or <−250 or ≥250 mL/year). To quantify which methods minimised error, we calculated mean squared error (MSE) comparing actual values to expected values from each modelling method for all individuals. Distributions of those MSE values across methods were compared with pairwise t-tests to examine statistical differences between the average MSE values. We further examined individual examples of common phenomena that generated extreme λ estimates within each method. We also sought to test whether the generated estimates reflected known associations between lung function decline and key clinical characteristics and outcomes. Univariate zero-inflated negative binomial regressions were performed to test the associations between individual-level rate of change estimates from each method and (1) the total number of all-cause hospitalisations and (2) the total number of COPD hospitalisations. Both outcomes were captured as concurrent with the individual's PFT timeline. The goal was to identify whether estimates generated by each method demonstrated an association between greater lung function decline and hospitalisations. Additionally, clinical characteristics at the time of first PFT, such as demographics, smoking status, baseline PFTs and comorbidities were compared across groups of λ, calculated using Huber regression in %/year. Huber estimates were categorised to allow for non-linearity across clinical characteristics. Pairwise t-tests or χ^2 tests, as appropriate, were performed comparing the middle group to the group with the greatest decline. Our analyses were conducted using Python, using Scikit-Learn[22] and Statsmodels,[23] and $p<0.05$ was considered statistically significant.

Patient and public involvement

As this study is an EHR-based study, designed to be able to leverage existing clinical data for both research and future patient care uses, it was not appropriate or possible to involve patients and/or the public in the design, conduct, reporting or dissemination of this study.

RESULTS

The JHHS study population consisted of 1417 patients with COPD with at least three outpatient PFTs over at least 1 year of follow-up (figure 1). Over one quarter of the study cohort had exactly three PFTs (n=406, 28.7%) with a median of five PFTs per individual (IQR 3–7). The median time between consecutive PFTs was 147 days (IQR 91–294 days) with over half of individuals (n=742, 52.4%) having a PFT within 120 days and 5% (n=71) within 28 days of their prior PFT. Individuals had an average age of 60±13.5 years, 60.4% were women, and 75% were ever smokers. The study population had a median FEV_1 at index PFT of 1.7 L (IQR 1.31–2.36 L) and median FEV_1/FVC ratio of 0.69 (IQR 0.59–0.76). Across each individual's follow-up time, there were a total of 4679 all-cause, and 216 COPD hospitalisations with a rate of 3.2 all-cause hospitalisations per person and 1.6 COPD hospitalisations per person.

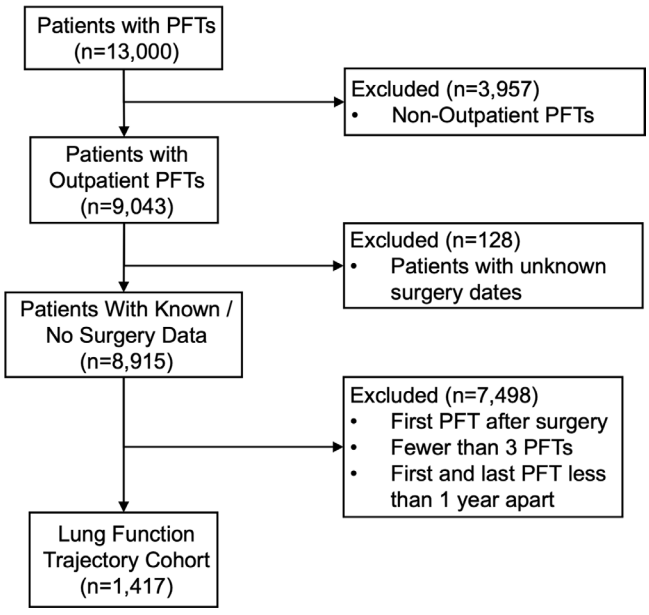

Figure 1 Flowchart of the JHHS cohort, where we exclude 1) PFTs recorded during any setting other than outpatient, 2) individuals for whom we know a transplant/surgical lung resection occurred without associated date data, 3) individuals for whom we do not have PFTs prior to a transplant/surgical lung resection, 4) individuals with fewer than three PFTs, 5) individuals for whom we have less than 1 year of PFT data. JHHS, Johns Hopkins Healthcare System; PFTs, pulmonary function tests.

Comparing estimated rates of lung function change

We compared distributions of estimated individual rate of lung function change, λ, across our simple (Total Change, Average Change), and regression (RANSAC, Quantile, Huber) methods, in both mL/year (figure 2A) and %/year (figure 2B). The trend of distributions for these methods was similar across both units: most individuals were estimated to lose approximately 30 mL or 2% per year, and around 10% of individuals were estimated to have rapid lung function decline at or below −200 mL/year or −20%/year.

The distribution of λ estimates was similar across all methods with the exception of Average Change, which yielded the greatest proportion of individuals with extreme value estimates. By contrast, Huber regression yielded more gradual estimates than the other methods,

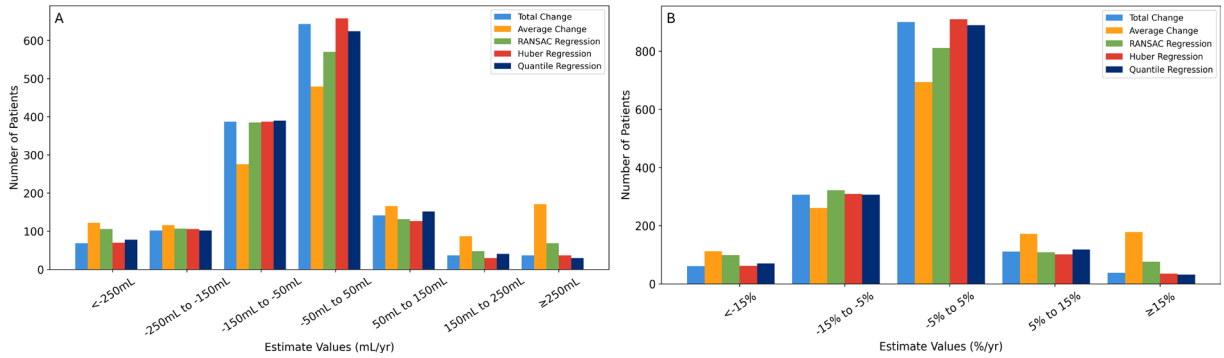

Figure 2 Estimated individual rates of lung function change (λ) for the JHHS cohort in mL/year (A) and %/year (B); values are binned. JHHS, Johns Hopkins Healthcare System; RANSAC, Random Sample Consensus.

minimising extreme values. All methods generated a similar median rate of change of approximately −30 mL/year or −2%/year with the exception of Average Change (online supplemental table 1). The minimum and maximum estimates are particularly relevant to the discussion of rapid lung function change. Average Change yielded a minimum rate of change of −3760 mL/year or −161.2%/year. Huber regression, by contrast, yielded the smallest extreme values with a minimum of −1089 mL/year or −40.2%/year (online supplemental table 1). Additionally, Huber regression estimates had the smallest coefficients of variation, coinciding with the decrease in extreme values and reduced variation compared with the other methods.

Lung function trajectory estimation method analysis

To characterise which method minimised prediction error, we compared distributions of MSE across methods (figure 3). In general, all the estimation methods had low (<0.1) MSE values for most individuals, but Average Change and RANSAC yielded the most instances with an MSE ≥0.1. Conversely, the Huber and Quantile regression estimates generated the fewest instances with higher error estimates. Comparing the average MSE by method, we found that Huber regression minimised MSE as compared with Total Change, Average Change and RANSAC but was not statistically different from Quantile regression (online supplemental table 2).

To better understand the observed estimates, we identified phenomena that appeared inconsistent with patient timelines. Examples are provided in figure 4. For simple timelines, such as figure 4A, methods yielded similar estimates; all λ values are within 2%/year since PFTs were evenly spaced and FEV_1 values changed gradually. Figure 4B highlights a common challenge with the non-regressive methods: the first FEV_1 was lower than the rest of the timeline, likely reflecting an atypical or erroneous value, and Total Change and Average Change yielded extreme estimates as a result. Figure 4C highlights another common phenomenon where significant FEV_1

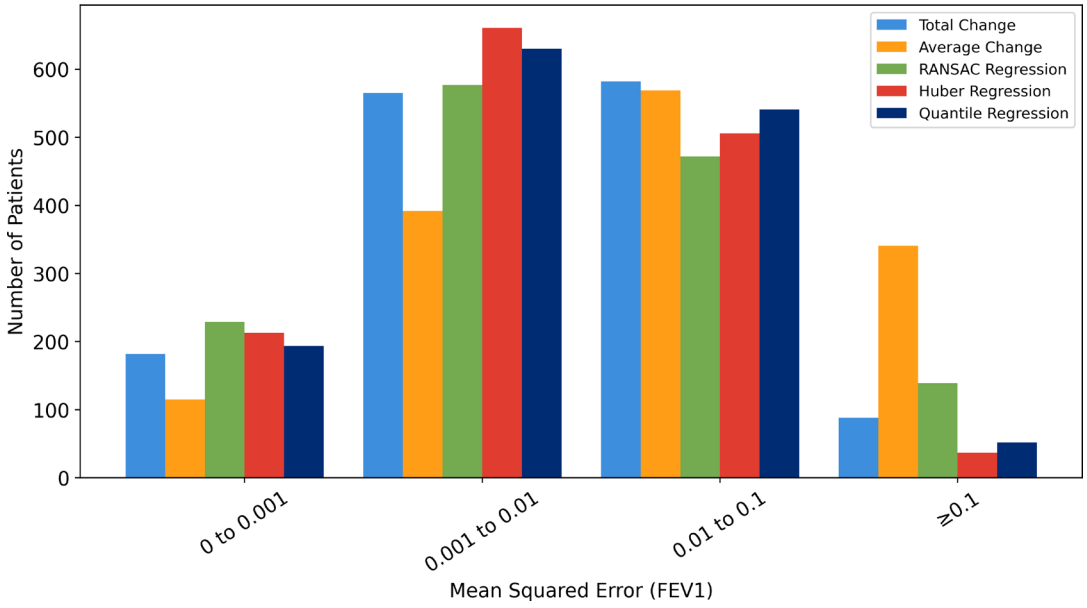

Figure 3 Distribution of mean squared error values for the JHHS cohort from %/year models by estimation method. FEV_1, forced expiratory volume in one second; JHHS, Johns Hopkins Healthcare System; RANSAC, Random Sample Consensus.

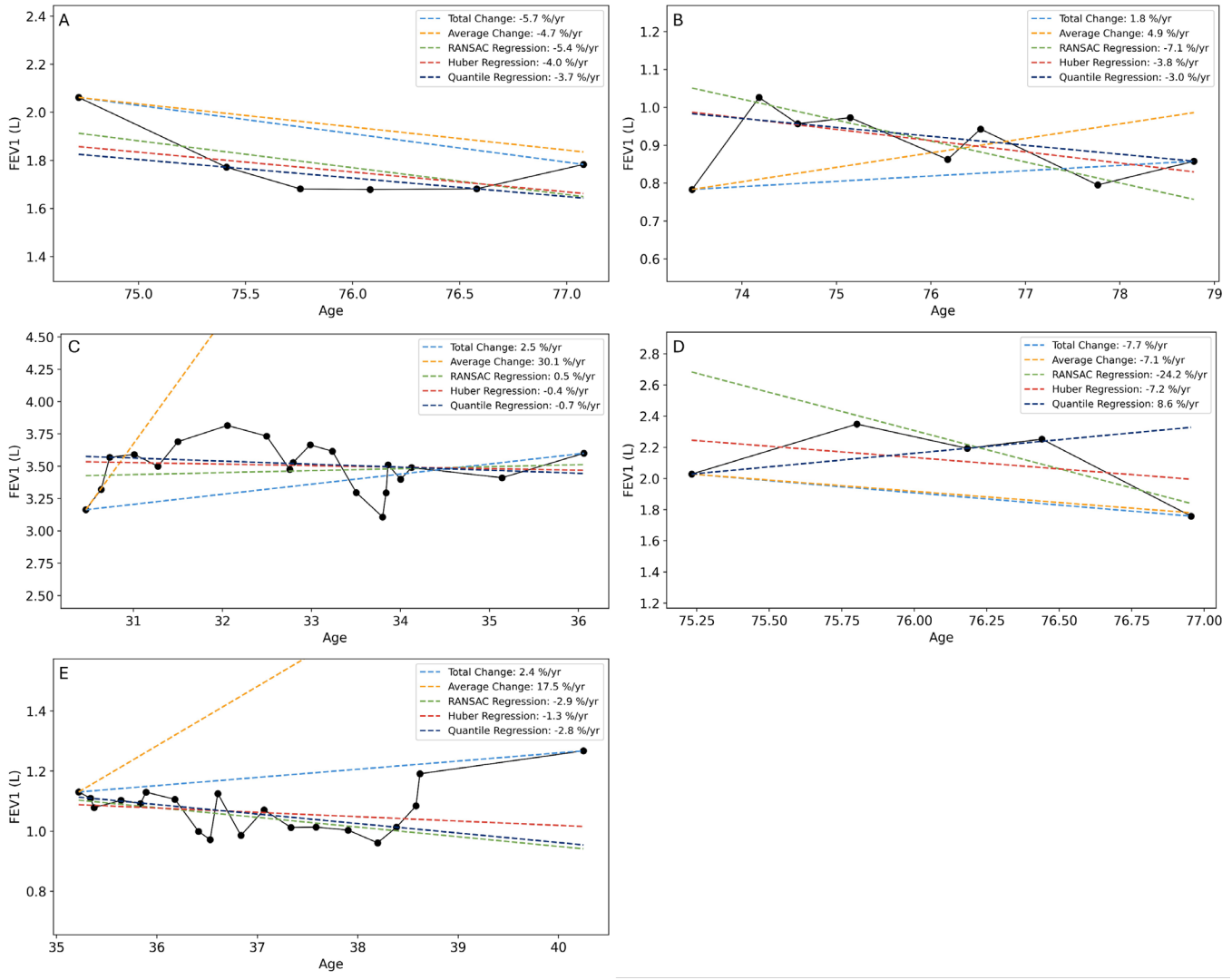

Figure 4 PFT Timelines (black) shown as FEV_1 value against an individual's age; estimated lung function trajectories are shown as dashed lines, where regression trajectories start at an optimised point. Panels A-E are examples of five individual patient timelines. FEV_1, forced expiratory volume in one second; PFT, pulmonary function test; RANSAC, Random Sample Consensus.

change over short time intervals led to extreme Average Change estimates; temporal proximity paired with even moderate FEV_1 change in lung function could yield poor estimates due to slope averaging.

Figure 4D,E provides examples of highlighting differences among the regression methods, which differ in their treatment of outliers. In our setting, Quantile regression estimates the median value against which to minimise difference and is thus generally robust to outliers. In figure 4D, the Quantile estimate indicated moderate improvement (λ^q=8.6%/year), which arose from a tight-fitting line along the first, third and fourth PFTs. While this makes sense mathematically, we have no reason to trust the accuracy of these particular FEV_1 values more or less than those excluded.

RANSAC regression, by contrast, explicitly values some PFTs over others via random sampling. In figure 4D, RANSAC likely excludes the first PFT and thus estimated rapid decline (λ^r=−24.2%/year) where Quantile estimated improvement. Because Quantile and RANSAC focused on different subsets of the timeline, they generated starkly different estimates. Huber regression, by contrast, estimated moderate lung function decline (λ^h =−7.2%/year), under the implicit assumption that any PFT may be in error but, taken together, define a general trajectory.

Finally, figure 4E highlights all of the aforementioned issues. Total Change entirely ignored the loss of lung function in the middle of the timeline. Average Change yielded rapid improvement, likely due to PFT temporal proximity. RANSAC regression appeared to ignore the final two PFTs, and Quantile regression found a line that fits the first and seventeenth PFTs while minimising the improvement in the latter portion of the timeline. Huber regression offered the middle estimate value (λ^h=−1.3%/

Table 1 Associations between estimated individual change in lung function and all-cause or COPD hospitalisations for the JHHS cohort

	All-cause hospitalisation		COPD hospitalisation	
	IRR (95% CI)	P value	IRR (95% CI)	P value
Total change	**0.982 (0.97 to 0.99)**	**<0.001**	0.982 (0.94 to 1.02)	0.358
Average change	1.001 (0.99 to 1.00)	0.461	1.005 (0.99 to 1.01)	0.143
RANSAC regression	1.000 (0.99 to 1.00)	0.905	1.003 (0.99 to 1.01)	0.05
Quantile regression	**0.980 (0.97 to 0.99)**	**<0.001**	0.970 (0.94 to 1.00)	0.085
Huber regression	**0.980 (0.97 to 0.99)**	**<0.001**	0.974 (0.94 to 1.01)	0.166

Bolded values represent statistically significant associations (p<0.05).
COPD, chronic obstructive pulmonary disease; IRR, incidence rate ratio; JHHS, Johns Hopkins Healthcare System; RANSAC, Random Sample Consensus.

year), accounting for both the decline and improvement in FEV_1.

Correlation with clinical outcomes

To better characterise the utility of each of these methods, we examined the clinical relevance of each method's λ (%/year) by evaluating univariate associations with all-cause and COPD hospitalisations that occurred between an individual's first and last PFTs (table 1). Total Change and Quantile and Huber regressions demonstrated significant associations with all-cause hospitalisations, noting a 2% decrease in hospitalisation risk with every 1% increase in λ. Although no method demonstrated significant associations with COPD hospitalisations, there was a trend towards lower risk with less rapid FEV_1 decline estimated by Total Change, Quantile and Huber.

Noting both the stability and clinical relevance of Huber regression in %/year, we chose to assess clinical characteristics at index PFT by groups defined by λ^h (table 2) to examine whether factors historically associated with lung function decline correlate with estimates. Individuals with more rapid lung function decline had a lower FEV_1 at index PFT and had lower baseline FVC and DLCO values. Individuals with faster lung function decline were more frequently current smokers, more commonly had comorbidities including coronary artery disease and diabetes and were more likely to have had an all-cause hospitalisation with a trend towards more COPD hospitalisation in the year prior to index PFT, consistent with expected predictors of lung function decline. We also noted that rapid improvers ($\lambda^h \geq 5\%$/year) were similar to rapid decliners with respect to most clinical characteristics.

Validation in external cohort

Given that much of the analysis and method selection examines extremes and relied on frequent patterns of outliers, we further attempted to validate our findings in an external cohort. The HAHN cohort had a smaller sample size with fewer PFTs and no available surgical data or data on inpatient versus outpatient PFTs (online supplemental figure 1). After exclusions, there were 124 individuals from the HAHN cohort, with a median 3 (IQR 3–4) PFTs per individual. Median time between successive PFTs was 396 days (IQR 244–666 days). Average baseline FEV_1 was similar to the JHHS cohort with a median of 1.6 L (IQR 1.3–2.0 L). However, there were fewer all-cause and COPD hospitalisations than the JHHS cohort during the follow-up time, with a total of 538 all-cause hospitalisations at a rate of 2.6/person, and 30 COPD hospitalisations for a rate of 1.4/person. Despite these cohort differences, distributions in FEV_1 change in HAHN were similar to those observed in JHHS across methods, with Average Change having the most dispersion and Huber having the least (figure 5). Distributions in MSE by method were also like those observed in JHHS, wherein all methods across most individuals had fairly minimal error; Average Change and RANSAC remained the methods with the most frequent higher MSE values (online supplemental figure 2). Huber regression continued to generate the lowest average MSE, though differences were only statistically significant from RANSAC (online supplemental table 3).

Total Change and Quantile and Huber regression estimates again demonstrated associations with all-cause hospitalisations in univariate models (table 3), with similar magnitude effects. In the HAHN cohort, all-cause hospitalisations decreased approximately 5% for every 1%/year increase in lung function change. Due to the low number of COPD hospitalisations in this cohort during follow-up time, this outcome could not be modelled. We also looked at clinical characteristics from the year prior to index PFT (online supplemental table 4). Although modest sample size limited comparisons, we still observed a trend towards increasing all-cause hospitalisations with more rapid lung function decline groups. In both cohorts, regardless of additional confounding from inpatient PFTs and surgical intervention, the Huber regression method appears to minimise noise in analysing real-world PFTs in a manner that is consistent with expected clinical characteristics and outcomes.

Table 2 Clinical characteristics for individuals at time of first PFT in the JHHS cohort, grouped by estimated individual rate of lung function change according to Huber regression (λ^h in %/year)

	λ < −10% (N=138)	−10% ≤ λ λ <−5% (N=233)	−5% <= λ λ<0% (N=629)	0% ≤ λ λ<5% (N=279)	5% ≤ λ (N=136)
Time between index and last PFT (years), median*	2.1	3.0	4.0	3.4	2.0
Age (years)	61.9±12.2	62.1±12.3	60.7±13.1	56.7±15.1	59.1±14.0
Female, n (%)*	82 (59.4%)	142 (60.9%)	395 (62.8%)	161 (57.7%)	75 (55.1%)
African American, n (%)*	37 (26.8%)	61 (26.2%)	169 (26.9%)	78 (28.0%)	39 (28.7%)
BMI (kg/m^2)	29.4±9.0	28.6±7.0	28.9±6.9	28.6±9.2	29.4±6.9
Current smoker, n (%)*	18 (13.0%)	28 (12.0%)	44 (7.0%)	20 (7.2%)	11 (8.1%)
FEV1 (L/s)*	1.7±0.8	1.7±0.8	1.9±0.8	1.9±0.8	1.5±0.7
FVC (L/s)*	2.5±0.9	2.6±1.0	2.9±1.0	2.9±1.1	2.5±1.0
DLCO (mL/min/mm Hg)*	10.5±5.4	12.2±6.0	13.6±5.5	13.3±5.6	12.3±6.0
Congestive heart failure, n (%)*	8 (5.8%)	26 (11.2%)	61 (9.7%)	27 (9.7%)	17 (12.5%)
Coronary artery disease, n (%)*	26 (18.8%)	41 (17.6%)	93 (14.8%)	35 (12.5%)	22 (16.2%)
Diabetes, n (%)*	28 (20.3%)	47 (20.2%)	118 (18.8%)	52 (18.6%)	30 (22.1%)
Sleep apnoea, n (%)*	25 (18.1%)	49 (21.0%)	109 (17.3%)	48 (17.2%)	34 (25.0%)
Pulmonary hypertension, n (%)	13 (9.4%)	27 (11.6%)	56 (8.9%)	26 (9.3%)	13 (9.6%)
Chronic kidney disease, n (%)*	13 (9.4%)	23 (9.9%)	56 (8.9%)	20 (7.2%)	12 (8.8%)
Patients w/ all-cause hospitalisation (year before first PFT), n (%)*	53 (38.4%)	78 (33.5%)	158 (25.1%)	79 (28.3%)	47 (34.6%)
Patients w/ COPD hospitalisation (year before first PFT), n (%)	4 (2.9%)	6 (2.6%)	10 (1.6%)	3 (1.1%)	1 (0.7%)

Values are given as number and percentage of cohort (n, %) or mean and standard deviation (μ, σ).
*Represents p value<0.001 for comparisons between λ< −10 and −5≤ λ<0% /year groups.
BMI, body mass index; COPD, chronic obstructive pulmonary disease; DLCO, diffusing capacity of the lung for carbon monoxide; FEV1, forced expiratory volume in one second; FVC, forced vital capacity; PFT, pulmonary function test.

DISCUSSION

In this EHR study of patients with COPD, we demonstrated that real-world PFT data are fraught with complexities that limit the utility of common methods for evaluating individual lung function trajectories. While most research studies of populations calculate absolute mL/year change using non-regressive methods, we identified that applying such methods in a real-world dataset generates inaccurate estimates. Rather, using a relative change metric and carefully handling outliers with regression methods to partially address uncertainty from clinical variability, testing quality and testing indication minimised erroneous estimation. When comparing methods, we identified Huber regression as clinically relevant and the least likely to generate implausible extremes. These findings offer insight into analytic methods that align

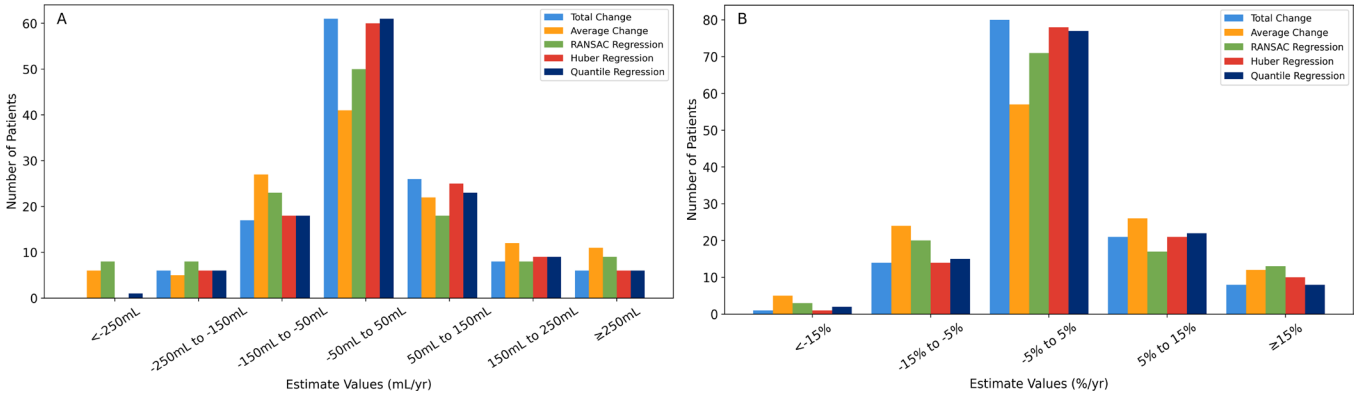

Figure 5 Estimated individual rates of lung function change (λ) for the HAHN cohort in (A) mL/year and (B) %/year; values are binned. HAHN, Highmark Allegheny Health Network; RANSAC, Random Sample Consensus.

Table 3 Associations between estimated individual change in lung function and all-cause hospitalisations in HAHN cohort

	All-cause hospitalisation	
	IRR (95% CI)	P value
Total change	**0.9498 (0.92 to 0.99)**	**0.006**
Average change	0.9989 (0.99 to 1.01)	0.772
RANSAC regression	0.9999 (0.98 to 1.02)	0.988
Huber regression	**0.9492 (0.91 to 0.98)**	**0.005**
Quantile regression	**0.9496 (0.92 to 0.98)**	**0.004**

Bolded values represent statistically significant associations (p<0.05).
HAHN, Highmark Allegheny Health Network; IRR, incidence rate ratio; RANSAC, Random Sample Consensus.

with clinical heuristics when interpreting longitudinal trends in lung function and offer a method to further evaluate clinical decision-making and management of COPD using EHR.

PFTs are vital to our understanding of COPD progression and prognosis.[24] National initiatives have prioritised the integration of PFT data into and across healthcare systems,[25] making an understanding of EHR PFT analytic complexities necessary. We identified two major challenges that must be addressed when analysing lung function trajectories from PFTs obtained in the clinical setting: timing and outliers.

Errors in estimation related to timing were the result of variable intervals between tests and variable number of tests. Research studies of lung function trajectories in COPD primarily rely on evenly spaced measurements[12 16 17] and do not frequently encounter mathematical issues with short intervals. In our cohort, however, this was a frequent problem that dramatically altered the distribution of estimated rate of change in FEV_1. Many research studies have relatively few tests with equal numbers across participants,[9 15 26–28] which limits the usefulness of regression methods. In a study examining three cohorts, the Framingham Offspring Cohort, Copenhagen City Heart Study and Lovelace Smokers Cohort, two spirometry measurements and a Total Change method were used to define rate of change.[15] In those instances, we would not expect regression methods to perform better than Total Change. Similarly, we found distribution and clinical relevance of Total Change was similar to Huber regression, likely because over 40% of patients had exactly three PFTs. As observed in examples, however, Total Change is susceptible to outlier first and last values, which are more likely to occur in real-world datasets. When additional PFTs are available, estimates can be more accurate when not limited to these two values. This was reflected in the extremes, where minimum values differed by approximately 10%/year between Total Change and Huber regression.

The challenges presented by having few tests are further compounded by disease severity, which predisposes calculation of an absolute change to be implausibly extreme. Individuals with lower absolute FEV_1 do not have the lung function to maintain the same absolute change in FEV_1 that they may have had with a higher FEV_1. As such, we found that a relative change, such as %/year was more compatible with real-world data in a COPD population.

Outlier values in the FEV_1 timelines of individuals may be associated with clinical indication for testing, acute illness and variable testing setting and quality. Using methods like Total Change and Average Change are, as mentioned previously, particularly susceptible to outliers and can amplify measurement error in FEV_1 to yield erroneous extremes. Strategies that implicitly or explicitly exclude outliers, like Quantile or RANSAC regression, discount informative outliers and are thus not ideal for estimating an individual's rate of change. This conceptually aligns with clinical practice; providers often examine flowsheets of all available FEV_1 values for a patient to estimate lung function trajectory and will weight values based on clinical judgement rather than assuming nonconforming PFTs are irrelevant. Additionally, most providers do not systematically examine the quality, clinical indication or testing setting for each test when assessing trajectory. To reflect this, Huber regression assumed that some PFTs may be in error by some margin, but we do not know *a priori* which PFTs are useful.

There are contexts when each method may yield appropriate results: the first and last PFT may accurately describe a trajectory, the PFTs may be evenly spaced and exhibit gradual PFT-to-PFT change, or a representative subset of PFTs may capture the overall trend. Huber regression, however, consistently yielded estimates that capture all of a timeline's PFTs without mathematical difficulty handling variable time intervals, and without assuming any individual FEV_1 is more useful than another.

We also examined whether each method's estimates concurred with historic observations. The median rate of change was approximately −30 mL/year or −2%/year across most methods, consistent with prior population estimates of lung function decline.[12 15 17 29] Furthermore, estimates from Total Change, Quantile and Huber were associated with all-cause hospitalisations and demonstrated trends towards associations with COPD hospitalisations consistent with previously observed associations between FEV_1 decline and hospitalisations.[13 14] Finally, we examined our best-performing method, Huber regression, with respect to clinical characteristics thought to predict lung function decline including smoking and prior exacerbations,[9 12 30–35] and Huber estimates were consistent with those trends in predictors. We also observed that patients with a lower baseline FEV_1 tended to have the highest rates of decline in %/year, similar to the 'horse-racing effect' observed in blood pressure studies.[36 37] This contradicts prior work in COPD, where individuals with mild–moderate

disease had the greatest decline in lung function.[38] It should be noted, however, that this study examined mL/year rather than %/year to remain consistent with the existing literature,[39] acknowledging that mild–moderate individuals starting at a higher FEV_1 will have a much larger absolute decrease in FEV_1 than severe individuals for the same %/year value.

Our analyses focused on extreme values to characterise clinically relevant subgroups defined thereby, such as rapid decliners. While most methods generated similar median estimates, they differed at the extremes, which were often implausible. Artefactual extremes resulting from the aforementioned challenges will therefore impact conclusions drawn on rapid decliners. We also note the prevalence of FEV_1 improvers. This group is not well-characterised in the literature and had a similar clinical characteristic profile to rapid decliners. While this prevalence may be attributable to medications or other successful management strategies, it may also indicate asthma overlap or a resilient phenotype that warrants further investigation.

This study has a few limitations. First, the COPD definition criteria were expansive and likely included individuals without COPD. Examination of analytic methods for repeated PFT measurements in a real-world diseased cohort, however, is applicable regardless of cohort inclusion specificity. Second, each method was evaluated by the frequency with which patterns that generate artefactual extremes occur. While it is possible that additional unobserved or infrequently observed complexities may bias the estimates, the robust sample size and external validation minimise that risk. In fact, even in our external validation cohort, without exclusion for inpatient PFTs or surgical interventions, we found that Huber regression performed well, further supporting our conclusions. Third, clinical characterisation did not include medication use or outpatient exacerbations, which could further explain some of the observed patterns. This is especially relevant for the individuals with improving lung function who had similar clinical characteristics to the individuals with declining slopes. However, the goal of this study was to define an individual's slope; future studies predicting an individual's lung function rate of change are warranted with inclusion of these characteristics. Finally, requiring at least three outpatient PFTs may have biased the population away from individuals who decline rapidly, but at least three PFTs were needed to adequately test simple versus regression methods. Although not the intention of this work, population-based approaches such as mixed effects modelling have been used in estimating lung function trajectories,[40–42] and it is a reasonable approach that optimises prediction by comparing individuals to a population of similar individuals. We did not include such an approach in our analyses, using instead modelling approaches of only an individual's data to offer applicability to clinical settings and avoid the challenge of identifying an appropriate reference population for PFT data.

In summary, we identified complexities in real-world PFT data that should be considered when analysing lung function trajectories. Our results suggest that regression method that allows for informative outliers, interpreted with a relative metric, offers the most consistent and reliable estimates of lung function change over time. Given the benefits of using EHR data to examine clinical decision-making and predictors of lung function decline, our results offer a method to conduct these analyses that align with clinical practice patterns. Together, our findings provide an analytic plan that is well-validated to support the use of EHR lung function trajectories in studies examining clinical outcomes in COPD.

Contributors AB and CC contributed to the conception, design of the study, data analysis, interpretation and preparation of this manuscript. ASG and MCM contributed to data analysis, interpretation and manuscript preparation. NP, AF, AS, RAW and NNH contributed to data collection, data interpretation and revision of the manuscript. All authors reviewed and approved the manuscript prior to submission for publication. AB is the guarantor for this work.

Funding This work is supported by grant funding from the Johns Hopkins inHealth Precision Medicine Initiative and collaboration research funding from Highmark Health/Allegheny Health Network. Additionally AB reports grant funding support from NHLBI K23HL153778.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Johns Hopkins Institutional Review Board (IRB00251249). Minimal risk study with waived consent was approved by the institutional review board.

Provenance and peer review Not commissioned; externally peer-reviewed.

Data availability statement No data are available.

ORCID iD

Aparna Balasubramanian http://orcid.org/0000-0002-1854-6672

REFERENCES

1 Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Br Med J* 1977;1:1645–8.
2 Anthonisen NR, Connett JE, Kiley JP, *et al*. Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. *JAMA* 1994;272:1497–505.
3 Buist AS. The US Lung Health Study. *Respirology* 1997;2:303–7.
4 Agusti A, Faner R. Lung function trajectories in health and disease. *Lancet Respir Med* 2019;7:358–64.
5 Calverley PMA, Anderson JA, Celli B, *et al*. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *N Engl J Med* 2007;356:775–89.
6 Vestbo J, Anderson JA, Brook RD, *et al*. Fluticasone furoate and vilanterol and survival in chronic obstructive pulmonary disease

with heightened cardiovascular risk (SUMMIT): a double-blind randomised controlled trial. *Lancet* 2016;387:1817–26.
7 Lipson DA, Barnhart F, Brealey N, *et al*. Once-Daily Single-Inhaler Triple versus Dual Therapy in Patients with COPD. *N Engl J Med* 2018;378:1671–80.
8 Rabe KF, Martinez FJ, Ferguson GT, *et al*. Triple Inhaled Therapy at Two Glucocorticoid Doses in Moderate-to-Very-Severe COPD. *N Engl J Med* 2020;383:35–48.
9 Celli BR, Thomas NE, Anderson JA, *et al*. Effect of pharmacotherapy on rate of decline of lung function in chronic obstructive pulmonary disease: results from the TORCH study. *Am J Respir Crit Care Med* 2008;178:332–8.
10 Morice AH, Celli B, Kesten S, *et al*. COPD in young patients: a pre-specified analysis of the four-year trial of tiotropium (UPLIFT). *Respir Med* 2010;104:1659–67.
11 Celli BR, Anderson JA, Cowans NJ, *et al*. Pharmacotherapy and Lung Function Decline in Patients with Chronic Obstructive Pulmonary Disease. A Systematic Review. *Am J Respir Crit Care Med* 2021;203:689–98.
12 Vestbo J, Edwards LD, Scanlon PD, *et al*. Changes in forced expiratory volume in 1 second over time in COPD. *N Engl J Med* 2011;365:1184–92.
13 Mannino DM, Davis KJ. Lung function decline and outcomes in an elderly population. *Thorax* 2006;61:472–7.
14 Marott JL, Ingebrigtsen TS, Çolak Y, *et al*. Lung Function Trajectories Leading to Chronic Obstructive Pulmonary Disease as Predictors of Exacerbations and Mortality. *Am J Respir Crit Care Med* 2020;202:210–8.
15 Lange P, Celli B, Agustí A, *et al*. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N Engl J Med* 2015;373:111–22.
16 Ross JC, Castaldi PJ, Cho MH, *et al*. Longitudinal Modeling of Lung Function Trajectories in Smokers with and without Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2018;198:1033–42.
17 Thomas ET, Guppy M, Straus SE, *et al*. Rate of normal lung function decline in ageing adults: a systematic review of prospective cohort studies. *BMJ Open* 2019;9:e028150.
18 Graham BL, Steenbruggen I, Miller MR, *et al*. Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. *Am J Respir Crit Care Med* 2019;200:e70–88.
19 Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 1981;24:381–95.
20 Koenker R, Bassett G. Regression quantiles. *Econometrica* 1978;46:33.
21 Huber PJ. *Robust statistics*. New York, NY: John Wiley & Sons, Ltd, 1981.
22 Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: machine learning in python. In: *machine learning in python*.
23 Seabold S, Perktold J. *Statsmodels: econometric and statistical modeling with python*. Austin, Texas, 2010:92–6.
24 Singh D, Agusti A, Anzueto A, *et al*. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: the GOLD science committee report 2019. *Eur Respir J* 2019;53:1900164.
25 McCormack MC, Bascom R, Brandt M, *et al*. Electronic Health Records and Pulmonary Function Data: Developing an Interoperability Roadmap. An Official American Thoracic Society Workshop Report. *Ann Am Thorac Soc* 2021;18:1–11.
26 Rosso A, Egervall K, Elmståhl S. Annual decline rate in FEV1s in community-dwelling older adults diagnosed with mild to moderate COPD. *NPJ Prim Care Respir Med* 2022;32:30.
27 Boueiz A, Xu Z, Chang Y, *et al*. Machine Learning Prediction of Progression in Forced Expiratory Volume in 1 Second in the COPDGene® Study. *Chronic Obstr Pulm Dis* 2022;9:349–65.
28 Drummond MB, Hansel NN, Connett JE, *et al*. Spirometric predictors of lung function decline and mortality in early chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012;185:1301–6.
29 Tashkin DP, Celli B, Senn S, *et al*. A 4-year trial of tiotropium in chronic obstructive pulmonary disease. *N Engl J Med* 2008;359:1543–54.
30 Anzueto A. Impact of exacerbations on COPD. *Eur Respir Rev* 2010;19:113–8.
31 Dransfield MT, Kunisaki KM, Strand MJ, *et al*. Acute Exacerbations and Lung Function Loss in Smokers with and without Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2017;195:324–30.
32 Kerkhof M, Voorham J, Dorinsky P, *et al*. Association between COPD exacerbations and lung function decline during maintenance therapy. *Thorax* 2020;75:744–53.
33 Zafari Z, Sin DD, Postma DS, *et al*. Individualized prediction of lung-function decline in chronic obstructive pulmonary disease. *CMAJ* 2016;188:1004–11.
34 Nishimura M, Makita H, Nagai K, *et al*. Annual change in pulmonary function and clinical phenotype in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012;185:44–52.
35 Casanova C, de Torres JP, Aguirre-Jaíme A, *et al*. The progression of chronic obstructive pulmonary disease is heterogeneous: the experience of the BODE cohort. *Am J Respir Crit Care Med* 2011;184:1015–21.
36 Peto R. The horse-racing effect. *Lancet* 1981;2:467–8.
37 Burrows B, Knudson RJ, Camilli AE, *et al*. The “horse-racing effect” and predicting decline in forced expiratory volume in one second from screening spirometry. *Am Rev Respir Dis* 1987;135:788–93.
38 Bhatt SP, Soler X, Wang X, *et al*. Association between Functional Small Airway Disease and FEV1 Decline in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2016;194:178–84.
39 Bhatt SP, Han MK. Reply: FEV_1 Decline in Patients with Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2017;195:1677.
40 Chen W, Sin DD, FitzGerald JM, *et al*. An Individualized Prediction Model for Long-term Lung Function Trajectory and Risk of COPD in the General Population. *Chest* 2020;157:547–57.
41 Bui DS, Lodge CJ, Burgess JA, *et al*. Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet Respir Med* 2018;6:535–44.
42 Taylor-Robinson D, Whitehead M, Diderichsen F, *et al*. Understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: a longitudinal study. *Thorax* 2012;67:860–6.