Samy Suissa, Pierre Ernst

Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital,
Departments of Epidemiology, Biostatistics and Medicine, McGill University, Montreal, Canada

Mega trials in COPD — clinical data analysis and design issues

Wielkie badania kliniczne w POChP — planowanie badań i analiza wyników

Abstract

The TORCH and UPLIFT randomised controlled trials have provided important data on the benefits of COPD treatments, but also some lessons in study design and data analysis that we will here review.

Firstly, it is fundamental that the study question be answerable by the study design. The question in the TORCH study was aimed at a comparison with 'usual care', but the placebo group was not 'usual care'. Secondly, TORCH and UPLIFT were among the first trials to follow the intent-to-treat principle, fundamental to avoid bias in randomised trials. However, this principle was followed for the mortality outcome, but not for lung function, so that the findings related to lung function decline are subject to bias from regression to the mean. Finally, a re-analysis of the TORCH study (performed to fully exploit the data as a 2×2 factorial trial) shows that a mortality benefit is entirely accounted for by the effect of the long-acting β-agonist salmeterol, with no effect attributable to the inhaled corticosteroid fluticasone component of the combination therapy. Together, these data suggest that long-acting bronchodilators, including anticholinergics such as tiotropium and beta-agonists, are associated with lower mortality of patients with COPD, but not inhaled corticosteroids. With COPD one of the major causes of morbidity and mortality worldwide, mega trials such as TORCH and UPLIFT are much needed, but must achieve the utmost scientific rigour in their design and analysis.

Key words: drug effectiveness, methods, chronic obstructive pulmonary disease, inhaled corticosteroids, anticholinergics, long-acting β-agonist

Pneumonol. Alergol. Pol. 2011; 79, 3: 227–231

Streszczenie

Randomizowane, kontrolowane badania nad lekami TORCH i UPLIFT dostarczyły istotnych informacji o korzyściach z ich stosowania u chorych na POChP, a także były lekcją dotyczącą wpływu planowania badań i analizy wyników na jakość uzyskanych wyników.

Celem badania TORCH było porównanie stosowanego leczenia z leczeniem rutynowym, jednakże grupa placebo nie otrzymywała rutynowego leczenia.

TORCH i UPLIFT były jednymi z pierwszych badań, w których analiza wyników oparta była na zasadzie *„intention-to-treat"*, bardzo ważnej dla uniknięcia wątpliwości w badaniach randomizowanych. Jednakże ta zasada obejmowała tylko analizę umieralności, ale już nie czynności płuc. Z tego powodu ta druga ulegała wpływowi zjawiska regresji do średniej. Natomiast dodatkowa analiza wyników badania TORCH wykazała, że korzystny wpływ na umieralność był spowodowany wyłącznie długodziałającym lekiem rozszerzającym oskrzela bez wpływu na obniżenie umieralności na POChP wziewnego steroidu w leczeniu skojarzonym. Łącznie, wyniki tych badań sugerują, że obniżenie umieralności na POChP jest związane z długodziałającymi lekami rozszerzającymi oskrzela, w tym antycholinergikami, takimi jak tiotropium i agonistami receptorów β, a nie z działaniem wziewnych kortykosteroidów. Przewlekła obturacyjna choroba płuc jest jedną z głównych przyczyn chorobowości i umieralności na świecie. Wielkie badania kliniczne, takie jak TORCH i UPLIFT, są potrzebne, ale powinny być prowadzone z najwyższą naukową starannością w planowaniu i analizie wyników.

Słowa kluczowe: skuteczność leku, metody, przewlekła obturacyjna choroba płuc, wziewne kortykosteroidy, antycholinergiki, długodziałające β-agoniści

Pneumonol. Alergol. Pol. 2011; 79, 3: 227–231

Correspondence addres: Samy Suissa M.D., Centre for Clinical Epidemiology, Jewish General Hospital, 3755 Cote Ste-Catherine, Montreal, Québec, Canada H3T 1E2, tel.: 514 340 7593, fax: 514 340 7564, e-mail: samy.suissa@mcgill.ca

Praca wpłynęła do Redakcji: 20.01.2011 r.

ISSN 0867–7077

Introduction

The TORCH and UPLIFT mega-trials have been study landmarks in the field of chronic obstructive pulmonary disease (COPD) therapeutics. They have advanced our knowledge as to the potential benefits of long-acting bronchodilators, including long-acting beta-agonists and the long-acting anticholinergic tiotropium, as well as of inhaled corticosteroids (ICS), in treating this disease [1, 2]. These mega-trials were important because the previous randomised controlled trials, that evaluated the effectiveness of inhaled corticosteroids in particular, reported results that were often contradictory and paradoxical [3, 4]. Generally, these former studies (which were smaller) found that inhaled corticosteroids either had no effect, or only minor beneficial effects on lung function as measured by FEV_1, the primary outcome measure, while several of the studies observed significant reductions in COPD exacerbation rates associated with ICS use. A meta-analysis of these trials even suggested a significant reduction in exacerbations of approximately 30% [5].

Much of this apparent meta-analysed benefit, however, resulted from an incorrect, unweighted, approach to the data analysis of these exacerbations, failing also to incorporate correction for overdispersion in their statistical analysis [6]. In addition, a pooled analysis of data from seven randomised trials involving more than 5,000 patients reported a significant 27% reduction in all-cause mortality associated with ICS use [7].

In view of this confusing background, the TORCH and UPLIFT mega trials were especially important to clarify the place of long-acting bronchodilators and inhaled corticosteroids in the treatment of patients with COPD.

In this paper, we review some key aspects of the design and statistical analysis of TORCH and UPLIFT, and interpret their results in light of the current recommendations for COPD management.

Study question and design

The *a priori* question under investigation is the foundation of a clinical trial, since it will impact directly on its design and analysis. In the TORCH trial, the authors: "hypothesised that the combination of the long-acting beta-agonist salmeterol and the inhaled corticosteroid fluticasone propionate would reduce mortality among patients with COPD, as compared with usual care" [1].

In the UPLIFT trial, the authors: "tested whether tiotropium would reduce the rate of decline in FEV_1 in patients with COPD who were permitted therapy other than other inhaled anticholinergic drugs, according to current COPD guidelines" [2].

Does the design of each study allow addressing its study question? One of the challenges in designing such studies is that the drugs being evaluated were already on the market; so some study patients may already be using the study drugs. Here, the study protocol required patients to stop this treatment prior to randomisation. For example, the TORCH trial involved 6,112 moderate to severe COPD patients randomised to one of four treatment groups (fluticasone, salmeterol, both, or a placebo) followed for three years. Any patients already using inhaled corticosteroids or long-acting beta-agonists prior to randomisation had to cease these medications. In particular, of the 1,524 patients randomised to a placebo, 22% had to stop inhaled corticosteroid therapy, 8% to cease taking long-acting beta-agonists, and 29% the combination of inhaled corticosteroids with a long-acting beta-agonist. Thus, nearly 60% of patients randomised to a placebo had to stop these two forms of treatment. Can we then claim that COPD patients who have been on maintenance therapy with a LABA and/or ICS, and who are made to stop these medications and replace them with placebos for three years represent "usual care"? This seems unlikely. So, it can be said that the design of the TORCH trial could not permit the study to answer its question, namely to assess the effect of combination salmeterol/fluticasone relative to "usual care".

In the UPLIFT study, 5,993 patients with COPD were randomly assigned to tiotropium (n = 2,987) or a placebo (n = 3,006) for four years. Patients were required to cease taking their anticholinergics at the time of randomisation, but could continue all other maintenance treatment during the four-year follow-up. Thus, in UPLIFT, 44% of the patients randomised to a placebo stopped their short-acting anticholinergic, which could be replaced by another rescue medication, while only 2% of the patients put on placebo were already on the long-acting anticholinergic tiotropium and had to stop it at randomisation. Consequently, we can say that the design of the UPLIFT trial generally permitted the answering of its study question; namely to assess the effect of adding tiotropium to the existing treatment of COPD patients.

In fact, the various previous trials of ICS in COPD were conducted, at least in part, among patients already using these medications before randomisation. The proportion of patients who were previous users of these drugs and were required to cease using them at the time of randomisation in

the placebo group varied from 26% in one trial [8] to as high as 77% in the OPTIMAL trial [9]. This unusual situation creates a challenge in interpretation. Indeed, among the patients who did not previously use inhaled corticosteroids, randomisation will lead to the desired comparison between patients initiating treatment with ICS, and similar patients who do not. On the other hand, among the patients who were previously regular users of inhaled corticosteroids, randomisation will in fact provide a comparison between patients who continue to use ICS (patients allocated to ICS treatment) and patients who discontinue ICS (patients allocated to a placebo). These different comparisons can lead to very different results [10].

Intent to treat: mortality

Before the TORCH and OPTIMAL studies, randomised trials of COPD therapy stopped patient follow-up at the time they discontinued the study drug. Thus, any outcome information arising after the patients had stopped treatment, but before the planned end of the study follow-up, was not collected. As such, the fundamental intent-to-treat analysis for such trials was not possible, since the data was truncated at the time of drug discontinuation.

While this problem may be trivial in other diseases, COPD trials characteristically demonstrate very high discontinuation rates, often occurring very early in the trial. Not following patients through to the end, and conducting the data analysis only until discontinuation of study drugs, will produce biased results if the reasons for discontinuation are associated with the outcome, and differ between treatments.

To avoid such bias, the TORCH trial correctly followed all patients up to the end of the three-year trial period to ascertain mortality, its primary outcome, even after discontinuation of the study medications. This was not done, however, for the secondary outcomes, including exacerbations, lung function and health status. This is of concern, because 44% of patients in the placebo group discontinued treatment, mostly during the first few months, compared to 34% in the combination therapy group. Thus, for the TORCH trial, the intent-to-treat results for mortality are valid; but the results describing the secondary outcomes, and in particular exacerbations, may be biased.

To illustrate this bias, a comparison can be made between the pooled analysis of seven major randomised trials involving 5,086 patients, that found a significant 27% reduction in all-cause mortality with ICS (hazard ratio 0.73; 95% CI: 0.55–0.96; p = 0.039), and the TORCH trial that found no reduction whatsoever with fluticasone [1, 7]. The seven trials that comprised the pooled analysis truncated patient follow-up when they discontinued the study drug. The pooled analysis in fact found no difference in mortality during the first nine months of follow-up, the time period where drop-outs were still rare, and thus most randomised patients were included in the mortality analysis, while the apparent benefit of ICS only became visible after nine months [11]. On the other hand, the TORCH trial that followed all patients up for three years found a hazard ratio of mortality for fluticasone relative to placebo of 1.06 (95% CI: 0.89–1.27; p = 0.53) [1]. Such a disagreement between two large mortality studies is probably a direct result of the follow-up process and a proper intent-to-treat analysis in the TORCH study.

Another aspect of the intent-to-treat analysis is the time period under study. In the TORCH study, the treatment was intended to be taken for three years and the analysis was based on all deaths occurring exactly during this period. Thus, deaths that occurred at three years plus one day were correctly not counted in the analysis. In the UPLIFT trial, the treatment was taken for four years (1,440 days) so that the proper intent-to-treat analysis should have been based on all deaths occurring exactly during this four-year period (hazard ratio 0.87; 95% CI: 0.76–0.99). However, the study design also involved a post-study follow-up where, at the end of the study, all patients were provided with (and asked to take) 40 μg of ipratropium four times daily and to return for a final assessment 30 days later. The authors incorrectly based their primary intent-to-treat analysis of mortality on the deaths occurring during the period of four years plus 30 days (1,470 days) which led to a hazard ratio of 0.89 (95% CI: 0.79–1.02).

Intent to treat: lung function

Decline in lung function over time is a fundamental measure of disease progression among patients with COPD. FEV_1 decline was in fact the primary outcome measure for the UPLIFT and many other randomised controlled trials evaluating whether pharmacological treatments could modify the natural history of COPD. The TORCH trial also considered this outcome and reported in a secondary paper that the yearly decline in FEV_1 was significantly slower with fluticasone, salmeterol, or both, compared to a placebo [12, 13].

The TORCH and UPLIFT trials were not designed for a full intent-to-treat analysis of lung func-

tion decline, but only measured FEV_1 until the patients discontinued treatment. For example, in the TORCH trial, of the 6,112 patients in the study, the lung function analysis involved only 5,343 subjects with at least one measurement of post-bronchodilator FEV_1 made up to twice yearly during follow-up. Consequently, 10,133 FEV_1 measurements were missing from a possible 36,672 measurements that the study could have yielded. This may cause bias from regression to the mean [14, 15].

Such a bias occurs for several reasons. Firstly, some patients are excluded altogether from the analysis. Nearly 18% of patients allocated to a placebo did not contribute a single FEV_1 value because they discontinued placebo before the first six-month visit when the first FEV_1 was measured. It is likely that these excluded patients would have had poorer FEV_1 values at that first visit had they been available to be measured. Thus, the slope of decline in the remaining subjects with better FEV_1 values at the first visit is probably exaggerated by regression to the mean. Secondly, discontinuing the follow-up of patients who have the initial FEV_1 value measured but are missing some subsequent values can also alter the slope of decline in FEV_1. In the TORCH study, the placebo patients who discontinued before the end of follow-up had a faster decline in FEV_1 (76 mL/year) than those completing the trial (54 mL/year) [12]. Here again, these slopes of decline may have been affected by regression to the mean.

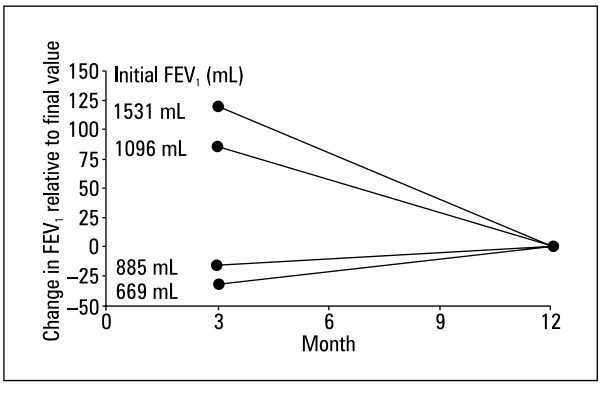

Figure 1. Depiction of regression to the mean: correlation between initial FEV_1 value (divided in quartiles) and the change in FEV_1 from months 3 to 12 (mean change in each quartile). Patients in the highest quartile of initial FEV_1 values show the largest decline (119 mL), while patients in the lowest quartile of initial FEV_1 values show an improvement of 32 mL

To illustrate the regression to the mean phenomenon, and its resulting bias, we used data on 322 subjects from the Canadian OPTIMAL trial, a three-arm randomised trial of 449 patients with moderate or severe COPD [9]. For illustration purposes, FEV_1 decline was simply measured as the difference in FEV_1 between the 12-month and three-month visits. The 322 patients had a mean FEV_1 of 1,131 mL at three months, with a change in FEV_1 from three to 12 months of 39 mL. Figure 1 depicts the regression to the mean phenomenon by showing that patients in the highest quartile of initial FEV_1 values (> 1,440 mL) have the largest decline (mean decline 119 mL), while the patients in the lowest quartile of initial FEV_1 values (< 770 mL) in fact show an improvement of 32 mL [15].

To illustrate this impact on the TORCH study results, we assumed that the 18% of patients who were missing from the TORCH placebo group were more likely to be the sicker patients with the lowest FEV_1. Thus, excluding the 18% of patients with the lowest FEV_1 at visit 1 (< 700 mL) from the 322 patients, the one-year rate of FEV_1 decline among the remaining subjects becomes 52.2 mL; a clear overestimate of the decline from the 39 mL from the complete data. This is because the 18% excluded patients have a mean *increase* in FEV_1 of 40.7 mL.

Factorial analysis of TORCH data

While the TORCH study aimed to compare the combination of fluticasone and salmeterol with a placebo, the study also included a fluticasone only and a salmeterol only arm [1]. This study was thus

Table 1. Factorial regression analysis of TORCH data to estimate the independent effects of fluticasone and salmeterol on the three-year incidence of all-cause mortality

	Medication allocated		Adjusted RR (95% CI)	p-value
	Yes (Deaths/N)	No (Deaths/N)		
Medication				
Fluticasone	439/3,067	436/3,045	1.00 (0.89–1.13)	0.9918
Salmeterol	398/3,054	477/3,058	0.83 (0.74–0.95)	0.0043

structured as a 2 × 2 factorial design of fluticasone (yes/no) and salmeterol (yes/no). However, TORCH was not analysed as a factorial trial, thus wasting much needed power and denying the reader important information about the independent contribution of each component of the combination [10, 16].

As mortality was the only outcome ascertained in a complete manner for a proper intent-to-treat analysis, mortality was used to perform the analysis corresponding to a 2 × 2 factorial design. This factorial analysis must be done using a regression model, in this case a generalised linear regression model with a binomial distribution, to estimate the three-year mortality rate ratio associated with fluticasone and salmeterol [10, 16]. The interaction term to assess whether there is synergy between the two drugs was found to be non-significant ($p = 0.32$) suggesting that a combination of fluticasone and salmeterol is not particularly more effective than the two components added independently. Table 1 presents the rates and the independent effects of fluticasone and salmeterol on mortality, namely adjusted for each other. While the salmeterol component is associated with a significant 17% reduction in mortality (rate ratio 0.83; 95% CI: 0.74–0.95; $p = 0.0043$), the fluticasone component provides no reduction whatsoever (rate ratio 1.00; 95% CI: 0.89–1.13; $p = 0.9918$) [10].

Conclusion

The randomised controlled trial is and will remain the fundamental tool to evaluate the benefit of COPD treatments. Its proper conduct, however, including the most rigorous study design and data analysis, is essential if it is to produce valid results.

The TORCH and UPLIFT trials have provided important lessons in this context. First and foremost, the study question must be answerable by the study design. We noted that the question in the TORCH study aimed at a comparison with 'usual care', but the placebo group treatment was not 'usual care'. TORCH and UPLIFT were among the first trials in COPD to follow the intent-to-treat principle, which is fundamental for randomised controlled trials to avoid bias. However, while this principle was followed for the mortality outcome, it was not followed for lung function decline, where patients were only measured until they discontinued study medications. As a result, the findings in both trials relating to lung function decline are subject to bias from regression to the mean.

Finally, the TORCH study, designed as a 2 × 2 factorial trial to assess the effects of an inhaled corticosteroid and a long-acting beta-agonist, should have exploited fully the data by using the corresponding data analysis. This factorial analysis shows that a mortality benefit is entirely accounted for by the effect of salmeterol, with no effect attributable to the inhaled corticosteroid component of the combination therapy.

As COPD is one of the major causes of morbidity and mortality worldwide, mega trials such as TORCH and UPLIFT are clearly needed, but must be designed and analysed with the utmost scientific rigour.

References

1. Calverley P.M., Anderson J.A., Celli B. et al. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. N. Engl. J. Med. 2007; 356: 775–789.
2. Tashkin D.P., Celli B., Senn S. et al. A four-year trial of tiotropium in chronic obstructive pulmonary disease. N. Engl. J. Med. 2008; 359: 1543–1554.
3. Highland K.B., Strange C., Heffner J.E. Long-term effects of inhaled corticosteroids on FEV1 in patients with chronic obstructive pulmonary disease. A meta-analysis. Ann. Intern. Med. 2003; 138: 969–973.
4. Sutherland E.R., Allmers H., Ayas N.T., Venn A.J., Martin R.J. Inhaled corticosteroids reduce the progression of airflow limitation in chronic obstructive pulmonary disease: a meta-analysis. Thorax 2003; 58: 937–941.
5. Alsaeedi A., Sin D.D., McAlister F.A. The effects of inhaled corticosteroids in chronic obstructive pulmonary disease: a systematic review of randomized placebo-controlled trials. Am. J. Med. 2002; 113: 59–65.
6. Suissa S. Statistical treatment of exacerbations in therapeutic trials of chronic obstructive pulmonary disease. Am. J Respir. Crit. Care Med. 2006; 173: 842–846.
7. Sin D.D., Wu L., Anderson J.A. et al. Inhaled corticosteroids and mortality in chronic obstructive pulmonary disease. Thorax 2005; 60: 992–997.
8. Szafranski W., Cukier A., Ramirez A. et al. Efficacy and safety of budesonide/formoterol in the management of chronic obstructive pulmonary disease. Eur. Respir. J. 2003; 21: 74–81.
9. Aaron S.D., Vandemheen K.L., 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.
10. Suissa S., Ernst P., Vandemheen K.L., Aaron S.D. Methodological issues in therapeutic trials of COPD. Eur. Respir. J. 2008; 31: 927–933.
11. Ernst P., Suissa S. Inhaled corticosteroids and mortality in COPD. Thorax 2006; 61: 735.
12. Celli B.R., Thomas N.E., Anderson J.A. et al. Effect of pharmacotherapy on rate of decline of lung function in chronic obstructive pulmonary disease: results from the TORCH study. Am. J. Respir. Crit. Care Med. 2008; 178: 332–338.
13. Suissa S. Medications to modify lung function decline in chronic obstructive pulmonary disease: some hopeful signs. Am. J. Respir. Crit. Care Med. 2008; 178: 322–323.
14. Bland J.M., Altman D.G. Some examples of regression towards the mean. Br. Med. J. 1994; 309: 780.
15. Suissa S. Lung function decline in COPD trials: bias from regression to the mean. Eur. Respir. J. 2008; 32: 829–831.
16. La Vecchia C., Fabbri L.M. Prevention of death in COPD. N. Engl. J. Med. 2007; 356: 2211–2212.