Concise Report

Regression to the mean using the disease activity score in eligibility and response criteria for prescribing TNF-α inhibitors in adults with rheumatoid arthritis

M. C. Greenwood, J. Rathi, A. J. Hakim, D. L. Scott and D. V. Doyle

Objectives. When patients with rheumatoid arthritis (RA) are selected to start TNF-α inhibitors on the basis of high disease activity scores (DAS), some of the fall in DAS will be due to regression to the mean (RTM). We have assessed the extent to which such RTM explains DAS improvements on TNF-α inhibitors in routine clinical practice.

Methods. We retrospectively evaluated DAS28 scores that had been recorded as part of routine assessment for two RA cohorts. (i) Thirty-five patients receiving TNF-α inhibitors who had been assessed when starting TNF-α inhibitors, 9–21 months prior and 1.5–6 months post-treatment. (ii) One hundred and seventy-seven clinic patients assessed twice, a year apart in the years immediately before the introduction of TNF-α inhibitors.

Results. In patients receiving TNF-α inhibitors, mean DAS fell 1.8 (95% confidence interval [CI] 1.3, 2.3) from the previous routine assessment. Twenty-four (69%) patients showed a fall in DAS of >1.2 from baseline but only 17 (49%) from the previous assessment. Regression analysis of results from the pre-biological era estimated that as much as 0.6 of the 1.8 apparent DAS response to TNF-α inhibitors might be accounted for by RTM.

Conclusions. Assessing change in DAS from commencement of biological therapy may overestimate response, due to the impact of RTM and fluctuation in disease. Adequacy of response might be better assessed by serial assessments and a wider range of patient-centred outcomes.

KEY WORDS: Disease activity score, Rheumatoid arthritis, Regression to the mean, TNF-α, Clinical guidelines.

Introduction

The use of TNF-α inhibitors to treat rheumatoid arthritis (RA) follows guidance from professional groups such as the British Society for Rheumatology Standards, Guidelines and Audit Working Group (SGAWG) and regulatory bodies such as the National Institute for Health and Clinical Excellence (NICE) [1]. UK patients starting biological treatment need to have two disease activity scores (DAS) 28 scores over 5.1 to show on-going active disease prior to starting treatment and also to show an adequate response indicated by a fall in DAS of at least 1.2 in order to receive continuing treatment; other countries use less stringent criteria.

The increasing use of observational data, particularly results from the British Society For Rheumatology Biologics Register (BSRBR), to assess and potentially justify the use of biological agents, needs to take into account the universal phenomenon of regression to the mean (RTM), which will be most marked when patients have been selected for high initial scores [2–4]. RTM in these circumstances partly reflects the relatively high measurement error in each component of DAS and partly reflects fluctuations in disease activity due to the remitting nature of RA. Consequently, patients selected for assessment at a time of high disease activity will inevitably show reductions in disease activity with time. An initial examination of plots of routinely assessed DAS scores suggested a pattern of particularly high scores at commencement of biological therapy (see examples in Fig. 1) and raised the question of whether the apparent response had been inflated by uncharacteristically high baseline scores. We have examined the effect of such RTM in patients receiving TNF-α inhibitors in routine clinical practice, investigating whether DAS was uncharacteristically high at commencement of therapy and assessing the magnitude of score reductions observed for patients with equally high DAS scores in the pre-biological era.

Patients and methods

The analyses made use of datasets derived from the regular assessment of DAS as part of the routine care for patients with RA attending the rheumatology clinic of a district general hospital, Whips Cross University Hospital (WX) and a teaching hospital, Kings College Hospital (KCH).

Patients receiving TNF-α inhibitors (dataset 1)

These comprised 35 patients (24 at WX and 11 at KCH) who had been assessed at commencement (baseline), 9–21 months prior to starting their first TNF-α inhibitor (previous routine assessment) and at 1.5–6 months into treatment (follow-up). Their mean age at the start of therapy was 58 yrs (range 31–78) and mean disease duration 17 yrs (range 3–45). Twenty-nine (83%) were women. Eighteen commenced infliximab, 15 etanercept and two adalimumab.

The previous routine assessment of DAS, unlike the score at baseline, had not been timed to coincide with disease symptoms or high disease activity so would provide an estimate of the group mean DAS score in the previous year.

Score changes on biological therapy, and the proportion achieving a 1.2 or more point reduction, were calculated as change from baseline and also as change from routine assessment.

Repeat assessments from the pre-biological era (dataset 2)

These comprised WX 177 patients with RA for whom DAS had been recorded on two occasions, 12 months apart, in the years...
shortly before the introduction of biological therapy. Their mean age at first assessment was 63 yrs (range 35-88) and disease duration 11 yrs (range 1-42). One hundred and twenty-eight (72%) were women (Fig. 2).

The least squares linear regression technique was applied to derive the line of best fit to predict mean follow-up score from initial score. This was used to estimate the mean annual score reduction that might have been anticipated for a group of patients with mean DAS equal to that of the patients starting TNF-α inhibitors, an estimate of the score reduction that might have otherwise been expected from pre-existing treatments and RTM.

The mean initial and follow-up DAS score of subgroups of patients with initial scores over 5.1, 5.5 and 6.0 were also calculated directly.

Ethical issues

The prospective collection of clinical data on patients receiving TNF-inhibitors was collected as part of the guidance for their use from the NICE. The collection of data on patients seen routinely was undertaken as part of routine clinical care. The analysis of the data from these two sources met the criteria for service evaluation as defined by the Central Office for Research Ethics Committees in the UK. Current UK guidance is that such service evaluations, which answer questions about the standards, obtained from treatment and overlap with clinical audit, do not require approval from Research Ethics Committees.

Results

Patients receiving TNF-α inhibitors (dataset 1)

Mean DAS was 5.5 (95% confidence interval 5.1, 6.0) at previous routine assessment, 6.4 (95% CI 6.0, 6.8) at baseline and 4.6 (95% CI 4.2, 5.0) on TNF inhibitors.

The mean DAS score reduction on biological treatment was 1.8 (95% CI 1.3, 2.3) when compared with baseline, and 0.9 (95% CI 0.4, 1.4) when compared with the previous routine assessment. A total of 24 (69%) patients achieved a reduction in DAS28 of >1.2 from baseline but only 17 (49%) achieved this level of improvement from the previous routine assessment.

Within the period from 9 to 21 months prior to baseline there was no evidence that the length of time between previous assessment and baseline influenced the analysis; there was no significant correlation between previous to baseline score difference and time between previous and baseline assessments (Pearson correlation coefficient 0.273, two-tailed sig. 0.112).

Repeat assessments from the pre-biological era (dataset 2)

Using dataset 2, the equation ‘Follow up DAS = 0.71 x initial DAS + 1.25’ was derived by least squares regression. This predicted, for any initial DAS score, the mean score that might have been anticipated after a year of routine care in the pre-biological era at WX hospital. The adjusted $r^2$ was 0.54. The RTM effect was evident in a slope of <1 (95% CI 0.61–0.81) and the intercept of >0 (95% CI 0.83–1.67). For a group of WX patients in the pre-biological era with initial scores of 6.4 (i.e. the baseline mean DAS seen at commencement of TNF-α inhibitors in dataset 1), a mean score 5.8 at follow-up would have been predicted, equating to a reduction of 0.6.

A subgroup of 40 WX patients had initial DAS scores of >5.1. Their mean DAS fell over the year by 0.4 from 5.8 to 5.4. For a subgroup of 24 with scores >5.5 and 11 with scores >6.0, the means fell by 0.7 and 0.9, respectively.

This dataset provided no evidence that the level of DAS fluctuation was related to disease duration and therefore no evidence that the disease duration differences between the two datasets would give rise to differences in the level of DAS fluctuation. The Pearson correlation coefficient for disease duration and absolute score difference was 0.012 (two-tailed sig. 0.871).

Discussion

There is a strong theoretical basis for the existence of the RTM effect when using the same measure to both select patients for treatment and assess benefit. These two separate analyses of data drawn from routine practice, though differing in approach, both demonstrate a significant RTM effect for DAS and raise concern about the advisability of using it to define both eligibility and response to treatment.

We found that mean DAS28 fell from 6.4 to 4.6 within 6 months of starting TNF-α inhibitors. The BSRBR reported similar data throughout the UK; mean initial DAS28 scores were 6.7 and they fell to 4.6 with therapy [5]. Experience in Sweden is broadly similar [6]. We also estimated that the mean DAS of patients with equally high initial scores would have fallen by 0.6 after a year of routine clinical care in the pre-biological era, equivalent to recent changes in placebo-treated cases in a very recent biological study reported by Emery et al. [7]; DAS scores fell a mean of 0.7 with placebo from an initial value of 6.8. Interestingly, Maini et al. [8] reported even larger falls in DAS28 with placebo treatment. On balance, we consider following the...
UK approach to treatment with TNF-\(\alpha\) inhibitors results in an overall mean fall in DAS28 of around 1.8, of which as much as 0.6 may be attributable to RTM.

These results raise two important issues. First, they impact on the question of switching TNF-\(\alpha\) inhibitor, an issue that has been the subject of considerable debate in updating NICE guidance for the use of TNF-\(\alpha\) inhibitor. The evidence base in support of switching a non-responder to an alternative TNF-\(\alpha\) inhibitor is incomplete and at least one report by Bennett and colleagues [9] shows that DAS28 falls by only 0.6, which is most likely entirely due to RTM. This issue remains highly controversial [10, 11] and it is an area in which further randomized controlled trials are needed, as the effect of RTM may make it difficult to interpret observational data.

Secondly, individual variability in the RTM effect limits the information which can be inferred from individual DAS28 score changes. This together with doubts as to the ability of the DAS28 formula to accurately reflect the disease activity of individual cases raises concern as to the advisability of using DAS28 to guide individual treatment decisions. DAS was originally developed using discriminant analysis and multiple regression to derive a formula that best reflected clinicians’ decisions to adjust treatment [12]; this was later extended to DAS28 [13] by Van der Cruyssen et al. [14]. However, a formula that gives a good overall fit for a group of patients will not necessarily reflect the particular balance of factors in an individual. Gardiner et al. [15] have pointed out that due to the logarithmic nature of its contribution to DAS, small changes of erythrocyte sedimentation rate (ESR) that would be considered clinically insignificant can make large changes in score. They suggest caution using DAS as the main response criterion in an individual patient when it is not representative of the clinical situation. Mäkinen et al. [16] have reported that the ESR had a low predicative power for remission while the predictive power of pain was high. Reliance on DAS28 alone for determining remission, or for eligibility or response to therapy may overemphasize the importance of ESR while excluding other important factors such as pain. A much wider range of outcomes such as pain, quality of life and fatigue should be taken into account.

Serial assessment of DAS pre- and post-treatment as suggested by Fransen et al. [17] and den Broeder et al. [18], considered together with other outcomes, may still prove useful in supplementing rather than replacing clinical judgement. However, in view of RTM and individual variability, it is inadvisable to determine individual patient treatment decisions using just one measure and simple cut-off points for eligibility and response.

**Rheumatology key messages**
- If patients with RA are selected for a new treatment when their disease is particularly active, the DAS-treatment response is likely to be inflated by a degree of spontaneous improvement.
- The fluctuating nature of RA disease activity and the measurement error inherent in DAS predispose to a RTM effect when patients have been selected for high initial scores.
- The adequacy of an individual’s response to biological therapy cannot be determined solely by DAS score change from commencement of therapy.

**Acknowledgements**

We are pleased to acknowledge financial support for this study from the Arthritis Research Campaign (Programme Grant S0682 and Integrated Clinical Arthritis Centre Grant P0572) and from National Health Service R&D Support Funding to Whips Cross Hospital and Kings College Hospital.

D.L.S. reports having been paid lecture fees by Wyeth. His department has received funding for clinical trials from and Roche and unrestricted clinical research grants from Amgen.

**References**