



OPINION

The INPULSIS enigma: exacerbations in idiopathic pulmonary fibrosis

Samy Suissa, Pierre Ernst

Department of Epidemiology and Biostatistics, Centre for Clinical Epidemiology, Jewish General Hospital, McGill University, Montreal, Canada

Correspondence to
Professor Samy Suissa, Centre for Clinical Epidemiology, Jewish General Hospital, 3755 Cote Ste-Catherine, H4.61, Montreal, Québec, Canada H3T 1E2; samy.suissa@mcgill.ca

Received 19 November 2014
Accepted 20 November 2014

INTRODUCTION

The recently published INPULSIS studies, two duplicate randomised trials evaluating nintedanib in the treatment of idiopathic pulmonary fibrosis (IPF) over 1 year, reported effectiveness in terms of declining lung function, but inconsistent results on exacerbations of IPF.^{1 2} Indeed, in the prespecified pooled analysis of the two trials, the risk of an investigator-reported acute exacerbation was lower by 36% with nintedanib compared with placebo but not statistically significant (HR, 0.64; 95% CI 0.39 to 1.05; $p=0.08$). On the other hand, another prespecified analysis found that, after adjudication of these same investigator-reported exacerbations, the risk of a confirmed acute exacerbation was lower by 68% and statistically significant (HR, 0.32; 95% CI 0.16 to 0.65; $p=0.001$). Such differences in results for the same outcome may puzzle readers and even possibly arouse suspicion of selective or incorrect reporting, especially when the findings straddle the magic 5% conventional level of statistical significance. We believe such differences have a statistical explanation which shows how one result is biased and the other valid.

In this paper, we describe the phenomenon of treatment effect dilution, which can lead to false non-significant effects, using the INPULSIS trials as an illustration. We also provide a contrasting example to this phenomenon using, as an illustration, the case of the risk of pneumonia in trials of inhaled corticosteroids in COPD.

ADJUDICATION OF EXACERBATIONS

An acute exacerbation of IPF in the INPULSIS trials needed to meet the following criteria: 'unexplained worsening or development of dyspnoea within the previous 30 days; new diffuse pulmonary infiltrates visualised on chest radiography, HRCT or both, or the development of parenchymal abnormalities with no pneumothorax or pleural effusion (new ground-glass opacities) since the preceding visit and exclusion of any known causes of acute worsening, including infection, left heart failure, PE and any identifiable cause of acute lung injury, in accordance with routine clinical practice and microbiological studies'.¹ In view of the complexity of applying this definition in routine practice, each investigator-reported exacerbation was verified by a central adjudication committee, blinded to the study-group assignment, and classified as a confirmed, suspected or not an acute exacerbation. This central adjudication exercise was surely desirable as the definition involved several

complex elements, and interpretation could easily vary across the 205 sites in 24 countries.

The INPULSIS trials involved 1061 IPF patients, of which 638 were randomised to nintedanib and 423 to placebo, all followed for 1 year. In all, 63 patients had at least one investigator-reported exacerbation, 31 under nintedanib and 32 under placebo, for a total of 69 exacerbations. In the adjudication process, two of these exacerbations had insufficient information and the remaining 67 were classified by the adjudication committee (table 1). Of the 67 adjudicated exacerbations, 31 were found not to be acute exacerbations, in other words, 'false' exacerbations. When comparing the two groups, the proportion of these false exacerbations was almost twice as high in the nintedanib group (59% of reported exacerbations) than in the placebo group (32% of reported exacerbations). At first glance, these major differences in the false rates give the appearance of an anomaly with the adjudication exercise, suggesting that it may have unduly favoured the nintedanib group by excluding more 'false' exacerbations and thus identifying fewer confirmed exacerbations in this group. This view of the data can be misleading.

STATISTICAL IMPACT OF ADJUDICATION

The assessment of the classification of exacerbations after adjudication shown in table 1 is based exclusively on the number of exacerbations. However, to properly evaluate the impact of misclassifying exacerbations also requires the denominator that generated these exacerbations, namely the number of patients in the trials. Table 2 displays the number of patients with exacerbations relative to the total number of patients, while providing our own calculation of the risks and risk ratios of exacerbations. It first shows that for investigator-reported acute exacerbations, the risk ratio comparing nintedanib with placebo is 0.64 (95% CI 0.40 to 1.04), with a p value of 0.07, while for the adjudicated exacerbations, the risk ratio is 0.40 (95% CI 0.20 to 0.81), with a significant p value of 0.01.

The most important information from this table, however, is the incidence of reported exacerbations adjudicated to be 'false' acute exacerbations, which is 3.0% in the nintedanib group compared with 2.8% in the placebo group. The similarity of these two false-positive rates in the active treatment and placebo groups confirms that the adjudication exercise was indeed valid in that patients in the trial had a roughly equal chance (around 3%) of showing signs of something that looked like an exacerbation but was in fact not so.

To cite: Suissa S, Ernst P. *Thorax* Published Online First: [please include Day Month Year] doi:10.1136/thoraxjnl-2014-206598

Table 1 Classification of the 69 investigator-reported exacerbations in the INPULSIS trials by the adjudication committee

Number (%)	Nintedanib	Placebo
Investigator-reported acute exacerbations	32	37
Confirmed/suspected exacerbation	12 (38)	24 (65)
Not an acute exacerbation	19 (59)	12 (32)
Insufficient data to adjudicate	1 (3)	1 (3)

Consequently, the analysis based on the investigator-reported acute exacerbations is not statistically significant ($p=0.07$) because of the inaccuracy introduced by the random addition of non-events to this outcome measure, leading to a dilution of the drug effect. On the other hand, the analysis of the adjudicated exacerbations is statistically significant ($p=0.01$) because it removed these false exacerbations equally (around 3%) in each group.

PNEUMONIAS IN COPD

The increased risk of pneumonia associated with inhaled corticosteroids in COPD is now well established from several randomised trials and meta-analyses.^{3 4} One of these, the Investigating New Standards for Prophylaxis in Reduction of Exacerbations (INSPIRE) trial, conducted a further investigation of all pneumonia adverse events.⁵ The trial involved 1323 patients with COPD randomised to the fluticasone–salmeterol combination ($N=658$) or tiotropium ($N=665$) and followed for 2 years. A total of 87 adverse events of pneumonia were reported, including 62 under fluticasone–salmeterol and 25 under tiotropium, an over twofold increase in the risk of pneumonia with fluticasone–salmeterol. Of the 87 reported pneumonia events, 64 had had a chest radiograph performed, of which 50 showed the presence of infiltrates consistent with the diagnosis of pneumonia.

This study offers another perspective into the phenomenon of treatment effect dilution by providing an example where the effect is not diluted. **Table 3** shows that, in contrast with **table 1**, the proportion of reported pneumonias that are confirmed by chest radiograph is equivalent in the two groups. **Table 4** displays the rates and rate ratios of pneumonias, comparing fluticasone–salmeterol with tiotropium. It shows that the rate ratio for the reported pneumonias is 2.3 (95% CI 1.5 to 3.7), with a p value of 0.0004, practically identical to that of the radiograph-confirmed pneumonias (rate ratio 2.4; 95% CI 1.3 to 4.5), with a p value of 0.0054.

Moreover, **table 4** also shows that the incidence rates of reported pneumonias that were not confirmed by chest radiograph and those for which no radiograph data were available

Table 3 Classification of the 87 investigator-reported pneumonia events in the Investigating New Standards for Prophylaxis in Reduction of Exacerbations (INSPIRE) trial by chest radiographic confirmation

Number (%)	Fluticasone–salmeterol	Tiotropium
Investigator-reported pneumonias	62	25
Chest radiograph confirmation	36 (58)	14 (56)
Not confirmed by chest radiograph	11 (18)	3 (12)
No chest radiograph	15 (24)	8 (32)

are also higher with fluticasone–salmeterol than with tiotropium. This systematic difference suggests that the reported pneumonias diagnosed clinically were likely true pneumonias that the chest radiograph may not have picked up but could have been seen with further imaging. It also refutes the notion that many of these pneumonia adverse events not confirmed radiologically could have been unresolved exacerbations.⁵ Indeed, had they instead been actual COPD exacerbations, the rate of these false pneumonias would have been equal between fluticasone–salmeterol and tiotropium, as was found for the confirmed COPD exacerbations in the trial (rate ratio 1.0; 95% CI 0.8 to 1.1).

CONCLUSION

The phenomenon of treatment effect dilution is evident from the INPULSIS trials of the effectiveness of treatment with nintedanib in reducing the risk of exacerbation in IPF. It explains the difference in findings from the analysis based on investigator-reported acute exacerbations, not statistically significant before adjudication, but statistically significant afterwards. We showed that such puzzling differences have a statistical explanation. Indeed, the inaccuracy introduced by randomly adding false exacerbations to the outcome will lead to a dilution of the drug's effectiveness as measured by the rate ratio. The adjudication exercise, blinded to treatment assignment, removed these false events, thus reducing the inaccuracy which distorted statistical significance.

This phenomenon is not only important in evaluating a drug's effectiveness, but also in assessing its risk. Indeed, if the adverse outcome includes events that are not genuine, the phenomenon will produce a diluted measure of risk that can lead to incorrectly concluding that a drug is safe when it is not.

Using the example of the increased risk of pneumonia with inhaled corticosteroids in COPD, we also showed that the adjudication exercise needs to be complete and accurate. In the INSPIRE trial, a significant proportion of reported adverse events of pneumonia did not have a chest radiograph, as this

Table 2 Risk and risk ratio of acute investigator-reported exacerbation and of adjudicated exacerbation in the INPULSIS trials comparing nintedanib with placebo

Number (%)	Nintedanib	Placebo	Risk ratio (95% CI)	p Value
Patients	638	423		
Patients with at least one investigator-reported acute exacerbation	31 (4.9)	32 (7.6)	0.64 (0.40 to 1.04)	0.07
Patients with a confirmed/suspected acute exacerbation*	12 (1.9)	20 (4.7)	0.40 (0.20 to 0.81)	0.01
Patients with a 'false' acute exacerbation	19 (3.0)	12 (2.8)		

*The numbers of patients were not provided in the paper and were estimated from the corresponding rates of confirmed/reported acute exacerbations among all exacerbations.

Table 4 Rate and rate ratio of investigator-reported and radiograph-confirmed pneumonia events in the Investigating New Standards for Prophylaxis in Reduction of Exacerbations (INSPIRE) trial

Number (rate per 100/year)	Fluticasone–salmeterol	Tiotropium	Rate ratio (95% CI)	p Value
Patients	658	665		
Patient-years	1011	945		
Investigator-reported pneumonias	62 (6.1)	25 (2.6)	2.3 (1.5 to 3.7)	0.0004
Radiograph-confirmed pneumonias	36 (3.6)	14 (1.5)	2.4 (1.3 to 4.5)	0.0054
Unconfirmed by radiograph	11 (1.1)	3 (0.3)		
No radiograph data	15 (1.5)	8 (0.8)		

was an a posteriori investigation into this adverse effect. Nevertheless, the risk with inhaled corticosteroids remained elevated for those where the chest radiograph did not show a presence of infiltrates, implying that the non-confirmed pneumonias diagnosed clinically were likely true pneumonias and not COPD exacerbations.

Randomised trials are essential to evaluate a drug's effectiveness, but methodological issues beyond randomisation can affect

their findings. We describe the phenomenon of treatment effect dilution which can be surmounted with a scientifically rigorous adjudication exercise. Generating an accurate measure of the outcome under study, uncluttered by false outcome events, will ensure an equally accurate estimate of the drug's effect that can be applied to better clinical practice.

Competing interests SS has participated in advisory board meetings or as speaker in conferences for AstraZeneca, Boehringer-Ingelheim, Novartis and Merck and has received research funding from Boehringer-Ingelheim.

Provenance and peer review Not commissioned; internally peer reviewed.

REFERENCES

- 1 Richeldi L, du Bois RM, Raghu G, *et al.* Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2071–82.
- 2 Thickett DR, Kendall C, Spencer LG, *et al.* Improving care for patients with idiopathic pulmonary fibrosis (IPF) in the UK: a round table discussion. *Thorax* 2014;69:1136–40.
- 3 Crim C, Calverley PM, Anderson JA, *et al.* Pneumonia risk in COPD patients receiving inhaled corticosteroids alone or in combination: TORCH study results. *Eur Respir J* 2009;34:641–7.
- 4 Singh S, Amin AV, Loke YK. Long-term use of inhaled corticosteroids and the risk of pneumonia in chronic obstructive pulmonary disease: a meta-analysis. *Arch Intern Med* 2009;169:219–29.
- 5 Calverley PM, Stockley RA, Seemungal TA, *et al.* Reported pneumonia in patients with COPD: findings from the INSPIRE study. *Chest* 2011;139:505–12.