



HAL
open science

Early non-response to certolizumab pegol in rheumatoid arthritis predicts treatment failure at one year. Data from a randomised Phase III clinical trial

Francis Berenbaum, Thao Pham, Pascal Claudepierre, Thibault de Chalus, Jean-Michel Joubert, Carine Saadoun, Lionel Riou França, Bruno Fautrel

► To cite this version:

Francis Berenbaum, Thao Pham, Pascal Claudepierre, Thibault de Chalus, Jean-Michel Joubert, et al.. Early non-response to certolizumab pegol in rheumatoid arthritis predicts treatment failure at one year. Data from a randomised Phase III clinical trial. *Joint Bone Spine*, 2017, 10.1016/j.jbspin.2017.01.011 . hal-01472840

HAL Id: hal-01472840

<https://hal.sorbonne-universite.fr/hal-01472840v1>

Submitted on 21 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Early non-response to certolizumab pegol in rheumatoid arthritis predicts treatment failure at one year. Data from a randomised Phase III clinical trial

Francis Berenbaum ^a, Thao Pham ^b, Pascal Claudepierre ^c, Thibault de Chalus ^d, Jean-Michel Joubert ^d, Carine Saadoun ^d, Lionel Riou França ^e, Bruno Fautrel ^f

^a INSERM UMRS 938, UPMC, Paris 6 University, Inflammation-Immunopathology-Biotherapy Department (DHU i2B), Department of Rheumatology, Assistance Publique-Hôpitaux de Paris, Saint-Antoine Hospital, 75012 Paris, France

^b Aix Marseille University, Rheumatology Department, 13284 Marseille, and AP-HM Sainte-Marguerite hospital, 13009 Marseille, France

^c AP-HP, Henri Mondor hospital, Rheumatology Department, and Paris Est Créteil University, Laboratoire d'Investigation Clinique (LIC) EA4393, 94010 Créteil, France

^d UCB Pharma, 92700 Colombes, France

^e Institut Phisquare, Fondation Transplantation, 75015 Paris, France

^f Paris 6 University-GRC UPMC-08; AP-HP, Rheumatology Department, 75005 Paris, and Groupe Hospitalier Pitié Salpêtrière, 75013 Paris, France

Corresponding author:

Francis Berenbaum

Rheumatology Department, Hôpital Saint-Antoine

184 rue du Faubourg Saint-Antoine

75012 Paris, France

Tel.: +33.1.49.28.33.60

Email: Francis.berenbaum@sat.aphp.fr

Abstract

Objectives. To compare different early clinical criteria of non-response determined at three months as predictors of clinical failure at one year in patients with rheumatoid arthritis starting therapy with certolizumab pegol.

Methods. Data were derived from a randomised Phase III clinical trial in patients with rheumatoid arthritis who failed to respond to methotrexate monotherapy. Patients included in this post-hoc analysis were treated with certolizumab pegol (400 mg *qd* reduced to 200 mg *qd* after one month) and with methotrexate. The study duration was twelve months. Response at three months was determined with the American College of Rheumatology-50, Disease Assessment Score-28 ESR, Health Assessment Questionnaire and the Clinical Disease Activity Index. The performance of these measures at predicting treatment failure at twelve months defined by the American College of Rheumatology-50 criteria was determined, using the positive predictive values as the principal evaluation criterion.

Results. 382 patients were available for analysis and 225 completed the twelve-month follow-up. At Week 52, 149 (38.1%) patients met the American College of Rheumatology-50 response criterion. Positive predictive values ranged from 81% for a decrease in Health Assessment Questionnaire- disability index score since baseline >0.22 to 95% for a decrease in Disease Assessment Score-28 score since baseline ≥ 1.2 . Sensitivity was $\leq 70\%$ in all cases. Performance of these measures was similar irrespective of the definition of treatment failure at 12 months.

Conclusions. Simple clinical measures of disease activity can predict future treatment failure reliably and are appropriate for implementing treat-to-target treatment strategies in everyday practice.

Keywords: Rheumatoid Arthritis, Anti-TNF, DMARDs (biologic), Disease Activity.

Accepted Manuscript

1. INTRODUCTION

Rheumatoid arthritis (RA) is a progressive destructive autoimmune disease of the joints with a prevalence of around 0.5% in the general population.[1-3] The disease course is typically characterised by episodic acute exacerbations separated by periods of relative disease stability. Recovery from exacerbations is rarely complete, resulting in progressive irreversible structural damage to the affected joints.[4]

The goals of treatment are to relieve symptoms, to prevent exacerbations and to slow structural joint damage. Long-term maintenance therapy is based on the use of disease-modifying antirheumatic drugs (DMARDs), a class that includes both synthetic drugs and biological agents. Considerable between-patient heterogeneity in response to DMARDs appears to exist, with ACR50 response rates in randomised clinical trials being generally less than fifty percent [5] and even lower in everyday clinical practice.[6] The recommended treatment strategy is to initiate treatment with DMARDs as early as possible after diagnosis, to monitor patients regularly using composite measures of disease activity and, in case of inadequate response, to switch to another DMARD in order to achieve or maintain tight control of disease activity.[7, 8] Such strategies have been shown to be highly effective.[9-13] Current practice guidelines recommend initiating treatment, typically with methotrexate, and re-evaluating the treatment at three-monthly intervals, and in case of non-response, switching to another synthetic DMARD or to a biological agent [7, 14] or moving to combination therapy [15] in order to achieve tight disease control.

In order to follow such a strategy, rheumatologists need to know the most appropriate clinical information to assess treatment response or failure. Although the American College of Rheumatology (ACR) [14] and the European League Against Rheumatism (EULAR) [7] both emphasise in their practice guidelines the importance of timely adaptation of treatment in case of inadequate response, neither provide any explicit guidance on the most appropriate criteria for defining treatment failure. In everyday practice, large improvements in disease status are unambiguous and the physician is reassured about maintaining the current treatment. However, if clinical evolution following treatment initiation is modest, physicians may adopt a strategy of 'watchful waiting' which may be detrimental to long-term prognosis rather than taking a more proactive approach.[16]

A number of studies have investigated the association between early measures of clinical outcome and long-term treatment response.[17-22] However, from the therapeutic perspective of clinical practice, the issue is not so much to predict treatment response but to predict treatment failure as accurately and as early as possible so that therapy can be adjusted accordingly and the overall long-term therapeutic outcome optimised. The sensitivity and specificity of a given predictive markers will not be the same with respect to treatment failure as with respect to treatment response. In particular, a marker that will predict treatment response with high sensitivity but with low specificity will predict failure with low sensitivity and high specificity. For this reason, it is important to identify reliable and practical markers of treatment failure.

To our knowledge, there is little information available on the comparative performance of potential early measures of treatment failure. A recent study of infliximab has suggested that

the combination of disease activity after six weeks and infliximab serum trough levels may be a useful as a predictor of treatment failure after six months of therapy.[23] A *post hoc* analysis of a study of certolizumab pegol [24] has also demonstrated that a change in DAS-28 score of <1.2 during the first three months of the study was predictive of failure to achieve low disease activity at one year. It remains to be demonstrated what is the most appropriate early clinical measure to determine treatment failure. The objective of the present study was to quantify and compare the positive predictive values (PPV) of different validated clinical criteria of non-response determined at three months with respect to clinical failure at one year in patients with RA starting biological anti-TNF α therapy with certolizumab pegol.

2. METHODS

Design

This study corresponds to a *post-hoc* analysis of data from the RAPID 1 [25] prospective, randomised, placebo-controlled, double-blind Phase III clinical trial, conducted between February 2005 and October 2006, which evaluated the efficacy and safety of certolizumab pegol in adult patients with RA. The methodology of this trial has been described in detail elsewhere [25] and is briefly summarised below.

The RAPID 1 trial included 982 adult patients fulfilling the 1987 ACR criteria for RA [26] who had been diagnosed at least six months previously and who were treated with methotrexate at a stable dose of ≥ 10 mg/week for at least two months, but had failed to respond adequately to methotrexate treatment for at least six months. Patients were randomised 2:2:1 to receive treatment with subcutaneous certolizumab pegol at an initial dosage of 400 mg given at weeks 0, 2, and 4, with a subsequent dosage of 200 mg or 400 mg given every 2 weeks, plus MTX, or placebo plus MTX. The planned treatment duration was one year. However, patients who failed to achieve a response according to the ACR20 criteria at weeks 12 and 14 were designated treatment failures and were withdrawn from the study at week 16. For patients who were considered to respond inadequately, the investigator could remove the patients from the study at any moment and switch them to open-label high maintenance dose (400 mg) certolizumab pegol.

Study population

The present *post-hoc* analysis was performed in the patients originally randomised to the 200 mg maintenance dose of certolizumab pegol, this being the approved dose regimen for the treatment of RA. All patients with no prior TNF α -inhibitor exposure were included in the analysis, irrespective of whether patients continued treatment throughout the duration of the study or not.

Outcome measures

In our analysis, outcome during the twelve weeks following initiation of treatment with certolizumab pegol and at one year was assessed. The choice of a period definition for the early non-response criterion was made in order not to categorise early responders who responded at some time before twelve weeks but failed to fulfil the response criterion at the twelve-week evaluation point as early non-responders, and the choice of the twelve-week period was decided based on the fact that clinical response is usually achieved within 12 weeks of treatment.[25] The choice of a point definition for treatment failure at one year is consistent with standard methods to determine symptomatic outcome in clinical trials in RA.[27] The assessment criteria were ACR 50 and ACR70 response rates,[28] change in DAS-28 score,[29] change in Health Assessment Questionnaire (HAQ) score,[30] and a simplified composite measure, the Clinical Disease Activity Index (CDAI).[31, 32] Early non-response was defined as the response criterion not being met at any of the evaluations up to the twelfth week of the study and treatment failure (late non-response) as the criterion not being met at the last evaluation at Week 52.

Statistical analysis

The analysis was performed on all patients randomised to the 200mg arm and had no prior TNF α -inhibitor exposure. Patients with efficacy data missing at Week 12 were excluded from the analyses. For the main analysis, efficacy data missing at the Week 52 visit were imputed using non-responder imputation (NRI). Patients who received rescue medication or who withdrew for any reason between Week 12 and Week 52 were considered non-responders. Sensitivity analyses were performed in which missing data were imputed using the last observation carried forward (LOCF) method or were not imputed at all (observed cases analysis). In the observed cases analysis, patients that left the clinical trial were excluded from the analysis, and only patients that completed the 52 weeks with valid measurements by week 12 and at week 52 for the respective criteria were considered.

The predictability of different early non-response measures by Week 12 with respect to treatment failure at Week 52 was assessed. The principal analysis was performed using failure to meet the ACR50 criteria at Week 52 as the definition of treatment failure, this being an outcome measure recommended in the European Medicines Agency's guidelines for clinical trials in RA.[27] Additional analyses were conducted using other definitions of treatment failure at Week 52. For each evaluation, sensitivity, specificity, PPV, negative predictive value (NPV), and accuracy were determined. The PPV, which in the present case represents the proportion of treatment failures (ACR50 response criterion at Week 52) who were identified as non-responders by Week 12, was considered the principal analysis criterion. The strength of the association was assessed with Matthew's correlation coefficient (MCC), generally considered a

good overall measure of accuracy.[33] A coefficient of 1 indicates a perfect prediction, a coefficient of 0 a no better than random prediction.

In order to assess whether the early non-response indicators were still predictive of subsequent treatment failure when taking into account the baseline characteristics of the patients, exploratory logistic regression models were fitted. The dependent variables were treatment failure (at week 52) using the ACR50, DAS-28 and CDAI criteria. The independent variables were early non-response (by week 12), age at baseline, DAS-28 at baseline and duration of disease at baseline. Models were fitted using a stepwise selection process with a probability threshold of 0.20. For any baseline characteristic retained in the model, the potential interaction with early non-response was also tested. Goodness of fit was assessed using the Hosmer and Lemeshow test.

3. RESULTS

Study population

Of the 393 patients who were originally randomised to the 200mg maintenance dose of certolizumab pegol, 11 were excluded since they had been previously exposed to a TNF α inhibitor. The study sample thus consisted of 382 patients randomised into the certolizumab pegol 200 mg maintenance dose group and with no prior exposure to TNF α inhibitor. Of these 382 patients, 96 (25.1%) were withdrawn from the study due to inadequate efficacy. These included 72 patients who failed to meet the ACR20 criterion at Week 12 and Week 14 and were thus withdrawn as specified in the protocol. In addition to these patients who discontinued the study for lack of efficacy, a further 41 patients withdrew from the study at some stage for other reasons, principally the occurrence of an adverse event (16 patients; 4.2%) or withdrawal of patient consent (15 patients; 3.9%). The characteristics of the study sample at inclusion are presented in Table 1. The majority of patients were women and the mean age of the sample was 51 years.

Treatment outcome

By Week 12, 49.3% of patients failed to meet the ACR50 criterion for response at least once between inclusion and Week 12 and were thus considered early non-responders. At Week 52, 61.5% of patients failed to meet this criterion and were thus considered treatment failures (Table 2).

Predictors of treatment non-response at Week 52

All early non-response criteria showed a high specificity of >80%, indicating that <20% of the patients who would fulfil the criterion for an ACR50 response at Week 52 would be incorrectly identified as non-responders at Week 12 (Table 3). Similarly, the positive predictive value was also high, >80%, indicating that >80% of patients identified as non-responders at Week 12 would fail to fulfil the ACR50 response at Week 52. The early non-response measures which performed best in terms of PPV were reduction in DAS-28 score ≤ 1.2 (95%) and CDAI score > 22 (93%) (Table 3). These were also the predictors which performed best with respect to specificity.

For all measures, sensitivity was poor to moderate ($\leq 70\%$) indicating that >30% of the patients who would eventually fail to fulfil the criterion for an ACR50 response at Week 52 would not be identified as non-responders at Week 12. The early non-response measures which performed best in terms of sensitivity were ACR50 response (70%) and CDAI score > 22 (40%). The NPV was also low (<65%), indicating that <65% of patients identified as responders at Week 12 would fulfil the ACR50 response criterion at Week 52. The MCC, which summarises overall

predictive power, ranged from 0.17 for HAQ-DI as the response criterion to 0.52 for the ACR50. The MCC was highest for the ACR50 (0.52) and the CDAI (0.39).

Sensitivity analyses were performed in which missing data were imputed by LOCF rather than NRI or not imputed (observed cases). With LOCF, the data obtained were qualitatively similar to the main analysis, and the estimates of PPVs in all cases within 5% of the estimates using NRI for data imputation (data not shown).

In the observed cases analysis, depending on the early non response measure, 216 to 219 patients had completed the 52 weeks treatment period and had valid measurements by Week 12 and at Week 52 for the considered criteria (i.e. approximately 164 patients were not considered in this analysis, representing 43% of the randomised set). In this analysis, the PPV estimates were systematically lower than in the NRI model (Table 4), although these still remained acceptable for the DAS-28 and the CDAI, especially if we consider that non responder patients that were withdrawn from the trial were not considered in this analysis.

In a next step, the performance of the early non-response measures was compared with respect to other definitions of treatment failure at Week 52 (Table 5). Specificity and PPVs remained high, irrespective of the definition of treatment failure chosen. In general, predictability was high when the same outcome measure was used for the early non-response measure and the treatment failure criterion. However, the relative performance of the early non-response measures with respect to PPV was the same regardless of the definition of treatment failure at Week 52.

Logistic regression analysis was performed in order to investigate patient characteristics at baseline and potential interactions affecting the predictability of treatment failure by the early non-response measures. Five different combinations of response criteria were evaluated in five separate models. These were early ACR50 non response *versus* ACR50 treatment failure, early DAS-28 non response *versus* DAS-28 treatment failure, early CDAI non response *versus* CDAI treatment failure, early DAS-28 non response *versus* CDAI treatment failure and early CDAI non response *versus* DAS-28 treatment failure. For all models, goodness of fit was high (Hosmer and Lemeshow test: $p > 0.3$). In all models, the early non-response criterion was significantly associated with treatment failure, with odds ratios ranging from 8 to 42. None of the baseline characteristics considered was consistently associated with treatment failure, although baseline DAS-28 score was retained in three logistic models and duration of disease at treatment initiation in one model. No evidence was obtained for an interaction between early non-response and baseline characteristics.

4. DISCUSSION

In this study, we demonstrate that the values of different early disease activity measures can be used to predict treatment failure in RA patients starting treatment with an anti-TNF α . Although all the measures performed acceptably in correctly predicting treatment failure, some differences were observed, notably with the HAQ-DI performing least well.

As has been observed previously by Curtis et al.,[19] several of the early measures tested are not entirely suitable for use in everyday clinical practice. For example, the ACR response criteria require computation of a panel of clinical and biological measures, which is time-consuming, and they are not systematically used in everyday clinical practice. The same is true to a lesser extent for the DAS-28.[34] In this context, simpler tools such as the SDAI or the CDAI have been proposed for following disease activity in routine practice.[31, 32, 35] It is thus pertinent that the CDAI performed well as an early non-response measure for subsequent treatment failure, with a PPV of 0.93 using the ACR50 at twelve months as the definition of treatment failure.

The primary analysis in this study was performed using ACR50 at twelve months as the definition of treatment failure. The rationale for this choice was that 12-month ACR50 was the primary outcome measure in the clinical trial from which the data were taken. However, this may not be the most relevant outcome for clinical practice. The ACR/EULAR guidelines have proposed a stringent consensus definition of remission for use in clinical practice,[36] which states that a patient must achieve a tender joint count ≤ 1 , a swollen joint count ≤ 1 and a patient global assessment ≤ 1 , or, as an index-based score, a CDAI ≤ 2.8 , to be considered in remission.

In a practice guideline on the use of rituximab in RA, Smolen et al. [37] suggested that a reduction in DAS-28 score of ≥ 1.2 (or equivalent measure) within four months should be regarded as a minimum improvement and that the target treatment response should be low disease activity characterised by a DAS-28 score ≤ 3.2 (or an SDAI < 11 or a CDAI < 10) together with optimisation of functional ability and quality of life. In this respect, the performance of the early non-response indicators evaluated was not specific for 12-month ACR, but was essentially similar with respect to the DAS-28 and CDAI definitions of treatment failure.

An original feature of this study was to compare different predictors of treatment failure, rather than predictors of treatment response. In the context of everyday practice, this is the most relevant information, as the clinician needs to identify as early as possible, and with confidence, which patients will fail to respond in the long term, so that treatment can be adapted. The PPV was considered the principal analysis criterion. The PPV informs the clinician on the probability that a person identified as a non-responder by Week 12 will truly represent a treatment failure at Week 52. This criterion depends on the prevalence of treatment failure at Week 52; the higher the prevalence, the higher the PPV. Assuming that the prevalence of treatment failure is higher in everyday clinical practice [6] than it is in clinical trials, the PPVs estimated in this study could be higher in a real life setting. The high PPVs observed in this study indicate that this is possible for between 80% and 100% of patients, depending on the measure. The specificity of these measures was also high, indicating that the number of patients who would be incorrectly identified at three months as treatment failures is low, ranging from 0% to 17% depending on the measure. If these patients are inappropriately identified as failures they will be moved to another treatment strategy and so will not be deprived of potentially effective

medication. On the other hand, the sensitivity of the early measures was low to moderate, being highest with the CDAI (40%) and the ACR50 (70%), suggesting that a substantial proportion of patients who will eventually fail treatment are not identified at three months. These patients thus remain at risk for exposure to ineffective treatment. Indeed, whatever the marker, the proportion of late failures is always higher than the proportion of early failures, which limits the sensitivity that can be achieved. Using a less strict criterion for early failure (eg CDAI >10) may increase the sensitivity of the measure by increasing the number of early failures, but this would be at the expense of reduced specificity. It should be noted that the use of NRI for handling missing data may artificially over-estimate the PPV values, since all patients for whom data was missing were considered to be treatment failures at study end. However, the majority of patients that were withdrawn from the study had inadequate efficacy (96 patients), which explains why we chose the NRI imputation method. Furthermore, another sensitivity analysis was performed in which missing data were imputed using LOCF, in which on-treatment response is taken into account, and this did not fundamentally change the findings of the study.

This study was performed in patients who had already been treated for a variable amount of time with methotrexate, had failed to respond adequately and were switched to certolizumab pegol 200 mg in the context of a randomised clinical trial. It will be very important to test early measures of treatment failure such as the CDAI in other patient populations. In particular, they should be evaluated in patients with early RA who start their first-line therapy, usually methotrexate. In addition, it should be noted that the analysis concerned a highly-selected population of patients enrolled in a Phase III trial, and it is possible that the criteria for

treatment failure identified here may perform differently in unselected patients in real-world conditions of treatment. This needs to be evaluated in future studies. Furthermore, it is possible that the time-course of clinical response may differ between certolizumab pegol and biological or synthetic DMARDs with different mechanisms of action or administration regimens, and the early non-response markers need to be evaluated in patients taking other DMARDs. It would also be important to test their performance with respect to long-term outcomes other than purely clinical ones, such as quality of life.

Until more studies have been performed in different patient populations, the implications of the present analysis for clinical practice should be considered with the utmost caution. However, we may tentatively propose to evaluate treatment outcome after three months using the CDAI in order to identify patients who will fail treatment and should be moved to another therapeutic strategy. This simple clinical measure can be determined during the consultation without the need for complementary biological tests. However, treatment response may not be maintained over time and certain patients identified as responders at three months may fail to respond at one year. This is reflected in the suboptimal sensitivity of the different markers evaluated in the study. For this reason, we would emphasise the need to re-evaluate the treatment response with the CDAI every three months and consider changing the treatment strategy if breakthrough disease activity is observed.

None of the early non-response measures evaluated in the study performed perfectly, and further work may be needed, notably to improve the sensitivity, so that more patients who will fail treatment can be identified early. Different thresholds of the CDAI could be tested, or a new

evaluation tool developed specifically to identify treatment failures, for example using decision-tree analysis.

In conclusion, this analysis of certolizumab pegol data has demonstrated that simple clinical measures of disease activity, with some of them easy to use in daily practice, can predict future treatment failure reliably and are appropriate for use to implement treat-to-target treatment strategies in clinical practice. However, it will be important to evaluate the early non-response measure chosen in a prospective study performed in a naturalistic treatment setting in order to validate its utility for helping patients achieve tight disease control and to identify the most appropriate cut-off thresholds.

Acknowledgments

The authors would like to thank FoxyMed (Paris, France) and Costello Medical Consulting for editorial support.

This publication was funded by UCB Pharma.

Conflicts of interest

FB is a consultant for UCB Pharma. TP is a consultant for UCB Pharma. PC received consulting fees, speaking fees and /or honoraria from Abbott, Wyeth-Pfizer, Roche, Chugai, Bristol-Myers Squibb, Schering-Plough, UCB Pharma, and was investigator for Roche Chugai, Sanofi Aventis, Abbott, Wyeth-Pfizer and BMS. TDC is an employee of UCB Pharma. JMJ is an employee of UCB Pharma, and holds UCB Pharma stock options. CS is an employee of UCB Pharma. LRF is a consultant for UCB Pharma. BF is a consultant for UCB Pharma.

References

1. Gabriel SE, Michaud K: Epidemiological studies in incidence, prevalence, mortality, and comorbidity of the rheumatic diseases. *Arthritis Res Ther* 2009, 11(3):229.
2. Guillemin F, Saraux A, Guggenbuhl P, Roux CH, Fardellone P, Le Bihan E, Cantagrel A, Chary-Valckenaere I, Euler-Ziegler L, Flipo RM et al: Prevalence of rheumatoid arthritis in France: 2001. *Ann Rheum Dis* 2005, 64(10):1427-1430.
3. Fautrel B, Cukierman G, Joubert JM, Laurendeau C, Gourmelen J, Fagnani F: Characteristics and management of rheumatoid arthritis in France: Analysis of a representative French national claims database resulting in an estimated prevalence of 0.35. *Joint Bone Spine* 2016;83:461-2.
4. Sokka T: Long-term outcomes of rheumatoid arthritis. *Curr Opin Rheumatol* 2009, 21(3):284-290.
5. Orme ME, Macgilchrist KS, Mitchell S, Spurdin D, Bird A: Systematic review and network meta-analysis of combination and monotherapy treatments in disease-modifying antirheumatic drug-experienced patients with rheumatoid arthritis: analysis of American College of Rheumatology criteria scores 20, 50, and 70. *Biologics* 2012, 6:429-464.
6. Hetland ML, Christensen IJ, Tarp U, Dreyer L, Hansen A, Hansen IT, Kollerup G, Linde L, Lindegaard HM, Poulsen UE et al: Direct comparison of treatment responses, remission rates, and drug adherence in patients with rheumatoid arthritis treated with adalimumab, etanercept, or infliximab: results from eight years of surveillance of clinical practice in the nationwide Danish DANBIO registry. *Arthritis Rheum* 2010, 62(1):22-32.
7. Smolen JS, Landewe R, Breedveld FC, Dougados M, Emery P, Gaujoux-Viala C, Gorter S, Knevel R, Nam J, Schoels M et al: EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. *Ann Rheum Dis* 2010, 69(6):964-975.
8. Gaujoux-Viala C, Gossec L, Cantagrel A, Dougados M, Fautrel B, Mariette X, Nataf H, Saraux A, Trope S, Combe B: Recommendations of the French Society for Rheumatology for managing rheumatoid arthritis. *Joint Bone Spine* 2014, 81(4):287-297.
9. Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, van Zeben D, Kerstens PJ, Hazes JM, Zwinderman AH, Roodman HK, Han KH, Westedt ML et al: Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. *Arthritis Rheum* 2005, 52(11):3381-3390.
10. Lard LR, Visser H, Speyer I, vander Horst-Bruinsma IE, Zwinderman AH, Breedveld FC, Hazes JM: Early versus delayed treatment in patients with recent-onset rheumatoid arthritis: comparison of two cohorts who received different treatment strategies. *Am J Med* 2001, 111(6):446-451.
11. Nell VP, Machold KP, Eberl G, Stamm TA, Uffmann M, Smolen JS: Benefit of very early referral and very early therapy with disease-modifying anti-rheumatic drugs in patients with early rheumatoid arthritis. *Rheumatology (Oxford)* 2004, 43(7):906-914.
12. Bakker MF, Jacobs JW, Verstappen SM, Bijlsma JW: Tight control in the treatment of rheumatoid arthritis: efficacy and feasibility. *Ann Rheum Dis* 2007, 66 Suppl 3:iii56-60.
13. Schipper LG, van Hulst LT, Grol R, van Riel PL, Hulscher ME, Fransen J: Meta-analysis of tight control strategies in rheumatoid arthritis: protocolized treatment has additional value with respect to the clinical outcome. *Rheumatology (Oxford)* 2010, 49(11):2154-2164.
14. Saag KG, Teng GG, Patkar NM, Anuntiyo J, Finney C, Curtis JR, Paulus HE, Mudano A, Pisu M, Elkins-Melton M et al: American College of Rheumatology 2008 recommendations for the use of

- nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis. *Arthritis Rheum* 2008, 59(6):762-784.
15. O'Dell JR, Mikuls TR, Taylor TH, Ahluwalia V, Brophy M, Warren SR, Lew RA, Cannella AC, Kunkel G, Pibbs CS et al: Therapies for active rheumatoid arthritis after methotrexate failure. *N Engl J Med* 2013, 369(4):307-318.
 16. Fautrel B, Verstappen SM, Boonen A: Economic consequences and potential benefits. *Best Pract Res Clin Rheumatol* 2011, 25(4):607-624.
 17. Aletaha D, Funovits J, Keystone EC, Smolen JS: Disease activity early in the course of treatment predicts response to therapy after one year in rheumatoid arthritis patients. *Arthritis Rheum* 2007, 56(10):3226-3235.
 18. Keystone EC, Curtis JR, Fleischmann RM, Furst DE, Khanna D, Smolen JS, Mease PJ, Schiff MH, Coteur G, Davies O et al: Rapid improvement in the signs and symptoms of rheumatoid arthritis following certolizumab pegol treatment predicts better longterm outcomes: post-hoc analysis of a randomized controlled trial. *J Rheumatol* 2011, 38(6):990-996.
 19. Curtis JR, Luijckens K, Kavanaugh A: Predicting future response to certolizumab pegol in rheumatoid arthritis patients: features at 12 weeks associated with low disease activity at 1 year. *Arthritis Care Res (Hoboken)* 2012, 64(5):658-667.
 20. Curtis JR, McVie T, Mikuls TR, Reynolds RJ, Navarro-Millan I, O'Dell J, Moreland LW, Bridges SL, Jr., Ranganath VK, Cofield SS: Clinical Response Within 12 Weeks as a Predictor of Future Low Disease Activity in Patients with Early RA: Results from the TEAR Trial. *J Rheumatol* 2013, 40(5):572-578.
 21. Curtis JR, Yang S, Chen L, Park GS, Bitman B, Wang B, Navarro-Millan I, Kavanaugh A: Predicting low disease activity and remission using early treatment response to antitumour necrosis factor therapy in patients with rheumatoid arthritis: exploratory analyses from the TEMPO trial. *Ann Rheum Dis* 2012, 71(2):206-212.
 22. Iannone F, Carlino G, Marchesoni A, Sarzi-Puttini P, Gorla R, Lapadula G: Early clinical response predicts low disease activity at one year in rheumatoid arthritis patients on treatment with certolizumab in real-life settings. An appraisal of the Italian registry GISEA. *Joint Bone Spine* 2016, 83(6):721-725.
 23. van den Bemt BJ, den Broeder AA, Wolbink GJ, van der Maas A, Hekster YA, van Riel PL, Benraad HB, van den Hoogen FH: The combined use of disease activity and infliximab serum trough levels for early prediction of (non-)response to infliximab in rheumatoid arthritis. *Br J Clin Pharmacol* 2013.
 24. van der Heijde D, Keystone EC, Curtis JR, Landewe RB, Schiff MH, Khanna D, Kvien TK, Ionescu L, Gervitz LM, Davies OR et al: Timing and magnitude of initial change in disease activity score 28 predicts the likelihood of achieving low disease activity at 1 year in rheumatoid arthritis patients treated with certolizumab pegol: a post-hoc analysis of the RAPID 1 trial. *J Rheumatol* 2012, 39(7):1326-1333.
 25. Keystone E, Heijde D, Mason D, Jr., Landewe R, Vollenhoven RV, Combe B, Emery P, Strand V, Mease P, Desai C et al: Certolizumab pegol plus methotrexate is significantly more effective than placebo plus methotrexate in active rheumatoid arthritis: findings of a fifty-two-week, phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. *Arthritis Rheum* 2008, 58(11):3319-3329.
 26. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS et al: The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988, 31(3):315-324.

27. European Medicines Agency: Draft guideline on clinical investigation of medicinal products other than non-steroidal anti-inflammatory drugs for treatment of rheumatoid arthritis. EMA, London; 2011.
28. American College of Rheumatology Ad Hoc Committee on Clinical Guidelines: Guidelines for the management of rheumatoid arthritis. *Arthritis Rheum* 1996, 39(5):713-722.
29. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL: Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995, 38(1):44-48.
30. Fries JF, Spitz P, Kraines RG, Holman HR: Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980, 23(2):137-145.
31. Aletaha D, Nell VP, Stamm T, Uffmann M, Pflugbeil S, Machold K, Smolen JS: Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. *Arthritis Res Ther* 2005, 7(4):R796-806.
32. Aletaha D, Smolen J: The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. *Clin Exp Rheumatol* 2005, 23(5 Suppl 39):S100-108.
33. Baldi P, Brunak S, Chauvin Y, Andersen CA, Nielsen H: Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* 2000, 16(5):412-424.
34. Schoels M, Aletaha D, Smolen JS, Bijlsma JW, Burmester G, Breedveld FC, Cutolo M, Combe B, Dougados M, Emery P et al: Follow-up standards and treatment targets in rheumatoid arthritis: results of a questionnaire at the EULAR 2008. *Ann Rheum Dis* 2010, 69(3):575-578.
35. Smolen JS, Breedveld FC, Schiff MH, Kalden JR, Emery P, Eberl G, van Riel PL, Tugwell P: A simplified disease activity index for rheumatoid arthritis for use in clinical practice. *Rheumatology (Oxford)* 2003, 42(2):244-257.
36. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, Aletaha D, Allaart CF, Bathon J, Bombardieri S et al: American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Arthritis Rheum* 2011, 63(3):573-586.
37. Smolen JS, Keystone EC, Emery P, Breedveld FC, Betteridge N, Burmester GR, Dougados M, Ferraccioli G, Jaeger U, Klareskog L et al: Consensus statement on the use of rituximab in patients with rheumatoid arthritis. *Ann Rheum Dis* 2007, 66(2):143-150.

Table 1: Patient characteristics at inclusion (n=382)

Age (mean \pm SD; years)	51.3 \pm 11.6
Gender (% female)	315 (82.5%)
Disease duration (mean \pm SD; years)	6.1 \pm 4.2
Methotrexate dose (mean \pm SD; mg/week)	13.6 \pm 4.4
No of tender/painful joints (median [interquartile range])	29 [21 – 39]
No of swollen joints (median [interquartile range])	20 [14 -28]
DAS-28 using the ESR (median [range])	7.0 [4.3 to 8.9]
HAQ-DI (mean \pm SD)	1.7 \pm 0.6
% RF positive (\geq 14 IU/ml)	79.0
CRP (median [range]; mg/l)	16.0 (1.0, 234)
ESR (median [range]; mm/h)	43.0 (5.0, 138)

DAS-28: disease activity score 28-joint assessment; HAQ: health assessment questionnaire; DI: disability index;

RF: rheumatoid factor; CRP: C-reactive protein; ESR: erythrocyte sedimentation rate;.

Table 2: Early and late treatment failure rates

Outcome measure	Early non-response definition	Missing data at Week 12	n	Non-responders by Week 12	Late failure definition	Non-responders at Week 52
ACR	ACR50 not met	11	371	183 (49.3%)	ACR50 not met	228 (61.5%)
DAS-28	Decrease in score since baseline ≤ 1.2	8	374	40 (10.7%)	Final score > 3.2	283 (75.7%)
HAQ-DI	Decrease in score since baseline < 0.22	12	370	57 (15.4%)	Final score > 0.5	293 (79.2%)
CDAI	Score > 22	11	371	99 (26.7%)	Final score > 10	238 (64.2%)

Data are presented using non-response imputation to handle missing data at Week 52. All 382 patients with no prior TNF α -inhibitor exposure were evaluated, with the exception of those for whom data was missing at Week ACR: American College of Rheumatology; DAS-28: disease activity score 28-joint assessment; HAQ: health assessment questionnaire; DI: disability index; CDAI: Clinical Disease Activity Index.

Table 3: Performance of different measures of early non-response as predictors of treatment failure at

Week 52 - Non responder imputed model

Early non-response measure	Specificity	PPV	Sensitivity	NPV	Accuracy	MCC
<i>NRI model</i>						
DAS-28 ($\Delta \leq -1.2$)	99% [94%;100%]	95% [82%;99%]	16% [12%;22%]	42% [37%;47%]	48% [42%;53%]	0.24
CDAI (score > 22)	95% [90%;98%]	93% [85%;97%]	40% [34%;47%]	50% [44%;56%]	61% [56%;66%]	0.39
ACR50 not met	83% [76%;89%]	87% [81%;91%]	70% [63%;76%]	63% [56%;70%]	75% [70%;79%]	0.52
HAQ-DI ($\Delta < -0.22$)	92% [86%;96%]	81% [68%;90%]	20% [15%;26%]	42% [37%;48%]	48% [43%;53%]	0.17

Treatment failure was defined by ACR50 at Week 52. Data are presented with their 95% confidence intervals. The different early non-response measures are presented in decreasing order of PPV. Non responder imputation; n=382. DAS-28: disease activity score 28-joint assessment; CDAI: Clinical Disease Activity Index. ACR: American College of Rheumatology; HAQ: health assessment questionnaire; DI: disability index.

Table 4: Performance of different measures of early non-response as predictors of treatment failure at

Week 52 - Observed cases analysis

Early non-response measure	Specificity	PPV	Sensitivity	NPV	Accuracy	MCC
Observed cases analysis – N						
DAS-28 ($\Delta \leq -1.2$) N=218	99% [94%;100%]	78% [40%;969%]	9% [4%;19%]	67% [60%;73%]	67% [60%;73%]	0.19
CDAI (score > 22) N=219	95% [90%;98%]	72% [50%;87%]	24% [15%;35%]	70% [63%;76%]	70% [64%;76%]	0.28
ACR50 not met N=219	83% [76%;89%]	66% [54%;77%]	62% [50%;73%]	80% [73%;86%]	76% [69%;81%]	0.46
HAQ-DI ($\Delta < -0.22$) N=216	92% [86%;96%]	42% [21%;66%]	11% [5%;21%]	67% [60%;73%]	65% [56%;71%]	0.05

Performance of different measures of early non-response as predictors of treatment failure at Week 52. Treatment failure was defined by ACR50 at Week 52. Data are presented with their 95% confidence intervals. The different early non-response measures are presented in decreasing order of positive predictive values. The observed cases analysis was performed on all patients with efficacy data available at both Week 12 and Week. DAS-28: disease activity score 28-joint assessment; CDAI: Clinical Disease Activity Index. ACR: American College of Rheumatology; HAQ: health assessment questionnaire; DI: disability index;

Table 5: Positive predictive values for different criteria for early non-response as predictors of other criteria of treatment failure

Criterion for non-response by Week 12	Criterion for treatment failure at Week 52		
	DAS-28 ≥ 3.2	HAQ-DI ≥ 0.5	CDAI ≥ 10
DAS-28 ($\Delta \leq -1.2$)	100% [89%;100%]	93% [78%;98%]	100% [89%;100%]
CDAI (score > 22)	99% [94%;100%]	97% [91%;99%]	97% [91%;99%]
ACR50 not met	93% [89%;96%]	92% [87%;96%]	86% [80%;91%]
HAQ-DI ($\Delta < -0.22$)	88% [76%;95%]	89% [78%;96%]	79% [66%;88%]

Treatment failure considered at Week 52. Data are presented with their 95% confidence intervals. Non responder imputation, n=382. DAS-28: disease activity score 28-joint assessment; CDAI: Clinical Disease Activity Index. ACR: American College of Rheumatology; HAQ: health assessment questionnaire; DI: disability index.