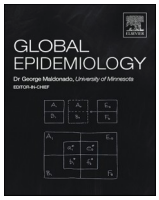
Global Epidemiology 8 (2024) 100165

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology



Estimating effects of aging and disease progression in current and former smokers using longitudinal models

Matthew Strand [a,b,*], Surya Bhatt [c], Matthew Moll [d], David Baraghoshi [a,b]

[a] Division of Biostatistics and Bioinformatics, National Jewish Health, Denver, CO, United States
[b] Department of Biostatistics and Informatics, Colorado School of Public Health, CU Anschutz, Aurora, CO, United States
[c] Lung Health Center, University of Alabama at Birmingham, Birmingham, AL, United States
[d] Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, United States

ARTICLE INFO

Keywords:
Longitudinal
Disease progression
Aging
Mixed models
Observational study
FEV_1

ABSTRACT

Objectives: To separate estimates of mean change in a health outcome into components of aging and disease progression for different severity groups of chronic obstructive pulmonary disease (COPD).
Study design and methods: A longitudinal model can be used to estimate mean change in a health outcome over time. Methods to separate this change into portions due to aging and disease progression are discussed, including conditions that allow for accurate estimation. Linear mixed models were used to estimate these changes for *forced expiratory volume in 1 s* (FEV_1) for various COPD severity and smoking groups using a large cohort (COPDGene) followed for over 10 years.
Results: Based on an analysis of 4967 subjects, age-related loss in FEV_1 was found to be about 1 % per year, consistent with published work. Excess average losses (those beyond natural aging) were significant for all severity groups (except nonsmokers), including those with smoking history but normal lung function. Subjects in higher severity groups tended to have less loss in FEV_1, but more relative loss, compared to baseline averages. Losses in FEV_1 that included both aging and disease progression ranged from 1 to 3 % over severity groups, with current smokers generally exhibiting greater mean losses in FEV_1 than former smokers.
Discussion: Effects of disease progression separate from aging can be estimated in observational studies, although care should be taken in order to make sure assumptions involving this separation are reasonable for a given study. This article demonstrates methods to estimate such effects using temporal changes in lung function for subjects in the COPDGene study.

Introduction

Longitudinal models are powerful tools that help researchers assess effects of progression of illness for observational studies of disease [1]. However, accurately estimating disease progression involves teasing out effects of natural aging. Self-selection and dropout inherent in observational studies propose some challenges in estimating effects of interest, although such studies provide real-life data that is relatively easy to obtain on a large scale compared with designed experiments. Longitudinal models will naturally have one or more predictors involving time and/or age. Here, we discuss the nature of such predictors when effects of both aging and disease progression are of interest, with an application involving changes in lung function over time for subjects with chronic obstructive pulmonary disease (COPD).

Forced expiratory volume in 1 s (FEV_1) naturally changes with age and decreases, on average, after approximately 25 years of age. These decreases can be magnified for those with an illness such as chronic pulmonary disease (COPD). Recent reports have estimated changes in FEV_1 over time for COPD [2,3], but with less focus on separation of aging and disease progression. Here we discuss methods to separate effects of aging and disease progression in depth, and then apply these methods to data from a large study of current and former smokers (COPDGene) graded by severity at baseline.

Models using time since disease onset

Fig. 1 shows a simple paradigm for disease progression and natural aging after age of peak health. The mean of the linear progression model

* Corresponding author at: Division of Biostatistics and Bioinformatics, National Jewish Health, United States.
E-mail address: strandm@njhealth.org (M. Strand).

https://doi.org/10.1016/j.gloepi.2024.100165
Received 5 April 2024; Received in revised form 22 August 2024; Accepted 25 September 2024
Available online 29 September 2024

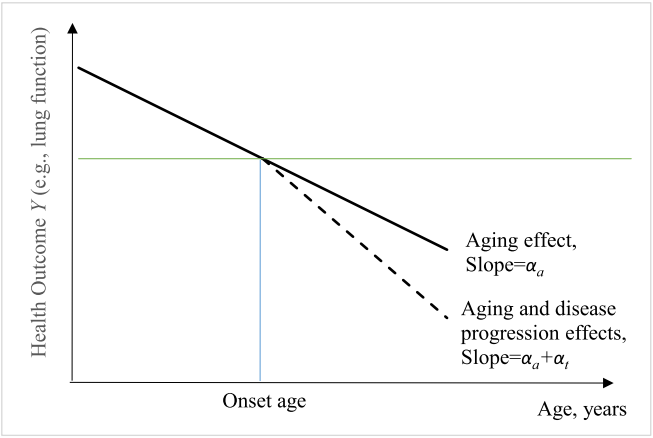


Fig. 1. Change in health outcome Y as a function of age and disease progression (after age of peak health). The solid line shows average progression over time in a population due to natural aging; the dashed line shows additional mean progression due to a particular illness; onset age is where the lines intersect.

for health outcome Y that accounts for both aging and disease progression can be expressed as

$$E(Y) = \begin{cases} \alpha_0 + \alpha_a Age & \text{when } Age < Onset\ age \\ \alpha_0 + \alpha_a Age + \alpha_t(Age - Onset\ age) & \text{when } Age \geq Onset\ age \end{cases} \quad (1)$$

In (1), (*Age – Onset age*) can be replaced with *Time since onset*; disease starts at *Time since onset* = 0 and the term with α_t kicks in when *Time since onset* ≥ 0. When time of disease onset is known, aging and disease progression effects can be estimated by including *Age* and *Time since onset*, both time varying, as predictors in the longitudinal model for *Y*. This piecewise regression model [4] can be fit by setting pre-disease times to 0 for the *Time since onset* variable. Outcome variables that are approximately normally distributed (after transformation, if necessary) can be fit using linear mixed models [5], which can easily be generalized for non-normal outcomes by employing generalized linear (mixed) models [6]. Repeated measures on subjects can be handled in mixed models by including either subject-level random effects and/or a non-simple error covariance structure. Model (1) can be generalized to account for nonlinearity. As an example, Supplement A includes a quadratic disease progression paradigm that has a smoother transition from healthy to disease states. Here, we primarily focus on the simpler case to illustrate the principles.

Time in study

In many observational studies, *Time since onset* is not known and *Time in study* is used instead. Including *Baseline age* and *Time in study* as predictors will allow for estimation of effects involving between-subject age differences and within-subject changes over time during the study, respectively [7,8], and possibly answer questions regarding aging versus disease progression. However, if only (time varying) *Age* is included, the effect being estimated is a weighted average of within- and between-subject effects. Morrell et al. argued that separation of effects might provide more accurate results since many observational studies have an age-dependent bias, e.g., when a subject just entering the study tends to be healthier than someone of the same age who has been in the study for a few years [9]. Thus, including two time-related predictors may even be helpful in estimating effects in groups for which no disease progression is expected.

Using baseline age versus time-varying age

It may seem more intuitive to include *Baseline age* and *Time in study* as predictors rather than *Age* and *Time in study*. However, the choice depends more on effects of interest. Also, the two approaches differ systematically, as the following equations illustrate, using the principle that *Age* = *Baseline age* + *Time in study*.

$$\begin{aligned} E(Y) &= \beta_0 + \beta_a Age + \beta_t Time\_in\_study \\ &= \beta_0 + \beta_a(Baseline\_age + Time\_in\_study) + \beta_t Time\_in\_study \\ &= \beta_0 + \beta_a Baseline\_age + (\beta_a + \beta_t) Time\_in\_study \end{aligned}$$

(Beta symbols are used here to distinguish it from the case where time since disease onset is known.) The coefficient β_a is the mean between-subject baseline age effect; $\beta_a + \beta_t$ is the (total) mean within-subject time effect, and β_t is the excess mean within-subject effect, i.e., time effects that extend beyond aging effects associated with baseline age. When *Age* and *Time in study* are included as predictors, the coefficients being estimated are β_a and β_t, respectively (e.g., [10,2]), whereas when *Baseline age* and *Time in study* are the predictors (e.g., [11–13]), they are β_a and $(\beta_a + \beta_t)$. It is easy to estimate any combination of these parameters from the same model using customized 'estimate' statements (e.g., in SAS or R). If there is no bias due to the enrollment process or dropout, then the estimate of β_a might accurately depict the aging effect, in which case what is leftover could be due to disease progression (estimate of β_t), but for many observational studies this may not hold. Researchers can see if estimates of β_a are consistent with aging effects reported in other studies. If so, separation of aging and disease progression may hold at least approximately.

Using 'Time in study' as a proxy for 'Time with disease'

Time in study is often somewhat arbitrary compared with a time variable of greater interest, *Time since disease onset*. But in some cases, the former can be used as a proxy for the latter. Mathematically, we are interested in cases where $\beta_a = \alpha_a$, $\beta_t = \alpha_t$ and $\beta_a + \beta_t = \alpha_a + \alpha_t$. These principles were tested by constructing *Time since onset* = *Time in study* + *constant* + *random error* + *f*(*Baseline age*), and fitting longitudinal models using Monte Carlo simulation (see Supplement B). When *f*(*Baseline age*) was set to 0 (*Time since onset* designed to be unrelated to *Baseline age*), no bias occurred in estimating α_a or α_t when using *Time in study* as the proxy; the addition of random error did not impact estimates, as expected, since it was Berkson-type error [14]. When *f*(*Baseline age*) was set to *constant*•(*Age at baseline*) (i.e., *Time since onset* linearly related to *Baseline age,* such as when older subjects entering a study tend to have had a disease for a longer period of time), bias occurred in estimating each of α_a and α_t, but the bias summed to 0 so that estimation of $\alpha_a + \alpha_t$ was still accurate (in absence of selection or dropout issues). The simulations also showed that when there was a hard change at time of disease onset (as in Fig. 1), estimation of parameters was easier when subjects were not changing disease status during the observation period. (A 'change point' was not added in the models being fitted since in practice, time of disease onset is unknown.)

Nonlinear effects

There are various ways that nonlinearity can be accounted for in modeling age and time effects in longitudinal models, including transforming model outcome or predictor variables (e.g., log-linear, linear-log or log-log models [15]), including time-varying predictors or including polynomial terms for time or age [1]. Models with quadratic effects for age and disease progression do allow for smoother, perhaps more intuitive changes between natural aging and disease progression. However, one advantage of using simple linear terms is that it is easier to separate time-varying age into between and within-subject components and express change in terms of slopes, whereas estimates of change in models with higher-order terms depend on age (or time). One may decide not to use the nonlinear model unless it clearly yields a better fit. Nevertheless, there may be cases where using the nonlinear approach is clearly necessary.

Selection bias and dropout in observational studies

Estimation of effects of interest (including separation of aging and disease progression) can be biased due to unintended selection bias at the beginning of a study, or due to non-ignorable dropout during the study. Some methods for handling *missing at random* (MAR) data may allow for reduction in bias of estimates, such as multiple imputation, inverse weighting, or by including certain predictors and relevant interaction terms in the model [16]. Although more difficult to deal with, data that tend to be *missing not at random* (MNAR) can employ techniques such as pattern mixture models and model selection techniques [16]. Another issue to deal with in many long-term studies of disease progression is how to handle those who die during the study [17]. When applying imputation techniques, it may not make sense to impute data for subjects after they die, particularly when the death is related to the disease being studied. Rather, it can be considered an alternative outcome. In such cases, reporting mortality rates jointly with estimates of progression of outcomes for study groups provides a more complete picture of the data and is particularly useful when comparing results between disease severity groups.

Application

Chronic obstructive pulmonary disease (COPD) is the fourth leading cause of death worldwide, and costs the U.S. health care system nearly $50 billion in direct costs per year [18]. Like other diseases, understanding expected progression of disease for subjects with different levels of severity of COPD is important in determining whom to target for therapy. The COPDGene study [19] was used to evaluate aging and disease progression estimates for *forced expiratory volume in 1 s* (FEV_1) for different severity groups. Groups were defined at baseline using FEV_1/FVC (where FVC = forced vital capacity) and percent of predicted FEV_1 using GOLD criteria [20], as follows: GOLD 0 = $FEV_1 \geq 80$ % of predicted and $FEV_1/FVC \geq 0.7$; PRISm (preserved ratio, impaired spirometry) $FEV_1 < 80$ % and $FEV_1/FVC \geq 0.7$; GOLD 1 through 4 all have $FEV_1/FVC < 0.7$, with $FEV_1 \geq 80$ % (GOLD 1), 50 % $\leq FEV_1 < 80$ % (GOLD 2), 30 % $\leq FEV_1 < 50$ % (GOLD 3) and $FEV_1 < 30$ % (GOLD 4). Baseline demographics are provided in Table 1. Our analysis included subjects with at least two measurements during the study, and subjects with changes in smoking status during the study were excluded.

Raw FEV_1 (in liters) was modeled on the natural log scale (yielding better fits than raw FEV_1 for nearly all groups) as a function of *Age*, *Time in study*, *sex*, *height*, *race* (White or Black), *body mass index* (*BMI*), *severity group* and *severity group*Time in study*. Our paradigm assumes one natural aging effect, and thus an interaction term between *Age* and *severity group* was not included in the model. Enrollment criteria included baseline age of 45 to 80 years and 10 pack-years of smoking history (except for the nonsmokers). Subjects had up to three measurements that were separated by 5-year intervals; sample sizes and number of total measurements are shown in Table 2. Due to use of log outcome, exponentiated fixed effects naturally have relative change interpretations (a linear model with respect to natural log FEV_1). Random intercepts and slopes for *Time in study* were included for subjects to account for repeated measures. The total within-subject effect was determined by including an 'estimate statement' for the same model fit (performed with SAS [v9.4, Cary, NC]). (An example of an estimate statement is given in the SAS macro program in Supplement B.) The between-subject baseline age effect (1.07 % loss per year) was very close to aging effects reported in the literature (e.g., see [21–26]). As expected, the estimate of β_t was not significant for nonsmokers ($p = 0.23$). Given no excess effects are expected for nonsmokers, the total mean within-subject effect ($\widehat{\beta}_a + \widehat{\beta}_t$) could also be used as an estimate for the aging effect from this group.

For other severity groups (see Table 2), given the 10 pack-year smoking history criteria for enrollment, some disease progression is expected for most if not all study subjects, even if minor, although the standard criteria for COPD is $FEV_1/FVC < 0.7$. Table 2 shows that FEV_1 tended to decline more with increasing severity, with GOLD 3 exhibiting the largest changes. The excess progression diminished for the GOLD 4 (highest severity) subjects, although approximately half of these subjects died before a P2 visit would have occurred (mortality rates at 5 and 11 years across severity groups for subjects started the study were 5 and

Table 1
Mean baseline characteristics of subjects, by baseline severity. Entries are mean (SD) unless otherwise noted.

Variable	NS ($n = 279$)	GOLD 0 (n = 2250)	PRISm ($n = 564$)	GOLD 1 ($n = 430$)	GOLD 2 ($n = 899$)	GOLD 3 ($n = 444$)	GOLD 4 ($n = 101$)
Height, cm	169.5 (9.4)	169.7 (9.3)	170.3 (9.5)	170.0 (9.6)	170.4 (9.6)	169.4 (9.9)	170.5 (9.2)
Female	56 %	51 %	57 %	45 %	47 %	46 %	41 %
White	87 %	67 %	64 %	81 %	77 %	82 %	87 %
FEV_1, liters	3.02 (0.76)	2.85 (0.67)	2.06 (0.48)	2.63 (0.65)	1.89 (0.51)	1.16 (0.29)	0.70 (0.17)
FEV_1/FVC	0.80 (0.05)	0.78 (0.05)	0.76 (0.05)	0.65 (0.04)	0.59 (0.08)	0.45 (0.09)	0.31 (0.05)
Age, years	60.2 (9.5)	58.1 (8.6)	58.4 (8.5)	63.0 (8.4)	62.9 (8.5)	64.2 (7.8)	64.0 (7.4)
Current smoker	0 %	44 %	50 %	41 %	37 %	21 %	10 %

Severity groups were defined at baseline using FEV_1/FVC and percent of predicted FEV_1as follows: GOLD 0 = $FEV_1 \geq 80$ % of predicted and $FEV_1/FVC \geq 0.7$; PRISm (preserved ratio, impaired spirometry) $FEV_1 < 80$ % and $FEV_1/FVC \geq 0.7$; GOLD 1 through 4 all have $FEV_1/FVC < 0.7$, with $FEV_1 \geq 80$ % (GOLD 1), 50 % $\leq FEV_1 < 80$ % (GOLD 2), 30 % $\leq FEV_1 < 50$ % (GOLD 3) and $FEV_1 < 30$ % (GOLD 4); NS = nonsmokers.

Table 2
Estimates of relative change per year for FEV_1, %, for COPD subjects graded by severity at baseline, with 95 % confidence intervals in parentheses unless otherwise noted, based on a linear mixed model fit (see text for model details). Estimates apply to subjects with fixed height and BMI over time.

Group*	Sample size (n), records (r)**	Baseline Mean FEV_1, liters (SD)	$100(exp(\widehat{\beta}_a)-1)$ (Between-subject baseline effect), % change per year	$100(exp(\widehat{\beta}_t)-1)$ (Within-subject excess effect), % change per year	$100(exp(\widehat{\beta}_a+\widehat{\beta}_t)-1)$ (Total within-subject effect), % change per year
Nonsmokers	n = 279, $r = 606$	3.02 (0.76)	−1.07 (−1.11, −1.02) (average for all groups)	−0.18 (−0.46, 0.11)	−1.24 (−1.52, −0.96)
GOLD 0	$n = 2250$, $r = 5717$	2.85 (0.67)		−0.51 (−0.60, −0.41)	−1.57 (−1.65, −1.49)
PRISm	$n = 564$, $r = 1382$	2.06 (0.48)		−0.41 (−0.59, −0.23)	−1.47 (−1.64, −1.30)
GOLD 1	n = 430, $r = 1092$	2.63 (0.65)		−1.16 (−1.35, −0.96)	−2.21 (−2.40, −2.02)
GOLD 2	$n = 899$, $r = 2193$	1.89 (0.51)		−1.55 (−1.70, −1.41)	−2.60 (−2.74, −2.46)
GOLD 3	$n = 444$, $r = 1051$	1.16 (0.29)		−2.07 (−2.28, −1.86)	−3.12 (−3.32, −2.91)
GOLD 4	n = 101, $r = 234$	0.70 (0.17)		−1.08 (−1.53, −0.63)	−2.13 (−2.57, −1.69)

* Severity groups are defined as in Table 1.
** One model was fit and interactions included in order to obtain estimates by severity group.

Table 3
Estimates of total within-subject relative change per year for FEV_1, %, for COPD subjects graded by severity at baseline and by smoking status (persistent current or former during study), with 95 % confidence intervals in parentheses, based on linear mixed model fits. Groups were as defined in Table 2; see text for model details.

Group	Current	Former	*p*-value*
GOLD 0	−1.80 (−1.93, −1.66), $n = 1000$	−1.39 (−1.50, −1.29), $n = 1250$	<0.0001
PRISm	−1.63 (−1.89, −1.36), $n = 284$	−1.31 (−1.54, −1.09), $n = 280$	0.08
GOLD 1	−2.73 (−3.06, −2.41), $n = 177$	−1.85 (−2.08, −1.62), $n = 253$	<0.0001
GOLD 2	−3.32 (−3.56, −3.08), $n = 334$	−2.18 (−2.35, −2.02), $n = 565$	<0.0001
GOLD 3	−2.90 (−3.39, −2.40), $n = 93$	−3.17 (−3.38, −2.95), $n = 351$	0.33
GOLD 4	−1.55 (−3.11, 0.03), n = 10	−2.19 (−2.63, −1.74), $n = 91$	0.45

* Comparing Current vs. Former using *t*-tests derived from the linear mixed model fits.

12 % for GOLD 0, 10 and 22 % for PRISm, 6 and 16 % for GOLD 1, 12 and 30 % for GOLD 2, 23 and 50 % for GOLD 3 and 49 and 74 % for GOLD 4). Dropout for reasons other than death could also impact results. Generally, the use of random intercept and slope for subjects is expected to minimize impacts of MAR-type missing data [16]. There was little or no improvement in goodness-of-fit by adding quadratic age/time terms to the models for the most part, perhaps due to the fact that a limited age range was studied (baseline age: range of 45 to 80 years, mean = 60.2 years, SD = 8.9 years, $n = 4967$ never, former or current smokers), and also because slopes for disease progression were allowed to vary by baseline severity group.

A separate model was fit without nonsmokers in order to examine progression by *smoke group* (persistent current or persistent former) in addition to *severity group*. This model contained the same predictors as previously described, plus *smoke group* and all interaction terms involving *smoke group*, *severity group* and *Time in study*. Random intercepts and slopes for *Time in study* for subjects were included, as before, with separate covariance parameters estimated for current and former smokers. Table 3 shows relative total within-subject estimates of change over time by smoking status (persistent current or former) and severity group. Current smokers in GOLD 0, 1 and 2 had significantly greater mean decreases in FEV_1 than former smokers, while a similar, marginally significant difference occurred for PRISm. GOLD 3 and 4 showed no significant mean difference between current and former smokers, however many of these subjects passed away before evaluation of progression could be determined. In addition, mortality rates were higher for Current versus Former smokers across GOLD groups. For example, the 5-year mortality rate was 29 % versus 21 % in Current and Former smokers for GOLD 3, respectively, and 56 % versus 47 % for GOLD 4.

Fig. 2 contains spaghetti plots for (a) nonsmokers and (b) GOLD 2 subjects, overlaid with mean between-subject (BS; blue) and total within-subject (WS; red) functions by age based on estimates from Table 2 for these two groups, averaged over other covariates in the model. As loglinear models were employed, slopes will be curved on the scale in original units; curves were placed to intersect at the mean of FEV_1 and age in each graph. The multiplicative change per year is exp $(\widehat{\beta}_a)$ for the BS function and $exp(\widehat{\beta}_a+\widehat{\beta}_t)$ for the WS function. For the nonsmoking controls (panel a), the curves are very similar and it is expected that they both capture natural aging effects. On the other hand, the total WS slope for GOLD 2 subjects (red line) in panel b is steeper than the BS slope (blue line), indicating an excess WS trend (difference between curves). As the aging effect is similar to that reported in other literature, this excess trend approximately reflects average disease progression, barring other potential sources of bias.

Discussion

Longitudinal observational studies allow for analysis of large amounts of data in real-world settings, but have inherent limitations. Results from the studies can still be quite informative, but should be qualified based on potential biases that may occur. Most notably, subjects are self-selected for participation and may dropout at any time, and thus average progression may not be representative of the population of interest. The biases can impact severity groups differently. For example,

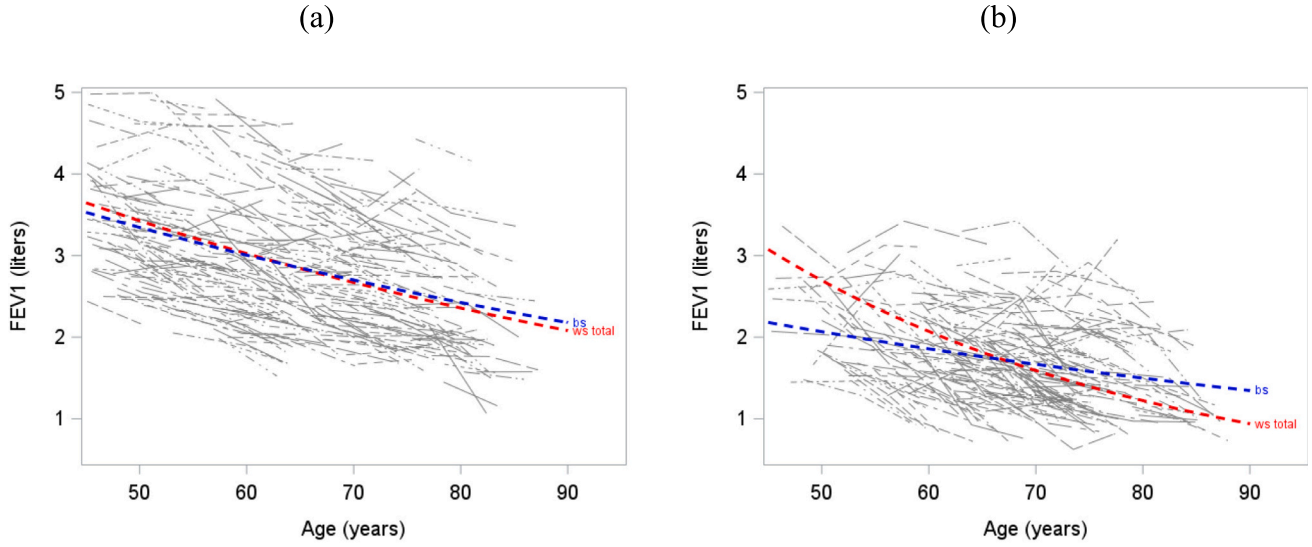


Fig. 2. Demonstration of within and between-subject effects. Spaghetti plots of observed data are shown in gray for (a) nonsmokers (b) GOLD 2 subjects (random sample of 20 %). Overlaid on the plots are the estimated average between-subject age effect (related to $exp(\widehat{\beta}_a)$ [bs] in blue) and the estimated average within-subject change over time (related to $exp(\widehat{\beta}_a+\widehat{\beta}_t)$ [ws total] in red). See text for more detail.

GOLD 4 subjects had notable slowing of FEV_1 change; however, they also by far had the highest mortality rate. Many subjects in this group cannot survive substantial decreases from their already relatively low levels of FEV_1, and thus pass away rather than continuing to progress in illness, leading to an attenuation of average progression for this group. This phenomenon may influence other groups as well, but probably not to the degree observed for GOLD 4.

The estimates shown in Table 2 suggest that changes in FEV_1 are in fact nonlinear if you consider a subject's entire history. For example, many subjects who fell into GOLD 4 at baseline likely at one time had normal spirometry, and thus progressed through earlier stages, even if just for a short time. On average, the earlier stages had less progression than later stages in terms of relative change. Models of absolute change [3] also show that mean change differ by GOLD stage, with highest changes in GOLD 1 or 2. Our attempts at including nonlinear terms generally did not yield improved fits, but this may be due to the fact that severity groups were allowed separate slopes over time, allowing progression to be modeled as approximately linear within GOLD stages. The models used for analysis here also allow for individual variation, as random terms for intercept and slope for time were included for subjects.

Including age and time in study as predictors in a longitudinal model allow for estimation of between- and within-subject time-related effects. Using time-varying age and time in study allow for estimation of between-subject age and excess within-subject effects, respectively, while using baseline age and time in study allow for estimation of between-subject age and total within-subject effects, respectively. These may specifically capture aging and disease progression effects, although for a given observational study, some investigation should be taken in order to determine if this assumption is reasonable. This can be carried out by verifying aging effects for an appropriate control group within the same study or from other reported literature. For the COPD application presented here, the baseline between-subject age effect was similar to natural aging effects in reported literature, suggesting that such a separation of effects was at least approximately reasonable.

Grant support

This work was supported by NHLBI U01 HL089897 and U01 HL089856. The COPDGene study (NCT00608764) is also supported by the COPD Foundation through contributions made to an Industry Advisory Committee comprised of AstraZeneca, Bayer Pharmaceuticals, Boehringer-Ingelheim, Genentech, GlaxoSmithKline, Novartis, Pfizer and Sunovion.

Credit authorship contribution statement

Matthew Strand: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Surya Bhatt:** Writing – review & editing. **Matthew Moll:** Writing – review & editing. **David Baraghoshi:** Writing – review & editing.

Declaration of competing interest

MS and DB: grant support via U01 HL089897, U01 HL089856 and the COPD Foundation.

MM: grant from NHLBI K08HL 159318-01A1; grant or contract from Bayer; consulting fees from Verona pharma, Sitka, TheaHealth, 2ndMD, Axon Advisors.

SB: grants or contracts for R01 HL151421, UH3HL155806, Nuivaira, Sanofi/Regeneron; consulting fees from Sanofi/Regeron, Boehringer Ingelheim, GSK; payment or honoraria from IntegrityCE.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gloepi.2024.100165.

References

[1] Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. John Wiley & Sons; 2012.
[2] Ragland MF, Strand M, Baraghoshi D, Young KA, Kinney GL, Austin E, et al. 10-year follow-up of lung function, respiratory symptoms, and functional capacity in the COPDGene study. Ann Am Thorac Soc 2022 Mar;19(3):381–8.
[3] Dransfield MT, Kunisaki KM, Strand MJ, Anzueto A, Bhatt SP, Bowler RP, et al. Acute exacerbations and lung function loss in smokers with and without chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2017 Feb 1;195(3):324–30.
[4] Naumova EN, Must A, Laird NM. Tutorial in biostatistics: evaluating the impact of 'critical periods' in longitudinal studies of growth using piecewise mixed effects models. Int J Epidemiol 2001 Dec 1;30(6):1332–41.
[5] Verbeke G, Molenberghs G, Verbeke G. Linear mixed models for longitudinal data. New York: Springer; 2000.
[6] McCulloch CE, Searle SR, Neuhaus JM. Generalized, linear, and mixed models. New York: John Wiley & Sons; 2001 Jan.
[7] Neuhaus JM, Kalbfleisch JD. Between-and within-cluster covariate effects in the analysis of clustered data. Biometrics 1998 Jun 1:638–45.
[8] Strand M, Nelson D, Grunwald G. Modeling between-subject differences and within-subject changes for long distance runners by age. J Quant Anal Sports 2018 Jun 27;14(2):81–90.
[9] Morrell CH, Brant LJ, Ferrucci L. Model choice can obscure results in longitudinal studies. J Gerontol Ser A 2009 Feb 1;64(2):215–22.
[10] Mroz MM, Maier LA, Strand M, Silviera L, Newman LS. Beryllium lymphocyte proliferation test surveillance identifies clinically significant beryllium disease. Am J Ind Med 2009 Oct;52(10):762–73.
[11] AlGhatrif M, Strait JB, Morrell CH, Canepa M, Wright J, Elango P, et al. Longitudinal trajectories of arterial stiffness and the role of blood pressure: the Baltimore longitudinal study of aging. Hypertension 2013 Nov;62(5):934–41.
[12] Luong G, Charles ST. Age differences in affective and cardiovascular responses to a negative social interaction: the role of goals, appraisals, and emotion regulation. Dev Psychol 2014 Jul;50(7):1919.
[13] Pearson JD, Morrell CH, Brant LJ, Landis PK, Fleg JL. Age-associated changes in blood pressure in a longitudinal study of healthy men and women. J Gerontol A Biol Sci Med Sci 1997 May 1;52(3):M177–83.
[14] Carroll RJ, Ruppert D, Stefanski LA. Measurement error in nonlinear models. 2nd ed. CRC press; 2006.
[15] Benoit K. Linear regression models with logarithmic transformations. Lond School Econ Lond 2011 Mar 17;22(1):23–36.
[16] Hogan JW, Roy J, Korkontzelou C. Handling drop-out in longitudinal studies. Stat Med 2004 May 15;23(9):1455–97.
[17] Kurland BF, Heagerty PJ. Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. Biostatistics 2005 Apr 1;6(2):241–58.
[18] Press VG, Konetzka RT, White SR. Insights about the economic impact of chronic obstructive pulmonary disease readmissions post implementation of the hospital readmission reduction program. Curr Opin Pulm Med 2018 Mar 1;24(2):138–46.
[19] Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGene) study design. COPD: J Chron Obstruct Pulmon Dis 2011 Feb 1;7(1):32–43.
[20] Patel AR, Patel AR, Singh S, Singh S, Khawaja I. Global initiative for chronic obstructive lung disease: the changes made. Cureus 2019 Jun;11(6).
[21] Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. Am J Respir Crit Care Med 1999 Jan 1;159(1):179–87.
[22] Kuster SP, Kuster D, Schindler C, Rochat MK, Braun J, Held L, et al. Reference equations for lung function screening of healthy never-smoking adults aged 18–80 years. Eur Respir J 2008 Apr 1;31(4):860–8.
[23] Ahmadi-Abhari S, Kaptoge S, Luben RN, Wareham NJ, Khaw KT. Longitudinal association of C-reactive protein and lung function over 13 years: the EPIC-Norfolk study. Am J Epidemiol 2014 Jan 1;179(1):48–56.
[24] Lange P, Parner J, Vestbo J, Schnohr P, Jensen G. A 15-year follow-up study of ventilatory function in adults with asthma. N Engl J Med 1998 Oct 22;339(17):1194–200.
[25] Liao SY, Lin X, Christiani DC. Occupational exposures and longitudinal lung function decline. Am J Ind Med 2015 Jan;58(1):14–20.
[26] Sherman CB, Xu X, Speizer FE, Ferris Jr BG, Weiss ST, Dockery DW. Longitudinal lung function decline in subjects with respiratory Symptoms1 3. Am Rev Respir Dis 1992;146:855–9.