

Real World Data: special section

Adding value to real-world data: the role of biomarkers

Darren Plant ^{1,2} and Anne Barton ^{1,2}

Abstract

Adding biomarker information to real world datasets (e.g. biomarker data collected into disease/drug registries) can enhance mechanistic understanding of intra-patient differences in disease trajectories and differences in important clinical outcomes. Biomarkers can detect pathologies present early in disease potentially paving the way for preventative intervention strategies, which may help patients to avoid disability, poor treatment outcome, disease sequelae and premature mortality. However, adding biomarker data to real world datasets comes with a number of important challenges including sample collection and storage, study design and data analysis and interpretation. In this narrative review we will consider the benefits and challenges of adding biomarker data to real world datasets and discuss how biomarker data have added to our understanding of complex diseases, focusing on rheumatoid arthritis.

Key words: rheumatoid arthritis, real-world data, registry, biomarkers, precision medicine

Rheumatology key messages

- Addition of biomarker information to real-world datasets is challenging but could capture clinically useful information.
- Biomarkers can detect pathologies present early in disease potentially paving the way for preventative interventions.
- Biomarkers can quantify important exposures underpinning intra-patient differences in disease outcomes to facilitate precision medicine.

Introduction

Real world datasets include information collected from individuals in routine clinical care with a particular condition or diagnosis or receiving a particular drug, as opposed to patients recruited to trials. Real world data can enhance patient management because the length of follow-up can be much longer than randomized control trials (RCTs); they include patients with complex issues, of the type seen in routine clinical settings, who are often excluded from RCTs due to the exclusion criteria; and cohort sizes tend to be larger than conventional RCTs. They provide health care professionals and researchers

with information on people with diseases, at the individual and group level, often over time enabling analysis of trends in treatment, risks of adverse events (e.g. infection) and disability.

Insight into the outcomes that are linked to particular disease trajectories can be obtained from clinical and demographic factors, disease and treatment history, and co-morbidity and risk exposures that are recorded in real world databases. However, these factors often only partially explain intra-patient differences in important clinical outcomes, suggesting that other factors contribute. Such factors could be biological, many of which can be measured directly or through a proxy; such biomarkers can correlate with stage of disease, response to treatment or adverse events [1]. Biomarkers may include serological, genetic, epigenetic or transcriptomic factors or might be captured in imaging data or in spectral maps of the proteome retrieved from relevant biological samples.

Advantages of biomarkers over routinely collected clinical data are that the markers are measured objectively and therefore not subject to observer bias, they can be measured reliably and precisely, and they are closer to the

¹Manchester Academic Health Science Centre, The University of Manchester, Arthritis Research UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research and ²Manchester Academic Health Science Centre, NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester, UK
Submitted 18 October 2018; accepted 28 February 2019

Correspondence to: Anne Barton, Arthritis Research UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, The University of Manchester, Manchester M13 9PT, UK.
E-mail: anne.barton@manchester.ac.uk

pathology and may provide mechanistic insight. Some biomarkers become so accepted that they become part of routine data collection; for example, anti-citrullinated peptide antibodies (ACPA) detected by anti-cyclic citrullinated peptide (anti-CCP) antibodies are now routinely tested because they define a subset of cases with RA with a more severe disease course [2].

In this review we will consider the advantages that addition of biomarker data to real world datasets brings, discuss how biomarker data have improved our understanding of disease outcomes in RA, and consider the challenges faced when interpreting biomarker discoveries or when planning to add biomarker data to real world data collections.

Key challenges that can be addressed using biomarker data

Once a patient presents with inflammatory arthritis, it is now widely accepted that the primary goal should be to control disease activity as quickly as possible. Indeed, real world data have shown that control of disease activity within the first 6 months of presentation of inflammatory arthritis can influence outcome 20 years later [3]. Over the past 20 years, significant advances in the management of RA have resulted in the development of new drugs that target various immune pathways and molecules, including cytokines (tumour necrosis factor (TNF) and interleukin-6), T cell signalling and B cell depletion [4]. Despite the large number of treatments now available to patients, it is still not clear which drug will be of benefit to an individual patient as each drug has a significant non-response rate. The challenge is to identify RA patients as early as possible before structural joint damage has accumulated and give them the most appropriate treatment for their disease. In an ideal world, robust biomarkers would be available to guide clinicians in the need for therapy (some patients will remit without treatment) and the treatments most likely to effectively control inflammation.

Disease development and progression

Biomarkers have provided a better understanding of the biological pathways underpinning an individual's risk of developing RA and subsequent disease progression. For example, genetic biomarkers anchor the relationship between the human leucocyte antigen (HLA) gene *HLA-DRB1* and RA development [5–8] implicating the activation of CD4-positive T cells as a critical step in the aetiology of the disease [9]. Indeed, large scale genome-wide genetic variant analysis, including over 100 000 individuals, has identified over 100 genetic regions outside the HLA region that are associated with RA development [6, 10] and many of these regions harbour genes that map to T cell activation pathways [11]. In other complex diseases, such as cardiovascular disease, information from variants across the genome has been used to construct genotypic risk scores, which identify those at highest risk of developing cardiovascular disease. Given that cardiovascular disease is common and primary prevention

strategies are established, the usefulness of a biomarker score associated with a tripling of risk is potentially clinically useful [12]. However, for RA where the risk of disease is lower, identifying a biomarker associated with a tripling of risk is not clinically useful. Indeed, *HLA-DRB1* genetic variants have long been established to confer that magnitude of risk but are not used to screen the population because carriage of the variants only increases risk from 1% to 3% (i.e. 97% risk of not developing RA).

The severity of joint destruction in RA is influenced by genetic factors, with an estimated heritability of 50% [13]. Some of the intra-patient differences in disease progression are explained by carriage of risk haplotypes at the *HLA-DRB1* locus, which are also strongly associated with RA susceptibility. Importantly, the DRB1 haplotypes that confer the greatest risk of disease development also identify patients with the most severe radiographic progression [14]. However, non-HLA genetic biomarkers of erosive disease have not been convincingly found, probably due to a lack of statistical power to detect subtle genetic effects.

Serological biomarkers of RA development and severity have been recognized for some time. A positive titre for anti-cyclic citrullinated antibody (ACPA) and rheumatoid factor (RF) can precede RA development by several years [15, 16] and are associated with a more severe disease course [2]. Both biomarkers perform individually well at classifying people with RA, compared with healthy controls. For example, in a recent large meta-analysis, the pooled sensitivity for ACPA and RF was 67% and 69% and the pooled specificity was 95% and 85%, respectively [17]. Indeed, ACPA and RF status now form part of the established classification criteria for RA [18] and are important enrichment biomarkers for defining RA cohorts for further study. The presence of other serological changes is also observed in the pre-symptomatic phase of RA; for example, a positive titre for anti-carbamylated (anti-CarP) antibodies precedes RA development in a subset of patients and is associated with a more severe disease course independently of ACPA [19–21].

Treatment outcomes

Over many years, high quality real world cohorts and registries of patients have been established to explore the long term safety and treatment outcomes for patients receiving medication for their RA. A number of important clinical factors are now known to correlate with clinical response to methotrexate and biologic drugs, including TNFi-inhibitors (TNFi). Gender and pre-treatment disability and activity [22, 23] explain some of the variation in response but these factors are not sufficiently predictive to be useful in a clinical setting, nor do they capture information about relevant target pathways or intra-patients differences in treatment exposure or adherence. Focus has therefore shifted to identifying treatment response biomarkers.

The search for reliable biomarkers of treatment response has led to an enormous effort in establishing large-scale biological sample collection in real world

TABLE 1 Examples of real world drug registries that include biomarker data in RA

Short name	Country	Type	Reference
BIOBADASER	Spain	Registry	[24]
BSRBR-RA/ BRAGGSS	UK	Registry	[22, 25]
DANBIO	Denmark	Registry	[26]
DREAM	Netherlands	Registry	[27]
SCQM	Switzerland	Registry	[28]
ARTIS	Sweden	Registry	[29]

ARTIS: Antirheumatic Therapies in Sweden; BRAGGSS: biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate; BSRBR-RA: British Society for Rheumatology Biologics Register for Rheumatoid Arthritis; DREAM: Dutch Rheumatoid Arthritis Monitoring; SCQM: Swiss Clinical Quality Management in Rheumatic Diseases.

cohorts (Table 1). This has paved the way for analysis of genomic and detailed clinical data [25], as well as large national [30] and international collaborations [31]. For example in the UK, the Medical Research Council and Arthritis Research UK jointly funded the MATURA (MAXimising Therapeutic Utility for Rheumatoid Arthritis) consortium [30] recognizing the need to maximize the interconnect between specialized groups with diverse expertise such as researchers, statisticians, medical practitioners and industry partners.

MATURA is still in progress but important findings from this research are beginning to emerge including the development of a classifier that is highly predictive of MTX non-response using genome-wide gene-expression data [32], the identification of genetic markers that are correlated with TNFi [33] and MTX [34] response, and the observation that genetic markers strongly correlate with the objective sub-components of the DAS28 i.e. serum C-reactive protein (CRP) level and the swollen joint count (SJC, out of 28 joints [35]), but not the tender joint count or patient's assessment of general health; traits that are not heritable are very difficult to model using biological factors such as biomarkers.

Genetic variants

Numerous genetic biomarkers of poor treatment outcomes in RA have been reported in the literature, but with modest confirmatory evidence between studies. Genetic markers with some evidence of replication in studies of TNFi response include variants at the *PTPRC*, *FCGR2A*, *TRAF1/C5*, *CHUK*, *IRAK3* and *NFKB1B* loci [36, 37]. However, none are sufficiently predictive of response alone to be clinically useful.

As mentioned above, large scale genome-wide genetic variant analyses have identified a great many genetic regions that are associated with RA development [6, 10]. Importantly, numerous genes within identified regions interact, at the protein level, with targets of approved treatments licensed for RA, and for other diseases [6],

indicating that robust genetic association studies may identify treatment response subgroups, or help to reveal drug mechanisms [38].

Epigenetic biomarkers

Epigenetic biomarkers such as DNA methylation and covalent histone modifications regulate gene expression through modulating accessibility of transcription factors to DNA [39]. Epigenetic marks are therefore a potentially important source of biomarkers of treatment response (reviewed elsewhere [39]). For example, the recently reported differential DNA methylation at the *LRPAP1* gene locus on chromosome 4 is correlated with response to the TNFi etanercept. Initially identified in 72 patient samples this finding was supported by genetic validation in a larger collection of 1204 TNFi treated patients [40]. Low density lipoprotein receptor-related protein 1 (LRPAP1) is highly expressed in mononuclear cells [41] and influences activity of transforming growth factor β [42], a potent anti-inflammatory cytokine [43].

To identify biomarkers of MTX response, Carini *et al.* used a recently developed DNA-array-based method to assessed 13 322 potential chromosome interactions relating to 123 genes with known importance to RA [44]. The authors identified a chromatin conformation signature consisting of five genomic regions (genes *FNAR1*, *IL-21R*, *IL-23*, *CXCL13* and *IL-17A*) that was able to detect non-responders to MTX with 90% sensitivity when validated in independent samples. Although further independent validation of this important finding is now required, it is attractive to consider how this biomarker signature/method could be implemented clinically. The method relies on uncomplicated sample processing (a small amount of whole blood is required) and makes use of an established laboratory technique (quantitative polymerase chain reaction).

Gene expression/transcriptomics

Gene expression profiling studies of TNFi response have reported few differences in pre-treatment samples between future EULAR good and non-responders, whereas marked differences in gene expression profile are observed in good responders when contrasting post-treatment samples with samples collected pre-treatment [33]. Therefore, early rather than pre-treatment biomarker profile may be identified in certain data types.

Drug levels and antidrug antibodies

Measuring drug levels in TNFi-exposed patients provides information as to whether a therapeutic circulating drug level is achieved [45, 46]; as expected, disease control is sub-optimal in patients with sub-therapeutic drug levels. Factors affecting drug levels include body mass index and presence of antibodies directed against the drug. A recent report [47] suggests that development of anti-TNFi antibodies in patients with Crohn's disease is associated with genetic markers within the HLA region (specifically HLA-DQA1*05), a finding that now requires replication. Importantly, studies also show that patients with high

circulating concentrations of TNFi are at increased risk of developing respiratory and skin infections, compared with patients with low circulating TNFi levels [48], suggesting that monitoring of drug levels provides clinically important information regarding under- or over-dosing patients.

Multi-omics

To achieve a more complete view of the mechanisms that underpin response to treatment in RA, it is likely that biomarkers present in a number of data types will need to be measured in the same individuals permitting a more holistic analysis approach. Recently a report by Tasaki *et al.* [49] showed the utility in a multi-omics approach to understanding treatment response in RA. By comparing blood transcriptomic, proteomic and immunophenotype profiles between patients, pre-and post-treatment, and healthy controls, the authors identified signatures associated with clinical remission. The authors were also able to make a distinction between clinical remission and molecular remission in patients who experienced successful treatment. Importantly, patients who achieved molecular remission, i.e. a biomarker profile more similar to healthy controls than to RA, achieved better long-term treatment outcomes, particularly if molecular remission was observed across more than one data type [49]. The study was small and the findings require further validation but the research highlights the importance of considering multiple data types as well as on-treatment sampling.

Challenges faced

It should be recognized that, whilst some clinically useful biomarkers have emerged from studies, many published biomarker studies never translate to the clinical setting for a range of reasons discussed below.

Logistical challenges

There is a great deal of discussion about storage and transportation of samples prior to biomarker discovery and some recommend that sample collection should follow international recommendations. This is difficult for real world biobanks associated with disease registries as sample collection is often opportunistic, when patients are being seen clinically. Undoubtedly, this could lead to false-positive reports of associations and, therefore, it is vital that positive associations are validated in independent data sets. The clinically useful biomarkers that have emerged, including autoantibodies and some proteomic biomarkers such as CRP, have proved themselves to be robust to collection, storage and transportation issues as they are relatively stable once collected. It is likely that biomarkers that translate successfully to the clinic in the future will either also need to be stable (hence the attraction of genetic variants as potential biomarkers) or be collected in a consistent, reproducible way to avoid generation of misleading data. One way to tackle this in real world data may be to test the stability of potential biomarkers [50] or to use repeated measures to smooth out the natural variation that occurs.

Limitations of outcome measure

Rather than being useful for prediction, assessment of on-treatment biomarker levels may provide a quantifiable measure for monitoring disease. Indeed serum CRP and the erythrocyte sedimentation rate (ESR), included in the DAS28, and the multi-biomarker disease activity (MBDA) score are already used to monitor disease activity [51, 52]. However, none are specific for RA and only correlate moderately well with synovitis [52, 53] or long-term outcomes [54].

The DAS28 encompasses both objective [SJC, CRP (or ESR)] and subjective (tender joint count and patient's assessment of general health) measures, the latter scores receiving higher weightings [51]. We have previously reported that psychological factors, such as depression and illness beliefs, correlate more closely with the subjective components than with the objective components [55] suggesting that the DAS28 also measures factors other than synovial inflammation, although the latter is the primary target for treatment. This finding has added to the debate of what we mean by 'treatment response'. There is an argument that, as many drugs were developed to specifically treat joint inflammation, whether an individual is classified as a responder should be based on how well joint inflammation is controlled. One recent study found that only the CRP and SJC were correlated with ultrasound-observed synovitis and a re-weighted score including only those components was better correlated with radiographic progression, compared with the DAS28 composite score or its individual sub-components [56].

Psychological factors, including high pre-treatment anxiety scores, have been reported to be associated with non-response to methotrexate, with the definition of non-response including patients who stopped their treatment before 6 months [23]. However, it is not clear whether the drug was effective at controlling synovitis; going forward, separating the concept of overall treatment response and effective control of synovitis will increase the likelihood of useful biomarkers being identified for the latter as biomarkers are more likely to reflect a biologically rather than psychologically driven process.

Inadequate adherence to treatment

Even with an optimized treatment response outcome that is reflective of synovial inflammation, the identification of biomarkers predictive of response would still not be straightforward as a number of factors can confound response prediction. We have previously reported that self-reported adherence to TNFi therapies was sub-optimal (27% of patients self-reporting non-adherence) and correlated with future response [58]. Therefore patients predicted to respond may be classified as non-responders because of non-adherence, reducing the accuracy and power of predictive biomarker studies. Inadequate adherence can be measured or monitored using appropriate biomarkers. For example, an HPLC-mass spectrometry assay for monitoring low dose MTX and its major metabolite, 7-OH-MTX, in urine samples from RA patients has recently been developed [59]. The assay can detect MTX

for up to 105 h after administration and 7-OH-MTX for up to 98 h, suggesting that this platform is suitable for assessing adherence to therapy in a clinical setting.

Lack of comparison group

RA is a relapsing–remitting disease in which patients experience moments of high and low disease activity as part of the natural disease course. Therefore a distinction that is difficult to make in real world datasets is whether disease activity has improved because of successful treatment, or because disease activity was initially very high and has more room to improve before the subsequent clinical visit. In a trial setting the issue is resolved by randomizing patients to one of two (or more) treatment arms but that is not possible in the real world. In statistics the phenomenon is referred to as regression towards the mean and must be considered when designing biomarker studies or when interpreting results from real world datasets. Ensuring matching of the baseline characteristics of the two comparator groups can be used for within-cohort studies or propensity score matching [60] can be used otherwise.

Specificity

Once biomarkers of clinical response to treatment have been identified, the next challenge will be to test if they are predictive of response to a specific treatment or if they are specific to a particular disease. Here numerous datasets are required, including patients exposed to alternative treatments, across different stages of disease and across different inflammatory diseases. In the UK, these datasets have already been established in highly successful national stratified medicine programmes in RA, psoriasis, systemic lupus erythematosus, Sjogren's syndrome, autoimmune hepatitis and primary biliary cirrhosis. The Immune-Mediated Inflammatory Disease Biobanks in the UK (IMID-Bio-UK) [61] is bringing together these datasets with the aim of harnessing biological samples, deeply phenotyped clinical cohorts and high quality biomarker data to address related and overlapping precision medicine questions that cannot be addressed exclusively in the individual collections.

Statistical challenges

Analysing real world databases, with or without biomarker data, exposes a number of potential biases that require consideration. A recent editorial published in *Nature* has highlighted statistical challenges that should be considered in designing and analysing studies. These include regression to the mean, natural variation, the selection of an appropriate outcome measure and using a continuous rather than a dichotomous outcome. Some of these issues are discussed in more detail in the review by Prof. Til Sturmer as part of this supplementary edition.

Interpreting findings from biomarker studies

A large number of biomarker studies have been published yet few markers have reached the clinic. When interpreting the findings of biomarker studies, therefore, a degree

of scepticism is healthy. Validation of findings is a prerequisite but even if replicated in independent datasets, a biomarker may not add sufficient information to make it clinically useful. How predictive a biomarker needs to be to be clinically useful will depend very much on the context in which it is used. For example, diagnostic accuracy is very important for tests that will determine whether treatment is given or withheld but a test with less stringent performance characteristics may be adequate where choices are being made between treatments, i.e. preferential prescribing (complementary diagnostics). Several studies are underway currently to establish recommended cut-offs for tests aimed at guiding selection of the first biologic therapy for RA, for example. Demonstration of clinical utility may need to be followed by health economic assessment to show that the cost of the test does not outweigh the benefit accrued. Helpfully, the OMERACT Consortium has developed guidelines for the clinical validation of biomarkers [50].

Summary and future perspective

In summary, sceptics would argue that, despite huge investment in biomarker research, few successes have emerged. We would argue that biomarker studies have added value to real world data collections in inflammatory arthritis and have already identified biomarkers that are being used or have the potential to be used in routine clinical practice. These include anti-CCP testing, which is already in use; drug level and antidrug antibody testing in patients being treated with TNFi, which is currently being reviewed by NICE; and anti-CarP antibody testing, which requires further assessment of clinical utility. A major goal for biomarker research is to help inform and personalize the choice of treatment recommended to patients to maximize clinical benefit. Before this goal can be realized in RA, a quantitative disease activity measure(s), which is easy to measure for all patients and is reflective of synovial inflammation, is first needed and will likely be discovered by well-designed biomarker experiments. For example, assessing blood-based biomarkers for correlation with MRI-observed synovitis might identify a biological surrogate for active joint-specific inflammation that is better than CRP alone and can be objectively measured in all patients. Other major obstacles to overcome include the inadequate control of confounding, which can partly be addressed by testing for drug adherence and anti-drug antibody titres and by more sophisticated methods of matching patient groups for comparison. Reproducibility of biomarker discoveries to date has been hampered by a lack of statistical power that can partly be addressed by increased funding for establishing and maintaining real world datasets and collection of biological samples as part of expected practice. Discovery of biomarkers of clinical importance will also likely require large collaborative efforts involving many academic groups and industry partners. The emergence of high dimensional datasets and advanced data-driven statistical methods, such as machine learning, offer the potential to develop accurate, robust and discriminative statistical classifiers of

important clinical outcomes in RA. It is possible that many small differences can be combined to develop scores that will, together, allow better prediction of clinically important outcomes; for example, genotypic risk scores have been proposed to aid identification of patients at the highest risk of developing cardiovascular disease [12]. However, underpowered and unreplicated studies will potentially suffer from over-fitting and poor model performance so any models developed need to be tested in a prospective clinical setting.

Funding: D.P. is funded by the NIHR Manchester BRC. We acknowledge support from Arthritis Research UK (grant ref 21754) and MRC/Arthritis Research UK stratified medicine award, MATURA (grant ref MR/K015346/1).

Disclosure statement: The authors have declared no conflicts of interest.

References

- Dalton WS, Friend SH. Cancer biomarkers—An invitation to the table. *Science* 2006;312:1165–8.
- Toes RE, van der Woude D. ACPA (anti-citrullinated protein antibodies) and rheumatoid arthritis. *Acta Reumatol Port* 36:205–7.
- Gwinnutt JM, Symmons DPM, MacGregor AJ *et al*. Twenty-year outcome and association between early treatment and mortality and disability in an inception cohort of patients with rheumatoid arthritis: results from the Norfolk Arthritis Register. *Arthritis Rheumatol* 2017;69:1566–75.
- National Institute for Health and Care Excellence (NICE). Rheumatoid arthritis in adults: management. Clinical Guideline CG79. London: NICE. <https://www.nice.org.uk/guidance/cg79/chapter/Recommendations#monitoring-rheumatoid-arthritis> (28 June 2018, date last accessed).
- Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987;30:1205–13.
- Okada Y, Wu D, Trynka G *et al*. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;506:376–81.
- Raychaudhuri S, Sandor C, Stahl EA *et al*. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 2012;44:291–6.
- Laufer VA, Chen JY. Integrative approaches to understanding the pathogenic role of genetic variation in rheumatic diseases. *Rheum Dis Clin North Am* 2017;43:449–66.
- Cope AP, Schulze-Koops H, Aringer M. The central role of T cells in rheumatoid arthritis. *Clin Exp Rheumatol* 2017;25(5 Suppl 46):S4–11.
- Eyre S, Bowes J, Diogo D *et al*. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet* 2012;44:1336–40.
- Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. *Nat Rev Rheumatol* 2013;9:141–53.
- Khera AV, Chaffin M, Aragam KG *et al*. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018;50:1219.
- Knevel R, Gröndal G, Huizinga TWJ *et al*. Genetic predisposition of the severity of joint destruction in rheumatoid arthritis: a population-based study. *Ann Rheum Dis* 2012;71:707–9.
- Viatte S, Plant D, Han B *et al*. Association of HLA-DRB1 haplotypes with rheumatoid arthritis severity, mortality, and treatment response. *JAMA* 2015;313:1645.
- Rantapää-Dahlqvist S, de Jong BAW, Berglin E *et al*. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. *Arthritis Rheum* 2003;48:2741–9.
- Brink M, Hansson M, Mathsson-Alm L *et al*. Rheumatoid factor isotypes in relation to antibodies against citrullinated peptides and carbamylated proteins before the onset of rheumatoid arthritis. *Arthritis Res Ther* 2016;18:43.
- Nishimura K, Sugiyama D, Kogata Y *et al*. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. *Ann Intern Med* 2007;146:797–808.
- Kay J, Upchurch KS. ACR/EULAR 2010 rheumatoid arthritis classification criteria. *Rheumatology* 2012;51(suppl 6):vi5–9.
- Gan RW, Trouw LA, Shi J *et al*. Anti-carbamylated protein antibodies are present prior to rheumatoid arthritis and are associated with its future diagnosis. *J Rheumatol* 2015;42:572–9.
- Brink M, Verheul MK, Rönneild J *et al*. Anti-carbamylated protein antibodies in the pre-symptomatic phase of rheumatoid arthritis, their relationship with multiple anti-citrulline peptide antibodies and association with radiological damage. *Arthritis Res Ther* 2015;17:25.
- Humphreys JH, Verheul MK, Barton A *et al*. Anticarbamylated protein antibodies are associated with long-term disability and increased disease activity in patients with early inflammatory arthritis: results from the Norfolk Arthritis Register. *Ann Rheum Dis* 2016;75:1139–44.
- Hyrich KL, Watson KD, Silman AJ, Symmons DPM; British Society for Rheumatology Biologics Register. Predictors of response to anti-TNF-therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. *Rheumatology* 2006;45:1558–65.
- Sergeant JC, Hyrich KL, Anderson J *et al*. Prediction of primary non-response to methotrexate therapy using demographic, clinical and psychosocial variables: results from the UK Rheumatoid Arthritis Medication Study (RAMS). *Arthritis Res Ther* 2018;20:147.
- Carmona L, Descalzo MA, Ruiz-Montesinos D *et al*. Safety and retention rate of off-label uses of TNF antagonists in rheumatic conditions: data from the

- Spanish registry BIOBADASER 2.0. *Rheumatology* 2011;50:85–92.
- 25 Potter C, Hyrich KL, Tracey A *et al.* Association of rheumatoid factor and anti-cyclic citrullinated peptide positivity, but not carriage of shared epitope or PTPN22 susceptibility variants, with anti-tumour necrosis factor response in rheumatoid arthritis. *Ann Rheum Dis* 2008;68:69–74.
- 26 Ibfelt EH, Jensen DV, Hetland ML. The Danish nationwide clinical register for patients with rheumatoid arthritis: dANBIO. *Clin Epidemiol* 2016;8:737–42.
- 27 Dream. <http://www.dreamregistry.nl/nl/> (9 August 2018, date last accessed).
- 28 Zufferey P, Dudler J, Scherer A, Finckh A. Disease activity in rheumatoid arthritis patients at initiation of biologic agents and 1 year of treatment: results from the Swiss SCQM registry. *Jt Bone Spine* 2013;80:160–4.
- 29 Wadström H, Eriksson J, Neovius M, Askling J; ARTIS Study Group. How good is the coverage and how accurate are exposure data in the Swedish Biologics Register (ARTIS)? *Scand J Rheumatol* 2015;44:22–8.
- 30 The MATURA Consortium. Matura: MAXimising Therapeutic Utility in Rheumatoid Arthritis. <http://www.matura.whri.qmul.ac.uk/> (19 September 2017, date last accessed).
- 31 Cui J, Stahl EA, Saevarsdottir S *et al.* Genome-wide association study and gene expression analysis identifies CD84 as a predictor of response to etanercept therapy in rheumatoid arthritis. *PLoS Genet* 2013;9:e1003394.
- 32 Plant D, Maciejewski M, Smith S *et al.* Profiling of Gene Expression Biomarkers as a Classifier of Methotrexate Nonresponse in Patients With Rheumatoid Arthritis. *Arthritis Rheumatol*. 2019. doi: 10.1002/art.40810.
- 33 Massey J, Plant D, Hyrich K *et al.* Genome-wide association study of response to tumour necrosis factor inhibitor therapy in rheumatoid arthritis. *Pharmacogenomics J* 2018;18:657–64.
- 34 Taylor JC, Bongartz T, Massey J *et al.* Genome-wide association study of response to methotrexate in early rheumatoid arthritis patients. *Pharmacogenomics J* 2018;18:528–38.
- 35 Prevoo ML, van Riel PL, van 't Hof MA *et al.* Validity and reliability of joint indices. A longitudinal study in patients with recent onset rheumatoid arthritis. *Br J Rheumatol* 1993;32:589–94.
- 36 Plant D, Prajapati R, Hyrich KL *et al.* Replication of association of the PTPRC gene with response to anti-tumour necrosis factor therapy in a large UK cohort. *Arthritis Rheum* 2012;64:665–70.
- 37 Bek S, Bojesen AB, Nielsen JV *et al.* Systematic review and meta-analysis: pharmacogenetics of anti-TNF treatment response in rheumatoid arthritis. *Pharmacogenomics J* 2017;17:403–11.
- 38 Nelson MR, Tipney H, Painter JL *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* 2015;47:856–60.
- 39 Plant D, Wilson AG, Barton A. Genetic and epigenetic predictors of responsiveness to treatment in RA. *Nat Rev Rheumatol* 2014;10:329–37.
- 40 Plant D, Webster A, Nair N *et al.* Differential methylation as a biomarker of response to etanercept in patients with rheumatoid arthritis. *Arthritis Rheumatol* 2016;68:1353–60.
- 41 Kolker E, Higdon R, Haynes W *et al.* MOPED: Model Organism Protein Expression Database. *Nucleic Acids Res* 2012;40(Database issue):D1093–9.
- 42 Boucher P, Li W-P, Matz RL *et al.* LRP1 functions as an atheroprotective integrator of TGFbeta and PDGF signals in the vascular wall: implications for Marfan syndrome. *PLoS One* 2007;2:e448.
- 43 Li MO, Wan YY, Sanjabi S, Robertson A-K, Flavell RA. Transforming growth factor-β regulation of immune responses. *Annu Rev Immunol* 2006;24:99–146.
- 44 Carini C, Hunter E, Scottish Early Rheumatoid Arthritis Inception cohort Investigators, Ramadass AS, Green J, Akoulitchev A *et al.* Chromosome conformation signatures define predictive markers of inadequate response to methotrexate in early rheumatoid arthritis. *J Transl Med* 2018;16:18.
- 45 Jani M, Chinoy H, Warren RB *et al.* Clinical utility of random anti-tumor necrosis factor drug-level testing and measurement of antidrug antibodies on the long-term treatment response in rheumatoid arthritis. *Arthritis Rheumatol* 2015;67:2011–9.
- 46 Jani M, Isaacs JD, Morgan AW *et al.* High frequency of antidrug antibodies and association of random drug levels with efficacy in certolizumab pegol-treated patients with rheumatoid arthritis: results from the BRAGGSS cohort. *Ann Rheum Dis* 2017;76:208–13.
- 47 Sazonovs A, Kennedy N, Moutsianas L *et al.* HLA-DQA1*05 is associated with the development of antibodies to anti-TNF therapy. *bioRxiv* 2018;410035.
- 48 Jani M, Dixon WG, Lunt M *et al.* OP0229 The association of biologic drug-levels with infection risk: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. *Ann Rheum Dis* 2018;77(Suppl 2):163–4.
- 49 Tasaki S, Suzuki K, Kassai Y *et al.* Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. *Nat Commun* 2018;9:2755.
- 50 Maksymowych WP, Landewe R, Boers M *et al.* Development of draft validation criteria for a soluble biomarker to be regarded as a valid biomarker reflecting structural damage endpoints in rheumatoid arthritis and spondyloarthritis clinical trials. *J Rheumatol* 2007;34:634–40.
- 51 DAS28. <https://www.das-score.nl/das28/en/> (12 July 2017, date last accessed).
- 52 Centola M, Cavet G, Shen Y *et al.* Development of a multi-biomarker disease activity test for rheumatoid arthritis. *PLoS One* 2013;8:e60635.
- 53 Hurnakova J, Zavada J, Hanova P *et al.* Serum calprotectin (S100A8/9): an independent predictor of ultrasound synovitis in patients with rheumatoid arthritis. *Arthritis Res Ther* 2015;17:252.
- 54 Hambardzumyan K, Bolce R, Saevarsdottir S *et al.* Pretreatment multi-biomarker disease activity score and

- radiographic progression in early RA: results from the SWEFOT trial. *Ann Rheum Dis* 2015;74:1102–9.
- 55 Cordingley L, Prajapati R, Plant D *et al.* Impact of psychological factors on subjective disease activity assessments in patients with severe rheumatoid arthritis. *Arthritis Care Res* 2014;66:861–8.
- 56 Hensor EMA, McKeigue P, Buch MH *et al.* O12 Validity of a2-component imaging-derived disease activity score (2C-DAS28) for improved assessment of synovitis in early rheumatoid arthritis. *Rheumatology* 2018;57(suppl_3):key075.194.
- 57 DAS28. EULAR response criteria. <https://www.das-score.nl/das28/en/difference-between-the-das-and-das28/importance-of-das28-and-tight-control/eular-response-criteria.html> (6 July 2017, date last accessed).
- 58 Bluett J, Morgan C, Thurston L *et al.* Impact of inadequate adherence on response to subcutaneously administered anti-tumour necrosis factor drugs: results from the Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate cohort. *Rheumatology* 2015;54:494–9.
- 59 Bluett J, Riba-Garcia I, Hollywood K *et al.* A HPLC-SRM-MS based method for the detection and quantification of methotrexate in urine at doses used in clinical practice for patients with rheumatological disease: a potential measure of adherence. *Analyst* 2015;140:1981–7.
- 60 Kihara M, Davies R, Kearsley-Fleet L *et al.* Use and effectiveness of tocilizumab among patients with rheumatoid arthritis: an observational study from the British Society for Rheumatology Biologics Register for rheumatoid arthritis. *Clin Rheumatol* 2017;36:241–50.
- 61 GtR. MICA: Immune-Mediated Inflammatory Disease Biobanks in the UK (IMIDBio-UK). <https://gtr.ukri.org/projects?ref=MR%2FR014191%2F1> (10 August 2018, date last accessed).