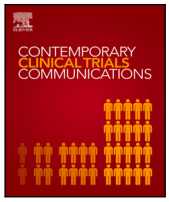
Contemporary Clinical Trials Communications 42 (2024) 101378

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc



Research paper

Baseline-dependent improvement in CF studies, plausibility of bias

Ellen Graham [a], Sonya L. Heltshe [b], Amalia S. Magaret [b,*]

[a] MS University of Washington, Department of Biostatistics, United States of America
[b] University of Washington, Departments of Pediatrics and Biostatistics Seattle Children's Research Institute, Center for Respiratory Biology and Therapeutics, United States of America

ARTICLE INFO

Keywords:
Cystic fibrosis
Single-arm
Clinical trial
Measurement error

ABSTRACT

Background: It has been commonly reported that therapeutic treatments in cystic fibrosis (CF) have ceiling effects, such that their efficacy is diminished for persons with high pre-treatment health (Montgomery et al., 2012 and Newsome et al., 2019). Floor effects have also been reported where decline is of lower magnitude in those with below-average pre-treatment health (Harun et al., 2016; Konstan et al., 2012 and Szczesniak et al., 2017). When measurement error is present, the statistical literature has warned of exaggerated or spurious associations between pre-treatment measures and subsequent change (Chambless and Davis, 2003 and Yanez et al., 1998). Measurement error, equivalently described as day-to-day variation, has been described to occur in CF outcome measurements such as forced expiratory volume in 1 s taken by spirometry (FEV_1pp) (Magaret et al., 2024; Stanojevic et al., 2020 and Thornton et al., 2023).

Methods: We conducted a simulation study to assess the potential for spurious floor or ceiling effects in studies of CF therapeutics. We considered uncontrolled or single-arm studies, and evaluated estimated association between pre-treatment FEV_1pp and treatment-induced change: post-versus pre-treatment.

Results: When day-to-day variation was present in FEV_1pp, at levels equivalent to those reported in large studies measuring spirometry both at home and in clinic, naive analytic approaches found spurious associations of change with baseline (Paynter et al., 2022 and Saiman et al., 2003). Type I error ranged from 31.9% to 98.3% for day-to-day variation as high as 3% to 15% relative to biological variation. Incorporating known day-to-day variation, the regression calibration approach corrected bias and controlled type I error (Chambless and Davis, 2003).

Conclusion: Exaggerated ceiling effects are possible. Further studies could provide meaningful confirmation of ceiling effects in CF, perhaps reducing day-to-day variation by incorporating multiple pre- and post-treatment measurements.

1. Introduction

Many observational and interventional studies in cystic fibrosis (CF) assessed change in continuous measures such as lung function as a primary outcome. In assessing lung function via spirometry as a primary outcome, change in percent predicted forced expiratory volume in 1 s (FEV_1pp) was often used; and it was often stratified by pre-treatment or starting FEV_1pp level in order to allow for the potential of differential benefit based on baseline health: ceiling or floor effects. A ceiling effect would occur when those with high pre-treatment FEV_1pp experience little improvement; a floor effect might occur when those with low pre-treatment FEV_1pp have little subsequent decline.

The presence of ceiling and floor effects have been widely accepted and cited as known phenomena in the realm of cystic fibrosis lung function [1–5]. Stratification on baseline has been a common, anticipated component of statistical analysis plans, often separating the efficacy of the drug on change in FEV_1pp by pre-treatment FEV_1pp above and below 70%. This stratification procedure was done during the licensure randomized clinical trials for ivacaftor, lumacaftor-ivacaftor efficacy, and elexacaftor-tezacaftor-ivacaftor, and was also performed during single-arm, open-label studies extending the indication to younger populations [6–12]. Observational studies and reviews have also tested for and identified ceiling and floor effects [1,3,5].

We have recently demonstrated, in the handful of studies in persons with cystic fibrosis where FEV_1pp has been measured repeatedly over short intervals, that substantial day-to-day variation in lung function is present [13]. The ratio of the day-to-day (or within-person) variability to the biological (or between-person) variability was between 3% and 16%, indicating that within-person variability is small but substantial compared to biological. This variation occurred during periods over

* Corresponding author.
E-mail addresses: graham18@uw.edu (E. Graham), sonya.heltshe@seattlechildrens.org (S.L. Heltshe), amag@uw.edu (A.S. Magaret).

https://doi.org/10.1016/j.conctc.2024.101378
Received 21 June 2024; Received in revised form 13 September 2024; Accepted 25 September 2024
Available online 5 October 2024

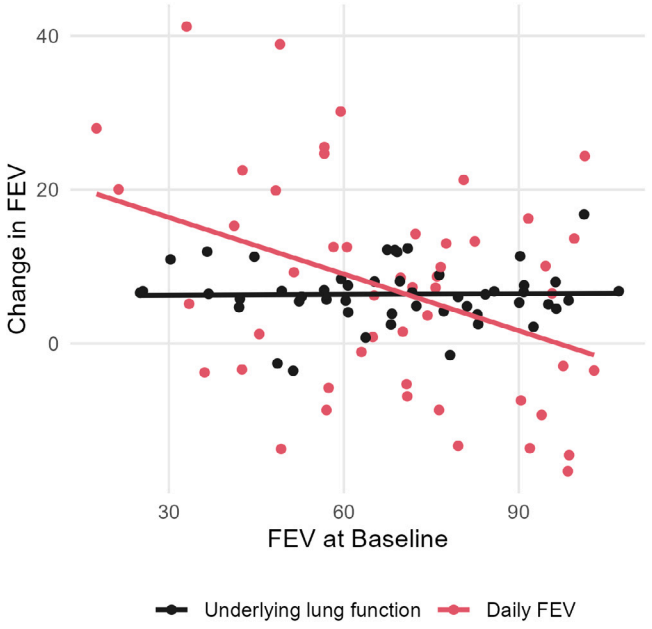


Fig. 1. Random variation induces inverse association of change with baseline. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

which no natural decline was anticipated, and during which no major fluctuations in clinical status (exacerbations) were observed.

In the context of day-to-day variation (or measurement error) the statistical literature has warned of spurious or exaggerated associations between baseline measures and subsequent change (floor or ceiling effects). Yanez et al. in 1998 and Chambless and Davis in 2003 proposed bias correction methods to avoid spurious ceiling effects [14,15].

1.1. Motivating example

We present a visual example of spurious ceiling effects, using simulated data. Data were simulated according to a random slope and intercept model, and parameters were selected using precedents from the e-ICE study and from cystic fibrosis modulator trials. Accordingly, underlying FEV_1pp averaged 70%, with biologic between-person standard deviation of 20% about the population mean [10,16]. On average FEV_1pp increased by 7% between the pre-treatment and follow-up observation, regardless of pre-treatment level, with a standard deviation of the change of 4% [17]. Day-to-day variation was added using standard deviations of 9% about the person-level means.

In Fig. 1, the black dots represented underlying true lung function as simulated, while the red dots represented daily measurements of FEV_1pp which are subject to the imposed random variation. A regression of the outcome of change over 1 month versus true underlying pre-treatment FEV_1pp reflected this true lack of association (black line). But after adding random day-to-day variation with a zero mean, we saw an observed association between change and pre-treatment FEV_1pp such that higher starting values appeared prone to less increase (pink line - an apparent ceiling effect).

With this example in mind, we conducted a simulation study to address the following questions: (1) Under realistic settings for outcomes related to cystic fibrosis, what is the potential to identify apparent spurious or exaggerated ceiling effects? (2) Are existing techniques to correct for the presence of day-to-day variation sufficient to answer questions of interest to the cystic fibrosis community? These questions have been explored in other contexts, such as diabetes, but not yet in cystic fibrosis [18]. We explore these questions using real and simulated data representing plausible measures of lung function in CF. We aimed to quantify potential errors in inference, building on our previous conceptual work [13].

2. Methods

2.1. Regression calibration

We briefly introduce the regression calibration approach of Chambless and Davis, which corrects the relationship between an outcome of interest when covariates may be measured with error. See (1) below, where $\boldsymbol{y}$ is a vector of outcomes for participant i at times $\boldsymbol{t} = (t_1, t_2, \ldots, t_k)'$, $\boldsymbol{z}$ are covariates measured without error and $\boldsymbol{x}$ are covariates measured *with error*.

$$\boldsymbol{y} = A + \boldsymbol{x}B_X + \boldsymbol{z}B_Z + \boldsymbol{\epsilon} \tag{1}$$

If $\boldsymbol{x}$ is measured with error, instead of observing $\boldsymbol{x}$ we observe $\boldsymbol{w} = \boldsymbol{x} + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is the measurement error.

Chambless and Davis demonstrated that substitution of the mismeasured covariate $\boldsymbol{w}$ in place of $\boldsymbol{x}$ in Eq. (1) induced bias in estimates of all coefficients A, B_x, and B_Z, similar to the apparent ceiling effect produced in Fig. 1; and they provided a correction to counter this bias [15]. See Appendix A for the form of the correction.

2.2. Research context

We applied the regression calibration approach to single-arm studies when outcome $\boldsymbol{y}$ was the observed change from baseline and primary covariate $\boldsymbol{w}$ was the observed baseline measurement, which varies from the true underlying $\boldsymbol{x}$. The correction approach required an estimate of response variation from a study in which the outcome was measured frequently, a source likely to be external to the study at hand.

For our purposes, *day-to-day variation* and *measurement error* were synonymous, and could mean any difference between the current measurement and the true, underlying health status. Variation in FEV over short intervals during healthy periods has been observed in CF [13,19,20]. Instrumentation error was one type of variation that is unlikely to be influential in CF. However, day-to-day variation in FEV_1pp about a stable *steady state* akin to a monthly average was anticipated and ubiquitous in CF due to fatigue, effort, time of day etc.

We applied the regression calibration approach to real and simulated studies of change, as change adjusted for baseline was a frequent analytic approach in randomized treatment trials [9,10], and also in single-arm studies, such as those expanding indications to younger age groups [8,11,12]. We focused on the potential bias in the single-arm setting, as bias in the estimate of change affects both arms of randomized trials and has low impact on estimates of differences between arms.

2.3. Simulation procedures

2.3.1. Data simulation model

First in simulation, we produced FEV_1pp measurements $\boldsymbol{u}$ at time points t using expression (2), a linear mixed effects model.

$$u_{it} = \beta_0 + \beta_1 * t + \beta_2 * Z_i + \beta_3 * Z_i * t + b_{0i} + b_{1i} * t + \epsilon_{it} \tag{2}$$

Here Z was a binary covariate such as the presence of a genetic mutation or previous history of a particular lung pathogen, t was time since treatment initiation, β_1 was the change for each increase in t, β_2 described potential associations of the outcome with Z, and β_3 the interaction of Z and t. Zs were simulated as independent Bernoulli(0.5) random variables, and $\epsilon \sim N(0, \sigma^2_\epsilon)$ represented the difference between the measured response and the underlying steady state. Biological variability was induced by including random intercepts b_{0i} and random slopes b_{1i}, where σ^2_{bio} was biological or between-participant variability,

Table 1
Simulation parameters.

Parameter	Values	Description and justification
β_0	70	Average FEV_1pp [10]
β_1	2.5	Improvement FEV_1pp in with treatment [10]
β_2	0 or 10	Difference in FEV_1pp for baseline covariate Z
β_3	0 or 1	Interaction of treatment and the baseline covariate
B_X	0 or −0.1	Ceiling effect: association of change with pre-treatment level
t	$(0, 1, 2, 4)'$	Times when FEV_1pp measured
σ_{bio}	20	Biological (between-person) standard deviation of FEV_1pp [24]
σ_δ	4	Standard deviation of change in FEV_1pp [10]
$\frac{\sigma_\epsilon^2}{\sigma_{\text{bio}}^2}$	0.03, 0.08, 0.15	Day-to-day variation of FEV_1pp as a ratio to biological [16,25]

and σ_δ^2 was variability in the change with t time. The random effects were simulated to be correlated using a bivariate normal:

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} \sigma_{\text{bio}}^2 & \sigma_{\text{bio}}^2 B_X \\ \sigma_{\text{bio}}^2 B_X & \sigma_\delta^2 + \sigma_{\text{bio}}^2 B_X^2 \end{pmatrix}\right) \quad (3)$$

The B_X in the above covariance matrix was the same B_X as in Eq. (1). When $B_X < 0$ a ceiling effect was created, where participants whose pre-treatment value was higher than average ($b_{0i} > 0$) tend to have subsequent follow-up values lower than average ($b_{1i} < 0$).

Note that there was no indicator for treatment arm, as we simulated single-arm studies in which all participants received the intervention.

2.3.2. Parameter values and their origins

We drew parameter values from clinical studies of people with cystic fibrosis. For example, with $\beta_1 = 2.5$, we saw a similar average improvement over four time intervals (10% at $t = 4$) to the 13.6% FEV_1pp improvement seen in Middelton et al. [10]. The degree of day-to-day variation $\frac{\sigma_\epsilon^2}{\sigma_{\text{bio}}^2}$ was chosen to represent best-case and worst-case from the AZM-0001, GOAL, PROMISE and eICE studies, including home and clinic spirometry [13,17,21,22]. These studies were conducted by the Therapeutics Development Network; and repeated, high-quality grade FEV1pp were obtained from home spirometry using protocolized procedures (best of three efforts) [23]. AZM-0001 and eICE were randomized trials of interventions around lung infection, and GOAL and PROMISE were/are observational studies assessing lung function following initiation of therapeutic modulator therapy.

We simulated data using all combinations of the values in Table 1, resulting in 24 different simulation settings. Sample size (n) for each scenario was determined by targeting 90% power with a 5% type 1 error rate to detect the interaction of change with pre-treatment level, under the assumption of no measurement error.

2.3.3. Scenarios to which simulations were applied

We considered 3 scenarios for estimating the response *change over time* of continuous measurements like FEV. The FEV measures were simulated using expressions (2) and (3). In each of the scenarios below, the models were fit both with (corrected) and without (naive) the regression calibration approach.

Model 1: Immediate Change, Mixed Effects Model: All of the change in FEV_1pp occurred between the time of treatment initiation and first follow-up measurement. When simulating data under this model, t in Eq. (2) was replaced with $I(t > 0)$. We then fit linear mixed effects models using change in FEV_1pp as the outcome and retaining pre-treatment FEV_1pp and binary characteristics Z as covariates:

$$u_{it} - u_{i0} = \gamma_0 + \gamma_1 u_{i0} + \gamma_2 Z_i + d_i \quad (4)$$

with d_i the individual-level random effect. The parameter γ_1 was of primary interest, as it associated change with pre-treatment level. As shown in Appendix B, in most cases $\gamma_1 \equiv B_X$ and $\gamma_2 \equiv \beta_3$. Recall that B_X and β_3 described association of change with pre-treatment level, and of change with covariates, in data generation steps (2) and (3).

Model 2: Linear Time Trend, Mixed Effects Model: In this setting FEV_1pp was anticipated to change linearly with time. Under this model, data were simulated from Eq. (2), and modeled using a similar form to Eq. (4) but scaled by time to estimate a single rate of change.

$$\frac{u_{it} - u_{i0}}{t} = \gamma_0 + \gamma_1 u_{i0} + \gamma_2 Z_i + d_i \quad (5)$$

Model 3: Linear Time Trend, Participant Level Regression: As an alternative to a linear mixed effects model, one may prefer to use a two-step linear regression procedure. First, the subject-specific linear model $u_{it} = \alpha_{i0} + \alpha_{i1} t$ was fit. Then the estimated slope was used as the response in simple linear regression, with no random effects:

$$\hat{\alpha}_{i1} = \gamma_0 + \gamma_1 u_{i0} + \gamma_2 Z_i \quad (6)$$

.

2.3.4. Analytic approach

Using both a naive approach (which performs regression on the mismeasured pre-treatment FEV without correction), and the regression calibration approach, we estimated operating characteristics over the 1000 simulated datasets per scenario. 1000 iterations were used to achieve 1.5% monte carlo standard error in computing power and type I error. We particularly focused on estimation of the association between pre-treatment level and change.

3. Results

3.1. Association of change with pre-treatment level, no additional covariates

3.1.1. Findings from naive model

Under study-based, plausible ranges for simulated measures such as treatment efficacy and between and within-person variability, bias in the magnitude of association between pre-treatment level and change ranged between 14% and 118%. The naive approach of estimating the association between pre-treatment level and change over time exaggerated B_X even for relatively small variation $\sigma_\epsilon^2/\sigma_{\text{bio}}^2$ (Table 2). For example, from the immediate change model results, when the ratio of error variance to between person variance was only 0.03 and there was no true relationship between pre-treatment level and change ($B_X = 0$), the naive model estimated $\hat{B}_X = -0.029$. Interpretation: for a participant whose true pre-treatment FEV_1pp was 40% above average, this estimate translated to a predicted FEV_1pp increase of 1.3%, for an absolute (relative) bias of −1.2% (−46%). In reality, all patients improved on average 2.5% FEV_1pp regardless of their true baseline FEV_1pp. As the relative error variance increased, the bias also increased such that at $\sigma_\epsilon^2/\sigma_{\text{bio}}^2 = 15\%$, the naive model predicted a FEV_1pp *decrease* of 2.6%, for an absolute bias of −5.1%, completely overturning the true improvement of 2.5%.

We note that for consistency we simulated an average change of $\beta_1 = 2.5\%$ per time point in all models. In the immediate change model (only), this is a smaller increase than what was observed in clinical trials (10% to 13% increase). In brief sensitivity analyses, a change of

Table 2
Measuring association of pre-treatment FEV_1pp with change over time B_X when other factors are fixed at $\beta_1 = 2.5$, $\beta_2 = \beta_3 = 0$.

Method	B_X	True value	$\frac{\sigma^2_\epsilon}{\sigma^2_{bio}}$	n	Bias	% Bias	MSE	Coverage	Rej
Immediate change model:									
Naive	0.00	0.00	0.03	200	−0.028	N/A	0.001	0.681	0.319
Naive	0.00	0.00	0.08	200	−0.074	N/A	0.006	0.178	0.822
Naive	0.00	0.00	0.15	200	−0.129	N/A	0.017	0.017	0.983
Adj.	0.00	0.00	0.03	200	0.001	N/A	0.000	0.944	0.056
Adj.	0.00	0.00	0.08	200	0.001	N/A	0.000	0.948	0.052
Adj.	0.00	0.00	0.15	200	0.004	N/A	0.000	0.909	0.091
Naive	−0.10	−0.10	0.03	48	−0.024	−24.4%	0.001	0.896	0.882
Naive	−0.10	−0.10	0.08	48	−0.066	−65.5%	0.004	0.734	0.900
Naive	−0.10	−0.10	0.15	48	−0.118	−118%	0.014	0.480	0.933
Adj.	−0.10	−0.10	0.03	48	0.003	2.9%	0.000	0.943	0.659
Adj.	−0.10	−0.10	0.08	48	0.005	5.4%	0.000	0.926	0.453
Adj.	−0.10	−0.10	0.15	48	0.008	7.5%	0.000	0.931	0.319
Linear time trend model:									
Naive	0.00	0.00	0.03	200	−0.018	N/A	0.000	0.812	0.188
Naive	0.00	0.00	0.08	200	−0.044	N/A	0.002	0.362	0.638
Naive	0.00	0.00	0.15	200	−0.076	N/A	0.006	0.071	0.929
Adj.	0.00	0.00	0.03	200	0.000	N/A	0.000	0.953	0.047
Adj.	0.00	0.00	0.08	200	0.000	N/A	0.000	0.939	0.061
Adj.	0.00	0.00	0.15	200	0.001	N/A	0.000	0.916	0.084
Naive	−0.10	−0.10	0.03	48	−0.014	−14.1%	0.000	0.919	0.940
Naive	−0.10	−0.10	0.08	48	−0.034	−34.4%	0.001	0.853	0.939
Naive	−0.10	−0.10	0.15	48	−0.065	−64.6%	0.004	0.644	0.965
Adj.	−0.10	−0.10	0.03	48	0.001	0.9%	0.000	0.948	0.834
Adj.	−0.10	−0.10	0.08	48	0.004	4.0%	0.000	0.933	0.668
Adj.	−0.10	−0.10	0.15	48	0.003	3.3%	0.000	0.924	0.543
Participant level regression model:									
Naive	0.00	0.00	0.03	200	−0.006	N/A	0.000	0.934	0.066
Naive	0.00	0.00	0.08	200	−0.015	N/A	0.000	0.810	0.190
Naive	0.00	0.00	0.15	200	−0.026	N/A	0.001	0.632	0.368
Adj.	0.00	0.00	0.03	200	0.000	N/A	0.000	0.949	0.051
Adj.	0.00	0.00	0.08	200	0.000	N/A	0.000	0.940	0.060
Adj.	0.00	0.00	0.15	200	0.000	N/A	0.000	0.952	0.048
Naive	−0.10	−0.10	0.03	48	−0.003	−3.1%	0.000	0.955	0.925
Naive	−0.10	−0.10	0.08	48	−0.007	−6.9%	0.000	0.938	0.926
Naive	−0.10	−0.10	0.15	48	−0.014	−13.8%	0.000	0.917	0.939
Adj.	−0.10	−0.10	0.03	48	0.000	0.0%	0.000	0.957	0.883
Adj.	−0.10	−0.10	0.08	48	0.001	1.1%	0.000	0.947	0.829
Adj.	−0.10	−0.10	0.15	48	0.000	0.2%	0.000	0.953	0.740

Bias = average distance between estimate and the parameter, MSE (mean squared error) = average squared distance between the estimate and the parameter, coverage = probability the true value was included in the confidence interval, and Rej (rejection percentage) = probability estimated p-value < 0.05, which could be either power (when the parameter is non-zero) or type I error (when the parameter equals zero). Note when $B_X = -0.10$ a true ceiling effect existed.

$\beta_1 = 10\%$ in the immediate change model results in similar percent bias (though larger absolute bias) with similar coverage and power.

Exaggerated associations were also observed when there were true, negative associations with pre-treatment level and change. With true $B_X = -0.10$, a participant with a pre-treatment FEV_1pp 40% above average should have a true change of 4.0% FEV_1pp less than the 2.5 average, so a predicted decrease of 1.5% (= 2.5% - 40*.10). But without accounting for the relative day-to-day variation of $\sigma^2_\epsilon/\sigma^2_{bio} = 15\%$, the naive approach estimated a larger predicted decrease of 6.2% (= 2.5% - 40*.218).

Using naive regression, coverage decreased with increasing day-to-day variation, and spurious significance rates increased. Type I error was as high as 0.983 in the presence of day-to-day variation of $\sigma^2_\epsilon/\sigma^2_{bio} = 15\%$, the same magnitude described for home spirometry from the eICE study [16].

Exaggerated and spurious associations between pre-treatment level and change persisted, though to a lesser degree, in both the linear time trend setting and the participant-level regression setting (lower sections, Table 2). Specifically, in participant-level regression, which displayed the lowest bias and false positive rates across a range of simulation settings, when relative day-to-day variation was moderate ($\sigma^2_\epsilon/\sigma^2_{bio} = 8\%$) or large ($\sigma^2_\epsilon/\sigma^2_{bio} = 15\%$) and there was no true floor/ceiling effect ($B_X = 0$), we saw increased rates of false significance at 19.0% and 36.8% respectively, which would incorrectly lead researchers to believe there is an association between pre-treatment FEV_1pp and change when in fact there is none.

3.1.2. Correction through the regression calibration approach

The regression calibration (corrected) model accurately identified the lack of association of pre-treatment FEV_1pp with change, controlling bias <10% in all cases, (<5% in most cases), and increased only modestly with increasing relative day-to-day variation.

Similarly using regression calibration, type I error was controlled at under 8%, and coverage was above 91%. In the linear time trend and participant-level regression settings (lower sections of Table 2), most of these values were even closer to the 5% and 95% targets. Power, or the rejection probability, decreased with increasing day-to-day variation, as variation naturally increased imprecision.

3.2. Association of change with pre-treatment FEV_1pp when other covariates present

To examine the impact of other covariates on the estimation in the presence of day-to-day variation, we included a non-zero coefficient ($\beta_3 = 1$) for the interaction between time and Z and let the main effect for Z be $\beta_2 = 10$. This means that FEV_1pp was 10% higher for those with $Z = 1$, and increased 1% faster over time. Fig. 2 focused on bias, and showed estimation of all parameters γ_0, γ_1 and γ_2 from Eqs. (4), (5) and (6). The magnitude of bias in association of change with pre-treatment FEV_1pp ($\gamma_1 \equiv B_X$) using the naive approach was just about identical to that found when no covariates were present (Table 2), indicating a persistent inflation of any ceiling effect.

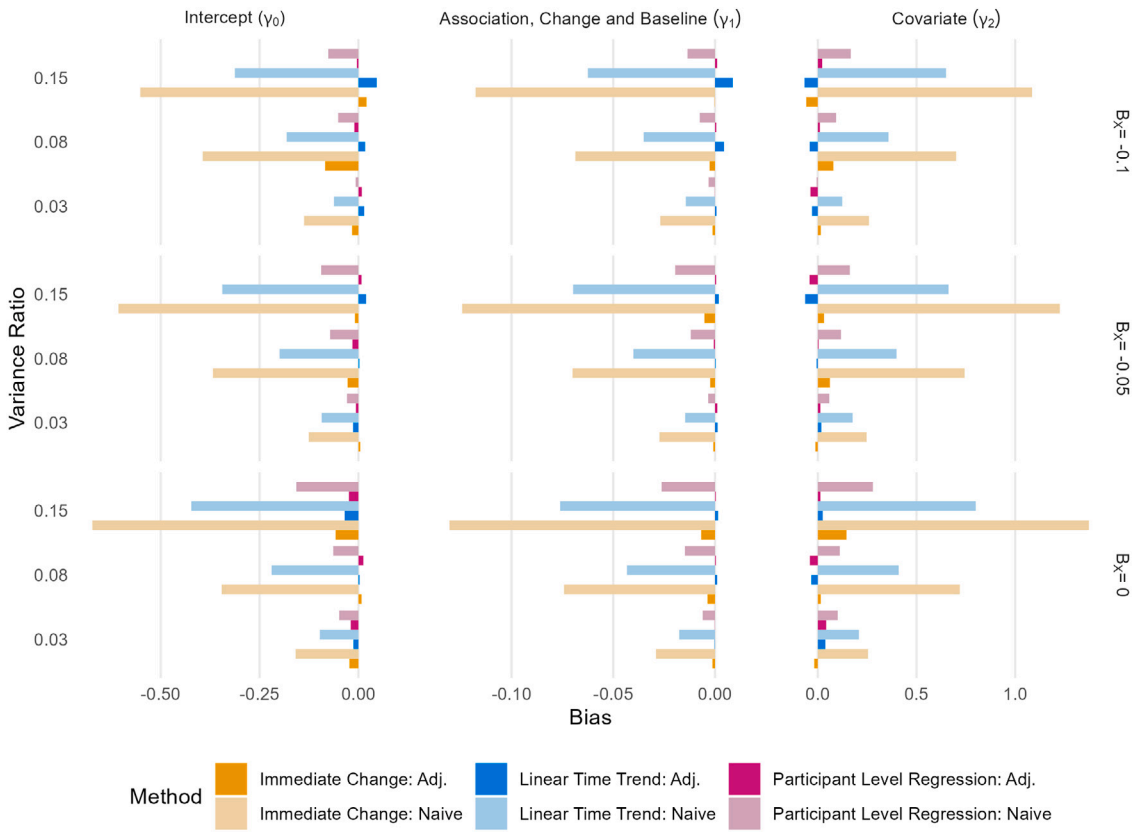


Fig. 2. Bias in estimating all parameters in the model in Eq. (1). γ_0, γ_1 and γ_2 refer to the coefficients in Eqs. (4), (5) and (6). The expected values are derived in Appendix B. Here, $\beta_1 = 2.5$ $\beta_2 = 10$, $\beta_3 = 1$.

Importantly for interpretation of clinical trial results, there was also bias in the intercept (γ_0) which estimates average change, or here, estimates the efficacy of the treatment under evaluation. Chambless showed that bias in the intercept was expected in the presence of measurement error, and could be quantified as $\gamma_0^* - \gamma_0 = (u_0\ Z)^T(I-\Sigma)\boldsymbol{\gamma}$, where Σ was a covariance matrix [15]. Bias in the above expression would equal zero if there was no measurement error, or if the expectations of both u_0 and Z are zero. Therefore, centering pre-treatment FEV_1pp and covariates could remove bias in the intercept, or average change. (In brief simulations not shown, we created a continuous, centered, covariate Z and showed that bias in γ_0 was close to zero even for naive regression.) A tailored expression for bias in all elements of γ was described in Appendix B.

While the naive approach had substantial bias in most settings we examined, the regression calibration correction consistently estimated all terms in the regression, including average change (γ_0) and association of change with the covariate Z (γ_2).

3.3. Dichotomized pre-treatment FEV_1pp

In practice, it has been common to analyze change versus categorical pre-treatment FEV. While the association of change with pre-treatment FEV_1pp was simulated linearly, pre-treatment FEV_1pp was entered into regression models using a threshold. This was done to estimate two distinct levels of change: among those with pre-treatment levels above 70% versus among those below. A FEV_1pp threshold of 70% was chosen as this is the mean FEV_1pp in our simulation study; and this threshold has also been a common choice in clinical studies in CF. The expectation of B_X was derived in Appendix B.

Just as in the previous section, bias in B_X was lowest for the participant-level regression model compared to the other two scenarios (results not shown). Table 3 showed that the naive analysis against

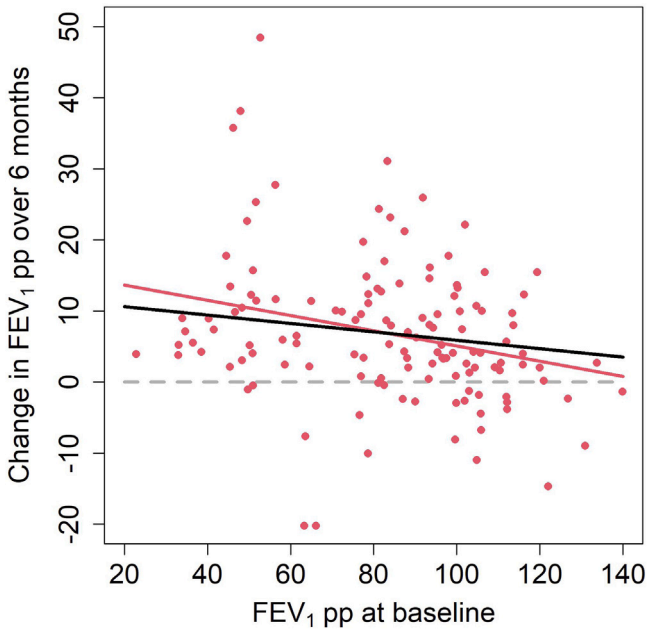


Fig. 3. Change in FEV_1pp over 6 months on ivacaftor versus pre-treatment FEV_1pp in GOAL study participants. The gray dashed line indicated no improvement. The pink solid line showed the linear regression fit assuming no measurement error, and the black solid line showed the fit assuming day-to-day variation at 5% of biological variation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Simulation results of estimating the association between dichotomized pre-treatment FEV_1pp measurement and change from pre-treatment FEV_1pp using the participant level regression model with parameters $\beta_1 = 2.5$, $\beta_2 = \beta_3 = 0$.

Method	B_X	True value	$\frac{\sigma^2_\epsilon}{\sigma^2_{bio}}$	n	Bias	% Bias	MSE	Coverage	Rej
Naive	0.00	0.000	0.03	200	−0.173	N/A	0.030	0.942	0.058
Naive	0.00	0.000	0.08	200	−0.504	N/A	0.254	0.894	0.106
Naive	0.00	0.000	0.15	200	−0.938	N/A	0.880	0.719	0.281
Adj.	0.00	0.000	0.03	200	0.017	N/A	0.000	0.938	0.062
Adj.	0.00	0.000	0.08	200	−0.009	N/A	0.000	0.949	0.051
Adj.	0.00	0.000	0.15	200	−0.040	N/A	0.002	0.956	0.044
Naive	−0.10	−3.192	0.03	48	−0.089	−2.8%	0.008	0.941	0.732
Naive	−0.10	−3.192	0.08	48	−0.399	−12.5%	0.159	0.945	0.761
Naive	−0.10	−3.192	0.15	48	−0.650	−20.4%	0.422	0.930	0.768
Adj.	−0.10	−3.192	0.03	48	0.106	3.3%	0.011	0.938	0.679
Adj.	−0.10	−3.192	0.08	48	0.101	3.2%	0.010	0.948	0.653
Adj.	−0.10	−3.192	0.15	48	0.266	8.3%	0.071	0.947	0.547

Bias = average distance between estimate and the parameter, MSE (mean squared error) = average squared distance between the estimate and the parameter, coverage = probability the true value was included in the confidence interval, and Rej (rejection percentage) = probability estimated p-value < 0.05, which could be either power (when the parameter is non-zero) or type I error (when the parameter equals zero). Note when $B_X = -0.10$ a true ceiling effect existed.

dichotomized pre-treatment FEV_1pp had a type I error rate of 28.1% when $B_X = 0$ and $\sigma^2_\epsilon/\sigma^2_{bio} = 15\%$. As when FEV_1pp was considered continuously, the corrected model reduced bias to under 8% and controlled type 1 error under 6%.

3.4. Re-analysis of GOAL study

We then estimated association of the change in FEV_1pp with pre-treatment FEV_1pp using real study data. GOAL was selected as it is one of two completed modulator studies for which we have access to data. Briefly, GOAL was a single-arm study of persons with CF and the G551D mutation who were age 6 and older, in which mechanisms and physiologic implications of ivacaftor were explored [17]. Ivacaftor had previously been approved as the first effective treatment to modulate (improve) the core genetic defect of CF [6]. Of 153 persons enrolled, 131 had FEV_1pp measured at 6 months. Initial FEV_1pp averaged 82.6% pre-ivacaftor and 90.1% at 6 months post-initiation of treatment, with a mean improvement of 6.7%.

Using naive regression to examine associations of change with pre-treatment FEV_1pp, change in FEV_1pp with ivacaftor was predicted to be −1.1% lower for each 10% increase in pre-treatment FEV_1pp (p=0.0018, Fig. 3). Persons with pre-treatment FEV_1pp of 40% had an expected FEV increase of 11.5%, while those with higher FEV_1pp (120%) had an expected increase of only 2.9%. After implementing regression calibration assuming 3%, 4%, and 5% relative day-to-day variation ($\frac{\sigma^2_\epsilon}{\sigma^2_{bio}}$), the significance of the association decreased to p = 0.025, p = 0.051, and p = 0.10, respectively. At 5% the magnitude of association was cut in half at −0.6% lower (versus −1.1%) for each 10% increase in pre-treatment level, and was non-significantly different from zero (black line in Fig. 3).

We could not estimate day-to-day variation in FEV_1pp directly from GOAL participants, as spirometry was performed only 3 times, and not proximally. However, our findings indicated that relative low day-to-day variation would impact conclusions regarding apparent ceiling effects. From Magaret [13] examining four recent studies of spirometry, relative day-to-day variation in FEV_1pp ($\frac{\sigma^2_\epsilon}{\sigma^2_{bio}}$) was 3% to 16% as high as between-person or biological variation.

Of interest, in all models the estimate for average change in FEV_1pp with ivacaftor (6.9%) matched closely what was found in the trial (6.7%). After including an additional covariate, the presence of the CFTR mutation F508del, this average efficacy estimate remained quite similar at 6.8% overall, at 5.0% in those with no F508, and 7.6% in those with at least one copy of F508. This apparent larger FEV improvement with F508del was not significant in any model (p ≈ 0.2).

4. Discussion

We presented simulations displaying the effect of day-to-day variation on analyses of change from pre-treatment level that were applicable to the setting of *single-arm or uncontrolled* studies within cystic fibrosis. Day-to-day variation was shown to introduce spurious and exaggerated associations between pre-treatment level and subsequent change, using standard modeling techniques. Bias occurred when pre-treatment level was evaluated continuously or when dichotomized. Implementing techniques to account for this variation, such as the regression calibration technique of Chambless et al. [15] was sufficient to correct for these biases and return to nominal type 1 error control. These findings were not applicable to randomized or controlled studies, in which the outcome is not change over time but *difference in change* over time between arms.

While these simulations were designed to represent realistic behavior of FEV_1pp trajectories in CF patients, the results are applicable to other outcomes such as a change in sweat chloride levels, CFQR score, or any analyses where the change from some continuous pre-treatment level value is of interest and where single measurements (subject to day-to-day variation) are taken. Assessment of the magnitude of the day-to-day variation relative to the biological or between-person variation can serve as a guide in determining whether correction approaches should be used. The relative magnitude of day-to-day variation should continue to be assessed as baseline health improves on modulators, and as instruction and management of home spirometry improves its accuracy.

Limitations of this work included focusing our simulation on a single outcome of FEV, and consideration of a limited set of scenarios. We only considered the methods of Chambless and Davis [15], when other approaches have been proposed for correcting for the effect of day-to-day variation in this scenario [26,27]. Furthermore, we only tested one kind of model misspecification, which was dichotomization of the baseline covariate. We also did not examine whether the regression calibration model performed well when measurement error was associated with its true value, as may be the case with FEV, where high values might have a higher (or lower) variability. We additionally assumed that the terms of these models can be interpreted as intended. There is an abundance of literature describing confounding in observational studies using change as the response variable when adjusting for baseline [28–31]. Researchers may consider whether their scientific questions can be answered by modeling change scores and adjusting for baseline, or whether alternative approaches would be preferable.

Above conceptual misunderstandings, the unsubstantiated claim of ceiling effects has two potentially serious consequences: excluding persons with certain baseline characteristics (1) from enrolling into studies, or (2) from being eligible for effective treatments. When we

assume that persons with high FEV_1pp can no longer benefit from treatment, we risk failing to provide real benefits from uniformly effective treatments. The assumption of the ceiling effects appears to have led to the exclusion of persons of high FEV_1pp from treatment trials of new therapeutics, with many clinical trials requiring intermediate pre-treatment level FEV_1pp levels (40% to 90% is common [7,9,10]).

Our findings lend further support to those of Helshe et al. who demonstrated that persons with no short-term (1 month) FEV improvement following initiation of ivacaftor were no less likely to have long-term (2 year) FEV improvement [32]. Of note, those with no recorded short-term improvement in that study had a higher initial FEV_1pp 91.7% relative to those who improved quickly (78.1%), potentially explaining the reason for the lack of improvement as regression to the mean following a random high initial FEV_1pp.

5. Conclusions

Recorded levels of FEV_1pp variation in clinical studies were sufficient to exaggerate potentially observed ceiling effects in uncontrolled studies, as well as potentially affect estimates of average treatment effect, depending on how covariates were modeled. Our simulations show that these effects can be mitigated through available statistical tools. Statistical analysis plans for single-arm trials of interventions should account for spurious findings when this variation is present. Ameliorative efforts could include mitigation of day-to-day variability by averaging pre- and post-treatment assessments over longer time periods, or incorporating correction for measurement error in analysis approaches, such as using methods by Yanez or Chambless [14,15,33].

CRediT authorship contribution statement

Ellen Graham: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Sonya L. Heltshe:** Writing – review & editing, Supervision. **Amalia S. Magaret:** Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank study participants from GOAL as well as the study investigator who gave permission for these data to be re-analyzed for this manuscript: Stephen Rowe. Funding was provided by grants from the Cystic Fibrosis Foundation.

Appendix A. Chambless correction

Here we provide the correction to the model (1) when measurement error was present in the outcome variable.

Under the model in Eq. (1), due to the effects of measurement error η, the expected value of $\boldsymbol{y}$ given covariates $\boldsymbol{w}, \boldsymbol{z}$ was not $A+\boldsymbol{x}B_X+\boldsymbol{z}B_Z$. Under the additional assumption of multivariate normality between $\boldsymbol{z}, \boldsymbol{x}$, and $\boldsymbol{\eta}$ normally distributed with only a single covariate z, we instead have

$$E(\boldsymbol{y}|\boldsymbol{w},\boldsymbol{z}) = A^* + \boldsymbol{z}(B_Z + \beta(1-\tilde{R})B_X) + \boldsymbol{w}\tilde{R}B_X$$

with $\beta = \frac{Cov(x,z)}{Var(z)}$, $\tilde{R} = \frac{Var(w)R-\beta^2Var(z)}{Var(w)-\beta^2Var(z)}$, $R = 1 - \frac{Var(\eta)}{Var(w)}$, and $A^* = A + \mu_z B_X \frac{\beta(1-R)}{1-Corr(z,x)^2R} + \mu_x B_X \left(1 - \frac{R(1-Corr(z,x))^2}{1-Corr(z,x)^2R}\right)$. Chambless and Davis showed that $\tilde{R}$ could alternatively be written as $\tilde{R} = \frac{Var(x|z)}{Var(w|z)}$, often called the reliability ratio in measurement error models, which was the ratio of the variability of measurement error of the baseline measurement to the biological variability. Chambless and Davis proposed an algorithm to correct this bias [15]. To obtain unbiased estimates of B_X and B_Z, the algorithm replaced $\boldsymbol{x}$ with $\boldsymbol{w}^* = \hat{\boldsymbol{w}}(\boldsymbol{z})(1-\tilde{R}) + \boldsymbol{w}\tilde{R}$ and y with $\boldsymbol{y}^* = \boldsymbol{y} - (\boldsymbol{w} - \boldsymbol{w}^*)\phi$ in Eq. (1). The specific forms of $\boldsymbol{y}^*$ in our contexts were described in Appendix B, Table 4. Here, $\phi = \frac{Cov(\boldsymbol{\eta},\epsilon)}{Var(\boldsymbol{\eta}|\boldsymbol{x},\boldsymbol{z})}$ and $\hat{\boldsymbol{w}}(\boldsymbol{z})$ was an estimate of $E(\boldsymbol{w}|\boldsymbol{z})$ obtained via linear regression of w on z.

Appendix B. Derivation of expected values

Here find outcomes and corrected outcomes when running the regression calibration method specifically for the three scenarios simulated (Table 4).

Subsequently, we derived the expected values of the γ parameters in Eqs. (4), (5) and (6), using the data generation parameters, the fixed covariates βs, B_X and random terms bs from Eqs. (2) and (3). While the derivation was only shown for two-time points, the results can be easily extended to additional time points. For simplicity, when baseline values were dichotomized, we only considered $\beta_2 = 0$. Here ϕ and Φ were the PDF and CDF of a standard normal random variable, respectively (see Table 5).

The proof follows. The relationship between random effects b_1 and b_0 was shown in Eq. (3). We used properties of multivariate normal distributions to rewrite b_1 as $b_1 = B_X b_0 + \tilde{b}_1$ with $\tilde{b}_1 \sim N(0, \sigma^2_{\tilde{\delta}})$ and $\tilde{b}_1 \perp (u_0, Z)$ and hence

$$E[b_1|f(u_0), Z] = E[B_X b_0 + \tilde{b}_1|f(u_0), Z] = B_X E[b_0|f(u_0), Z]$$

for any function f.

Also, taking Eq. (2) at time $t = 1$ and subtracting when $t = 0$, we obtained $u_1 - u_0 = \beta_1 + \beta_3 Z + b_1$, as the remaining terms canceled. Below, we began with the expectation of this difference. First, when continuous values are used (no threshold):

$$\begin{aligned}
E(u_1 - u_0|u_0, Z) &= \beta_1 + \beta_3 Z + E(b_1|u_0, Z) && \text{(Eq. (2))}\\
&= \beta_1 + \beta_3 Z + B_X E(b_0|u_0, Z) && \text{(Eq. (3))}\\
&= \beta_1 + \beta_3 Z + B_X E[u_0 - (\beta_0 + \beta_2 Z)|u_0, Z] && \text{(Eq. (2))}\\
&= \beta_1 + \beta_3 Z + B_X u_0 - B_X(\beta_0 + \beta_2 Z)\\
&= (\beta_1 - B_X\beta_0) + B_X u_0 + (\beta_3 - B_X\beta_2)Z.
\end{aligned}$$

Hence $\gamma_0 = (\beta_1 - B_X\beta_0)$, $\gamma_1 = B_X$, and $\gamma_2 = (\beta_3 - B_X\beta_2)$.

We also demonstrated when baseline values were dichotomized at a threshold ν. (For simplicity, we only considered the case when $\beta_2 = 0$.)

$$\begin{aligned}
E(u_1 - u_0|I(u_0 > \nu), Z) &= \beta_1 + \beta_3 Z + E(b_1|I(u_0 > \nu), Z)\\
&= \beta_1 + \beta_3 Z + B_X E[b_0|I(u_0 > \nu)]\\
&= \beta_1 + \beta_3 Z + B_X E[b_0|I(u_0 > \nu) = 0] +\\
&\qquad B_X \left(E[b_0|I(u_0 > \nu) = 1]\right.\\
&\qquad \left. - E[b_0|I(u_0 > \nu) = 0]\right) I(u_0 > \nu)\\
&= \beta_1 + \beta_3 Z + B_X E[b_0|b_0 + \beta_0 \le \nu] +\\
&\qquad B_X \left(E[b_0|b_0 + \beta_0 > \nu]\right.\\
&\qquad \left. - E[b_0|b_0 + \beta_0 \le \nu]\right) I(u_0 > \nu)\\
&= \left(\beta_1 - B_X\sigma_{\text{bio}}\frac{\phi(\nu-\beta_0)}{\Phi(\nu-\beta_0)}\right) +\\
&\qquad B_X\sigma_{\text{bio}}\left(\frac{\phi(\nu-\beta_0)}{1-\Phi(\nu-\beta_0)} + \frac{\phi(\nu-\beta_0)}{\Phi(\nu-\beta_0)}\right)\\
&\qquad \times I(u_0 > \nu) + \beta_3 Z
\end{aligned}$$

Where the last line followed from the properties of the truncated normal distribution. When $\nu = \beta_0$ as in our simulations, $\gamma_0 = \beta_1 - B_X\sigma_{\text{bio}}\sqrt{\frac{2}{\pi}}$, $\gamma_1 = 2B_X\sigma_{\text{bio}}\sqrt{\frac{2}{\pi}}$, and $\gamma_3 = \beta_3$.

Table 4
Outcomes and corrected outcomes used in model fitting.

Model	Response y_{it}	$\mathrm{Cov}(\epsilon,\eta)Var(\epsilon)^{-1}$	Modeled response y^*_{it}
Immediate Change	$u_{it}-u_{i0}$	-1	$y_{it}-w^*_i$
Linear Time Trend	$\frac{u_{it}-u_{i0}}{t}$	$-1/t$	$y_{it}-w^*_i/t$
Participant Level Regression[a]	$\hat{\alpha}_{1i}=\sum_{t=0}^{T} c_t u_{it}$	c_0	$c_0 w^*_i+\sum_{t=1}^{T} c_t y_{it}$

[a] Let $M_i=\begin{pmatrix}\mathbf{1}_T & \boldsymbol{t}\end{pmatrix}$ be the data matrix for $u_{it}=\alpha_{i0}+\alpha_{i1}t$. We then let $c_t=[(M'M)^{-1}M']_{1t}=\sum_{t=0}^{T}\frac{t_t-\bar{t}}{\sum_{t=0}^{T}(t_t-\bar{t})^2}u_{it}$ be the coefficient of u_{it} when calculating $\hat{\alpha}_{i1}$.

Table 5
Expected values of regression parameters under different data analysis procedures.

Baseline value u_0	γ_0	γ_1	γ_2
Entered as continuous	$\beta_1-B_X\beta_0$	B_X	$\beta_3-B_X\beta_2$
Dichotomized at ν	$\left(\beta_1-B_X\sigma_{\mathrm{bio}}\frac{\phi(\nu-\beta_0)}{\Phi(\nu-\beta_0)}\right)$	$B_X\sigma_{\mathrm{bio}}\left(\frac{\phi(\nu-\beta_0)}{1-\Phi(\nu-\beta_0)}+\frac{\phi(\nu-\beta_0)}{\Phi(\nu-\beta_0)}\right)$	β_3

References

[1] Michael W. Konstan, Jeffrey S. Wagener, Donald R. Vandevanter, David J. Pasta, Ashley Yegin, Lawrence Rasouliyan, Wayne J. Morgan, Risk factors for rate of decline in FEV1 in adults with cystic fibrosis, J. Cyst. Fibros. 11 (5) (2012) 405–411.

[2] A.B. Montgomery, T. Abuan, M.A. Yeager, Regulatory aspects of Phase 3 endpoints for new inhaled antibiotics for cystic fibrosis patients with chronic *Pseudomonas aeruginosa* infections, J. Aerosol. Med. Pulm Drug Deliv. 25 (4) (2012) 198–203.

[3] Sabariah Noor Harun, Claire Wainwright, Kerenaftali Klein, Stefanie Hennig, A systematic review of studies examining the rate of lung function decline in patients with cystic fibrosis, Paediatr. Respir. Rev. 20 (2016) 55–66.

[4] Rhonda Szczesniak, Sonya L. Heltshe, Sanja Stanojevic, Nicole Mayer-Hamblett, Use of FEV1 in cystic fibrosis epidemiologic studies and clinical trials: A statistical perspective for the clinical researcher, J. Cyst. Fibros. 16 (3) (2017) 318–326.

[5] S.J. Newsome, R.M. Daniel, S.B. Carr, D. Bilton, R.H. Keogh, Investigating the effects of long-term dornase alfa use on lung function using registry data, J. Cyst. Fibros. 18 (1) (2019) 110–117.

[6] Bonnie W. Ramsey, Jane Davies, N. Gerard McElvaney, Elizabeth Tullis, Scott C. Bell, Pavel Dřevínek, Matthias Griese, Edward F. McKone, Claire E. Wainwright, Michael W. Konstan, Richard Moss, Felix Ratjen, Isabelle Sermet-Gaudelus, Steven M. Rowe, Qunming Dong, Sally Rodriguez, Karl Yen, Claudia Ordoñez, J. Stuart Elborn, A CFTR Potentiator in Patients with Cystic Fibrosis and the G551D Mutation, New Engl. J. Med. 365 (18) (2011) 1663–1672.

[7] Claire E. Wainwright, J. Stuart Elborn, Bonnie W. Ramsey, Gautham Marigowda, Xiaohong Huang, Marco Cipolli, Carla Colombo, Jane C. Davies, Kris De Boeck, Patrick A. Flume, Michael W. Konstan, Susanna A. McColley, Karen McCoy, Edward F. McKone, Anne Munck, Felix Ratjen, Steven M. Rowe, David Waltz, Michael P. Boyle, Lumacaftor–Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del CFTR, New Engl. J. Med. 373 (3) (2015) 220–231.

[8] Carlos E. Milla, Felix Ratjen, Gautham Marigowda, Fang Liu, David Waltz, Margaret Rosenfeld, VX13-809-011 Part B Investigator, Lumacaftor/ivacaftor in patients aged 6-11 years with cystic fibrosis and homozygous for F508del-CFTR, Am. J. Respir. Clin. Care Med. 195 (7) (2017) 912–920.

[9] Harry G.M. Heijerman, Edward F. McKone, Damian G. Downey, Eva Van Braeckel, Steven M. Rowe, Elizabeth Tullis, Marcus A. Mall, John J. Welter, Bonnie W. Ramsey, Charlotte M. McKee, Gautham Marigowda, Samuel M. Moskowitz, David Waltz, Patrick R. Sosnay, Christopher Simard, Neil Ahluwalia, Fengjuan Xuan, Yaohua Zhang, Jennifer L. Taylor-Cousar, Karen S. McCoy, VX17-445-103 Trial Group, Efficacy and safety of the elexacaftor plus tezacaftor plus ivacaftor combination regimen in people with cystic fibrosis homozygous for the F508del mutation: a double-blind, randomised, phase 3 trial, Lancet (London, England) 394 (10212) (2019) 1940–1948.

[10] Peter G. Middleton, Marcus A. Mall, Pavel Dřevínek, Larry C. Lands, Edward F. McKone, Deepika Polineni, Bonnie W. Ramsey, Jennifer L. Taylor-Cousar, Elizabeth Tullis, François Vermeulen, Gautham Marigowda, Charlotte M. McKee, Samuel M. Moskowitz, Nitin Nair, Jessica Savage, Christopher Simard, Simon Tian, David Waltz, Fengjuan Xuan, Steven M. Rowe, Raksha Jain, Elexacaftor–Tezacaftor–Ivacaftor for Cystic Fibrosis with a Single Phe508del Allele, New Engl. J. Med. 381 (19) (2019) 1809–1819.

[11] Mark Chilvers, Jane Davies, Carlos Milla, Simon Tian, Zifei Han, Alexandra Cornell, Caroline Owen, Felix Ratjen, Long-term safety and efficacy of lumacaftor-ivacaftor therapy in children aged 6-11 years with cystic fibrosis homozygous for the F508del CFTR mutation: a phase 3, open-label, extension study, Lancet Respir. Med. 9 (7) (2021) 721–732.

[12] Edith T. Zemanick, Jennifer L. Taylor-Cousar, Jane Davies, Ronald L. Gibson, Marcus A. Mall, Edward F. McKone, Paul McNally, Bonnie W. Ramsey, Jonathan H. Rayment, Steven M. Rowe, Elizabeth Tullis, Neil Ahluwalia, Chenghao Chu, Thang Ho, Samuel M. Moskowitz, Sabrina Noel, Simon Tian, David Waltz, Tanya G. Weinstock, Fengjuan Xuan, Claire E. Wainwright, Susanna A. McColley, A Phase 3 Open-Label Study of Elexacaftor/Tezacaftor/Ivacaftor in Children 6 through 11 Years of Age with Cystic Fibrosis and at Least One F508del Allele, Am. J. Respir. Crit. Care Med. 203 (12) (2021) 1522–1532.

[13] Amalia S. Magaret, Ellen Graham, Lindsay J. Caverly, Elizabeth A. Cromwell, Alex Paynter, Margaret Rosenfeld, Christina S. Thornton, Christopher H. Goss, Impact of day-to-day variation in FEV1 on measures of change: a conceptual description, J. Cyst. Fibros. 23 (5) (2024) 943–946.

[14] N.D. Yanez, R.A. Kronmal, L.R. Shemanski, The effects of measurement error in response variables and tests of association of explanatory variables in change models, Stat. Med. 17 (22) (1998) 2597–2606.

[15] Lloyd E. Chambless, Vicki Davis, Analysis of associations with change in a multivariate outcome variable when baseline is subject to measurement error, Stat. Med. 22 (7) (2003) 1041–1067.

[16] Noah Lechtzin, Nicole Mayer-Hamblett, Natalie E. West, Sarah Allgood, Ellen Wilhelm, Umer Khan, Moira L. Aitken, Bonnie W. Ramsey, Michael P. Boyle, Peter J. Mogayzel Jr., Ronald L. Gibson, David Orenstein, Carlos Milla, John P. Clancy, Veena Antony, Christopher H. Goss, eICE Study Team, Home monitoring of patients with cystic fibrosis to identify and treat acute pulmonary exacerbations eICE study results, Am. J. Respir. Clin. Care Med. 196 (9) (2017) 1144–1151.

[17] Steven M. Rowe, Sonya L. Heltshe, Tanja Gonska, Scott H. Donaldson, Drucy Borowitz, Daniel Gelfond, Scott D. Sagel, Umer Khan, Nicole Mayer-Hamblett, Jill M. Van Dalfsen, Elizabeth Joseloff, Bonnie W. Ramsey, GOAL Investigators of the Cystic Fibrosis Foundation Therapeutics Development Network, Clinical mechanism of the cystic fibrosis transmembrane conductance regulator potentiator ivacaftor in G551D-mediated cystic fibrosis, Am. J. Respir. Crit. Care Med. 190 (2) (2014) 175–184.

[18] Angus G. Jones, Mike Lonergan, William E. Henley, Ewan R. Pearson, Andrew T. Hattersley, Beverley M. Shields, Should Studies of Diabetes Treatment Stratification Correct for Baseline HbA1c? PLoS One 11 (4) (2016) e0152428.

[19] S. Stanojevic, N. Filipow, F. Ratjen, Paediatric reproducibility limits for the forced expiratory volume in 1 s, Thorax 75 (10) (2020) 891–896.

[20] C.S. Thornton, A.S. Magaret, L.A. Carmody, L.M. Kalikin, R.H. Simon, J.J. LiPuma, L.J. Caverly, Quantifying variation in home spirometry in people with cystic fibrosis during baseline health, and associations with clinical outcomes, J. Cyst. Fibros. (2023).

[21] L. Saiman, B.C. Marshall, N. Mayer-Hamblett, J.L. Burns, A.L. Quittner, D.A. Cibene, S. Coquillette, A.Y. Fieberg, F.J. Accurso, P.W. Campbell, Azithromycin in patients with cystic fibrosis chronically infected with *Pseudomonas aeruginosa*: a randomized controlled trial, JAMA 290 (13) (2003) 1749–1756.

[22] A. Paynter, U. Khan, S.L. Heltshe, C.H. Goss, N. Lechtzin, N.M. Hamblett, A comparison of clinic and home spirometry as longtudinal outcomes in cystic fibrosis, J. Cyst. Fibros. 21 (1) (2022) 78–83.

[23] Therapeutics Development Network Coordinating Center, https://www.cff.org/researchers/therapeutics-development-network.

[24] Margaret Rosenfeld, Donald R. VanDevanter, Clement L. Ren, Eric P. Elkin, David J. Pasta, Michael W. Konstan, Wayne J. Morgan, Investigators of Coordinators of the Epidemiologic Study of Cystic Fibrosis, Decline in lung function does not predict future decline in lung function in cystic fibrosis patients, Pediatr. Pulmonol. 50 (9) (2015) 856–862.

[25] Nicole Hamblett, A Prospective Study to Evaluate Biological and Clinical Effects of Significantly Corrected CFTR Function (The PROMISE Study), 2022, Clinical trial registration NCT04038047, clinicaltrials.gov.

[26] L. Harrison, D.T. Dunn, H. Green, A.J. Copas, Modelling the association between patient characteristics and the change over time in a disease measure using observational cohort data, Stat. Med. 28 (26) (2009) 3260–3275.

[27] Julia Braun, Leonhard Held, Bruno Ledergerber, Swiss HIV Cohort Study, Accounting for baseline differences and measurement error in the analysis of change over time, Stat. Med. 33 (1) (2014) 2–16.

[28] Frederic M. Lord, A paradox in the interpretation of group comparisons, Psychol. Bull. 68 (5) (1967) 304–305.

[29] Stephen Senn, Change from baseline and analysis of covariance revisited, Stat. Med. 25 (24) (2006) 4334–4344.
[30] Guanghan F. Liu, Kaifeng Lu, Robin Mogg, Madhuja Mallick, Devan V. Mehrotra, Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? Stat. Med. 28 (20) (2009) 2509–2530.
[31] M. Maria Glymour, Jennifer Weuve, Lisa F. Berkman, Ichiro Kawachi, James M. Robins, When is baseline adjustment useful in analyses of change? An example with education and cognitive change, Am. J. Epidemiol. 162 (3) (2005) 267–278.
[32] Sonya L. Heltshe, Steven M. Rowe, Michelle Skalland, Arthur Baines, Manu Jain, Ivacaftor-treated patients with cystic fibrosis derive long-term benefit despite no short-term clinical improvement, Am. J. Respir. Clin. Care Med. 197 (11) (2018) 1483–1486.
[33] N. David Yanez, Richard A. Kronmal, Lynn R. Shemanski, Bruce M. Psaty, Cardiovascular Health Study, A regression model for longitudinal change in the presence of measurement error, Ann. Epidemiol. 12 (1) (2002) 34–38.