EDITORIAL

Lung function decline in COPD trials: bias from regression to the mean

S. Suissa*,#,¶

The decline in lung function over time is a fundamental measure of disease progression among patients with chronic obstructive pulmonary disease (COPD). As a result, forced expiratory volume in one second (FEV1) decline has been used as a central outcome measure for many randomised controlled trials evaluating whether pharmacological treatments could modify the natural history of COPD. These trials, mostly focussed on assessing the benefit of inhaled corticosteroids, and their meta-analyses highlighted the complexities of analysing data from repeated FEV1 measurements over time, particularly in the context of studying COPD patients who discontinue follow-up early and in large numbers. This has lead to contradictory results and divergent conclusions [1–3]. At present, the recent paper from the Towards a Revolution in COPD Health (TORCH) trial reports that the yearly decline in FEV1 was significantly slower with fluticasone, salmeterol or both compared with placebo [4].

A common misconception in the interpretation of these studies is that the effect of the study drug on FEV1 decline is likely to be greater than the data suggest. The rationale is that more patients receiving placebo were discontinuing the study drugs and the patients who dropped out early had a steeper decline in lung function than those who remained. Consequently, it has generally been believed that the resulting analysis "actually minimises the differences observed in the rate of FEV1 decline" [4].

To clarify this misconception, a fundamental aspect of the design and statistical analysis of these trials is addressed, namely bias from regression to the mean resulting from the absence of an authentic intent-to-treat approach. It is described based on the TORCH trial and illustrated using data from the Canadian Optimal randomised trial.

BIAS FROM REGRESSION TO THE MEAN
Most of the trials to date are not designed for a full intent-to-treat analysis of lung function decline. The TORCH trial, for example, only measured FEV1 until the patients discontinued treatment. It involved 6,112 moderate-to-severe COPD patients randomised to one of four treatment groups including fluticasone, salmeterol, both or placebo, followed for 3 yrs. However, the lung function analysis involved only 5,343 subjects with 26,539 measurements of post-bronchodilator FEV1 made twice a year during follow-up to estimate the rate of FEV1 decline between 6 months and 3 yrs after randomisation. Consequently, 10,133 measurements were missing so that the pure intent-to-treat analysis was not possible. These missing measurements are likely to not be a random sample of all possible 36,672 measurements that the study could have yielded, i.e. they are not missing at random, which can generate two forms of bias from regression to the mean [5].

The first form of the bias results from excluding some subjects altogether. Nearly 18% of patients allocated to placebo did not contribute a single FEV1 value because they discontinued placebo before the first 6-month visit when the initial FEV1 was measured. In contrast, only 9% of patients allocated to combination therapy did not make it to this first 6-month visit. It is generally accepted that these excluded patients
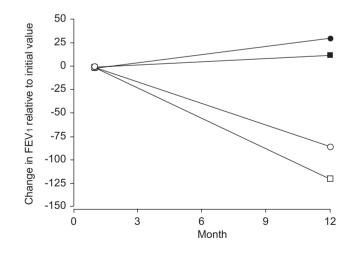

FIGURE 1. Depiction of regression to the mean. Correlation between initial forced expiratory volume in one second (FEV1; divided in quartiles) and the change in FEV1 from visit 1 to 4 (mean change in each quartile). Patients in the highest quartile of initial FEV1 values have the largest decline (119 mL) while the patients in the lowest quartile of initial FEV1 values show an improvement of 32 mL. ●: initial FEV1 <770 mL; ■: initial FEV1 770–1030 mL; ○: initial FEV1 1040–1440 mL; □: initial FEV1 >1442 mL.

*McGill Pharmacoepidemiology Research Unit, Jewish General Hospital, Depts of #Epidemiology and Biostatistics, and ¶Medicine, McGill University, Montreal, QC, Canada.

CORRESPONDENCE: S. Suissa, Division of Clinical Epidemiology, Royal Victoria Hospital, 687 Pine Avenue West, Ross 4.29, Montreal, QC, Canada H3A 1A1. Fax: 1 5148431493. E-mail: samy.suissa@clinepi.mcgill.ca

TABLE 1 Estimation of decline in forced expiratory volume in one second (FEV1) among 322 subjects across four visits during a 1-yr follow-up from the Canadian Optimal trial

	Subjects n	Mean FEV1 at visit 1[#]	Slope in FEV1 decline over all four visits mL·yr⁻¹	
			Mean	p-value
All subjects	322	1131	38.6	0.001
Stratification according to FEV1 at visit 1				
Highest 82% ⩾700 mL	264	1248	52.2	0.0002
Lowest 18%[¶] <700 mL	58	600	-40.7	0.020
Highest 91% ⩾630 mL	293	1190	44.1	0.0006
Lowest 9%[¶] <630 mL	29	532	-52.5	0.054

[#]: in mL; [¶]: typically excluded subjects with no data on FEV1 as in the TORCH study: 18% from the placebo group and 9% from the combination treatment group.

would have had the worse FEV1 values at the first visit had they been available to be measured. Thus, the slope of decline in the remaining subjects with better FEV1 values at the first visit may have been affected by regression to the mean. This phenomenon is illustrated below.

The second form of this bias results from discontinuing the follow-up of patients who have the initial FEV1 value measured but are missing some subsequent values; these measurements are also unlikely to be missing at random. In the TORCH study, the placebo patients who discontinued before the end of follow-up had a faster decline in FEV1 (76 mL·yr⁻¹) than those completing the trial (54 mL·yr⁻¹) [4]. Here again, these slopes of decline may have been affected by regression to the mean.

ILLUSTRATION

Data from the Canadian Optimal study, a three-arm randomised trial of 449 patients with moderate or severe COPD, are used to illustrate the bias [6]. Measurements of post-bronchodilator FEV1 were made at randomisation (visit 0) and at 4, 20, 36 and 52 weeks thereafter during the 1-yr follow-up (visits 1–4). The 322 subjects who had measurements of FEV1 for all visits were used in this illustration. FEV1 decline was measured as of visit 1 and the rate of FEV1 decline was estimated in two ways: 1) as the difference in FEV1 between visits 4 and 1; and 2) using measurements from all four visits with a mixed linear regression model accounting for

within-subject correlation, with the slope standardised to a 1-yr time span.

The 322 patients had a mean FEV1 of 1,131 mL at visit 1, with a change in FEV1 from visit 1 to 4 of 38.7 mL. There was a significant correlation of 0.33 between this decline and the FEV1 measure at visit 1, which indicates that patients with the highest initial FEV1 have the greatest decline, while the patients with the lowest initial FEV1 have the lowest decline. Figure 1 depicts this correlation by showing that patients in the highest quartile of initial FEV1 values (>1,440 mL) have the largest decline (mean decline 119 mL), while the patients in the lowest quartile of initial FEV1 values (<770 mL) in fact show an improvement of 32 mL.

Table 1 shows that the overall 1-yr decline in FEV1 estimated using measurements from all four visits was 38.6 mL (p=0.001). It shows that if the 18% of patients with the lowest FEV1 at visit 1 (<700 mL) are excluded, in keeping with the hypothesis that patients in poorer health are more likely to leave the study, the 1-yr rate of FEV1 decline among the remaining subjects becomes 52.2 mL, a clear overestimate of the decline. If only the 9% of patients with the lowest FEV1 at visit 1 (<630 mL) are excluded, analogous to the lower exclusion rate with combination therapy in TORCH, the 1-yr rate of FEV1 decline among the remaining subjects is 44.1 mL, still an overestimate but less marked. Conversely, the two groups of excluded subjects have a mean increase in FEV1 of 40.7 and 52.5 mL, respectively.

TABLE 2 Estimation of decline in forced expiratory volume in one second (FEV1) among 322 subjects across four visits during a 1-yr follow-up from the Canadian Optimal trial

	Subjects n	Mean FEV1 at visit 2[#]	Change in FEV1 visits 1–2[#]	Slope in FEV1 decline visits 1–4[¶]	p-value
All subjects	322	1111	19.7	38.6	0.001
Stratification according to mean FEV1 at visit 2					
Highest 80% ⩾740 mL	264	1238	11.8	50.3	0.0004
Lowest 20%[+] <740 mL	58	610	50.6	1.7	0.93

[#]: in mL; [¶]: in mL·yr⁻¹; [+]: typical subjects with an initial FEV1 value but values missing during follow-up.

Table 2 addresses the issue of subjects with a FEV_1 value at visit 1, but values possibly missing thereafter. Here, it is found that if the 20% of patients who have the lowest FEV_1 at visit 2 (<740 mL) are excluded, the 1-yr rate of FEV_1 decline among the remaining subjects becomes 50.3 mL. These excluded subjects had, between the first two visits, a decline in FEV_1 of 50.6 mL over the initial 4-month period but no significant decline over all four visits (1.7 mL).

CONCLUSION

The intent-to-treat principle for randomised controlled trials is fundamental to avoid bias. Selection bias can occur when patients are permitted to exit the study when they discontinue the study medication. A more severe form of this bias occurs when randomised patients are excluded from the analysis altogether.

The TORCH study performed an authentic intent-to-treat analysis for the outcome of mortality by assessing the survival status of all 6,112 patients randomised for the entire 3-yr follow-up period. However, the recent TORCH analysis of lung function was not a true intent-to-treat analysis since it was based on only 5,343 out of the 6,112 patients with randomised FEV_1 values. Moreover, twice as many patients randomised to placebo were excluded altogether compared with the combination therapy group, while many patients discontinued early with no further FEV_1 measurements. This paper shows that such differential exclusion rates can introduce selection bias if the basis for exclusion is associated with the outcome, in this case decline in FEV_1.

The current illustration can help interpret the TORCH findings. Assume, for instance, that the FEV_1 decline in the TORCH study was 38.6 mL·yr^{-1} equally for all treatment arms. The exclusion of 18% of placebo patients, most likely with the lowest initial FEV_1 value, can result in a mean FEV_1 decline of 52.2 mL among the remaining subjects, compared with the exclusion of 9% in the treated patients which changes the decline to 44.1 mL. Thus, instead of comparing 38.6 with 38.6 mL·yr^{-1}, a study based on such incomplete groups would mistakenly compare 44.1 with 52.2 mL·yr^{-1} and infer that the declines are different. In addition, among those with an initial FEV_1 value, the TORCH study placebo patients who discontinued before the end of follow-up had a decline of 76 mL·yr^{-1} compared with 54 mL for those completing the trial. In the present illustration, the patients most likely to discontinue had a 4-month decline in FEV_1 of 50.6 mL (extrapolated to 152 mL over 1 yr) compared with a 1-yr FEV_1 decline among the remaining subjects of 50.3 mL, both quite different from the true 1-yr decline of 38.6 mL. Incidentally, this second form of the bias may explain the different initial spikes in FEV_1 seen just after randomisation in these trials. Such bias is not eliminated by advanced techniques of data analysis, such as mixed linear regression, that account for within-subject correlation and variable numbers of FEV_1 measurements per subject.

While the TORCH study may provide some evidence that pharmacological therapy could modify the decline in lung function, it could also reflect an artificial effect of the phenomenon of regression to the mean. Indeed, the decline was estimated after excluding patients, in a context where the patients with the best initial forced expiratory volume in one second value generally have the greatest decline in forced expiratory volume in one second and those with the poorest initial forced expiratory volume in one second value have the lowest decline, or even an increase, in forced expiratory volume in one second. In such a context, the impression that the effect of the study drug on forced expiratory volume in one second decline is greater than the data would suggest is a misconception; the effect of the drug could even be nil. The regression to the mean phenomenon can clearly lead to bias in randomised trials of chronic obstructive pulmonary disease treatment, thus, proper attention to the intent-to-treat principle becomes crucial to avoid this bias and provide valid data. This will hopefully be addressed in the upcoming Understanding Potential Long-term Impacts on Function with Tiotropium (UPLIFT) trial [7].

REFERENCES

1 Highland KB, Strange C, Heffner JE. Long-term effects of inhaled corticosteroids on FEV_1 in patients with chronic obstructive pulmonary disease. A meta-analysis. *Ann Intern Med* 2003; 138: 969–973.

2 Sutherland ER, Allmers H, Ayas NT, Venn AJ, Martin RJ. Inhaled corticosteroids reduce the progression of airflow limitation in chronic obstructive pulmonary disease: a meta-analysis. *Thorax* 2003; 58: 937–941.

3 Soriano JB, Sin DD, Zhang X, *et al.* A pooled analysis of FEV_1 decline in COPD patients randomized to inhaled corticosteroids or placebo. *Chest* 2007; 131: 682–689.

4 Celli B, Thomas NE, Anderson JA, *et al.* Effect of pharmacotherapy on rate of decline of lung function in COPD: results from the TORCH study. *Am J Respir Crit Care Med* 2008; 178: 332–338.

5 Bland JM, Altman DG. Some examples of regression towards the mean. *BMJ* 1994; 309: 780.

6 Aaron SD, Vandemheen KL, Fergusson D, *et al.* Tiotropium in combination with placebo, salmeterol, or fluticasone-salmeterol for treatment of chronic obstructive pulmonary disease: a randomized trial. *Ann Intern Med* 2007; 146: 545–555.

7 Decramer M, Celli B, Tashkin DP, *et al.* Clinical trial design considerations in assessing long-term functional impacts of tiotropium in COPD: the UPLIFT trial. *COPD* 2004; 1: 303–312.