Assessing prognosis and prediction of treatment response in early rheumatoid arthritis: systematic reviews

Rachel Archer,* Emma Hock, Jean Hamilton, John Stevens, Munira Essat, Edith Poku, Mark Clowes, Abdullah Pandor and Matt Stevenson

School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK

*Corresponding author r.archer@sheffield.ac.uk

Declared competing interests of authors: John Stevens is a shareholder of GlaxoSmithKline plc (GSK House, Middlesex, UK), AstraZeneca plc (Cambridge, UK) and Shire Pharmaceuticals Group plc (Shire plc, Dublin, Republic of Ireland).

Published November 2018
DOI: 10.3310/hta22660

Scientific summary

Treatment response in early rheumatoid arthritis

Health Technology Assessment 2018; Vol. 22: No. 66
DOI: 10.3310/hta22660

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Rheumatoid arthritis (RA) is a chronic and debilitating disease that can lead to increasing disability, pain and irreversible joint damage. Symptoms include pain, morning stiffness, swelling and tenderness of joints, loss of mobility, warmth of the peripheral joints and fatigue. RA is associated with a reduced quality of life and substantial direct and indirect costs resulting from treatment and reductions in productivity. Treatments for RA include conventional disease-modifying antirheumatic drugs (cDMARDs) and biologic disease-modifying antirheumatic drugs (bDMARDs). Key outcomes in RA include measures of joint destruction (e.g. radiographic progression), disease activity [as assessed via, for example, the Disease Activity Score-28 (DAS28)] and disability [as assessed via, for example, the Health Assessment Questionnaire (HAQ) scores].

Health-care professionals need to be able to determine at an early stage of disease which patients may experience a worse prognosis in order to inform effective disease management and avoid pharmacological overtreatment of patients. There is currently no clear consensus on which of the available tests and assessment tools used in RA provide the best assessment of prognosis in people newly diagnosed with RA and whether or not patient or disease characteristics can predict how well patients will respond to different drug treatments.

This report was commissioned by the National Institute for Health Research Health Technology Assessment programme as project number 14/151/08.

Objectives

The objectives of this work were to undertake systematic reviews to determine the:

- use of selected tests and assessment tools in the evaluation of prognosis in patients with early RA
- potential of selected tests and assessment tools as predictive markers of treatment response in patients with early RA.

Methods

Two related systematic reviews were undertaken. The systematic reviews were informed by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (www.prisma-statement.org/) and current good practice in prognostic reviews advocated by the Cochrane Prognosis Methods Group (http://methods.cochrane.org/prognosis/). A final protocol for this assessment was registered on PROSPERO (CRD42016042402).

Review 1 (clinical prediction models)

Prognostic research involves the study of the relationship between future outcomes among people with a given baseline health state in order to improve health. A prognostic model is a formal combination of multiple predictors from which the probability of a specific event can be estimated for individual patients.

Searches of electronic databases (e.g. MEDLINE, EMBASE, The Cochrane Library, Web of Science Conference Proceedings; searched to September 2016), registers, relevant websites, hand-searching of reference lists of included studies and key systematic reviews were conducted and contact was made with experts.

Primary studies describing the development, external validation and impact of eligible clinical prediction models in adult patients with early RA (defined as being within 2 years of the onset of symptoms) were eligible for inclusion. The prognostic variables considered in the assessment were informed by the phase 1 scoping searches and agreed following discussion between the review team and clinical advisors. The prognostic variables selected for inclusion in review 1 were anticitrullinated protein/peptide anti-bodies (ACPAs), rheumatoid factor, erosions/joint damage assessed via X-ray, C-reactive protein (CRP) levels, erythrocyte sedimentation rate (ESR), swollen joint count (SJC), DAS28, early RA untreated for $\geq$ 12 weeks following the onset of symptoms, smoking status and HAQ scores. Eligible outcomes were joint damage as assessed on radiographs, DAS28 and HAQ scores.

Data extraction was informed by the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS). Assessment of the study quality characteristics of clinical prediction modelling studies was informed by criteria included in the Prediction model study Risk Of Bias Assessment Tool (PROBAST).

Data on the predictive performance of included clinical prediction models were described in a narrative synthesis, presented separately for internal and external validation studies. Evidence synthesis using meta-analysis was considered for external validation studies.

Review 2 (prediction of treatment response)

Searches for evidence were undertaken using data sources, as described for review 1.

Review 2 included evidence on the interaction between baseline covariates and treatment on salient outcomes in adult patients with early RA. The response to cDMARDs and bDMARDs was studied. Eligible studies involved at least 6 months' treatment duration (with the exception of 12 weeks for certolizumab pegol). Eligible predictive variables were the same as for review 1, with the addition of body mass index (BMI) and vascularity of synovium assessed using power Doppler ultrasound (PDUS). The outcomes selected for inclusion were the same as for review 1, with the addition of selected definitions of response/remission [European League Against Rheumatism (EULAR) response; remission as a DAS28 of $<$ 2.6, a Disease Activity Score of $<$ 1.6 and/or American College of Rheumatology/EULAR remission].

Studies of predictive variables were assessed by criteria informed by the Quality in Prognosis Studies (QUIPS) tool.

A formal meta-analysis was not performed as, for specific outcome measures and potential treatment effect modifiers, there were no studies that shared any treatments in common. Results were presented with regard to assessing the predictive ability of baseline patient and/or disease characteristics according to different treatments by study.

Results

Review 1 (clinical prediction models)

Twenty-eight studies that investigated the use of assessment tools and tests in the evaluation of a prognosis in early RA patients were identified. These included 22 model development studies and one combined model development/external validation study that reported a total of 39 clinical prediction models for the outcomes of radiographic joint damage, DAS28 and HAQ score. An additional five external validation studies, which tested the performance of eight clinical prediction models for radiographic joint damage outcomes, were also included.

Included studies varied in terms of the methods applied to develop the clinical prediction models, for example, in the strategies used to select or reject candidate predictors from the final model and in the handling of continuous predictors. Several studies presented a 'matrix model', and continuous variables

were frequently categorised to allow this presentation format. For models developed using randomised controlled trial data with patients assigned to alternative treatment strategies, the model development generally failed to assess interactions between predictors and treatment group and so did not generate truly treatment-specific models.

Model development studies varied in the reporting of predictive performance. A key measure of model predictive performance, the *c*-statistic, was presented from internal validation in 8 of the 23 model development studies; sensitivity and specificity (eight studies), accuracy (seven studies), positive predictive value and/or negative predictive value (12 studies) were also reported. Of the eight studies that reported *c*-statistics from internal validations, *c*-statistics for radiographic progression outcomes ranged between 0.63 [with Degboé *et al.* (Degboé Y, Constantin A, Nigon D, Tobon G, Cornillet M, Schaeverbeke T, *et al.* Predictive value of autoantibodies from anti-CCP2, anti-MCV and anti-human citrullinated fibrinogen tests, in early rheumatoid arthritis patients with rapid radiographic progression at 1 year: results from the ESPOIR cohort. *RMD Open* 2015;**1**:e000180) predicting a Δ Sharp/van der Heijde score (SHS) of ≥ 5 at 1 year] and 0.87 [with Houseman *et al.* (Houseman M, Potter C, Marshall N, Lakey R, Cawston T, Griffiths I, *et al.* Baseline serum MMP-3 levels in patients with rheumatoid arthritis are still independently predictive of radiographic progression in a longitudinal observational cohort at 8 years follow up. *Arthritis Res Ther* 2012;**14**:R30) predicting a Δ SHS of ≥ 10.5 at 8.2 years]. Two studies predicting HAQ also generated *c*-statistics from internal validation {0.78 [Dirven *et al.* (Dirven L, Visser K, Klarenbeek NB, Ewals JA, Han KH, Peeters AJ, *et al.* Towards personalized treatment: predictors of short-term HAQ response in recent-onset active rheumatoid arthritis are different from predictors of rapid radiological progression. *Scand J Rheumatol* 2012;**41**:15–19) HAQ ≥ 1 at 3 months] to 0.82 [Bansback *et al.* (Bansback N, Young A, Brennan A, Dixey J. A prognostic model for functional outcome in early rheumatoid arthritis. *J Rheumatol* 2006;**33**:1503–10) HAQ ≥ 1.5 at 5 years]}. However, even if consistent approaches had been used for internal validation, comparing the performance of clinical prediction models that have been internally validated in different populations would still be limited, because good discriminative ability in the population used to develop the model would be expected. External validation is required to provide an objective comparison.

For the eight models that were externally validated, predictive performance varied considerably. Five clinical prediction models [Syversen, Swedish Farmacotherapy (SWEFOT), Étude et suivi des polyarthrites indifférenciées récentes (ESPOIR), multibiomarker disease activity and Study Of New-Onset Rheumatoid Arthritis (SONORA)] were externally validated only in one population per outcome definition. Three clinical prediction models [Active controlled Study of Patients receiving Infliximab for the treatment of Rheumatoid arthritis of Early onset (ASPIRE) CRP, ASPIRE ESR and Behandelings Strategie (BeSt)] were externally validated using the same outcome definition in more than one population. The results of the random-effects meta-analysis indicated that the most favourable performance across external validations was for the BeSt model [*c*-statistic 0.72, 95% confidence interval (CI) 0.20 to 0.96], followed by ASPIRE ESR (*c*-statistic 0.62, 95% CI 0.44 to 0.78) and ASPIRE CRP (*c*-statistic 0.55, 95% CI 0.13 to 0.91). However, there is considerable heterogeneity for all three models, with the wide CIs suggesting substantial uncertainty in the expected predictive performance in a new sample of patients. The 95% CIs of the pooled estimates contain 0.5 for all three clinical prediction models, indicating that there is limited confidence that the performance of the models is better than would be expected by chance.

The inconsistent results generated by the clinical prediction models on external validation indicate that there is heterogeneity in the populations in which the models are being tested that is not explained by the currently proposed models. However, the meta-analysis was limited by the small number of available external validation studies. The synthesised estimates are indicative of performance in the observed studies, but cannot be used to provide a definitive conclusion about the performance in future studies or to explore the reasons for the heterogeneity between studies.

Despite the identification of 23 model development studies and six external validations (including the combined model development/external validation study), uncertainty remains over the optimal prediction model(s) for use in clinical practice. There were limitations identified in the methods used to develop the

clinical prediction models in many of the development studies, including the handling of continuous predictors and failure to assess interactions. It is therefore likely that the most clinically useful prediction model may contain predictors from across more than one of the reviewed clinical prediction models and/or consider alternative handling of key predictive variables.

Review 2 (prediction of treatment response)

Review 2 identified 12 primary studies with which to assess the prediction of treatment response according to baseline covariates. The covariates examined included ACPA status, smoking status, erosions, rheumatoid factor status, CRP levels, ESR, SJC, BMI and vascularity of synovium on PDUS. The outcomes examined included erosions/radiographic progression, disease activity, physical function and DAS28 remission.

There was statistical evidence to suggest that ACPA status, SJC 28 and PDUS status at baseline may be treatment effect modifiers, but not necessarily that they are prognostic of response for all treatments. Most of the results were subject to considerable uncertainty and were not statistically significant. In general, there was insufficient evidence that the effect of treatment depended on baseline characteristics.

Conclusions

Review 1 (prognostic models)

No single clinical prediction model can currently be recommended in preference to any other for use in clinical practice on the basis of uncertainties and limitations in the available evidence. The optimal prediction model(s) may include variables (e.g. biomarkers/genetic tests) that are not routinely or currently available. Any practical and cost implications associated with their use would need to be evaluated before future implementation.

Review 2 (prediction of treatment response)

There was limited evidence with which to assess whether or not specific baseline variables can predict differential effects according to the treatment administered. Nevertheless, the available evidence suggested that some baseline variables do affect relative treatment effects and that not all baseline variables may be prognostic of response for all treatments.

The effects of covariates were rarely assessed in single models adjusting for all covariates and with the inclusion of interaction terms with treatment. Although there was statistical evidence to suggest that some baseline covariates affect treatment response differentially, the results were subject to considerable uncertainty and there was generally insufficient evidence that the effect of treatment depended on baseline characteristics. This may be a real effect or may be because studies lacked statistical power to detect interaction effects. In future analyses, the true effect of baseline variables should be evaluated in single multivariable models, adjusting for all relevant covariates and interactions with treatment.

Suggested research priorities

Review 1 (prognostic models)

Recommendations for further research include:

- collaborative research, including the use of individual participant data, for further (1) development/internal validation and (2) external validation of optimal clinical prediction model(s) to demonstrate predictive performance and generalisability
- adherence to good model development and reporting standards of future clinical prediction model studies
- research to investigate the effects on patient outcomes (and the cost-effectiveness) of the use in clinical practice of optimal internally and externally validated model(s).

Review 2 (prediction of treatment response)

Recommendations for further research include the following:

- Clinical prediction models should be developed and validated with respect to individual treatments.
- The assessment of treatment by covariate interactions should follow good statistical practice: subgroup analyses should be avoided, categorising continuous baseline covariates should be avoided and the interactions between treatments and baseline variables should be specifically modelled.
- The results of multivariable analyses presented in published reports should include estimates of the main effects of covariates and any interaction effects together with their standard errors and covariances for secondary research purposes.

Study registration

This study is registered as PROSPERO CRD42016042402.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.513

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 14/151/08. The contractual start date was in June 2016. The draft report began editorial review in August 2017 and was accepted for publication in February 2018. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).