Erratum to:

Therapeutic monitoring of TNF-alpha inhibitors in rheumatoid arthritis (DAR 17/10/02) (date 7 January 2019)

Page 1:

- Provision of 3 author names in full (Andy to Andrew, Rich to Richard)
- Removal of Dr Martin Pitt from the author list as he had not reviewed the final repot prior to submission. Acknowledgment for contribution made to the development of the report added to the acknowledgments on Page 2

Page 2:

- Update to “Declared competing interests of the authors”. An updated conflict of interest form was received from Dr Richard C Haigh after the report was submitted on 7 January, and Dr Meghna Jani changed to Dr Jani at the author’s request
- Addition of Martin Pitt to the “Acknowledgments”
- Revision of author list following removal of Martin Pitt as a co-author in the section “This report should be referenced as follows”

Page 3:

- Removal of Martin Pitt from “Contributions of authors” to reflect changes made to author list (see also Page 1 and Page 2)
- Correction of author contribution for Andrew Salmon to better reflect his contribution to the diagnostic appraisal report (DAR)

Abstract: Pages 4 to 5

- Revisions made to the abstract to align with corrections and/or clarifications made in the main report (see below)

Scientific Summary: Pages 6 to 38

- **Pages 8-13:** Revisions made due to the revisions made to Section 2 (removal of studies from the clinical effectiveness section
- **Page 20:** Correction to text
- **Page 22:** Correction of utility values reported in the table
- **Page 24 to 26, 28 and 31 to 32:** Corrections to results following corrections made in the model (e.g. utility value, corrections in calculations)

- **Page 33 to 34 and Page 37:** Corrections due to removal of included studies from the clinical effectiveness review and also due to corrections in the model

Plain English Summary: Pages 39 to 40

- SECTION REPLACED
    - Edited to fit the word count and style specified by the National Institute for Health Research (NIHR)

Section 1 Background

- **Page 68:**
    - Correction to listed variations/kits for RIDASCREEN, addition of code numbers

Section 2 Assessment of clinical effectiveness

- **Pages 81 to 85, 87 to 89, 91 to 92, 95 to 99, 103 to 116:**
    - In response to a comment from NICE Technical Team regarding the included observational studies, the seven observational studies were reviewed by two additional reviewers and discussed within the broader Team. The AG therefore excluded six of the seven studies as follows: Senabre Gallego (2017); Chen (2016); Inciarte-Mundo (2016); Lopez-Casla (2013); Rosas (2015); and Paredes (2015 and 2016). Reasons for exclusion are provided within the report. Given the paucity of evidence, broader evidence was used to inform model parameters.

Section 3 Systematic review of existing cost-effectiveness evidence

- **Page 121:** The citation in Table 31 has been changed from Ucar 2017 to Arango 2017
- **Page 129 and Page 130:** Figure 5 was redrawn to align with the reported algorithm. The text in the preceding paragraph was also edited to align with changes in Figure 5
- **Page 133 to 134:** Edits to clarify the results reported by Gavan 2017 were made

Section 4 Independent economic assessment

- **Page 140 to 142:** updated due to changes in Section 2
- **Pages 145 to 150**
    - Removal of duplicate text proceeding Table 42 (word error)
    - Table 42 correction of utility values; footnotes added to clarify duration of follow-up in the model

    - Section 4.1.3.2 “Primary and secondary non-responders”: removal of text from this section to align with changes in Section 2
- **Page 151 to 152 and 155:** Clarifications regarding modelling of flares in the AG model
- **Page 152:** addition of text comparing patients characteristics from the INGEBIO study and BSRBR registry; clarification of assumptions related to time in remission
- **Page 160 to 161:** Removal of duplicate text in Section 4.1.9 (word error)
- **Pages 163 to 167:**
    - Page 163 footnote added to Table 50 to note that Solymbic not available in EU
    - Page 165 removal of duplicate text from Section 4.1.9.5 (word error)
    - Pages 166 to 167 removal of duplicate text (word error)
- **Pages 171 to 174**
    - Page 171 CIC mark up added
    - Section 4.1.9.1.9 corrections to text and Table 54 and Table 55 to align with information provided by manufacturer
- **Pages 185 to 186** Removal of duplicate text (word error)
- **Page 189**
    - Correction to Table 66
    - Clarification on utility values for mixed disease states (remission/LDA; LDA/active disease)
- **Page 196:**Correction to the explanation regarding the derivation of utilities
- **Page 197 to 201 and 204 to 209**
    - Results updated based on corrections made in model (corrections to utilities and errors in formulae in the model)

Section 5 Discussion

- **Pages 210 to 215:** Text edited to align text with revisions made to Section 2 and Section 3.

Section 6 Conclusions

- **Pages 218 and 219:** Text edited to align text with revisions made to Section 2 and Section 3.

Appendix 2: Included and excluded studies

- **Pages 277 to 278:** Removal of seven observational studies from the list of included studies (studies added to the list of excluded studies table [see below, page 325])

(see Section 2 edits above for rationale). Included studies were placed in alphabetical order by study design.

- **Page 329:** 7 observational studies removed from the included studies table were added to this table of excluded studies (see Section 2 edits above for rationale).

Appendix 3: Quality Assessment

- **Pages 330 to 409:** Section updated to reflect edits made in Section 2 (see Section 2 edits above for rationale)

Therapeutic monitoring of TNF-alpha inhibitors in rheumatoid arthritis

Produced by

Peninsula Technology Assessment Group (PenTAG)
University of Exeter Medical School
South Cloisters
St Luke's Campus
Heavitree Road
Exeter
EX1 2LU

Authors

Irina Tikhonova, *Senior Research Fellow*[1]
Huiqin Yang, *Senior Research Fellow*[1]
And***rew*** Salmon, *Associate Research Fellow*[2]
Segun Bello, *Postdoctoral Research Associate*[1]
Jaime Peters, *Senior Research Fellow*[1]
Sophie Robinson, *Information Specialist*[1]
Mohsen Rezaei, *Research Fellow*[1]
Sophie Dodman, *Research Assistant*[1]
Helen Coelho, *Research Fellow*[1]
Andriy Kharechko, *Postdoctoral Research Associate*[1]
Rich***ard*** C Haigh, *Consultant Rheumatologist*[3]
Meghna Jani, *NIHR Clinical Lecturer in Rheumatology*[4]
Timothy J McDonald, *Consultant Clinical Scientist and Clinical Associate Professor*[3,5]
Martin Hoyle, *Associate Professor of Health Technology Assessment*[1]
Stuart Logan, *Cerebra Professor of Paediatric Epidemiology*[5]

[1] Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School
[2] PenCHORD, University of Exeter Medical School
[3] Royal Devon and Exeter hospital
[4] University of Manchester
[5] University of Exeter Medical School

Correspondence to

Dr Irina Tikhonova
Peninsula Technology Assessment Group (PenTAG)

University of Exeter Medical School
South Cloisters
St Luke's Campus
Heavitree Road
Exeter
EX1 2LU
[redacted]

Date completed 07/01/2019

Source of funding This report was commissioned by the NIHR HTA Programme as project number 17/10/02.

Declared competing interests of the authors Dr Jani declared receipt of speaker fees from Grifols. Dr Haigh reported grants from Pfizer and personal fees from Pfizer outside the submitted work. The remaining authors listed declared no competing interests.

Acknowledgments We would like to thank Rob Anderson for his support in preparation of the protocol and in the finalisation of the report, Ruben Mujica Mota for providing comments on the report, Jo Thompson-Coon for overall support of the project, Martin Pitt for his contribution to the PenTAG independent economic assessment, and Louise Crathorne for support during the finalisation of the project (all from the University of Exeter Medical School). We are grateful to Sean Gavan and Katherine Payne from the University of Manchester for their advice on economic assessment of TNF testing; Grant Smith, Nicola Finch, Leanne Brown, Keith Oldfield and Rob Storey from the Royal Devon & Exeter NHS Foundation Trust for their advice on costing. We acknowledge the excellent administrative support of Sue Whiffin and Jenny Lowe.

Rider on responsibility for report The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows: Tikhonova I, Yang H, Salmon A, Bello S, Peters J, Robinson S, Rezaei M, Dodman S, Coelho H, Kharechko A, Haigh R***C***, Jani M, McDonald TJ, Hoyle M, Logan S. Therapeutic monitoring of TNF-alpha inhibitors in rheumatoid arthritis: 2019, Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School (Report for NICE).

Contributions of authors

Segun Bello	Assisted with tasks on development of the protocol for the *Assessment of clinical effectiveness* section, screened records for the clinical-effectiveness systematic review, conducted data extraction for included studies, conducted study quality assessment, contributed to drafting and editing of the *Assessment of clinical effectiveness* section
Helen Coelho	Wrote the *Background* section of the report
Sophie Dodman	Conducted a review of literature to inform costs, contributed to the writing of the *Resources and costs* section
Richard C Haigh	Provided clinical advice and contributed to the editing of the report
Martin Hoyle	Contributed to the development of the NICE Scope and Protocol and was the Project Director till 29 June, 2018
Meghna Jani	Provided clinical advice, academic input regarding therapeutic drug monitoring and contributed to the editing of the report
Andriy Kharechko	Conducted a review of literature to inform utilities, contributed to the writing of the *Utilities* section
Stuart Logan	Director and Guarantor of the report
Timothy McDonald	Provided clinical advice and contributed to the editing of the report
Jaime Peters	Contributed to the development of the PenTAG independent economic assessment, including screening titles and abstracts for an additional systematic review supporting the PenTAG independent economic assessment; contributed to the writing and editing of the report
Mohsen Rezaei	Contributed to the cost-effectiveness systematic review as the first reviewer, drafted the *Systematic review of existing cost-effectiveness evidence* and *Independent economic assessment* sections, contributed to the writing of the model code
Sophie Robinson	Developed and conducted the literature searches for the systematic reviews of clinical effectiveness and cost effectiveness, and for the additional systematic review for comparator tests; developed and conducted the searches for the literature review of health utilities; contributed to the writing, editing and proofreading of the report
Andrew Salmon	Contributed to the cost-effectiveness systematic review as the second reviewer, screening of titles and abstracts for the additional systematic review supporting the PenTAG independent economic assessment; contributed to the writing of the *Systematic review of existing cost-effectiveness evidence* and *Independent economic assessment* sections; checked the working of the simulation model between the submission deadline, and erratum/addendum phase
Irina Tikhonova	Provided overall project management; led the health-economic and modelling team; contributed to the writing of the *Abstract, Scientific summary*, *Systematic review of existing cost-effectiveness evidence*, *Independent economic assessment, Discussion* and *Conclusion* sections, and the model code; collated the report
Huiqin Yang	Led the clinical effectiveness of the project, and the writing of the *Abstract* and *Assessment of clinical effectiveness* sections; contributed to the writing of the *Scientific summary, Discussion* and *Conclusion* sections of the report

Please refer to the International Committee for Medical Journal Editors (ICMJE) Uniform Requirements for Manuscripts submitted to Biomedical Journals see http://www.icmje.org/

PAGES 4 to 5

Abstract

Background

Rheumatoid arthritis (RA) is a chronic systemic autoimmune disease, primarily causing inflammation, pain and stiffness (synovitis) in the joints. Those with severe disease may be treated with biological disease modifying anti-rheumatic drugs (bDMARDs), including TNF-α inhibitors. Monitoring response to these treatments typically involves clinical assessment and the use of response criteria (DAS28 or EULAR).

Commercial enzyme-linked immunosorbent assay (ELISA) tests can also now be used to detect and measure drug concentrations and drug antibody levels in the blood. These tests may inform whether adjustments to treatment are required, or help clinicians to understand the reasons for treatment non-response or a loss of response.

Methods

A systematic review was conducted to identify studies reporting the clinical and cost-effectiveness of using ELISA tests to measure drug levels and anti-drug antibodies for monitoring response to TNF-α inhibitors (adalimumab [ADL], etanercept [ETN], infliximab [IFX], certolizumab pegol [CTZ], and golimumab [GLM]) in people with RA who had achieved treatment target (remission or low disease activity [LDA]), or in those with primary non-response or a secondary non-response to treatment.

An economic analysis was conducted to estimate health and economic outcomes of adding TNF-α inhibitors testing to usual practice to guide treatment decisions. The costs and resource use were considered from the perspective of the NHS and Personal Social Services. No discounting was applied to costs or effects due to the short-term time horizon used. Sensitivity analyses explored the effect of different uncertainties on the economic outcomes.

Results

Two studies (in four publications) were identified. One was a non-randomised trial (the INGEBIO study) that compared TDM with standard care but had serious limitations in relation to the NICE scope: one-third of participants had RA, analyses were mostly not by intention-to-treat, follow-up was only 18 months, and, there was no explicit algorithm for guiding clinicians in using the test results to inform treatment.One observational study was

identified but was of minimal value in informing whether ELISA test-based monitoring is clinically effective or not.

The exploratory economic analyses were inconclusive and show considerable uncertainty in the cost-effectiveness of TDM of TNF-alpha inhibitors in RA. Different outcome data from the same study produced opposite conclusions on the cost-effectiveness of Promonitor testing in people receiving ADL who are in remission/LDA. Results based on the longer follow-up outcomes suggested that monitoring is more costly and produces fewer QALYs than standard care. Of the sensitivity analyses conducted, only the impact of monitoring on the rate of flares impacted substantially on the results. Exploratory analyses of using Promonitor to monitor patients in remission/LDA receiving ETN or INF showed the same results as that for ADL.

Conclusions

There is limited valid and applicable research evidence, and much uncertainty in relation to key potential drivers of the effectiveness and cost-effectiveness of using ELISA-based testing for the TDM of TNF-α inhibitors in RA and no firm conclusions are possible.

Funding

This report was commissioned by the NIHR HTA Programme as project number 17/10/02.

Word count: 487 words

extraction and quality assessment were undertaken by one researcher and checked by a second. If RCTs had been identified, the Cochrane Risk of Bias tool for RCTs would have been used. The risk of bias was assessed using the Cochrane (ROBINS-1) tool for non-randomised studies with adaptations as appropriate.

For clinical effectiveness outcomes, mean differences, relative risks (RRs), odds ratios (ORs) or incidence rate ratios (IRRs) with 95% confidence intervals (CIs) were extracted from comparative studies, where reported. Quantitative methods of synthesis were not possible given the lack of studies and data were synthesised narratively (i.e. using text and tabulated information).

Assessment of cost-effectiveness

A systematic review of published economic evaluations of using ELISA tests relative to the alternatives and standard care was undertaken:

- To gain insights into the key drivers of cost-effectiveness of TNF testing.
- To get an overview of the alternative modelling approaches that have been adopted to evaluate the use of therapeutic drug monitoring in people with RA.
- To provide a summary of the findings of previous relevant cost-utility, cost-effectiveness, and cost-benefit studies.

Eligible studies for inclusion to the systematic review were selected according to the following criteria:

- The inclusion criteria for population and interventions were as in the clinical-effectiveness systematic review.
- The following types of economic evaluations were included: cost-utility analyses, cost-effectiveness analyses, cost-benefit analyses, cost-consequence analyses and cost-minimisation analyses. Systematic reviews of economic studies were also included.

Screening was done independently by two reviewers. Disagreements between reviewers were resolved by discussion. All references considered to meet the inclusion criteria by either reviewer at the title and abstract stage were included for full-text screening.

Due to the lack of RCTs identified in the clinical effectiveness systematic review, additional literature searches to identify studies evaluating *any tests* used to monitor anti-TNF-α treatment of people with rheumatoid arthritis were conducted.

The choice of modelling approach was primarily driven by the availability and quality of the evidence identified in the clinical effectiveness systematic review and the additional searches. Due to the limited evidence identified in the clinical-effectiveness systematic review, the multifactorial nature of decisions to adjust treatments in people with RA and the recent changes in the biologics market, which contributed to the uncertainty in the prices of biologics and their uptake within the UK, a simplified modelling approach (threshold analysis), was chosen. Cost-utility analyses were also conducted assuming the estimates of the cost of testing derived from Jani and colleagues (2016). The costs and resource use were considered from the perspective of the NHS and Personal Social Services. No discounting was applied to estimated costs and quality-adjusted life-years (QALYs) due to a short-term time horizon adopted in the study. Clinical outcomes from the INGEBIO study were used in all the analyses. Costs considered in the economic evaluation included the costs of testing, the cost of treatments received by people with RA, and healthcare costs. The costs of testing comprised those of the test kits, staff time to perform the tests and staff training, the cost of the testing service and sample transport. The costs were obtained from the British National Formulary (BNF) and the NHS Reference Costs, from the documents provided by the test manufacturers, and published and unpublished sources. Quality-adjusted life-years (QALYs) were used as as the outcome measure of health benefit. A review of HRQoL studies informed the selection of health states' utilities, and disutilities for flares and AEs. Various sensitivity analyses were conducted to explore the effect of structural and parametric uncertainties on the economic outcomes. Probabilistic sensitivity analysis was deemed inappropriate because of a very substantial variation in clinical practice with respect to treatment, drug dose-tapering and flare management strategies in people with RA. The effect of such variations on the economic outcomes was explored in one-way deterministic sensitivity analyses and a number of scenario analyses.

Results

Clinical effectiveness

No studies met the systematic review's original inclusion criteria. Due to this, the review criteria were broadened to include studies where people with RA make up less than 70% of the study population.

Two studies (reported in four publications) were, therefore, included in the systematic review. Three abstracts (Arango and colleagues 2017; Gorostiza and colleagues 2016; Ucar and colleagues 2017) reported the same non-randomised controlled trial (the INGEBIO study), but this was conducted in people being treated for a range of diseases including RA.

The other study included a historical control group prior to the introduction of treatment monitoring with ELISA tests.

The INGEBIO study used Promonitor ELISA kits to monitor drug levels and/or anti-drug antibody levels and the study by Pascual-Salcedo (2013) used Sanquin ELISA kits. No studies were identified evaluating the following ELISA testing kits: IDKmonitor, LISA-TRACKER, RIDASCREEN and MabTrack.

Table 1 summarises which treatments and ELISA kits were used in the studies selected in the clinical-effectiveness systematic review.

Table 1: Clinical effectiveness evidence relevant to specific combinations of TNF-α inhibitors and test kits from the NICE scope

		Promonitor	IDKmonitor	LISA-TRACKER	RIDASCREEN	Sanquin*
ADL	*drug*	✓[1]	X	X	X	✓[2]
	antibody	✓[1]	X[3]	X	X	✓[2]
ETN	*drug*	X	X	X		✓[2]
	antibody	X	X	X		✓[2]
IFX	*drug*	X	X[3]	X	X	✓[2]
	antibody	X	X	X	X	✓[2]
GLM	*drug*	X	X	X		X
	antibody	X	X	X		
CTZ	*drug*			X		X
	antibody			X		

Key:

X Indicates availability of a test to measure drug or antibody level in people treated with the specified TNFi and that no studies have been identified in the clinical-effectiveness systematic review, reporting on using therapeutic drug monitoring for the specified test kit and TNFi.

✓ Indicates availability of a test to measure drug or antibody level in people treated with the specified TNFi and that at least one source for the specified combination of the test kit and TNFi has been identified in the clinical-effectiveness systematic review.

ADL: adalimumab; CTZ: certolizumab pegol; ETN: etanercept; GLM: golimumab; IFX: infliximab; TNFi: tumour necrosis factor inhibitor

Notes:

* The type of Sanquin test kits used in these studies (MabTrack or those used by Sanquin Diagnostic Services) was not reported.

[1] Arango and colleagues 2017, Ucar and colleagues 2017 and Gorostiza and colleagues 2016 (INGEBIO)

[2] Pascual-Salcedo and colleagues 2013

[3]Indicates that a test for total anti-drug antibodies is also available (total anti-drug antibodies include both unbound, i.e. free, antibodies and those bound to TNF-α inhibitor)

In the study that used the Sanquin testing service, the type of kits was not reported. Neither study specified the type of testing (*reflex* or *concurrent* testing).Both studies included individuals in remission, with the INGEBIO study also including individuals with low disease activity (at baseline). Both studies included mixed populations, with 37% and 49% of participants having RA in the INGEBIO and Pascual-Salcedo and colleagues (2013) studies,

respectively. The INGEBIO study had a (*mixed disease*) population of 169 patients, whereas the Pascual-Salcedo and colleagues (2013) study was smaller (43 patients across the mixed disease population).

The included studies measured drug levels and/or anti-drug antibody levels in people treated with ADL, ETNand/or IFX. There were no studies identified in people treated with CTZ and/or GLM. Both studies were judged to be at moderate risk of bias.

Comparative controlled evidence – one study

Three abstracts (Arango and colleagues, 2017; Gorostiza and colleagues, 2016; Ucar and colleagues, 2017) reported the same non-randomised controlled trial (the INGEBIO study), which focused on the population who had achieved treatment target (remission or LDA) and remained clinically stable for at least six months. ADL and anti-ADL antibody levels were measured using Promonitor ELISA kits. Monitoring test results were revealed to physicians in the intervention arm. Physicians did not follow any test-based treatment algorithm for the management of people with RA and they used their best judgements to optimise treatment doses. Such monitoring test results were not revealed to physicians in the control arm. This reflected standard care in Spain where treatment decisions were based on clinical judgements without knowledge of drug levels and anti-drug antibodies.

This trial recruited a mixed disease population of 169 participants, including 63 people with RA. The other participants in the study had psoriatic arthritis (PsA) and ankylosing spondylitis (AS). The results of the total mixed population were extracted as the study's authors were not willing to provide the separate results for the cohort of participants with RA. Therefore, there is limited generalisability of findings from this mixed disease population to the target RA population of this technology assessment.

The findings from this trial (Ucar and colleagues 2017) showed that, at 18-month follow-up, the rate of flares per patient-year was 0.463 for the intervention group and 0.639 for the control group, with a statistically non-significant rate difference of -0.176 (95% CI -0.379 to 0.0289). There was a non-significant reduction in risk of flare in the intervention group compared with the control group (incidence rate ratio (IRR) 0.7252, 95% CI 0.4997 to 1.0578). Median time to first flare was 145 days for participants in the intervention group and 136.5 days for participants in the control group. The intention-to-treat (ITT) analysis from the abstract by Gorostiza and colleagues (2016) showed that at 34-week follow-up, 67.5% (54/80) in the intervention group remained in remission while 64.0% (32/50) in the control group remained in remission, with the difference in proportions of 3.5% (95% CI -13.3% to 20.3%; p=0.68).

In terms of dose adjustment due to monitoring response, the finding by Arango and colleagues (2017) showed that aADL dose was tapered (i.e. reduced) in 35 participants of the intervention group (35.7%) and in 18 participants of the control group (34.6%). This trial (Ucar and colleagues, 2017) also reported that participants' HRQoL (EQ-5D-5L) measures were higher in the intervention group at all visits compared with the control group (further details were not reported). However, statistically significant results were only observed at Visit 2 (p=0.001) and Visit 3 (p=0.035), and no further details were reported.

Overall, the findings of this controlled study showed that there was a non-significant reduction in risk of flare in the intervention group compared with the control group. HRQoL measures were higher in the intervention group at all visits compared with the control group, with statistically significant results being observed at two visits. However, there was an imbalance at baseline in disease severity between the intervention and control groups, and a lack of adjusting for this baseline imbalance in the analysis of clinical outcomes. Also, there were higher attrition rates for some outcomes. On top of the limited applicability to populations with RA, and the lack of full description of methods (in abstracts), these deficiencies resulted in serious risk of bias associated with the findings.

Observational evidence - one study

One observational study, including a mixed-disease population of 43 individuals, was identified evaluating the effect of TDM (with ELISA-based testing) on clinical outcomes in people with RA who had achieved remission.

This study used Capture ELISA (Sanquin) in people with arthritis receiving IFX, ADL, or ETN, and included a historical control (i.e. comparing pre- and during-TDM practice). Changes in disease activity and in the direction and magnitude of therapeutic dose were reported. The study showed a non-significant reduction in the mean DAS28 score following the implementation of TDM at seven-year follow-up (pre-TDM: mean 2.51 [SD 0.85] vs.during-TDM: 2.31, [SD 0.52]; p=0.061). This study also showed statistically significant reductions in weekly mean dose per patient by each anti-TNF and increases in mean interval of administration for each anti-TNF following the implementation of TDM; unfortunately, these results related to the wider study sample with more than half of participants with PsA rather than RA.

Cost effectiveness

Cost-effectiveness systematic review

In the cost-effectiveness systematic review, five studies relevant to the decision problem were found: two were reported as abstracts (Pascual-Salcedo and colleagues, 2013 and

Ucar and colleagues, 2017), two as full-text journal article (Krieckaert and colleagues, 2015 and Laine and colleagues, 2016) and a PhD thesis (Gavan, personal communication, 6 August, 2018). Furthermore, only two (out of six) TNF testing kits from the NICE scope (Promonitor and Sanquin) and three (out of five) TNF inhibitors (ADL, ETN, IFX) were considered in the selected studies (Table 2).

anti-TNFs and similar performance of the the Promonitor test kits used for measuring the drug and antibody levels of different TNF inhibitors. Estimates of the cost of testing were based on Jani and colleagues (2016) and clinical advice.

Exploratory analysis for those in remission/LDA

The clinical and economic effect of ADL tapering in people with RA in remission/LDA was evaluated in the INGEBIO study (Arango and colleagues, 2017, Ucar and colleagues, 2017 and Gorostiza and colleagues, 2016).

The clinical outcomes reported in Ucar and colleagues (2017) were incorporated in our economic analysis to estimate the cost of drug and antibody testing, at which the addition of ELISA testing to usual clinical practice would result in zero NMB.

Since the patent for the adalimumab originator product, Humira®, has expired in October 2018, and the acquisition costs for the ADL biosimilars were not known to the AG at the time of writing, in the threshold analysis, the annual acquisition cost was varied from £1,000 to £9,187 per patient-year; the latter represents the annual cost of Humira® assuming the dose of 40 mg every two weeks delivered by subcutaneous injection using pre-filled pen and the NHS indicative price from the BNF.

The other major assumptions were as follows, with further details in Table 3:

- ADL dose tapering is implemented by increasing the interval between doses from two to three weeks (i.e. by spacing doses).
- Dose is tapered in a proportion of people in each arm at the start of simulation.
- Some people may flare after reducing the dose of their TNF inhibitors (Bykerk and colleagues, 2016).
- The full dose of ADL is restored *in all people* on tapered doses when they flare.
- Treatment wastage is £370 per patient-year in people on full dose; it is reduced proportionally to the reduction in treatment dose.
- ADL is self-administered (usually at home), and, therefore, the administration cost is zero.
- The costs associated with flare management are:
    - £423 per flare for diagnostic investigations
    - £68 per month for treatment (excluding the cost of DMARDs)
- The annual per-patient costs of managing remission and LDA/active disease health states are £11,409 and £18,889 respectively.
- The cost of managing an AE is £1,622.

Assumption	**Estimate**	**Source**	**Relevant table/sections in the report**
Administration cost for Humira® (ADL) (per patient-year)[2]	£0	Clinical advice	Section 4.1.10.1.7
Cost of flare management [3, 4]	£423/per flare	Cost of diagnostic investigations (Maravic and colleagues, 2005	Section 4.1.10.1.19
	£68/month	Monthly cost of treatment (excluding DMARDs) (Maravic and colleagues, 2005)	Section 4.1.10.1.19
Cost of managing health states (per patient-year) [5]			
Remission	£11,409	Barbieri and colleagues (2005), Radner and colleagues (2014), National Schedule of Reference Costs 2017-18	Section 4.1.10.1.16
LDA/active disease	£18,889		Section 4.1.10.1.16
Cost of managing AEs (per infection)	£1,622[6]	TA375[4]	Section 4.1.10.1.20
Utilities			
Remission	0.718	Estimated from HAQ scores for different HAQ bands reported by Radner and colleagues (2014)	Section 4.1.10.2.1
LDA/active disease	0.568[7]		Section 4.1.10.2.1
Disutility of flare	0.140	Markusse and colleagues, 2015	Section 4.1.10.2.2
Disutility of AEs	0.156	TA375, Oppong and colleagues (2013)	Section 4.1.10.2.3
Flare rate			
Intervention	0.463	Ucar and colleagues 2017	Section 4.1.9.1.1
Control	0.639	Ucar and colleagues 2017	Section 4.1.9.1.1
Mean time to first flare (days)			
Intervention	208.07	Derived from Kaplan-Meier estimates (from the INGEBIO study) of time to first flare, provided by Ucar and colleagues (personal communication, 9 September, 2018)	Section 4.1.9.1.3
Control	189.32		Section 4.1.9.1.3
Flare duration (days)[8]	7	TA375	Section 4.1.9.1.2

SCIENTIFIC SUMMARY PAGE 24 to 26

Mortality associated with RA was not modelled, and no discounting was applied to the costs and outcomes due to the short-term time horizon of about 18 months adopted in this study.

All costs were inflated to 2017-18 prices using the healthcare price index.

Results: adalimumab and Promonitor

Threshold analysis

The results of the threshold analysis, assuming the Promonitor test kit is used to monitor people with RA in remission/LDA recieving originator ADL (Humira®) are presented in Table 4 and Figure 1. Figure 1 shows the annual cost of ELISA-based testing at which TDM would become cost-effective at the two WTP thresholds used in NICE decision making for the range of ADL acquisition costs of £1,000–£9,187. Since the data reported in Arango and colleagues (2017)[12] are for a longer follow-up than that reported in Ucar and colleagues (2017), the results using the two different reports of the outcomes of the INGEBIO study are presented.

If the results of Ucar and colleagues (2017) are used, then with the current price of originator ADL, testing would need to be cheaper than £430 per year in order for TDM to be judged as cost-effective at the cost-effectiveness threshold of £20,000 per QALY gained. Using the results presented in Arango and colleagues (2017), however, there would be no cost of testing at which testing becomes cost-effective (because using these outcomes testing was estimated to be both more costly and less effective than standard care).

Table 4: Threshold value for the cost of testing

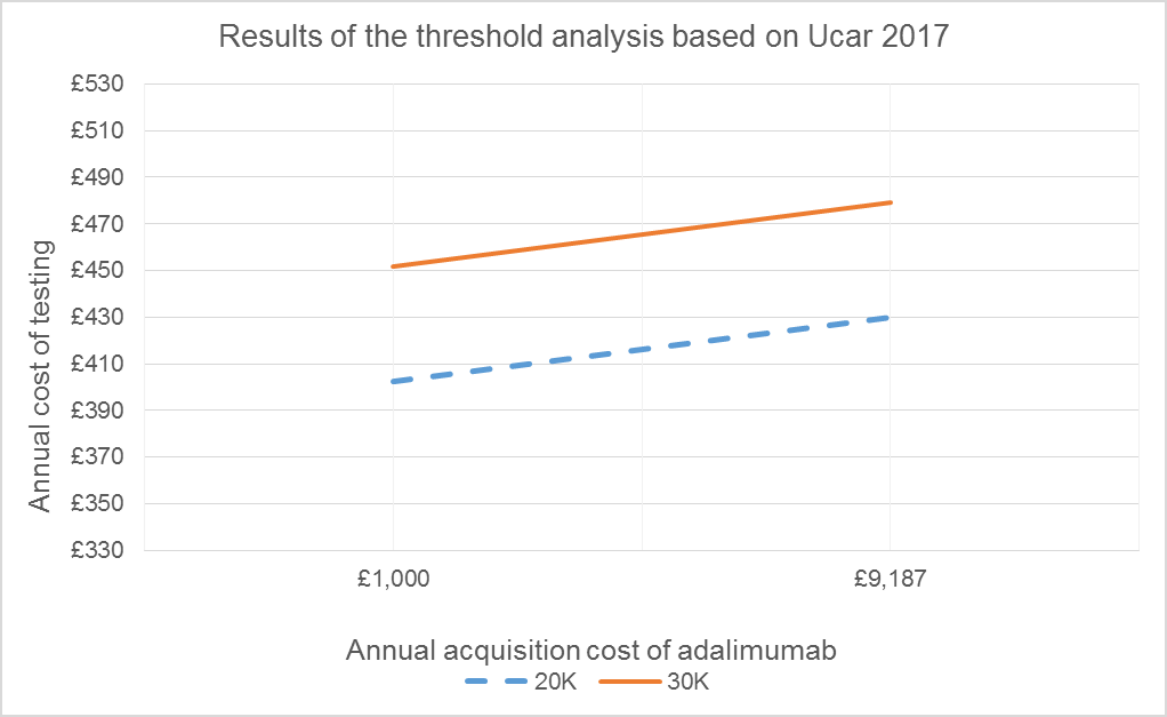
ICER threshold	Results based on INGEBIO study, Ucar and colleagues 2017	Results based on INGEBIO study, Arango and colleagues 2017
£20,000	£430	-£200
£30,000	£479	-£246

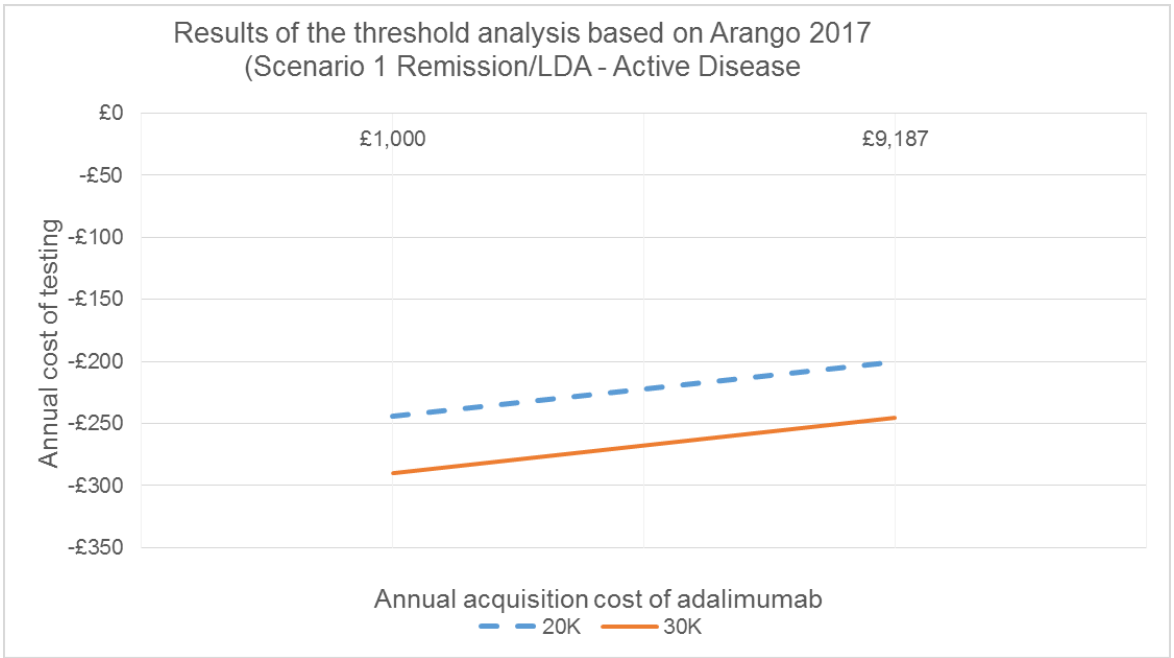
Key: ICER: incremental cost-effectiveness ratio

Such differences in the results are due to differences in the mean duration of remission (as reported in Ucar 2017) and remission/LDA (Arango 2017) between the control and intervention arms. Arango reported a longer duration of remission/LDA in the control group than in the intervention group (475.2 versus 460.2 days), while Ucar and colleagues 2017

reported a longer duration in the intervention group (344 versus 329 days in the control group).

Figure 1: Results of the threshold analyses Arango and colleagues (2017) and Ucar and colleagues (2017)





Source: Ucar and colleagues (2017) and Arango and colleagues (2017)

These results are inconclusive for two reasons. First, because they are in opposite directions and, second because they are based on very small and uncertain differences in outcomes

(QALY differences of less than 0.01). The negative value of the cost of testing at which NMB equals zero means that, when using the trial results as presented in Arango and colleagues (2017), there are no (positive) values of the cost of testing at which it would be a cost-effective option.

Cost-utility analysis

The incremental QALYs and incremental costs for testing versus standard care strategy are shown in Table 5, assuming:

- regular testing is undertaken in people with RA in remission/LDA treated with Humira® and tested using Promonitor
- the costs of testing are as in Jani and colleagues (2016)
- the frequency of testing is *one test per patient-year* and
- that testing of drug and antibody levels is done *concurrently (singlet dilution) in a UK laboratory.*

The outcome data were derived from two reports of the INGEBIO study, Ucar 2017 and Arango 2017.

Table 5: Cost-effectiveness results in patients in remission/LDA treated with Humira® and tested using Promonitor

	Intervention arm	Control arm	Intervention vs. control
Based on Ucar and colleagues (2017)			
QALYs (mean)	0.924	0.918	0.007
Total costs (mean)	£32,178	£32,438	-£260
ICER (Cost / QALY gained)			ICER not relevant - Intervention dominates standard care
Based on Arango and colleagues (2017)			
QALYs (mean,)	0.947	0.954	-0.007
Total costs (mean)	£36,284	£35,923	£361
ICER (Cost / QALY gained)			ICER not relevant - Standard care dominates Intervention

Key: ICER: incremental cost effectiveness ratio; QALY: quality-adjusted life year; vs: versus
Note: The postage was £4 per parcel.

As with the threshold analyses, these results are inconclusive for two reasons. First, because they are in opposite directions and, second because they are based on very small and uncertain differences in outcomes (QALY differences of less than 0.01). Furthermore, it

is not possible to argue that either the analysis based of Ucar and colleagues (2017) or that based on Arango and colleagues (2017) is more valid than the other – they both have significant weaknesses. The follow-up in Arango and colleagues (2017) is over a longer time

Table 6: Additional sensitivity analyses (people in remission/low disease activity)

Sensitivity analysis	Assumptions	ICER		Source (relevant sections)
		Ucar and colleagues (2017)	**Arango and colleagues (2017)**	
Impact of flares only (health states and AEs are not included)	Only flares contribute to differential costs and QALYs	ICER = £72,645/QALY	ICER = £8,804/QALY	Scenario C (people in remission, Gavan 2017, Section 3.3.2.3)
Tapering strategy	Spacing: reduction of ADA dose to 40mg every 4 weeks	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	2nd dose reduction Exeter biologic clinic recommendations (Appendix 4)
Treatment wastage	No wastage	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Assumption
Flare duration, days	19	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Weighted average based on Bykerk and colleagues (2014)[13 16 16 16 16 16 16] and clinical advice
Proportion of flared patients in whom full dose is restored	55%	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Bykerk and colleagues (2014) and clinical advice
	0%	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Clinical advice
Utilities[2]				
Remission	0.496	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Estimated from HAQ scores reported in TA375 (Fig. 94, p.366) (Section 4.1.10.2.1)
LDA/active disease	0.302			

SCIENTIFIC SUMMARY PAGE 31 to 32

In all but one sensitivity analysis the intervention dominated standard care when data from Ucar and colleagues (2017) was used, and it was dominated by standard care if data were derived from Arango and colleagues (2017). When the impact of *flares only* was modelled (i.e. health states and AEs were not included), the ICERs in the analyses by Ucar and colleagues and Arango and colleagues were £72,645/QALY and £8,804/QALY, respectively (Table 6).

The first four deterministic sensitivity analyses (Table 7) used estimates from Arango and colleagues (2017) only, as it was expected that results based on data from Ucar and colleagues (2017) would be similar to those in the major analysis, i.e. the intervention would dominate standard care. Changing these parameters had no impact on the findings, standard care was estimated to dominate the intervention in all analyses. A sensitivity analysis was conducted in which the ADL acquisition cost was varied: data from Ucar and colleagues the intervention was estimated to dominate standard care and using data from Arango and colleagues standard care was estimated to dominate the intervention.

Table 7: One-way deterministic sensitivity analyses

Parameter	Assumption	ICER	Source
Percentage of people in whom the biologic was tapered	+20% in the intervention arm and -20% in the control arm	ICER not relevant - Standard care dominates Intervention	Arango and colleagues (2017)
Flare rate	-20% in the intervention arm, +20% in the control arm	ICER not relevant - Standard care dominates Intervention	Arango and colleagues (2017)
Differential time in remission	+10% in the intervention arm, -10% in the control arm of the differential time in remission	ICER not relevant - Standard care dominates Intervention	Arango and colleagues (2017)
Costs of managing health states	- 20%	ICER not relevant - Standard care dominates Intervention	Arango and colleagues (2017), Radner and colleagues (2014), Barbieri and colleagues (2005)

Key: ICER: incremental cost-effectiveness ratio

Sensitivity analyses were conducted and are presented in the main cost-effectiveness results section of the report.

Other scenario analyses considered but not conducted due to no or low quality clinical data were: analysis of testing in the context of primary or secondary non-response; analysis for non-responders who did not adhere to treatment with biologic therapies, including switching to intravenously administered IFX.

Probabilistic sensitivity analysis was deemed inappropriate because of a very substantial variation in clinical practice with respect to disease management in people with RA in England.

Results: etanercept and infliximab and Promonitor

The cost-effectiveness of TNF testing in people treated with Etanercept (originator and biosimilar) and Infliximab (biosimilar) using the Promonitor test kits was explored in scenario analyses. Enbrel is the most expensive originator product considered in this assessment, while biosimilars Erelzi and Flixabi/Renflexis have the lowest acquisition costs among the TNF inhibitors administered via subcutaneous and intravenous routes, respectively.

In those analyses, it was assumed, based on clinical advice (and a lack of evidence to the contrary), that the clinical effectiveness of the different TNF inhibitors is likely to be the same, and the clinical effectiveness estimates from Ucar and colleagues (2017) were adopted, with all assumptions, except acquisition and administration costs, as in Table 3.

The information on the actual costs to the NHS of the TNF inhibitors was not available to the AG at the time of writing, and therefore the list prices of the biologics were assumed. The results are presented in Table 8.

Table 8: Cost-effectiveness results for the other tests and TNF inhibitors: people in remission/LDA

Treatment		ICER	
	Cost per year (£)	**Ucar and colleagues (2017)**	**Arango and colleagues (2017)**
ETN			
*Enbrel®**	9,327	ICER not relevant - Intervention dominates standard care (ICER -£38,247 (total costs -£261; total QALYs 0.007])	ICER not relevant - Standard care dominates Intervention (ICER £53,203 (total costs £360; total QALYs -0.007])
Erelzi	8,394	ICER not relevant - Intervention dominates standard care (ICER -£37,597 [total costs -£256; total QALYs 0.007)	ICER not relevant - Standard care dominates Intervention (ICER -£54,351 [total costs £368; total QALYs -0.007)

IFX[1]			
Flixabi/ Renflexis (no wastage)	5164	ICER not relevant - Intervention dominates standard care (ICER -£36,580 [total costs -£249; total QALYs 0.007)	ICER not relevant - Standard care dominates Intervention (ICER -£56,144 [total costs £380; total QALYs -0.007)

Key: ETN: etanercept; ICER: incremental cost-effectiveness ratio; IFX: infliximab

Notes:

It was assummed that blood samples would be sent for testing to UK laboratories, and the postage of £4 (per small parcel) was applied[14].

* The originator (or reference) product

[1] IFX administration cost was assumed to be 283 per injection (Section 4.1.10.1.7).

Source: Ucar and colleagues (2017)[8] and Arango and colleagues (2017)[12]

Discussion

Strengths and limitations

Clinical effectiveness

A comprehensive literature search was undertaken to identify both published and unpublished studies. These included electronic searches of a variety of bibliographic databases as well as screening of clinical trial registers and conference proceedings to identify unpublished studies. The review process followed recommended methods to minimise the potential for error and/or bias. The quality of included studies was assessed and accounted for when interpreting the review results. Appropriate synthesis methods were employed taking into account the heterogeneity of study characteristics.

In terms of limitations, non-English-language studies were excluded. Only two relevant studies were identified for the evaluation of clinical effectiveness of TDM based on ELISA testing in the target populations. No evidence was found in relation to clinical effectiveness of therapeutic drug monitoring in people with RA who had experienced a primary non-response or a secondary non-response. Only one relevant non-randomised controlled trial that was conducted in a mixed disease population (including only 37% people with RA) was identified. No studies were identified evaluating the following ELISA kits: IDKmonitor, LISA-TRACKER, RIDASCREEN and MabTrack. There was considerable clinical heterogeneity between the two included studies.

Only in the INGEBIO study, also included in the systematic review of cost-effectiveness studies, was an ELISA test-guided treatment compared against standard care/monitoring. In this study, however, physicians were not obliged to follow any test-based treatment algorithm but could use testing to alter doses, based on their judgement, in patients from the intervention arm. Moreover, the study was reported in abstracts only, and the reported outcomes may not be directly relevant to the NHS clinical practice since the study was conducted in Spain. Therefore, an additional systematic literature review to identify RCTs evaluating *any tests* used to monitor anti-TNF-α treatment of people with RA was conducted to support the economic assessment. However, no relevant sources were identified.

Cost effectiveness – systematic review evidence

A systematic review of published economic evaluations of using ELISA tests relative to the alternatives and standard care was undertaken to help inform the type and structure of the independent economic assessment. The results of this review indicate limited existing evidence on the cost-effectiveness of therapeutic drug monitoring in people with RA. Despite a comprehensive search of the literature, only five studies were identified. Two (out of five)

TNF testing kits from the NICE scope (Promonitor and Sanquin) and three (out of five) TNF inhibitors (ADL, ETN, and IFX) have been assessed in the selected studies. The systematic review was also limited by reporting as two (out of five) selected studies were reported as abstracts. These studies therefore mainly informed the planning of the independent model-based analysis.

Cost effectiveness – model-based analysis

Despite substantial weaknesses in the clinical effectiveness evidence base, a simple model was developed to estimate the cost-utility of ELISA test-based monitoring for people with RA taking bDMARDS.

The analyses conducted are inconclusive and suggest considerable uncertainty in the cost-effetciveness of therapeutic monitoring of TNF-alpha inhibitors in RA. Data from two reports of the same study produced very different conclusions on the cost-effectiveness of Promonitor testing in people receiving ADL who are in remission/LDA. The results based on the longer follow-up (Arango and colleagues 2017[12]) suggested that monitoring is more costly and produces fewer QALYs than standard care.

Of the sensitivity analyses conducted, only one impacted the results: the assumption that the rate of flares alone changes as a consequence of monitoring. The ICERs in the analyses by Ucar and colleagues (2017) and Arango and colleagues (2017) were £72,645/QALY and £8,804/QALY, respectively.

Exploratory analyses of using Promonitor to monitor patients in remission/LDA receiving ETN or IFX were undertaken, and showed the same results as that for ADL: using the longer follow-up (Arango and colleagues 2017) monitoring is more costly and produces fewer QALYs than standard care.

The main effectiveness evidence in the model was from the poorly reported INGEBIO study (a non-randomised controlled trial from Spain, where <40% of participants had RA), heavily supplemented by input parameters from other studies and expert advice. The results of the economic analysis should therefore be viewed as exploratory and highly speculative. For example, although the INGEBIO study only evaluated testing using Promonitor ELISA kits, for those in remission/LDA treated with Humira® (ADL), with further assumptions these results have been used to estimate the threshold testing costs at which TDM would become cost-effective with people taking other TNF inhibitors (and taking either originator products or biosimilars.

results of changes in therapeutic dose from the historically controlled study by Pascual-Salcedo (2013), were for a mixed population (including RA, PsA and/or AS).

Cost effectiveness

Outcomes from the INGEBIO study were also used in the economic analysis for people in remission/LDA. It was a pragmatic trial, and therefore it is likely that the results could be generalisable to routine practice settings. However, the generalisability to UK clinical practice settings of the findings in the INGEBIO study (in Spain) and therefore the economic results remain uncertain.

Since findings from the mixed population considered in the INGEBIO study might not be generalisable to the RA population, and the quality of this trial was judged to be at serious risk of bias, the economic results presented here should be considered with caution.

Due to the severe paucity of relevant studies, not all test kits and TNF inhibitors from the NICE scope could be modelled using reported clinical outcomes considered in this study. It is therefore not clear whether and to what extent the economic estimates obtained for people treated with ADL are applicable to people treated receiving the other anti-TNF treatments.

According to NHS England documentation, some originator manufacturers have offered discounts, changing the potential for cost savings for the NHS. Therefore, the list prices of TNF inhibitors assumed in the analyses reporting ICERs might not adequately reflect the actual costs of the biologics to the NHS in the coming years.

Conclusions

The findings from this assessment demonstrate very limited evidence on the effect of TDM based on ELISA tests for optimising anti-TNF therapies in people with RA, either in those who had achieved remission or LDA, or in those who had experienced a primary non-response or a secondary non-response.

In relation to clinical effectiveness, limited data were identified evaluating TDM in the target populations. One non-randomised trial compared TDM with standard care (the INGEBIO study) had serious limitations in relation to the NICE scope: only one-third of the participants had RA, many of the analyses were not by intention-to-treat, follow-up was for only 18 months, there was no explicit algorithm for guiding clinicians in how the results of testing should change treatment (e.g. tapering), and the study was only reported in three abstracts. In addition, a single observational study was also identified but was of minimal value in informing whether ELISA test-based monitoring is clinically effective or not.

Plain English Summary

Rheumatoid arthritis (RA) is a long-term condition that causes pain, swelling and stiffness in the joints. The symptoms usually affect the hands, feet and wrists. Those with severe disease may be treated with drugs called TNF-α inhibitors (adalimumab, etanercept, infliximab, certolizumab pegol, and golimumab). Some people improve on these drugs, whereas others improve initially and then lose response. One cause of lost response is that individuals develop antibodies against the drug which cancel out the effect of treatment.

Tests have been developed to measure the level of drug and the level of antibodies against these drugs in blood samples. Monitoring in this way could allow for treatment to be adjusted in response to the test outcomes to optimize benefit for the patient, and also to help clinicians to better understand the reasons for non-response or a loss of response to treatment.

The aim of this study was to find out whether it would be clinically effective (i.e. good for patients and their families) and cost effective (i.e. a good use of limited NHS resources) to use these tests to monitor drug levels and anti-drug antibodies as a means of assessing treatment response in people with RA who are controlled, or have not responded or have lost response.

There was limited published evidence on the effect of using these tests to measure levels of drugs or antibodies to drugs to inform treatment decisions. A simple mathematical model drew on evidence from one poorly reported study and was heavily supplemented by input parameters from other studies and expert advice. Results from the model were inconclusive and suggest considerable uncertainty in the cost effectiveness.

Given substantial uncertainty, the results presented in this study should be considered with caution.

Word count: 285

Technologies	Company	Variations/kits	Drug/antibodies assessed
		LISA-TRACKER anti-golimumab (LTG003)	Free[1] anti-GLM antibodies
		LISA-TRACKER Duo adalimumab (LTA005)	Total[2] ADL
		LISA-TRACKER Duo certolizumab (LTC005)	Total[2] CTZ
		LISA-TRACKER Duo etanercept (LTE005)	Total[2] ETN
		LISA-TRACKER Duo Infliximab (LTI005)	Total[2] IFX
RIDASCREEN	R-Biopharm	RIDASCREEN ADM monitoring (G09043)	Free[1] ADL
		RIDASCREEN anti-ADM antibodies (G09044)	Free[1] antibodies to ADL
		RIDASCREEN IFX monitoring (G09041)	Free[1] IFX (Remicade®, Remsima®, Inflectra®)
		RIDASCREEN anti-IFX antibodies (G09042)	Free[1] antibodies to IFX
MabTrack ELISA kits	Sanquin	MabTrack level adalimumab M2910	Free[1] ADL
		MabTrack ADA adalimumab M2950	Free[1] antibodies to ADL
		MabTrack level infliximab M2920	Free[1] IFX (Remicade®, Remsima®, Inflectra®)
		MabTrack ADA infliximab M2960	Free[1] antibodies to IFX

Key: ADL: adalimumab; CTZ: certolizumab pegol; ETN: etanercept; GLM: golimumab; IFX: infliximab

Notes:

[1] Free TNF-α inhibitor is drug that is unbound to antibody, and free anti-drug antibodies are those that are unbound to drug.

[2] Total anti-drug antibodies include both unbound (free) antibodies and those bound to TNF-α inhibitor.

SECTION 2: PAGES 81 to 85

2.2.4 Critical appraisal strategy

One reviewer independently assessed the quality of included studies in terms of risk of bias. If RCTs had been identified, the Cochrane Risk of Bias tool for RCTs would have been used.[56] The Cochrane (ROBINS-1) tool was used for non-randomised studies with adaptations as appropriate.[57] The Cochrane (ROBINS-1) tool was used to assess the quality of uncontrolled observational studies with adaptations as appropriate, although the tool was primarily designed for non-randomised controlled studies. The risk of bias of included studies was taken into account when interpreting results. The quality assessment was checked by another reviewer. Any disagreements were resolved by consensus.

2.2.5 Methods of data synthesis

Given the clinical heterogeneity associated with interventions, outcomes and length of follow-up and the methodological heterogeneity identified (e.g. different study designs), quantitative synthesis was not possible and clinical effectiveness data were synthesised in a narrative fashion. Publication bias could not be investigated because quantitative synthesis was not possible.

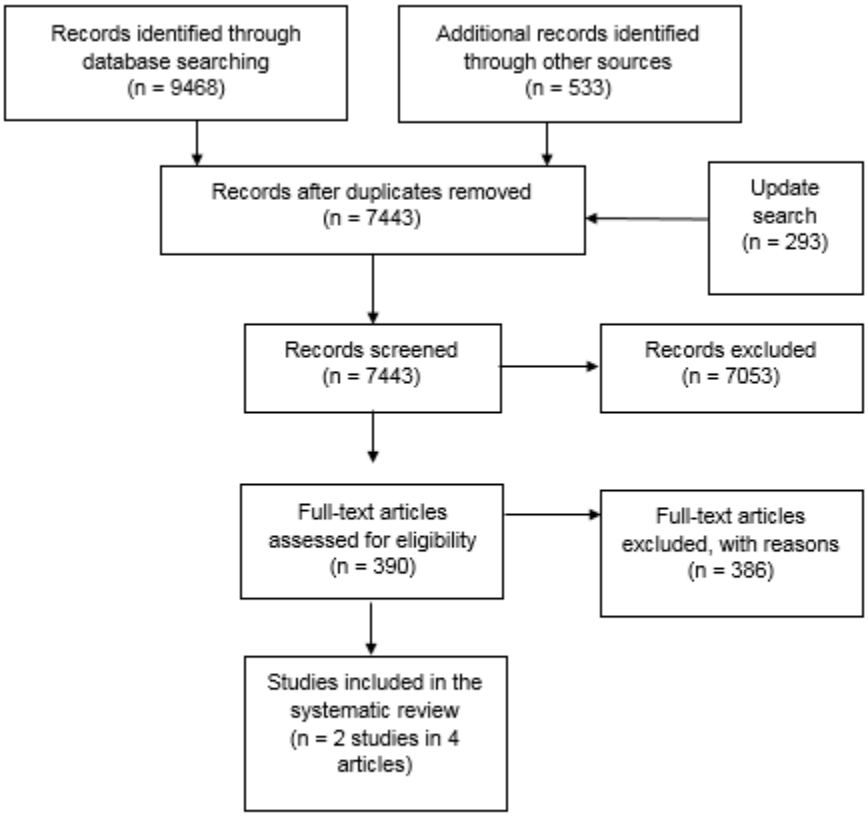
2.3 Clinical effectiveness results

The next section provides information on the quantity of research available, including characteristics and risk of bias of the included studies. This is then followed by the results section with clinical effectiveness of therapeutic drug monitoring by using ELISA tests in people with RA who were treated with TNF-α inhibitors.

2.3.1 Quantity and quality of research available

The literature searches of bibliographic databases identified 7,443 references. After initial screening of titles and abstracts, 390 were considered to be potentially relevant and were ordered for full paper screening. In total, two studies reported in four articles[11,15,59,64] were included in the systematic review of clinical effectiveness of using ELISA tests for monitoring response to TNF-α inhibitors in people with RA. Both included trials with linked citations are presented in *Appendix 2.* Figure 2 shows a flow diagram outlining the screening process with reasons for exclusion of full-text papers.

Figure 2: Flow diagram of study inclusion process for the clinical-effectiveness review



One study was reported in three abstracts, with considerable overlap in data and reporting. The paper with the most up-to-date and complete data was selected for the data extraction.

Number and type of studies excluded

A list of full-text papers that were excluded along with the reasons for their exclusions is given in Appendix 2. These papers were excluded because they failed to meet one or more of the inclusion criteria in terms of the type of study design, participants, interventions or outcomes being reported.

2.3.2 Assessment of clinical effectiveness

2.3.2.1 Characteristics of included studies

The characteristics of included studies are presented in Table 14 and Table 15.[11,15,59,64] Both studies recruited people with RA who had achieved treatment target (remission or LDA). One study reported in three abstracts used Promonitor ELISA kits for monitoring drug levels and/or anti-drug antibody levels.[11,15,64] One study[59] used Sanquin ELISA kits to measure drug levels and/or anti-drug antibody levels of three anti-TNFs (IFX, ADL, and ETN for the

treatment of RA. The type of Sanquin test kit used in this study was not reported. The two included studies were conducted in Spain. Neither study reported funding sources.

Non-randomised controlled studies

Three abstracts[11,15,64] were identified reporting the same non-randomised control trial (the INGEBIO study). In this trial, monitoring testing results of drug levels and anti-drug antibodies were revealed to physicians in the intervention arm. The monitoring test results were not revealed to physicians in the control arm. This reflected standard care in Spain where treatment decisions were based on clinical judgements without the knowledge of drug levels and anti-drug antibodies. Given that this was a pragmatic trial, it is likely that the findings could be generalisable to routine practice settings. For standard care in the control arm, clinicians did not follow any national guideline for the management of people with RA as there were no national guidelines for monitoring in Spain at the time of the study. Clinicians used their best judgements to optimise treatment doses. This trial recruited a mixed population of 169 people with RA (n=63), PsA (n=54) and ankylosing spondylitis (n=52) recruited from three sites in Spain. The study focused on the population who had achieved treatment target (remission or LDA) and remained clinically stable for at least six months.

The included abstracts reported a sample size of people with RA ranging from 54 to 63 at baseline. The abstracts by Ucar and colleagues (2017) and Arango and colleagues (2017) reported results on the basis of 18-month follow-up. The abstract by Gorostiza and colleagues[64] reported results only based on 34-week follow-up. This trial reported the following relevant clinical outcomes: change in disease response, dose adjustment due to monitoring response (e.g. proportion of participants tapered), and participants' HRQoL outcomes.

The median duration of disease at baseline among participants in the three abstracts[11,15,64] ranged from 117 to 124 months. All participants were treated with adalimumab (ADL) 40 mg (via subcutaneous injection). ADL and anti-adalimumab antibody (ADAb) levels were measured using Promonitor-ADL and Promonitor-ANTI-ADL (Grifols-Progenika). The frequency of testing in this trial was once every two to three months. There were a total of eight visits during the trial period (details were not provided).

Observational study

One observational study[59] assessed the clinical effectiveness of using ELISA tests for monitoring response to TNF-α inhibitors in people with RA. The study recruited people who had achieved treatment target (remission or LDA) and had a historical control. This

observational study reported the following relevant clinical outcomes: change in disease activity, and change in direction and magnitude of therapeutic dose.

The observational study used Sanquin ELISA kits for measuring drug levels for three anti-TNFs (IFX, ETN, ADL). The sample size was 43. The study measured only the anti-TNF drug levels.[59] It was unclear whether drug trough levels were assessed in thestudy.

The included studies did not report other outcomes such as number of inconclusive results, time to result, frequency of treatment switch to an alternative biologic, rates of hospitalisation and rates of surgical interventions.

The observational study measured drug levels and/or anti-drug antibody levels in participants treated with ADL, ETN and IFX. No studies were identified in which participants were treated with certolizumab pegol and golimumab. No studies reporting on the use of ELISA testing in people with RA receiving biosimilar products were identified. No relevant studies (including both controlled trials and observational studies) were identified that assessed other eligible ELISA kits including IDKmonitor ELISA kits, LISA-TRACKER ELISA kits, RIDASCREEN ELISA kits and MabTrack ELISA kits.

SECTION 2 PAGE 87 to 89

Table 15: Characteristics of the included studies – Observational study

Study	Study date	Location	Study design	Population	Description of tests	Frequency of Measuring	Sample size	Length of follow-up
Pascual-Salcedo 2013[59]	2006-2012	Spain	Historically controlled study	Remission/LDA	Drugs: IFX, ADL, ETN Capture ELISA (Sanquin, Amsterdam)	NR	43	7 years

Key: ADL: adalimumab; ELISA: enzyme linked immunosorbent assay; ETN: etanercept; IFX: infliximab; LDA, low disease activity; NR, not reported

2.3.2.2 Baseline characteristics of included studies

Baseline characteristics of included studies are presented in Table 16 and Table 17. The mean age of participants enrolled in the Ucar abstract was 53.59 years but the mean age for participants was not reported for the observational study. The proportion of females was at least 75% of the total population in each study. The mean disease duration of RA was 17 years in the observational study.

The definition of remission/LDA was described as DAS28 score <3.2 in the observational study but not reported in the non-randomised controlled study (see Table 16 and Table 17). Both studies used one or more anti-TNF therapies (ADL, IFX, or ETN) for the treatment of RA. The mean treatment duration for participants receiving anti-TNF inhibitors was six years in the observational study but not reported in the non-randomised controlled trial.

Only methotrexate was reported as a co-therapy in the non-randomised controlled trial while none were reported in the observational study.

SECTION 2 PAGE 91 to 92

Table 17: Baseline characteristics – observational study

Study/ Author	Mean age (yrs)	Sample size	Definition of Remission (DAS28)	Definition of LDA (DAS28)	Definition of flare (DAS28)	% male	Mean disease duration (yrs)	Mean time on biologic (yrs)	Co-therapies	Anti-TNF received	Dose manipulation
Pascual-Salcedo 2013[59]	NR	43	<3.2[a]	<3.2[a]	NR	NR	17.52 (SD 9.38)	5.85 (SD 1.33)	NR	ADL; ETN; IFX (doses NR)	Optimization strategy (adjusting drug dose according to clinical activity)

Key: ADL: adalimumab; ETN: etanercept; IFX: infliximab; LDA: low disease activity; NR: not reported; SD: standard deviation; r; yrs: years

Notes:

[a] Grouped as 'remission or LDA

relating to risk of bias were assessed for each *individual* study: confounding, selection, group classification, co-interventions, missing data, outcome measurement and selective outcome reporting. The quality assessments on the basis of all relevant domains for each study and of specific outcomes are presented in Appendix 3. Table 19 and Table 20 present the quality assessment of included studies.

Table 19 presents the quality assessment of the non-randomised controlled study (the INGEBIO study).[11.15.64] This non-randomised controlled study was judged to be at serious risk of bias. There was an issue of *baseline imbalance* in the proportions of participants with remission and LDA between the intervention and control groups: 73.4% of participants were in remission at baseline in the intervention group while 83.3% of participants were in remission at baseline in the control group. The remaining participants (i.e., 26.6 % of participants in the intervention group and 16.7% of participants in the control group) had achieved LDA at baseline. Furthermore, there was a lack of adjustment for this baseline imbalance variable in the analysis of clinical outcomes. These deficiencies resulted in serious risk of bias associated with the findings.

Table 21 presents the attrition rates for each outcome of the non-randomised controlled study (the INGEBIO study).[11.15.64] As seen in Table 21 there were high attrition rates for three outcomes (proportions of participants who remained in remission, proportions of participants who changed from LDA to remission and proportions of participants who received dose tapering). The attrition rates ranged from 11.2% to 30.8%, which can lead to attrition bias. Furthermore, there were unbalanced attrition rates in these outcomes between the intervention and control groups.

Table 20 presents the quality assessment of the observational study.[55] The study had a historical control group and was judged to be at moderate risk of bias because there was non-contemporaneous control bias due to the use of historical control in this study. It should be noted that the same group of participants were assessed during the first period (i.e. the historical control where TDM was not introduced) and the second period (where TDM was implemented). Attrition rates are shown in Table 22.

Overall, the non-randomised controlled study[11.15.64] was judged to be at serious risk of bias while the observational study was judged to be at moderate risk of bias.

Table 19: **Risk of bias in included studies – The INGEBIO non-randomised controlled study**

Studies	Confounding (differential prognosis between groups)	Selection	Group Classification	Co-intervention	Missing data	Outcome measurement	Selective outcome reporting	Overall risk of bias
Arango 2017[15]	Serious	Low	Low	NI	Serious	Moderate	Low	Serious
Gorostiza 2016[64]	Serious	Low	Low	NI	Serious	Moderate	Low	Serious
Ucar 2017[11]	Serious	Low	Low	NI	NI	Moderate	Low	Serious

Risk of bias judgement: low/moderate/serious/critical/NI NI: no information

Table 20: **Risk of bias in included studies – observational study**

Studies	Confounding	Selection	Group Classification	Co-intervention	Missing data	Outcome measurement	Selective outcome reporting	Overall risk of bias
Pascual-Salcedo 2013[59]	Moderate	Moderate	Moderate	NI	NI	Moderate	Low	Moderate

Key: Risk of bias judgement: low/moderate/serious/critical/NI NI: no information

Table 21: **Attrition in the INGEBIO non-randomised controlled study**

Outcome	Baseline population		Follow-up population		Percent attrition		
	IG	CG	IG	CG	IG	CG	Overall
Disease flare	109	60	Unclear	Unclear	Indeterminate	Indeterminate	Indeterminate
% remaining in remission	109	60	71	46	34.9 (109-71/109)	23.3 (60-46/60)	30.8 (169-117/169)
% change from LDA to remission	29	10	28	7	3.4 (29-28/29)	30.0 (10-7/10)	11.4 (39-35/35)
ADL tapering	109	60	98	52	10.1 (109-98/109)	13.3 (60-52/60)	11.2 (169-150/169)
HRQoL	109	60	Unclear	Unclear	Indeterminate	Indeterminate	Indeterminate

Key: ADL: adalimumab; CG: control group; HRQoL: health-related quality of life; IG: intervention group; LDA: low diease activity

Table 22: Attrition in the observational study (Pascual-Salcedo 2013)

Outcomes	Study	Baseline population	Follow-up population	Percent attrition
Mean DAS scores	Pascual-Salcedo 2013[55]	43	NI	Indeterminate
Weekly mean drug dose	Pascual-Salcedo 2013[55]	43	NI	Indeterminate
Mean interval of drug administration	Pascual-Salcedo 2013[55]	43	NI	Indeterminate

Key: DAS: disease activity score NI: no information

SECTION 2 PAGE 103 to 116

Table 24: **Health-related quality of life outcomes**

Study	Population	IG (N)	CG (N)	Length of follow-up (months)	Outcome measure	Relative measure[a]	P-value (IG vs CG) Visit 2	P-value (IG vs CG) Visit 3
Ucar 2017[11]	Remission/ LDA	109	60	18	Health-related quality of life (EQ-5D-5L)	Higher in IG throughout follow-up[b]	0.001	0.035
Arango 2017[15]	Remission/ LDA	98	52	18	Health-related quality of life (EQ-5D-5L)	Higher in IG throughout follow-up[b]	0.001	0.035

Key: CG: control group; IG: intervention group; LDA: low disease activity

Notes:

[a] All data included a mixed population of 169 patients (rheumatoid arthritis (RA), psoriatic arthritis (PsA) and ankylosing spondylitis)

[b] No specific number of patients for results specified

2.3.3.2 Observational study

The observational study[64] evaluated the effect of using ELISA tests for monitoring response to TNF-α inhibitors in people with RA.

The study included participants who had achieved treatment target (remission or LDA). [55] The study had a historical control and was judged to be at moderate risk of bias (see Section 2.3.2.4).

2.3.3.2.1 Change in disease activity

The observational study[64] evaluated the effect of TDM on change in disease activities at the duration of follow-up of seven years. This study recruited people with RA who had achieved remission or LDA. The study did not report relevant information on the duration of remission/LDA.[64] The sample size was 43. Table 26 presents the results of changes in disease activity.

The study by Pascual-Salcedo and colleagues (2013)[64] had a historical control (i.e. the first period where TDM was not introduced). The findings showed that the mean DAS28 score of participants was 2.51 (standard deviation (SD) 0.85) during historical control period. Compared with the historical control, there was a non-significant reduction in the mean DAS28 score (mean 2.31, SD 0.52) at seven-year follow-up during the second period where TDM was introduced (p=0.061).

Overall, the finding from the historically controlled study[64] showed that TDM was associated with a non-significant reduction in mean DAS28 scores at seven-year follow-up compared

with the historical control. It should be noted that the quality of data was judged to be at moderate risk of bias, which compromises the reliability of the findings.

Table 26: Change in disease activity

Study	Study design	Populations (e.g. remission)	Sample size	Missing data (at follow-up)	Length of follow-up	Outcome measure	Findings
Pascual-Salcedo 2013[59]	Historically controlled study	Remission/ LDA	43	NR	7 years	Mean DAS28 score	1st period: 2.51 (SD 0.85) 2nd period*: 2.31(SD 0.52), p=0.061

Key: DAS28: disease activity score in 28 joints; LDA: low disease activity

Notes:

*Therapeutic drug monitoring was introduced in the second period

2.3.3.2.3 Change in direction and magnitude of therapeutic dose

The observational study[59] evaluated the outcome of change in direction and magnitude of therapeutic dose in people with RA who had achieved remission or LDA. The study recruited participants who had achieved remission or LDA but did not report relevant information on duration of remission or LDA. The sample size of the study was 43. Table 27 presents the results of change in direction and magnitude of therapeutic dose. It should be noted that the results from the study by Pascual-Salcedo et al (2013)[59] on the change of therapeutic dose were presented for the mixed population (including 43 people with RA and 45 people with PsA). Therefore, there was limited generalisability of findings from this mixed population to the target RA population.

The findings from the study by Pascual-Salcedo and colleagues (2013)[59] demonstrated that, compared with the historical control (i.e. the first period where TDM was not used), there were statistically significant reductions in the weekly mean dose per participant by each drug during the second period following the introduction of TDM. For participants receiving IFX, a statistically significant reduction in the weekly mean dose per participant during the second period was observed (mean 0.42 mg/kg/week, SD 0.12), compared with the first period (mean 0.51 mg/kg/week, SD 0.14) ($p<0.001$). For participants receiving ADL, a statistically significant reduction in the weekly mean dose per participant during the second period was also observed (mean 15.52 mg/week, SD 4.81) for the second period, compared with the first period (mean 19.19 mg/week, SD 3.72) ($p<0.001$). Similarly, for participants receiving ETN, there was a statistically significant reduction in the weekly mean dose per participant during the second period (mean 35.04 mg/kg/week, SD13.37) for the second period, compared with the first period (mean 42.09 mg/kg/week, SD 13.25) ($p=0.009$).

The findings from the study by Pascual-Salcedo and colleagues (2013)[59] further showed that, compared with the historical control, there was a statistically significant increase in the mean interval of administration for each drug during the second period where TDM was implemented. For participants receiving IFX, a significantly increased mean interval of administration was observed during the second period (mean 9.7 weeks, SD1.44), compared with the first period (mean 8.52 weeks, SD 1.43) ($p<0.001$). For participants receiving ADL, a significantly increased mean interval of administration was also observed during the second period (mean 2.95 weeks, SD1.58), compared with the first period (mean 2.19 weeks, SD 0.58) ($p=0.007$). Likewise, for participants receiving ETN, a significantly increased mean interval of administration was observed during the second period (mean 1.61 weeks, SD 0.91), compared with the first period (mean 1.09 weeks, SD 0.27) ($p=0.004$).

Overall, the limited data from the observational study showed that TDM for optimisation of anti-TNF therapies was associated with reductions in therapeutic dose of anti-TNFs in people with RA who had achieved remission or LDA. This would be expected to lead to cost saving associated with TDM. However, the reliability of findings may be compromised by the poor quality of data being identified.

Table 27: Changes in number, direction and magnitude of dose

Study	Study design	Population (e.g. remission)	Sample size	Missing data (at follow-up)	Length of follow-up	Outcome measure	Findings
Pascual-Salcedo 2013[59]	Historically controlled study	Remission/ LDA	43	NR	7 years	Weekly mean dose per person by drug (1st period vs. 2nd period)*	IFX (mg/kg/week): 0.51 (SD 0.14), 0.42 (SD 0.12) (p<0.001) ADL (mg/week): 19.19 (SD 3.72), 15.52 (SD 4.81) (p<0.001) ETN (mg/week): 42.09 (SD 13.25), 35.04 (SD 13.37) (p=0.009)
						Mean interval of administration by drug (weeks) (1st period vs. 2nd period)*	IFX: 8.52 (SD 1.43), 9.7 (SD 1.44) (p<0.001) ADL: 2.19 (SD 0.58), 2.95 (SD 1.58) (p=0.007) ETN: 1.09 (SD 0.27), 1.61 (SD 0.91) (p=0.004)

Key: ADL: adalimumab, ETN: etanercept, IFX: infliximab; LDA: low disease activity; NR: not reported; SD: standard deviation; vs.: versus

Notes:

*These results were from a mixed population of 43 people with RA and 45 people with PsA

2.3.4 Discussion

This systematic review has identified two studies (reported in four publications)[8,12,55,56] that evaluated the effect of TDM on clinical outcomes in people with RA who had achieved remission or LDA. Three articles[11,15,64] reported the same non-randomised controlled trial (the INGEBIO study). The remaining study was an observational study evaluating the impact of TDM.

Both studies recruited people with RA who had achieved remission or LDA. The non-randomised controlled trial used Promonitor ELISA kits for monitoring drug levels and/or anti-drug antibody levels. The observational study[59] used Sanquin ELISA kits to measure drug levels and/or anti-drug antibody levels. It was unclear whether these tests were performed at the centralised testing service. The included study measured drug levels and/or anti-drug antibody levels in participants who were being treated with ADL, ETN and/or IFX. No studies were identified in participants treated with certolizumab pegol or golimumab. No studies were identified evaluating eligible ELISA kits including IDKmonitor ELISA kits, LISA-TRACKER ELISA kits, RIDASCREEN ELISA kits and MabTrack ELISA kits.

Comparative controlled evidence

Three abstracts[11,15,64] were identified reporting the same non-randomised controlled trial (the INGEBIO study), which focused on the population who had achieved treatment target (remission or LDA). In this trial, ADL and anti-ADAb levels were measured using Promonitor-ADL and Promonitor-ANTI-ADL (Grifols-Progenika). This trial recruited a mixed population of 169 participants including a cohort of 63 people with RA. The results of the total mixed population were reported in the review as the authors were not able to provide the results for the subgroup of people with RA.

The findings from this non-randomised controlled trial (the INGEBIO study) showed that there was a non-significant reduction in risk of flare in the intervention group compared with the control group. In particular, participants' HRQoL outcomes were higher in the intervention group at all visits compared with the control group, with statistically significant results being observed at two visits. However, as the quality of this trial was judged to be at serious risk of bias, the results should be interpreted with caution. Ideally, randomising participants is required to minimise the risk of bias for the study findings.

Evidence from observational studies

One observational study was identified evaluating the effect of TDM on clinical outcomes in people with RA who had achieved remission or LDA. The study[59] had a historical control.

Change in disease activity

The observational study (Pascual-Salcedo and colleagues 2013),[59] evaluated the effect of TDM on change in disease activities at duration of follow-up of two to seven years, with a sample size of 43. The study focused on people who had achieved remission or LDA. Overall, the finding from the historically controlled study (Pascual-Salcedo and colleagues 2013) showed that TDM was associated with a non-significant reduction in mean DAS28 scores at seven-year follow-up compared with the historical control (where TDM was not introduced). It should be noted that the quality of data was judged to be at moderate risk of bias, which has compromised the reliability of the findings.

Change in direction and magnitude of therapeutic dose

The observational study[59] evaluated the outcome of changes in direction and magnitude of therapeutic dose in people with RA who had achieved remission or LDA. The sample size of the study was 43.

Overall, the limited data from the observational study showed that TDM for optimising anti-TNF therapies was associated with reductions in therapeutic dose of anti-TNFs in people with RA who had achieved remission or LDA. This would be expected to lead to cost saving associated with TDM. Where statistically significantly results were observed, these results may be clinically significant. However, the reliability of the findings may be compromised by the poor quality of data being identified.

2.3.4.1 Reliability of the findings

The non-randomised controlled study[11,15,64] was judged to be at serious risk of bias (see Section 2.3.2.4). In this trial, there was an issue of baseline imbalance in disease severity between the intervention and control groups. Furthermore, there was a lack of adjusting for this variable in the analysis of clinical outcomes. There were higher attrition rates for some outcomes, which can lead to attrition bias. These deficiencies resulted in serious risk of bias associated with the findings. Therefore, the results should be interpreted with caution.

The historically controlled observational study was judged to be at moderate risk of bias (see Section 2.3.2.4), because there was non-contemporaneous control bias due to the use of a historical control. It should be noted that the same group of participants were assessed during the first period (the historical control where TDM was not introduced) and the second period (where TDM was implemented). However, the observational study had a small sample size. Therefore, the overall poor quality of included studies compromises the reliability of the findings.

2.3.4.2 Generalisability of the findings

Given that both studies were conducted in Spain, the findings from these studies may have limited generalisability to the UK setting due to variations in clinical practice and health policies between different countries. Furthermore, the findings from the non-randomised controlled trial (the INGEBIO study) and the results of changes in therapeutic dose from the study by Pascual-Salcedo and colleagues (2013)[59] were presented for a mixed population. Therefore, there was limited generalisability of findings from the mixed population (including RA, PsA and/or ankylosing spondylitis) to the target RA population.

2.3.4.3 Implications for future research

One ongoing Norwegian multicentre RCT was identified (the NOR-DRUM Study)[66] that evaluates the effect of TDM in people with RA in remission compared with standard care (see Section 2.3.2.3). This ongoing trial will provide further useful data on the impact of TDM in the target population.

Further controlled trials with a large sample size (especially RCTs) are required to assess the impact of using Promonitor ELISA tests for monitoring anti-TNF therapies in people with RA who had achieved remission or LDA.

No studies were identified that assessed other eligible ELISA kits including IDKmonitor ELISA kits, LISA-TRACKER ELISA kits, RIDASCREEN ELISA kits and MabTrack ELISA kits.

Therefore, future large RCTs are required to assess the impact of using those ELISA tests for monitoring anti-TNF therapies in people with RA who had achieved remission or LDA. More robust evidence is also needed to evaluate the impact of using Sanquin tests for monitoring anti-TNF therapies in this population.

Future RCTs are warranted to evaluate the clinical effectiveness of using ELISA tests for monitoring anti-TNF therapies in people with RA who had experienced a primary non-response or a secondary non-response.

There were no studies identified for patients who were being treated with certolizumab pegol and golimumab. Future RCTs are required to assess the clinical effectiveness of using ELISA tests for monitoring such anti-TNF therapies in the target populations.

2.3.4.4 Conclusions

In relation to clinical effectiveness, limited data were identified evaluating TDM in the target populations. One non-randomised trial compared TDM with standard care (the INGEBIO study) had serious limitations in relation to the NICE scope: only one-third of the participants

had RA, many of the analyses were not by intention-to-treat, follow-up was for only 18 months, there was no explicit algorithm for guiding clinicians in how the results of testing should change treatment (e.g. tapering), and the study was only reported in three abstracts. In addition, one observational study was also identified but was of minimal value in informing whether ELISA test-based monitoring is clinically effective or not.

Table 31: **Observational cost-effectiveness studies of therapeutic drug monitoring tests in people with rheumatoid arthritis**

Study	Population	Setting	Test	TNF-α inhibitor	Study design	N	Time-frame	Outcome	Cost measures	Results	Comments
Arango 2017 (INGEBIO)[15]	People with RA,PsA and AS, treated with ADL who remained clinically stable for at least 6 months	Clinic, Spain	Trough ADL and ADAb measured by Promonitor-ADA and Promonitor anti-adalimumab (Progenika) at 8 time points	ADL	Non-randomised controlled trial	109 participants in IG and 60 in CG, of which 30 and 33 people with RA, respectively	18 months	DAS28, BASDAI, BASFI and HAQ-DI, days with active disease	Average cost of ADL per patient-year	Mean QALY were 1.145 and 1.076 during follow-up period per person in IG and CG, respectively; the average cost of Humira® (ADL) per patient-year was 10,664.54€ vs 9,856.45€ (-808.08€, 8% savings) in the CG and IG, respectively (the results reported for the mixed population)	Data is reported for all participants and is not reported by disease subgroup. People with rheumatic diseaase have better quality of life, lower risk of flares and incur lower treatment costs if management is complemented with ELISA testing.

SECTION 3 PAGE 129

Table 34: Cost-effectiveness results reported in Krickaert and colleagues (2015)

Perspective	Costs		QALYs		ICER
	intervention	control	intervention	control	
Societal	€15,466,869	€18,028,517	591.65	587.81	−€ 646,266
Healthcare provider	€13,607,067	€16,153,357	591.65	587.81	−€ 666,541

Key: ICER, incremental cost effectiveness ratio; QALYs, quality-adjusted life years.
Note: Cost year was not reported.

Probabilistic sensitivity analysis around the base case scenario predicted that ELISA testing would dominate usual care in 72% of scenarios. Scenario sensitivity analyses around; e.g. the drug level cut-offs used, or the definitions of a good EULAR response, showed that ELISA testing is generally cost-saving, although some scenarios reported loss of QALYs.

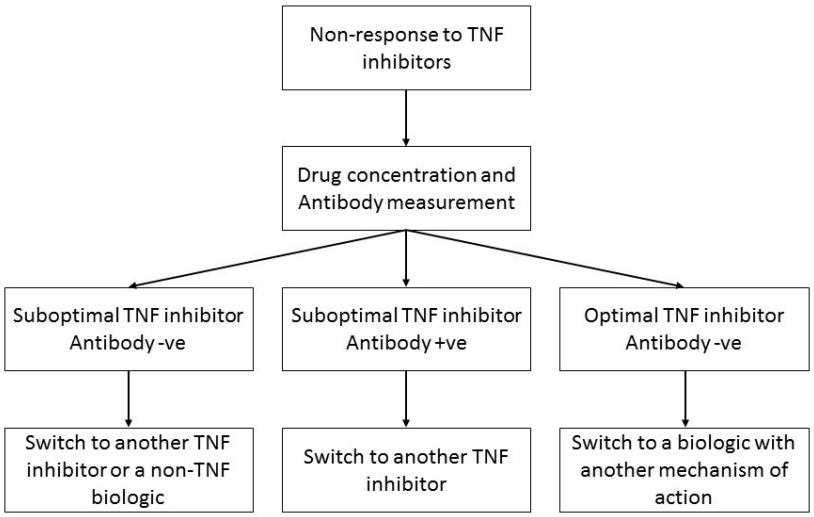
3.3.2.2 Laine and colleagues 2016

Laine and colleagues (2016) conducted a cost-effectiveness study in Finland. The intervention involved assessment of drug and anti-drug antibody levels in people with RA treated with ADL or IFX. The data on drug and anti-drug antibody levels were taken from the clinical sample registry of United Medix Laboratories Ltd in Helsinki, Finland, which included 486 and 1,137 samples from participants on ADL and IFX, respectively. The drug levels were measured using ELISA, while antibody level was assessed using radioimmunoassay. All measurements of antibody and ADL drug levels were outsourced to Sanquin Diagnostic Services (Amsterdam, Netherlands). Approximately half of IFX drug level was measured by the United Medix Laboratories using Promonitor test kit (Progenica, Derio, Spain).

Clinical management decisions based on the test results followed the algorithm proposed by Vincent et al (2013)[75] (Figure 5). Possible treatment decisions included switching to another TNF-α inhibitor or switching to a bDMARD with a different mechanism of action.

SECTION 3: PAGE 130

Figure 5: Algorithm for interpretation of test results



Source: Vincent 2013

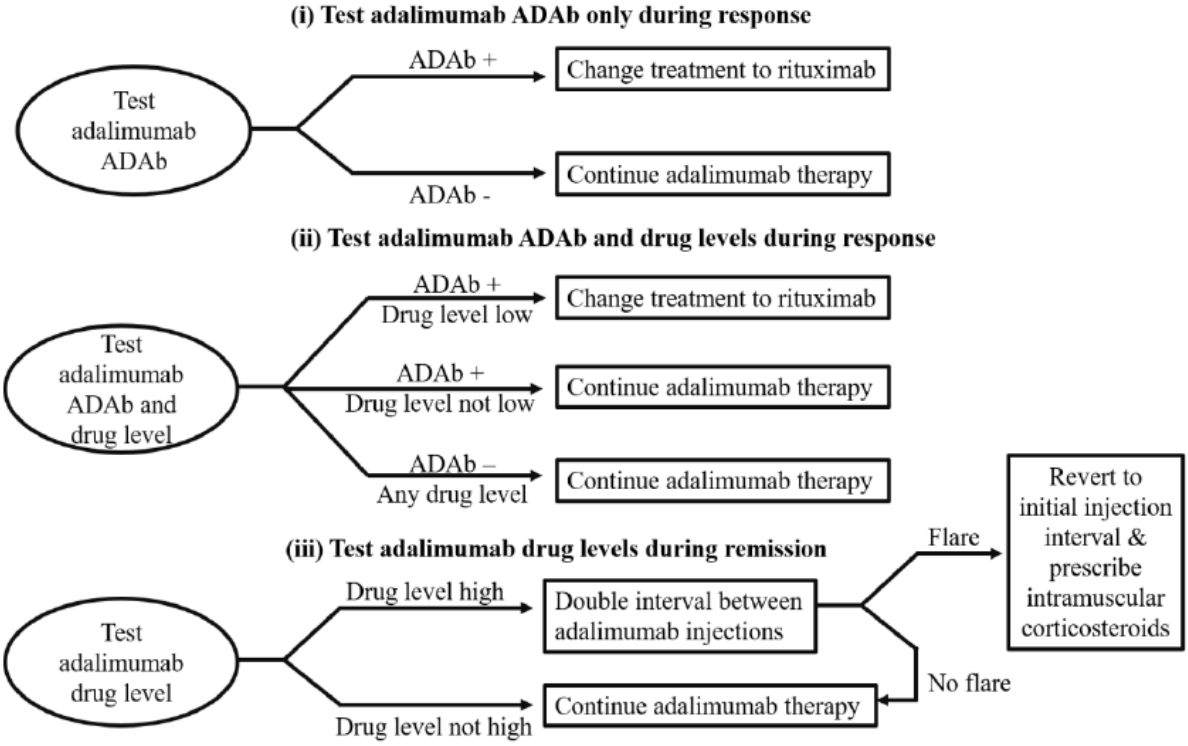
The economic impact of clinical decision-making was modelled in a short-term (three to six months) scenario with 100 hypothetical non-responders. They were compared with a non-testing scenario in which the same participants were managed only by clinical judgment in routine practice. The outcome measures were the changes in the probability of undergoing periods of sub-optimal treatment, and the cost-effectiveness of routine monitoring compared to clinical judgement only. An inappropriate clinical decision was defined to lead to ineffective treatment for at least three to six months. The authors justified this time period based on the typical follow-up visit frequencies of people with RA treated with biologics in Finland (no data sources were provided). This meant that all participants in the control arm experienced a three-month delay in receiving appropriate treatment. This delay was estimated to cost €1,471 for every month, which included the drug cost estimate per month of subcutaneous TNF-α inhibitor (€1,140), travel and lost working and leisure time costs for a laboratory visit (€17.4), costs of the possible standard safety-related laboratory tests (€6.8), travel and lost working and leisure time cost for a follow-up visit to an outpatient specialist clinic (€66.6), and specialist visit (€240.6). Long-term efficacy-related costs were not modelled. The cost of resource use was valued according to the national unit costs inflation adjusted to the year 2013.

SECTION 3 PAGE 133 to 134

population with RA in England using summary attributes of patients from the British Society for Rheumatology Biologics Register – Rheumatoid Arthritis (BSRBR-RA).

One of the test strategies considered in Gavan (2017)[35] was monitoring of drug and antibody levels in participants responding to treatment in order to avoid the harm associated with secondary non-response. Another possible test strategy was dose adjustment in patients in remission, informed by the results of TNF testing. Figure 7 shows the algorithm used in the Gavan study for management decisions in participant in whom TDM was performed.

Figure 7: Algorithm for test interpretation used in Gavan (2017)



Source: Gavan (2017)[32]

Utilities were calculated based on mapping the HAQ score from the BSRBR-RA using a quadratic mapping algorithm estimated previously for the NICE TA195 by Malottki and colleagues (2011).[14] Costs included the costs of treatment, hospitalisations and testing. Quantities of resource utilisation were derived from published sources (Stevenson and colleagues, 2016[50] Jani and colleagues, 2016[53]), unit costs were taken from the NHS reference costs 2015-2016 and BNF (accessed 8 April 2016).

Based on the 12 strategies modelled (Table 35), Gavan (2017)[35] concluded that routine use of ADL testing was cost-effective compared to current practice, but was unlikely to be cost-

effective relative to dose reduction (without testing) for people in remission (Strategy 11). Compared to current practice, Strategies 1-6 and Strategy 8 were estimated to be cost-effective. Strategies 9 and 10 were estimated to be less costly, but produced fewer QALYs compared to current practice. Strategy 7 was dominated by current practice, i.e. current practice was associated with fewer costs and greater QALYs than Strategy 7. In the incremental analysis, all strategies except three (Strategies 1, 3 and 11) were shown to be dominated or extendedly dominated by another strategy, i.e. another strategy or combination of strategies was cheaper and produced more QALYs. Of the three remaining strategies, Strategies 1 (adalimumab antibody [ADAb] and drug level testing every three months) and 3 (ADAb and drug level testing every three months, drug level test in remission after two years) were not cost-effective compared to Strategy 11 (no testing, just half dose in remission after two years) at a willingness to pay of £20,000 - £30,000 per QALY gained: ICER for Strategy 1 vs 11 was £38,575, ICER for Strategy 3 vs 11 was £37,043. Since strategy 11 consists of dose reduction after two years for people in remission, the analysis of the chosen strategies therefore suggests that ADL testing may not be cost-effective compared to dose reduction alone.[35]

3.4 Quality of identified cost-utility studies

Table 36 shows the results of assessing the included studies against the Consensus Health Economic Criteria (CHEC).[68] Methodological quality of included modelling studies assessed using the Philips checklist[69] is addressed in Table 37.

.

SECTION 4: PAGES 140 to 142

4. Independent economic assessment

4.1 Methods

4.1.1 Summary of available evidence

A systematic review of economic evaluations was completed (refer to Section 3). Table 38 provides an overview of those tumour necrosis factor-alpha (TNF-α) treatments and enzyme linked immunosorbent assay (ELISA) kits from the NICE scope which were considered in the studies identified in the cost-effectiveness systematic review.

A systematic review of clinical effectiveness evidence was conducted (refer to Section 2). Table 39 summarises which combinations of treatments and ELISA kits were used in the included studies.

Table 39. Clinical effectiveness evidence relevant to specific combinations of TNF-α inhibitors and test kits from the NICE scope

		Promonitor	IDKmonitor	LISA-TRACKER	RIDASCREEN	Sanquin*
ADL	drug	✓[1]	X	X	X	✓[2]
	antibody	✓[1]	X[3]	X	X	✓[2]
ETN	drug	X	X	X		✓[2]
	antibody	X	X	X		✓[2]
IFX	drug	X	X[3]	X	X	✓[2]
	antibody	X	X	X	X	✓[2]
GLM	drug	X	X	X		X
	antibody	X	X	X		
CTZ	drug			X		X
	antibody			X		

Key:

X Indicates availability of a test to measure drug or antibody level in people treated with the specified TNFi and that no studies have been identified in the clinical-effectiveness systematic review, reporting on using therapeutic drug monitoring for the specified test kit and TNFi.

✓ Indicates availability of a test to measure drug or antibody level in people treated with the specified TNFi and that at least one source for the specified combination of the test kit and TNFi has been identified in the clinical-effectiveness systematic review.

ADL: adalimumab; CTZ: certolizumab pegol; ETN: etanercept; GLM: golimumab; IFX: infliximab; TNFi: tumour necrosis factor inhibitor

Notes:

* The type of Sanquin test kits used in these studies (MabTrack or those used by Sanquin Diagnostic Services) was not reported.

[1] Arango and colleagues 2017,[15] Ucar and colleagues 2017[11] and Gorostiza and colleagues 2016[64] (INGEBIO)

[2] Pascual-Salcedo and colleagues 2013[59]

[3] Indicates that a test for total anti-drug antibodies is also available (total anti-drug antibodies include both unbound, i.e. free, antibodies and those bound to TNF-α inhibitor)

As shown in Table 39 no clinical-effectiveness evidence related to IDKmonitor, LISA-TRACKER, RIDASCREEN and MabTrack tests has been identified. In studies which used Sanquin test kits, the type of kits was not reported. For two drugs from the NICE scope, golimumab (GLM) and certolizumab pegol (CTZ), no studies were found that investigated the use of therapeutic drug monitoring (TDM) in people with RA treated with TNF-α inhibitors. In those studies where antibody testing was conducted, the type of testing (reflex or concurrent testing) was not reported. In the clinical-effectiveness systematic review, no studies reporting on the use of ELISA testing in people with RA receiving biosimilar products were identified.

Both the included studies included people in remission or with low disease activity (LDA) (refer to Section 2.3.3 for further details).

In both studies the study populations were mixed, with 37% of people with RA in the INGEBIO study and 49% in Pascual-Salcedo and colleagues (2013).[59] Moreover, populations considered in the selected studies were relatively small, with the only exception being the INGEBIO study which had a (mixed disease) population including 169 participants (Section 2.3.2.1).

In the INGEBIO STUDY which compared test versus no-test treatment strategies (Section 2.3.3.1), physicians were not obliged to follow any therapeutic algorithm based on TDM results but could use testing to alter doses based on their clinical judgement in participants from the intervention arm. The study was conducted in Spain. The longest follow-up of 18 months was reported by Arango and colleagues (2017).[15] Some of the clinical outcomes are shown in Table 40.

Table 40: Clinical outcomes and follow-up period from Ucar and colleagues (2017) and Arango and colleagues (2017)

Outcome	Ucar and colleagues 2017		Arango and colleagues 2017	
	IG	CG	IG	CG
Proportion of patients with tapered dose, %	35.8%	36.7%	35.7%	34.6%
Rate of flares per patient-year	0.463[1]	0.639[1]	0.463[1]	0.639[1]
Mean duration of remission (Ucar 2017) or remission/LDA (Arango 2017), days	344	329	460.2	475.2
Mean follow-up, days	499	505	530.8	544.6

Key: CG: control group; IG: intervention group; LDA: low disease activity

Note:

[1]The rate of flares per patient-year reported in Ucar and colleagues 2017 is the same as in Arango and colleagues 2017 (even though these sources reported outcomes for different follow-up periods)

Source: Ucar and colleagues (2017)[11] and Arango and colleagues (2017)[15]

The authors (Arango and colleagues [2017] and Ucar and colleagues [2017]) reported the mean cost of adalimumab (ADL) treatment per patient-year, and mean quality-adjusted life years (QALYs) (based on EQ-5D-5L) accrued over the observed period in the intervention and control arms. Since the study was reported in the abstract form only, it was not clear how the mean QALYs were estimated.

4.1.1.1 Search for additional effectiveness evidence

Studies identified by the searches conducted for the clinical effectiveness review but not considered eligible for inclusion in the review (e.g. studies reporting correlations between drug/antibody levels and therapeutic outcomes, and/or studies reporting only drug/antibody levels before and after dose reductions) were used to inform the model where appropriate.

Due to the lack of randomised controlled trial (RCT) evidence on the effectiveness of the tests that are defined within the NICE scope, a systematic review of the literature was conducted to identify RCTs evaluating any tests used to monitor anti-TNF-α treatment in people with RA. The aim of this search was to identify any evidence on the effectiveness of any strategies of treatment monitoring that could be used to inform scenario analyses for the modelling.

Searches were carried out in MEDLINE, MEDLINE In-Process, Embase, The Cochrane Library and Web of Science. Searches were limited to RCTs and carried out in October 2018. The search strategy is provided in Appendix 4.

A total of 1,418 hits were identified and independently screened by two reviewers versus the inclusion criteria shown in Table 41. No relevant papers were identified.

Table 41: Inclusion criteria

Criteria	Specification
Population	As for the clinical-effectiveness systematic review (see Section 2.2.1.1)
Interventions	Any test outside of the scope for monitoring patients receiving TNF-α inhibitors (ADL, ETN, IFX, CTZ, GLM).
Comparator	Current practice (i.e. no testing)
Outcomes	As for the clinical-effectiveness systematic review (see Section 2.2.1.4)

analysis as quantitative synthesis of evidence related to people in remission/LDA identified in the clinical-effectiveness systematic review, was not possible (Section 2.2.5).

Since the patent for the ADL originator product, Humira®, expired in October 2018, and the acquisition costs for the ADL biosimilars were not known to the AG at the time of writing (Table 50), the annual acquisition cost was varied from £1,000 to £9,180 per patient-year in the threshold analysis. The latter represents the annual cost of Humira® assuming the dose of 40 mg every two weeks delivered by subcutaneous injection using a pre-filled pen and the NHS indicative price from the British National Formulary (BNF) (Table 50).

The other major assumptions were as follows (see also Table 42):

- ADL dose tapering is implemented by increasing the interval between doses from two to three weeks (i.e. by spacing doses).
- Dose is tapered in a proportion of people in each arm at the start of simulation.
- Some people may flare after reducing the dose of their TNF inhibitors (Bykerk and colleagues, 2016).
- The full dose of ADL is restored *in all people* on tapered doses when they flare.
- Treatment wastage is £370 per patient-year in people on full dose; it is reduced proportionally to the reduction in treatment dose.
- ADL is self-administered (usually at home), and, therefore, the administration cost is zero.
- The costs associated with flare management are:
    - £423 per flare for diagnostic investigations
    - £68 per month for treatment (excluding the cost of DMARDs)
- The annual per-patient costs of managing remission and LDA/active disease health states are £11,409 and £18,889 respectively.
- The cost of managing an AE is £1,622.
- The utilities for remission and LDA/active disease health states are 0.71***8*** and 0.5***6***8, respectively.
- The disutility of flare is 0.140.
- The duration of flare is seven days.
- The rates of AEs in people on full and tapered doses are 3/100 and 2/100 patient-years, respectively.
- The duration of AE is 28 days.
- The time horizon is defined by the follow-up in Ucar and colleagues (2017).

Table 42: Model assumptions in the analyses with people in remission/low disease activity

Assumption	Estimate	Source	Relevant table/ sections in the report
Dose tapering strategy	Spacing: from 40 mg of ADL every 2 weeks to 40 mg every 3 weeks	1st dose reduction in the Exeter biologic clinic recommendations (Appendix 5)	Section 4.1.10.1.5
Proportion of patients on tapered dose:			
Intervention	35.8%		Table 37
Control	36.7%		Table 37
Proportion of flared patients in whom the full dose is restored	100%	Exeter biologic clinic recommendations	Appendix 5
Mean duration of remission (days)			
Intervention	344		Table 37
Control	329		Table 37
Mean follow-up (days)	505	As in the control arm (Ucar and colleagues, 2017)[1]	Table 37
Acquisition costs (per patient-year): Humira®			
Full dose[2]	£9,187	BNF	Section 4.1.10.1.3
Tapered dose	£6,125	BNF, Exeter biologic clinic recommendations	Appendix 5
Flared patients[3]	£9,187	BNF, Exeter biologic clinic recommendations	Appendix 5
Treatment wastage on full dose (per patient-year)	£370	Clinical advice	Section 4.1.10.1.6
Administration cost for Humira® (ADL) (per patient-year)[3]	£0	Clinical advice	Section 4.1.10.1.7
Cost of flare management[4,5]	£423/per flare	Cost of diagnostic investigations (Maravic and colleagues, 2005[9])	Section 4.1.10.1.19
	£68/month	Monthly cost of treatment (excluding DMARDs) (Maravic and colleagues, 2005[9])	Section 4.1.10.1.19

Assumption	Estimate	Source	Relevant table/ sections in the report
Cost of managing health states (per patient-year)[6]			
Remission	£11,409	Barbieri and colleagues (2005),[67] Radner and colleagues (2014),[10] National Schedule of Reference Costs 2017-18[68]	Section 4.1.10.1.16
LDA/active disease	£18,889		Section 4.1.10.1.16
Cost of managing AEs (per infection)	£1,622[7]	TA375[4]	Section 4.1.10.1.20
Utilities			
Remission	0.718	Estimated from HAQ scores for different HAQ bands reported by Radner and colleagues (2014)[10]	Section 4.1.10.2.1
LDA/active disease	0.568[8]		Section 4.1.10.2.1
Disutility of flare	0.140	Markusse and colleagues, 2015[69]	Section 4.1.10.2.2
Disutility of AEs	0.156	TA375,[4] Oppong and colleagues (2013)[7]	Section 4.1.10.2.3
Flare rate			
Intervention	0.463	Ucar and colleagues 2017[8]	Section 4.1.9.1.1
Control	0.639	Ucar and colleagues 2017[8]	Section 4.1.9.1.1
Mean time to first flare (days)			
Intervention	208.07	Derived from Kaplan-Meier estimates (from the INGEBIO study) of time to first flare, provided by Ucar and colleagues (personal communication, 9 September, 2018)	Section 4.1.9.1.3
Control	189.32		Section 4.1.9.1.3
Flare duration (days)[9]	7	TA375[4]	Section 4.1.9.1.2
Rate of AEs			
Patients on full ADL dose	3/100 patient-years	Senabre Gallego and colleagues (2017)[5]	Section 4.1.9.2.1
Patients on reduced ADL dose	2/100 patient-years[10]	Singh and colleagues (2015)[6]	Section 4.1.9.2.1

Assumption	Estimate	Source	Relevant table/ sections in the report
Duration of AE (days)	28	TA375,[4] Oppong and colleagues (2013)[7]	Section 4.1.9.2.2

Key: ADL: adalimumab; AE: adverse event; BNF: British national Formulary; HAQ: health assessment questionnaire; HAD: high disease activity; MDA: moderate disease activity; OR: odds ratio; PPP: purchasing power parities; RA: rheumatoid arthritis; RCTs: randomized controlled trials; TA: technology appraisal

Notes:

[1] The length of follow-up in the control arm (505 days) was used as the time horizon in the economic analyses, which was slightly longer than follow-up in the intervention group (499 days) (refer to Section 4.1.6).

[2] Assuming 40 mg every two weeks by subcutaneous injection using pre-filled pen, and NHS indicative price from the BNF.

[3] The mean time to first flare was estimated from additional evidence (Kaplan-Meier curves for time to first flare) from the INGEBIO study provided by Ucar and colleagues (2007)[11] (poster, personal communication).

[4] The estimates were derived from the costs of managing a flare in a hypothetical person with a 10-year history of RA in the French setting. The costs were converted to pound sterling based on PPP and inflated to 2017-18 prices using the healthcare price index (Section 4.1.10.1.1).

[5] The estimates from Maravic and colleagues (2005)[4] do not include the cost of rheumatology appointments.

[6] The costs of managing health states were included by HAQ-dependency, i.e. by assigning an annual cost to mutually exclusive HAQ intervals. The proportion of patients in different health states were derived from Radner and colleagues (2014)[6]

[7] The estimate of £1,479 per patient-year from the source was inflated to 2017-18 prices using the healthcare price index (Section 4.1.10.1.11).

[8] The estimate was computed from a weighted average HAQ score for the LDA, MDA, HDA health states reported by Radner and colleagues (2014)[6] and mapped to EQ-5D values following Malottki and colleagues (2011)[14] (Section 4.1.10.2.1).

[9] This estimate was used for calculation of QALYs only since it was assumed that the ADL dose in people with flares is switched back to the full dose indefinitely.

[10] Based on OR=1.31 for standard-dose biologics in people with RA reported by Singh and colleagues (2015).[13] The OR estimate was obtained in a Bayesian network meta-analysis (using a binomial likelihood model) of 11 published RCTs (n=4,788) to assess the risk of serious infections in anti-TNF-biologic-experienced people with RA.

Of note, in the primary analysis, QALYs were estimated based on heath-state utilities as well as disutilities of flares and AEs. It was assumed (based on Smolen and colleagues, 2017,[76]) that people in any health state (i.e. in remission, LDA and active disease) can experience flares (Section 4.1.9.1).). In the INGEBIO study, flare rates in the intervention and control arms were not stratified further according to dose (full or tapered). Therefore, the same rate of flares was applied to all patients within each arm (see Section 4.1.4 for further details on how flares were modelled). Utilities for the mixed disease population in the INGEBIO study were assumed to be the same as those for the population of people with RA only since no evidence on HRQoL directly relevant to the population considered in INGEBIO has been identified.

When modelling the effect of AEs on HRQoL and costs, the Assessment Group (AG) adopted the approach used in TA375 - it was assumed that only serious adverse events (serious infections in particular) would carry a significant cost and disutility burden (p. 381, TA375[8]). This was supported by the opinion of our clinical experts.

Mortality associated with RA was not modelled and no discounting was applied to the costs and outcomes due to the short-term time horizon of about 18 months.

4.1.3.2 Primary and secondary non-responders

Due to the lack of relevant data identified in the systematic review of clinical effectiveness, the cost-effectiveness of TNF-testing in non-responders could not be evaluated.

4.1.4 Model structure

The cost of TNF testing, under which the treatment strategy based on test results and clinical judgement has zero NMB, was estimated in the following way:

$$Total\ cost\ of\ testing = ICER\ threshold * \Delta QALYs - \Delta Costs,$$

where $Total\ cost\ of\ testing$ was comprised of the cost of the expected resource use and costs associated with testing patient samples to monitor drug trough and anti-drug antibody levels (refer to Section 4.1.10.1.8); and, $ICER\ threshold$ represents the *NICE* cost-effectiveness threshold of either £20,000 per QALY or £30,000 per QALY gained. Threshold analyses were conducted for both thresholds.

The costs incurred in each arm were estimated as follows:

$$Costs = acquisition\ cost + administration\ costs + cost\ of\ managing\ health\ states + cost\ of\ managing\ flares + costs\ of\ managing\ adverse\ events$$

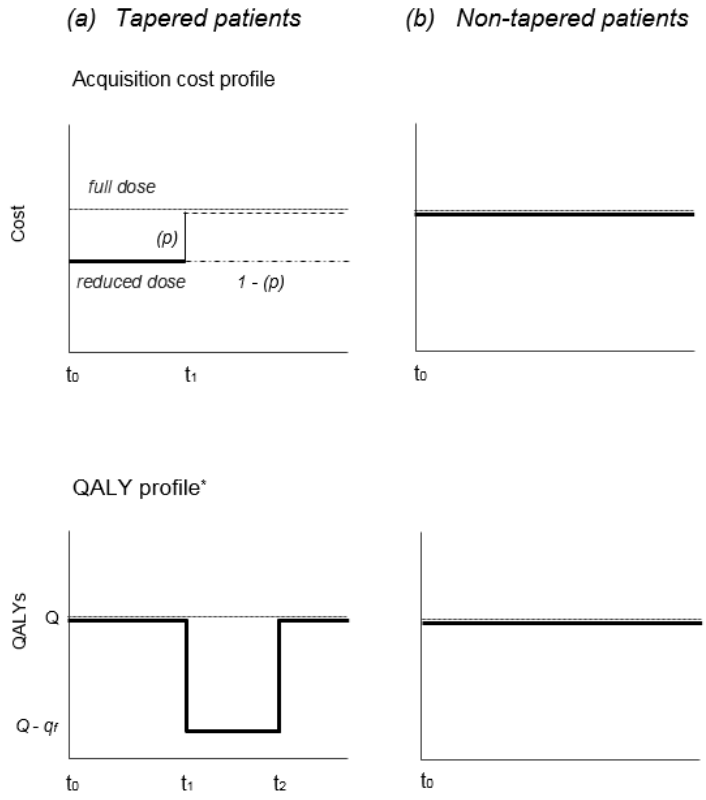
QALYs were derived as follows:

$$QALYs = duration\ of\ remission * utility\ score\ for\ remission + duration\ of\ active\ disease * utility\ score\ for\ active\ disease - average\ duration\ of\ flare * rate\ of\ flare * disutility\ of\ flare - average\ duration\ of\ adverse\ event * rate\ of\ adverse\ event * disutility\ of\ adverse\ event$$

SECTION 4 PAGE 151-152

Based on Gavan (2017),[35] Figure 8 shows a graphical illustration of the cost and QALY profile depending on whether the dose is tapered or not. The figure shows changes in the acquisition cost and QALYs *due to flares* over time. Note that, for the sake of clarity, the other components of the total costs and QALYs, considered in the analysis, are not depicted here.

Figure 8: Acquisition costs and QALY change in tapered and non-tapered patients due to flare (Gavan 2017)



Notes:
$(t_1 - t_0)$ is the time on tapered dose; (approximated by the mean time to first flare estimated from Kaplan-Meier curves for time to first flare, reported in an additional source provided by Ucar and colleagues on request from the AG): $(t_2 - t_1)$ is the duration of flare; q_f is the disutility of flare.
* Change in QALYs due to flare
Source: Based on Gavan (2017)[35]

In this scenario, all patients had their drug levels tested at (t_0), and in some patients the dose was tapered (Figure 8 (a)). A proportion of patients on tapered doses were assumed to flare at t_1, which prompted treatment to revert to its original dose in a proportion of patients (p) while in other patients (1-p) the dose remained the same (i.e. tapered). In those patients who flared the disutility of flare, q_f, was applied for the duration of flare, $(t_2 - t_1)$. In non-tapered patients (Figure 8 (b)), the acquisition cost was based on the cost of the full dose of adalimumab. It was assumed that these people do not experience flares.

The impact of flares on costs and QALYs was modelled based on the approach used by Gavan (as shown on Figure 8 (a)); i.e. flares were modelled in patients on full and tapered doses because flare rates reported in Ucar and colleagues (2017) and Arango and colleagues (2017), were not stratified by treatment dose. The estimates of the mean time to first flare were used to model the time when dose in flared patients was restored to full (which impacted on the drug acquisition costs and wastage), while the rates of flares were used to estimate the costs of flare management and reduction in QALYs due to flares in different arms. The time to first flare was arm-specific; these estimates were derived from Kaplan-Meier curves (KM) for time to first flare (reported in an additional source provided by Ucar and colleagues on request from the AG) using the area under the curve (AUC) method. It was also assumed (based on Smolen and colleagues, 2017[76]) that flares could occur *in any health state* (i.e. in remission, LDA and active disease) (Section 4.1.9.1).

1.1.1 Population

The population modelled were people in remission/LDA at baseline. Table 43 presents baseline characteristics of participants included in the INGEBIO study, used for model parameterisation, along with the characteristics of people with RA, *responding to biologics,* from the British Society for Rheumatology Biologics Register (BSRBR).

Table 43: Patient baseline characteristics

Study	# RA patients	Age	% females	Disease duration, years	Treatment history	Concomitant treatments	Disease state
INGEBIO	Mixed pop: 63 people with RA of total 169 participants	53.6[1]	42%[1]	Median=10	NR	MTX[2] – 76.7%	77% people in remission, 23% LDA (at baseline)
BSRBR data for responders[3]	10,186	56	76.3%	Mean=13 (years at the time of initiation of 1st biologic)	Mean=3.90 (previous DMARDs)	*NR*[4]	30.6% – good responders

Key: BSRBR: British Society for Rheumatology Biologics Register; DMARDs: disease modifying anti-rheumatic drugs; LDA: low disease activity; MTX: methotrexate; NR: not reported; RA: rheumatoid arthritis; TNF: tumour necrosis factor

Notes:

[1] Weighted average across treatment arms

[2] Patients concurrently receiving anti-inflammatories or MTX are more likely to respond to anti-TNF (Dennison and colleagues, 2016)[77]

[3] Table 189 (TA375,[8] p.367). Of note, as stated in the source, the BSRBR database contained a very small number of MTX-naïve patients at the time the analysis was performed.

[4] As stated in TA375,[8] (p. 354) the BSRBR database contains a very small number of MTX-naïve patients.

Patients in the BSRBR were slightly older, on average, compared to patients in the INGEBIO study, and were more likely to be female.

Subgroups

People with RA can be grouped according to three clinical scenarios: primary non-response, secondary non-response and remission. However, with regards to particular characteristics, there are no subgroups for which the tests are expected to significantly vary. Therefore, no subgroups were considered in this assessment.

Duration of remission

The mean time in remission in both arms was reported in Ucar and colleagues (2017) and Arango and colleagues (2017). Importantly, the comparator arm in these sources had longer follow-up periods compared to the control arm, and the length of follow-ups in the intervention arms were used as the time horizons in the economic analyses using data from Ucar and colleagues (2017) and Arango and colleagues (2017) data. However, the estimates of the mean duration in remission in the intervention arm were not adjusted since the Kaplan-Meier curves for time in remission were not available to the AG. The results reported in this assessment are therefore likely to overestimate the cost-effectiveness of the intervention under consideration. However, the small difference in the length of follow-up between the two groups (around 1%), would mean this difference is expected to be small.

SECTION 4 PAGE 155

Table 45: Flare duration in the BRASS study

	Estimates		
Duration, days	<7	7–13	≥14
Proportion of patients, %	57%	14%	30%

Key: BRASS: Brigham RA Sequential Study
Source: Bykerk and colleagues, 2014[16]

The estimate of seven days was adopted in the primary analysis. This was consistent with the estimate used in TA375.[50] In a scenario analysis, the effect on the results of a longer duration of flare, 19 days, was examined. This estimate was a weighted average of data reported in the BRASS study[16] (Table 45), and the estimates provided by Dr Meghna Jani.

4.1.8.1.3 Time to first flare

Arango and colleagues (2017)[15] and Ucar and colleagues (2017)[11] reported *median time* to first flare observed in the intervention and control arms of the INGEBIO trial. However, according to the NICE *Guide to the Methods of Technology Appraisal,*[78] *mean estimates* should be utilised in economic analyses of health interventions.

The mean estimates for the intervention and control arms were derived from Kaplan-Meier (KM) curves for the time to first flare estimated in the INGEBIO study and reported in an additional source (a poster presentation*:* provided by Ucar and colleagues (2007)[11] to the AG on request in September, 2018) by using the area under the curve (AUC) approach. The KM estimates were available for 300 days (Figure 9), and were extrapolated for the duration of follow-up reported in Ucar and colleagues (2017)[11] (Table 40) and Arango and colleagues (2017)[15] (Table 40). Since the proportions of participants on tapered dose in the intervention and control groups levelled at around 240 days after dose tapering, it was assumed that these proportions remained the same until the end of the observational periods in Ucar and colleagues (2017)[11] and Arango and colleagues (2017)[15] (i.e. no parametric model fitting was performed). Estimates of the mean time to first flare are shown in Table 46.

In clinical practice flares have been observed in tapered and non-tapered patients, with an increased risk of flares in tapered patients. However, in the economic analysis flares were modelled in all patients as the reported flare rates in the intervention and control arms of the INGEBIO study were not stratified by dose. This is a limitation of the study.

SECTION 4 PAGE 160 to 161

Table 48: Comparison of effect estimates from consistency and inconsistency models from Singh and colleagues (2015)

Comparison	Consistency Model*	Inconsistency Model*
LD biologic +/- traditional DMARD vs. SD biologic +/- traditional DMARD, OR (95% CrI)	0.71 (0.50,1.01)	0.7 (0.27,1.68)

Key: CrI: credible interval; DMARDs: disease modifying anti-rheumatic drugs; LD, low dose; OR, odds ratio; SD, standard dose; vs.: versus

Note:

* Dias and colleagues (2011; updated April 2014)[85] NICE technical support document 4.

Source: Appendix 10b, Singh and colleagues (2015)[13]

4.1.8.2.2 Duration of serious adverse events

In TA375[50] serious infections were assumed to persist for 28 days. This estimate was adopted in the primary analysis.

4.1.9 Model parameters

Parameter values assumed in the primary analysis for people in remission/LDA are shown in Table 42. The derivation of the parameter values is detailed in the following sections.

4.1.9.1 Resources and costs

Costs considered in the economic evaluation included the costs of testing, the costs of treatments received by people with RA, and healthcare costs. The costs of testing comprised those of the test kits, staff time to perform test and staff training, the cost of testing service and sample transport. Costs were obtained from the British National Formulary (BNF),[86] NHS Reference Costs,[7] from documents provided by test manufacturers, and published and unpublished sources.

4..9.1.1 Conversion to GBP

Where conversion from other currencies to GBP was required, IMF purchasing power parity was used to convert within the year (e.g. from 2010 EUR to 2010 GBP), after which inflation was applied. The Campbell and Cochrane Economic Methods Group (CCEMG) – EPPI-Centre Cost Converter was used for the purchasing power parity (PPP) conversion.[87]

4.1.9.1.2 Inflation to 2017-18 prices

Unit costs were inflated to 2017–18 prices by inflating to 2015–16 prices using the Hospital and Community Health Services Pay and Prices Index (Table 49), and then to 2017–18 prices using the average increase in the index for the previous three years (from 2013–14 to 2015–16), with the average rate of 1.1% per annum.

Table 49: Hospital and Community Health Services combined pay and prices inflation indices

Year	Pay and prices (%)
2008-09	3.9
2009-10	0.6
2010-11	3.0
2011-12	2.1
2012-13	1.7
2013-14	1.1
2014-15	0.9
2015-16	1.3

Source: HCHS pay and price inflation, Pay and Price Series - Department of Health[88]

SECTION 4 PAGE 163 to 167

ADL; for people prescribed IFX biosimilars (Inflectra® and Remsima®) are given; and, a biosimilar (Benepali®) is used in some people prescribed ETN (Dr Meghna Jani, personal communication).

Although the NICE guidance recommends that people with RA patients receive the anti-TNF treatment with the lowest acquisition and administration costs, in practice other non-cost factors such as patient characteristics, hospital characteristics and changes in regional rheumatology clinical guidelines may influence treatment selection (Gavan, 2017).[35]

4.1.9.1.4 Drug acquisition costs

Annual acquisition costs of the TNF-inhibitors, assumed in the cost-utility analyses, were estimated using list prices (in accordance with NICE guidelines[78]) and assuming adherence to standard dosing regimen for each drug (Table 50).

Table 50: Acquisition costs of biologics

Biologic	Dosing regimen	Cost per dose	Cost per year	Additional cost in Year 1
ADL	40 mg, every 2 weeks. In non-responsive patient dose may be increased to 40 mg, once weekly.			
Humira®*		£352.14	£9,187.08	
Amgevita®		NR		
Cyltezo®		NR		
Imraldi®		NR		
Solymbic®[7]		NR		
Hyrimoz®		NR		
Halimatoz®		NR		
ETN	50 mg, once weekly			
Enbrel®*		£178.75 (25 mg/ 0.5 ml)	£9,326.92	
Benepali/ Brenzys		£164	£8,557.29	
Erelzi		£160.88	£8,394.23	
Lifmior®		NR		
CTZ	Loading dose: 400 mg, at Weeks 0, 2, and			
Cimzia®*		£357.50	£ 9,326.92	£1,072.50[2]

Biologic	Dosing regimen	Cost per dose	Cost per year	Additional cost in Year 1
	4. Maintenance dose: 200 mg every 2 weeks[1]			
GLM	50 mg once per month, on the same date each month.[3]			
Simponi®*		£762.97	£9,155.64[4]	
IFX	3 mg/kg at Week 0, 2 and 4. Then 3 mg/kg every 8 weeks.[5]			
Remicade®*		£419.62 per vial (100 mg powder for concentrate for solution for infusion vials), 2 or 3 vials per administration	£5,747.48 (assuming no vial wastage), £8,210.69 (assuming full vial wastage)	£1,982.70
Inflectra® & Remsima®[6]		£377.66 (100 mg powder for concentrate for solution for infusion vials)	£5,172.76 (assuming no vial wastage), £7,389.66 (assuming full vial wastage)	£1,784.44
Flixabi®/ Renflexis®		£377.00 (100 mg powder for concentrate for solution for infusion vials)	£5,163.72 (assuming no vial wastage), £7,376.75 (assuming full vial wastage)	£1,781.33
Zessly®		NR		
Ixifi®		NR		

Key: NR: not reported; PAS: patient access scheme

Notes:

* Originator/ reference products

[1] Once clinical response is confirmed, 400 mg every four weeks may be considered

[2] Assuming no PAS arrangement

[3] Body weight up to 100 kg, 50 mg once per month, on the same date each month. Body weight exceeding 100 kg, initially 50 mg once a month (one the same date) for three to four doses, if treatment response is inadequate dose may be increased to 100 mg once a month.

[4] Based on standard dosing regimen for patient weighing less than 100 kg

[5] If treatment response is inadequate after 12 weeks, dose may be increased in 1.5 mg/kg increments every eight weeks to a maximum dose of 7.5 mg/kg every eight weeks. Alternatively, intervals between doses may be reduced, to a minimum dosing interval of 3 mg/kg every four weeks

[6] Cost per year was calculated assuming patient weight of 70 kg (as in TA375[50])

[7] Not available in the EU

The annual costs of ADL, ETN, GLM and CTZ were estimated from the price of solution for injection *pre-filled pens* since these biologics are administered subcutaneously and can therefore be self-administered. Consultation with clinical experts confirmed that all the TNF

inhibitors considered in this study except IFX are usually self-administered by people with RA at home.

Consistent with acquisition cost calculations in TA375,[50] the cost per annum of IFX was estimated using average weight of 70 kg.[8] IFX is administered intravenously (Section 4.1.10.1.7).

TA375[50] reports that the manufacturers of GLM provide the 100 mg dose at the same price as the 50 mg dose under a patient access scheme (PAS) arrangement, this discount does not affect the annual costs presented in Table 50 as these are based on the assumption that a patient weighs less than 100 kg.

Of note, the acquisition costs of the cheapest available pens for each drug are equivalent to the cost of the cheapest available dose, therefore annual acquisition costs for the self-administration route are equivalent to acquisition costs for biologics administered during outpatient visits.

The estimates for the additional acquisition costs for the first year (the last column of Table 50) are presented for information only. They were not used in any analyses since the population in this assessment are people experienced in biologics.

4.1.9.1.5 Dose tapering

According to EULAR recommendations for the management of RA with synthetic and bDMARDs,[93] tapering of bDMARDs should be considered in people in persistent remission after having tapered GC, especially if this treatment is combined with a conventional synthetic DMARD. In this context, tapering means reduction of dose (e.g. reducing etanercept 50 mg to 25 mg/ week (Smolen and colleagues, 2013[94]) or extension of interval between applications, ‘spacing’ (e.g. increasing the interval between ADL injections to 10 days rather than one week as in the Exeter biologic clinic recommendations described in Appendix 5).

The AG is aware that there is no gold standard on how dose tapering should be performed. Studies evaluating dose tapering have used different approaches. In clinical practice, dose tapering varies extensively depending on clinical opinion. For example, according to Exeter biologic clinic recommendations, when tapering the ADL dose, the dose should be reduced by one-third to 40 mg every three weeks and further reduced at three months to 40 mg every four weeks in people with LDA or remission (Appendix 5).

In the primary analysis, the assumption of reducing the dose by one-third (the first dose reduction in the Exeter biologic clinic recommendations, Appendix 5) was implemented (Table 42), while the assumption of halving the dose (the second dose reduction) was explored in a sensitivity analysis (Table 73).

4.1.9.1.6 Wastage

The dose tapering strategy suggested in the Exeter biologic clinic recommendations (Appendix 5) is spacing. Therefore, when this tapering strategy is used, there is no wastage of the self-administered drugs due to partial use of the dose in pre-filled pens. Clinical advice (Dr Richard C Haigh, Royal Devon and Exeter NHS Foundation Trust) indicated that wastage of IFX due to partial use of vials is usually avoided.

In the primary analyses, however, wastage of £370 per patient-year was incorporated (Table 42). This estimate was based on a survey conducted at the Royal Devon and Exeter NHS Foundation Trust (Dr Richard C Haigh, personal communication). It was derived from data from 119 people with RA treated with biologics, and included missed doses and over-supply (defined as a delivery of treatment even if >4 weeks supply was available at home). It was assumed that £370 per year was wasted, on average, in people on a full dose of a TNF inhibitor, while in people on a tapered dose, wastage was reduced proportionally to the reduction in treatment dose. In scenario analyses for other biologics, the treatment wastage was assumed to be proportionate to the drug acquisition price. The effect on the outcome of no wastage was explored in a sensitivity analysis (Table 70).

4.1.9.1.7 Drug administration costs

ADL, ETN, GLM, and CTZ are usually self-administered via subcutaneous injection using a pre-filled pen. In this scenario, there is no administration cost for delivery. Alternatively, these drugs may be administered by a district nurse. The average administration cost assumed in TA375[8] (which was based on an estimate reported in TA247[95]) was £2.61 (cost

SECTION 4 PAGE 171 to 174

At the Exeter Clinical Laboratory (Royal Devon and Exeter NHS Foundation Trust) which conducts approximately 80% of testing for monitoring biologics in the UK, the fully recovered cost of TNF testing (staff, reagents, consumables, over-heads and depreciation of equipment) is [redacted] per test (Dr Timothy McDonald, personal communication); this includes testing of both drug and antibody levels and covers all the components in Phase 2 reported in Jani and colleagues (2015)[53] (Table 53). Of note, at the Exeter Clinical Laboratory only IDKmonitor test kits are currently used for clinical services.

Dr Timothy McDonald advised us that laboratories which conduct TNF testing have previously negotiated arrangements with the manufacturers of bDMARDS to cover the cost of biological monitoring, including assays and personnel costs. However, based on advice from Dr Meghna Jani, that might vary by geographical area and only for certain biologics (e.g. newer biosimilars).

4.1.9.1.9 Assay costs provided by the manufacturers

The cost of reflex and concurrent testing for each assay were derived from information request documents submitted by the manufacturers of the test kits (Table 54). Many of the manufacturers of the test kits also offer price discounts (which depend on the uptake of testing) for test kits used in singlet or duplicate, concurrent or reflex testing with different number of tests per year.

For the economic analyses, the cost used for Promonitor was provided by the manufacturer was used (Table 54). Additional cost-utility analyses were conducted for each level of discount and each type of testing. The resulting ICERs are not presented in the report due to the fact that they are very uncertain. However, the results are available in the model developed by the AG.

4.1.9.1.10 Processing costs

In addition to assay costs, the cost of testing also depends on processing costs, such as administration and laboratory personnel time. In the study conducted by Jani and colleagues (2016),[53] it was assumed that during the pre-testing phase (Phase 1, Table 53), one outpatient appointment with a consultant rheumatologist is required to discuss the need for testing, followed by an appointment with a phlebotomist or clinical support worker to obtain blood trough levels. Regarding the testing phase (Phase 2, Table 53), it was assumed that hospital laboratories would have the basic materials required to conduct ELISA, so several resource use items were excluded from the micro-costing exercise. This study reported that

additional costs associated with laboratory personnel time processing the samples would be incurred during the testing phase (Table 53).

Finally, Jani and colleagues (2016)[53] reported that the treatment decision stage requires interpretation of results by a consultant rheumatologist, discussion of the results with patients via a telephone call, and lastly a letter outlining results and treatment decision (Phase 3, Table 53).

Table 54: Analysis of samples: assay costs per sample

Test		Number of wells	Number of controls per assay	Singlet testing of patient samples			Duplicate testing of patient samples			Sources and comments
				Number of samples analysed per assay	Cost per assay	Cost per sample	Number of samples analysed per assay	Cost per assay	Cost per sample	
IDK Monitor[a]	Drug levels monitoring	96	2 controls, 6 standards (tested once/in duplicate)	88/80	£855.00	£9.72/ £10.69	40	£855.00	£21.38	Costs per assay: Biohit Costs per sample (duplicate/ singlet testing of controls and standards): Biohit Costs per sample (singlet testing of samples, duplicate testing of controls and standards): calculated IFU recommends all patient samples and controls should be run in duplicate.
	Anti-drug antibody monitoring	96	2 controls; 1 standards (tested once/in duplicate)	93/90	£775.00	£8.33/ £8.61	45	£775.00	£17.22	Costs per assay: Biohit; Costs per sample (duplicate/ singlet testing of controls and standards): Biohit Costs per sample (singlet testing of samples, duplicate testing of controls and standards): calculated IFU recommends all patient samples and controls should be run in duplicate.

Test		Number of wells	Number of controls per assay	Singlet testing of patient samples			Duplicate testing of patient samples			Sources and comments
				Number of samples analysed per assay	Cost per assay	Cost per sample	Number of samples analysed per assay	Cost per assay	Cost per sample	
Promonitor[b]	Drug level monitoring	96	2 controls; 6 standards (tested once/in duplicate)	88/80	£700.00	£7.95/ £8.75	40	£700.00	£17.50	Costs per assay: Grifols Costs per sample calculated IFU states all patient samples and controls could be run in singlicate.
	Anti-drug antibody monitoring	96	2 controls; 6 standards (tested once/in duplicate)	88/80	£700.00	£7.95/ £8.75[e]	40	£700.00	£17.50	Costs per assay: Grifols Costs per sample: calculated IFU states all patient samples and controls could be run in singlet.
RIDASCREEN[a]	Drug level monitoring	96	2 controls; 6 standards (tested once/in duplicate)	88/80	£565.00	£6.42/ £7.06	40	£565.00	£14.13	Costs per assay: Biopharm Costs per sample: calculated IFU recommends all samples and controls should be run in duplicate.
	Anti-drug antibody monitoring	96	2 controls; 6 standards (tested once/in duplicate)	88/80	£775.00	£8.81 / £9.69	40	£775.00	£19.38	Costs per assay: Biopharm Costs per sample: calculated IFU recommends all samples and controls should be run in duplicate.

Test		Number of wells	Number of controls per assay	Singlet testing of patient samples			Duplicate testing of patient samples			Sources and comments
				Number of samples analysed per assay	Cost per assay	Cost per sample	Number of samples analysed per assay	Cost per assay	Cost per sample	
LISA-TRACKER[a]	Drug monitoring	48[d]	1 control; 5 standards (tested once/in duplicate)	42/36	£836.77	£19.92 / £23.24	24	£836.77	£34.87	Costs per assay: Cambridge Life Sciences (UK distributor) Cost per sample: calculated IFU indicates all samples and controls could be run in singlicate.
	Anti-drug antibody monitoring	48[d]	1 control; 5 standards (tested once/in duplicate)	42/36	£836.77	£19.92 / £23.24	24	£836.77	£34.87	Costs per assay: Cambridge Life Sciences (UK distributor) Cost per sample: calculated IFU states all samples and controls could be run in singlet.
MabTrack [a]	Drug monitoring	96	2 controls; 6 standards (tested once/in duplicate)	88/80	€1259.50	€14.31 / €15.74	40	€1259.50	€31.49	Cost per assay: Sanquin Cost per sample: calculated IFU recommends duplicate testing of samples, but singlet of controls and standards.
	Anti-drug antibody monitoring	96	2 controls; 2 standards (tested once/in duplicate)	92/88	€847.90	€9.21 / €9.64	44	€847.90	€19.27	Cost per assay: Sanquin Cost per sample: calculated IFU recommends duplicate testing of

Test		Number of wells	Number of controls per assay	Singlet testing of patient samples			Duplicate testing of patient samples			Sources and comments
				Number of samples analysed per assay	Cost per assay	Cost per sample	Number of samples analysed per assay	Cost per assay	Cost per sample	
										samples, but singlet of controls and standards.
Sanquin Diagnostics	Adalimumab/ Infliximab drug monitoring	N/A	N/A	8[f]	€50	€6.25	4	€50	€12.50	Diagnostic service, cost per assay: Sanquin Cost per sample: calculated
	Adalimumab/ Infliximab antibody monitoring	N/A	N/A	8[f]	€50	€6.25	4	€50	€12.50	Diagnostic service, cost per assay: Sanquin Cost per sample: calculated
	Certolizumab/ golimumab/ etanercept drug monitoring	N/A	N/A	8[f]	€90	€11.25	4	€90	€22.50	Diagnostic service, cost per assay: Sanquin Cost per sample: calculated
	Certolizumab/ golimumab/ etanercept antibody monitoring	N/A	N/A	8[f]	€90	€11.25	4	€90	€22.50	Diagnostic service, cost per assay: Sanquin Cost per sample: calculated

Key: ADl = adalimumab; IFX = infliximab; NR= not reported; TBC = to be confirmed

Notes:

[a] Costs exclude VAT

[b] Cost inclusive of VAT

[c] Unclear whether cost includes or excludes VAT

[d] In the information request documents, the manufacturer provided only the cost of the 48-well assay, the cost of the 96-well assay was not provided as the manufacturer reported that this assay is rarely purchased

[e] Cost provided by the manufacturer, used in the model

[f] Request for information states that the cost of an 8 serial dilution is €50, the ERG assumes that analysis of 8 patient samples costs €50/€90

4.1.9.1.11 Cost of sample transport

One of the very minor cost components considered by Jani and colleagues (2016)[53] was "Transport, receipt and storage of sample" which was £2.22 (2015 prices) per batch of 40 samples (refer to Table 1 in Jani and colleagues 2016).[53]

Blood samples are received at the Exeter Clinical Laboratory (Royal Devon and Exeter NHS Foundation Trust) as small parcels via Royal Mail. Clinical advice from Dr Timothy McDonald indicated that it is extremely unlikely that samples would be sent to Sanquin Diagnostic Services in the Netherlands as the transportation cost would be higher compared to that within the UK. Postage costs are approximately £4 per parcel shipped within the UK and approximately £10 per parcel shipped to Sanquin Diagnostic Services.[17] Moreover, sending samples abroad would lead to a longer turnaround time and take expertise out of the NHS (Timothy McDonald, personal communication, December 2018).

Therefore, in all analyses relevant to MabTrack and Sanquin Diagnostic Services, the cost of sample transport of £10 was applied, while for all the other tests the postage of £4 per parcel was assumed (i.e. it was assumed that parcel would be posted to a laboratory within the UK).[17]

4.1.9.1.12 Frequency of testing

Rosas and colleagues (2015)[58] reported the total number of drug and anti-drug antibody monitoring tests in RA patients in remission over a two-year period (94 tests in 45 patients), which is approximately one test per patient per year.

Dr Meghna Jani, confirmed that in people in remission/under routine follow up, TNF testing may be conducted once a year. However, if tapering is performed on the basis of drug level, a clinician would typically check drug levels at least every six-months to ensure that the level has not dropped too low.

In the primary analysis, one TNF test per patient-year was assumed, while six-monthly testing was modelled in a sensitivity analysis (Table 73).

SECTION 4 PAGE 185 to 186

(Table 42).

4.19.1.20 Cost of managing adverse events

In TA375, the weighted average cost of serious infection in RA patients was estimated to be £1,479 based on relevant NHS costs (NHS reference costs schedules 2010-11), weighted by inpatient activity (TA375 report, p. 311). Relevant HRG codes were identified based on Lekander and colleagues (2010). Conservatively the without complications and contraindications HRG costs were used.

The average cost inflated to 2017–18 prices using the healthcare price index (Section 4.1.10.1.1) was £1,622 (per infection). This cost was incorporated in our analysis (Table 42).

4.1.9.2 Health related quality of life

A review of health-related quality-of-life (HRQoL) studies was conducted to inform the selection of utilities for the economic analysis. Utilities of remission/LDA and active disease health states, and disutilities for flares and serious adverse events (e.g. severe infections) identified in the review are described below.

4.1.9.2.1 Health state utility values

Ucar and colleagues (2017)[11] and Arango and colleagues (2017)[15] (INGEBIO) provided results on the average duration of remission in the intervention and control arms. However, none of the sources reporting INGEBIO study provided definitions of remission.

sensitivity analysis, when the disease states were assessed according to the CDAI and DAS28 indices (Table 59).

The HAQ scores were mapped to the EQ-5D values using the same formula as in Section 4.1.10.2.2 and presented in Table 66.

Table 66: EQ-5D scores for the states of disease activity according to the SDAI, CDAI and DAS28

Type of index	Remission		LDA		MDA/HDA	
	Mean	Range	Mean	Range	Mean	SD
SDAI	0.718	(0.565, 0.804)	0.635	(0.432, 0.796)	0.483	(0.222, 0.694)
CDAI	0.72	(0.573, 0.804)	0.626	(0.415, 0.794)	0.486	(0.229, 0.694)
DAS28	0.70	(0.532, 0.804)	0.666	(0.477, 0.804)	0.483	(0.226, 0.691)

Key: CDAI: Clinical Disease Activity Index; DAS28: Disease Activity Score for 28 joints; HDA: high disease activity; LDA: low disease activity; MDA: moderate disease activity; SDAI: Simplified Disease Activity Index

Ucar and colleagues (2017) reported mean duration of remission in the intervention and control arms. In the economic analysis based on this source, the SDAI value for remission, 0.718, was applied (Table 66). The utility value for a mixed disease state (LDA/active disease) was approximated by a weighted average of the estimates for LDA and MDA/HDA, 0.568.

As the health states in Arango and colleagues (2017) were defined differently to those in Ucar and colleagues (2017) (remission/LDA and active disease), in analyses based on Arango and colleagues (2017), the SDAI value of 0.483 for MDA/HDA was used as the utility value for active disease health state, and a weighted average of the estimates for remission and LDA, 0.665, was used to approximate the utility value for the mixed disese state.

HSUVs obtained from HAQ scores reported in TA375[50] (as described in the next section) were assumed in a scenario analysis.

The AG is aware of several algorithms for converting the HAQ score to utility in RA and that the estimates of utilities may vary when different mapping algorithms are used.[104] To address this uncertainty, HSUVs were estimated using a quadratic equation proposed by Malottki and colleagues (2011) and used in TA375 to map HAQ to EQ-5D scores.[14,50] In TA375 a comparison of published relationships between utility and HAQ was conducted.[50] Three of the eight studies in the comparison in TA375 reported data from the UK. Of the three studies, Bansback and colleagues (2007)[105] included data for UK and Canadian

patients and Kobelt and colleagues (2002)[106] included data for UK and Swedish patients and were therefore not considered relevant for the purposes of this analysis. Hurst and colleagues (1997)[107] included only people with RA in Scotland. Malottki and colleagues (2011)[14] used the data set from Hurst and colleagues (1997)[107] to estimate the coefficients of their mapping equation and therefore there is little difference between the two sources.

HSUVs estimated from HAQ by EULAR response category

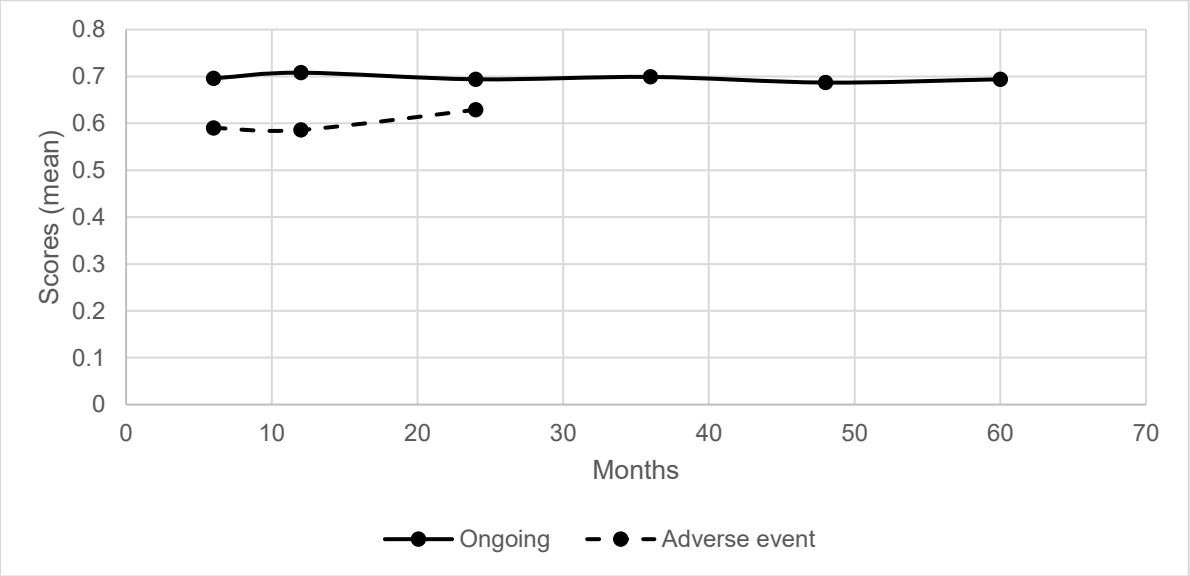
In TA375,[50] the model was based on EULAR response category (good/moderate/none) for consistency with NICE guidance on biologics in RA and to align more closely to UK clinical practice in terms of the assessment of response to therapies.

SECTION 4 PAGE 196 to 201

People with RA also demonstrated lower utility gain upon termination of the therapy independently of the reason for withdrawal of treatment compared to the other sub-populations.

Figure 15 provides another possible way of estimating the disutility of adverse events as the weighted average difference between the plotted values for patients who remained in ongoing treatment for the whole duration of the study, and those who eventually withdrew due to an AE, which in this case (for three observations at six, 12 and 24 months' timepoints) is 0.106. The value for utility loss of 0.156 estimated in Oppong et al. (2013) for England was based on an expected average disease duration of 28 days, while the effect of AEs observed in Gülfe et al. (2016) lasted longer than four weeks. This may explain the difference between the two curves in the figure.

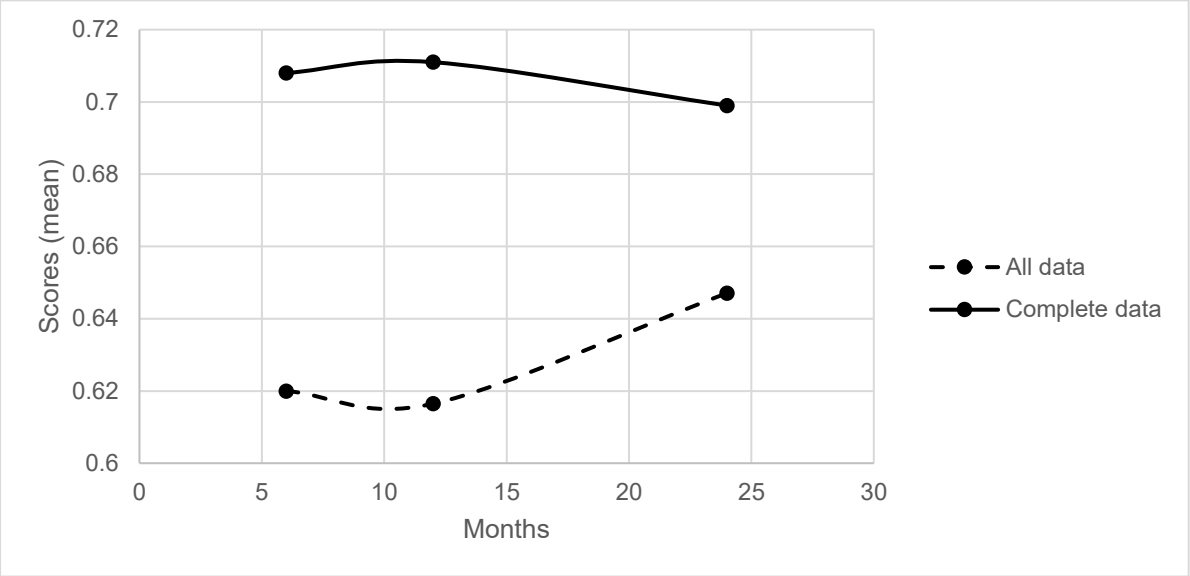
Figure 15: EQ-5D during follow-up upon withdrawal from treatment



Key: PsA: psoriatic arthritis; RA: rheumatoid arthritis; SpA; spondyloarthritis
Source: Gülfe and colleagues, 2010

While using all available data increases the generalisability of the study, it also leads to lower improvement estimates as compared to using data only for those participants for whom complete follow-up information is available from all visits (incomplete data sets may be caused by withdrawal from treatment, for example) (Figure 16).

Figure 16: EQ-5D for people with rheumatoid arthritis (all participants versus participants with complete data) during the first and second lines of anti-TNF-α treatment



Key: PsA: psoriatic arthritis; RA: rheumatoid arthritis; SpA; spondyloarthritis
Source: Gülfe and colleagues, 2010

4.1.9.2.5 Mortality

Whilst there is evidence of an association between HAQ improvement and reduced mortality risk, the impact of TNF testing on mortality was not considered due to the short-term time horizon adopted in this study and a relatively small difference in the mean duration of remission (Ucar and colleagues, 2017)[11] and remission/LDA (Arango and colleagues, 2017)[15] across the treatment arms in the INGEBIO study.

4.1.10 Checking the model for wiring errors

The Excel code was checked in the following ways: all calculations were performed by one person and checked by another person, and the reasonableness of outputs given extreme input values was checked.

4.2 Cost effectiveness results

4.2.1 Adalimumab and Promonitor

4.2.1.1 Threshold analysis

The results of the threshold analysis, assuming the Promonitor test kit is used to monitor people with RA in remission/LDA recieving originator ADL (Humira®) are presented in Table 70 and Figure 17.

Figure 17 shows the annual cost of ELISA-based testing at which TDM would become cost-effective at the two WTP thresholds used in NICE decision making for the range of ADL acquisition costs of £1,000–£9,187. Since the data reported in Arango and colleagues (2017)[152] are for a longer follow-up than that reported in Ucar and colleagues (2017)[11], the results using the two different reports of the outcomes of the INGEBIO study are presented.

If the results of Ucar and colleagues (2017)[11] are used, then with the current price of originator ADL, testing would need to be cheaper than £430 per year in order for TDM to be judged as cost-effective at the cost-effectiveness threshold of £20,000 per QALY gained. Using the the results presented in Arango and colleagues (2017);[15] however, there would be no cost of testing at which testing becomes cost-effective (because using these outcomes testing was estimated to be both more costly and less effective than standard care).

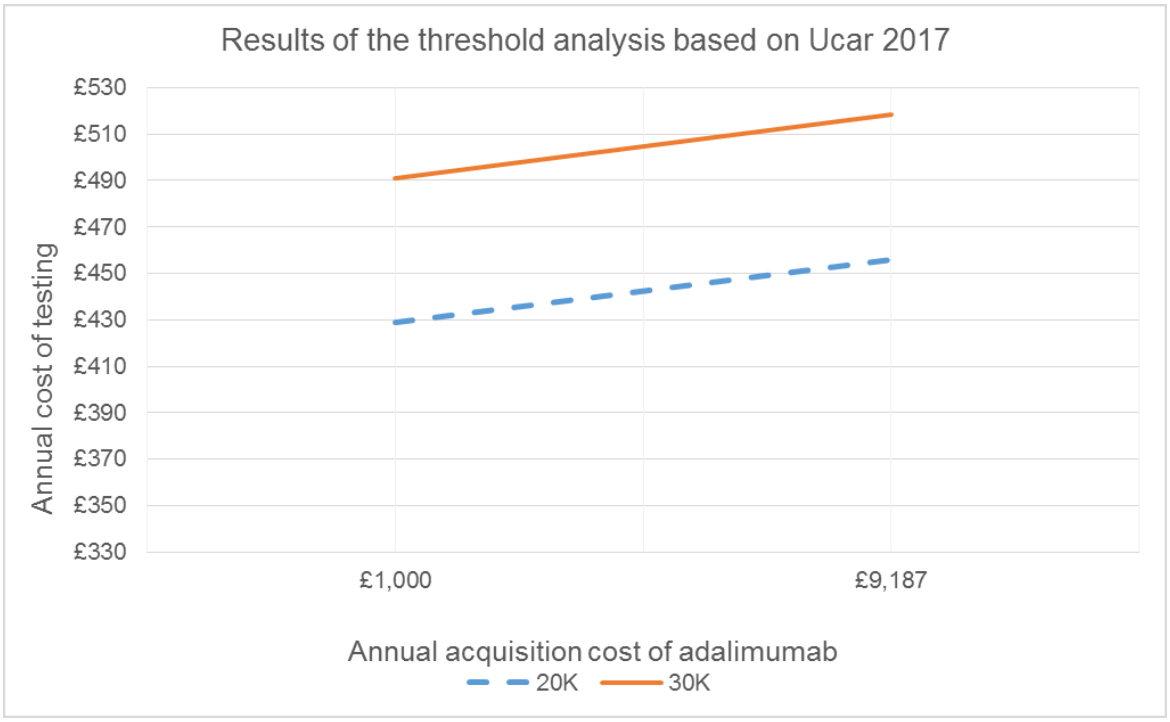
Table 70: **Threshold value for the cost of testing at which NMB is zero**

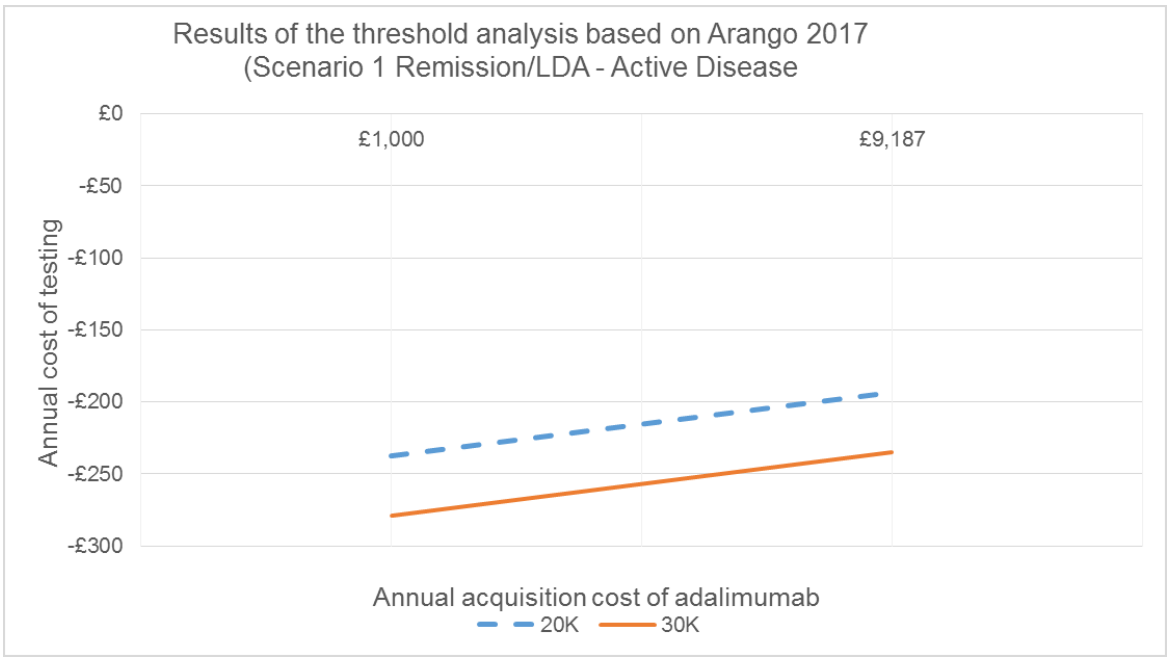
ICER threshold	Results based on INGEBIO study, Ucar and colleagues 2017	Results based on INGEBIO study, Arango and colleagues 2017
£20,000	£430	-£200
£30,000	£479	-£246

Key: ICER: incremental cost-effectiveness ratio; NMB, net monetary benefit
Source: Ucar and colleagues (2017)[11] and Arango and colleagues (2017)[15]

Such differences in the results are due to differences in the mean duration of remission (as reported in Ucar 2017) and remission/LDA (Arango 2017) between the control and intervention arms. Arango reported a longer duration of remission/LDA in the control group than in the intervention group (475.2 versus 460.2 days), while Ucar and colleagues 2017 reported a longer duration in the intervention group (344 versus 329 days in the control group).

Figure 17: Results of the threshold analyses using Ucar and colleagues (2017) and Arango and colleagues (2017)





Source: Ucar and colleagues (2017)[11] and Arango and colleagues (2017)[15]

These results are inconclusive for two reasons. First, because they are in opposite directions and, second because they are based on very small and uncertain differences in outcomes (QALY differences of less than 0.01). The negative value of the cost of testing at which NMB equal zero means that, when using the trial results as presented in Arango and

colleagues (2017),[15] there are no (positive) values of the cost of testing at which it would be a cost-effective option.

4.2.1.2 Cost-utility analysis

The incremental QALYs and incremental costs for testing versus standard care strategy are shown in Table 71, assuming:

- regular testing is undertaken in people with RA in remission/LDA treated with Humira® and tested using Promonitor
- the costs of testing are as in Jani and colleagues (2016)
- the frequency of testing is one test per patient-year and
- that testing of drug and antibody levels is done concurrently (singlet dilution) in a UK laboratory.

The outcome data were derived from two reports of the INGEBIO study, Ucar and colleagues, 2017 and Arangoand colleagues, 2017.

As can be seen from Table 71, the main cost components are drug acquisition and the costs of managing health states. The main differences in costs between the intervention and control arms are the costs of managing health states and flares, and the cost of phlebotomy appointment. The main QALY components are those for the health states. The differences in QALYs for flares and AEs between the intervention and control arms are very small.

Table 71: Cost-effectiveness results in patients in remission/LDA treated with Humira® and tested using Promonitor

	Intervention arm	Control arm	Intervention vs. control
Based on Ucar and colleagues (2017)			
Costs			
Drug acquisition	£12,078	£12,120	-£42
Drug admin	£0	£0	£0
Drug wastage	£486	£488	-£2
Cost of managing health states	£19,071	£19,379	-£307
Cost of flare management	£281	£388	-£107
Cost of managing AEs	£64	£64	£0
Cost of phlebotomy appointment	£162	£0	£162
Other costs of testing	£30	£0	£30
Cost of sample transport	£6	£0	£6
Total costs (mean)	£32,178	£32,438	-£260

	Intervention arm	Control arm	Intervention vs. control
QALYs			
Remission	0.676	0.647	*0.029*
LDA/active disease	0.250	0.274	-0.023
Flares	-0.002	-0.002	-0.001
AEs	-0.0005	-0.0005	-0.000
Total QALYs (mean)	0.924	0.918	0.007
ICER (Cost / QALY gained)			*ICER not relevant - Intervention dominates standard care*
Based on Arango and colleagues (2017)			
Costs			
Drug acquisition	£13,075	£13,149	-£74
Drug admin	£0	£0	£0
Drug wastage	£527	£530	-£3
Cost of managing health states	£22,112	£21,757	£355
Cost of flare management	£303	£418	-£115
Cost of managing AEs	£69	£70	£0
Cost of phlebotomy appointment	£162	£0	£162
Other costs of testing	£30	£0	£30
Cost of sample transport	£6	£0	£6
Total costs (mean)	£36,284	£35,923	£361
QALYs			
Remission/LDA	0.838	0.865	-0.027
Active disease	0.112	0.092	0.020
Flares	-0.002	-0.003	-0.001
AEs	-0.001	-0.001	-0.000
Total QALYs (mean)	0.947	0.954	-0.007
ICER (Cost / QALY gained)			ICER not relevant - Standard care dominates Intervention

Key: ICER: incremental cost effectiveness ratio; QALY: quality-adjusted life year; vs: versus
Note: The postage was £4 per parcel

As with the threshold analyses, these results are inconclusive. First, because they are in opposite directions and, second because they are based on very small and uncertain differences in outcomes (QALY differences of less than 0.01). Furthermore, it is not possible to argue that either the analysis based of Ucar and colleagues (2017)[11] or that based on Arango and colleagues (2017)[15] is more valid than the other – they both have significant weaknesses (refer to Section 2). The follow-up in Arango and colleagues (2017)[15] is over a

longer time horizon (545 days in the control arm) that Ucar and colleagues (2017)[11] (505 days

SECTION 4 PAGE 204 to 209

Table 73: Sensitivity analyses (people in remission/low disease activity)

Sensitivity analysis	Assumptions	ICER		Source (relevant sections)
		Ucar and colleagues (2017)	Arango and colleagues (2017)	
Impact of flares only (health states and AEs are not included)	Only flares contribute to differential costs and QALYs	ICER £72,645 (incr costs £47; incr QALYs 0.001)	ICER £8,804 (incr costs £6; incr QALYs 0.001)	Scenario C (people in remission, Gavan 2017,35 Section 3.3.2.3)
Tapering strategy	Spacing: reduction of ADA dose to 40 mg every 4 weeks	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	2nd dose reduction Exeter biologic clinic recommendations (Appendix 5)
Treatment wastage	No wastage	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Assumption
Flare duration, days	19	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Weighted average based on Bykerk and colleagues (2014)[16] and clinical advice
Proportion of flared patients in whom full dose is restored	55%	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Bykerk and colleagues (2014)[16] and clinical advice
	0%	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Clinical advice
Utilities[2]				
Remission	0.496	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Estimated from HAQ scores reported in TA375[50] (Fig. 94, p.366) (Section 4.1.10.2.1)
LDA/active disease	0.302			

Sensitivity analysis	Assumptions	ICER		Source (relevant sections)
		Ucar and colleagues (2017)	Arango and colleagues (2017)	
Disutility of flare	0.085	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Minor B type of utility (Table 69, Section 4.1.10.2.2)
	0.116	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Major B type of utility, (Table 69, Section 4.1.10.2.2)
Frequency of testing (tests/year)	2	ICER not relevant - Intervention dominates standard care	ICER not relevant - Standard care dominates Intervention	Rosas and colleagues 2015,[58]r clinical advice (Section 4.1.10.1.20)
Cost of testing	Refer to Table 56 for the cost of testing			
Duplicate concurrent testing with initial phlebotomy appointment		In all analyses, ICER not relevant -Intervention dominates standard care	In all analyses, ICER not relevant - Standard care dominates Intervention	Jani and colleagues (2015)[114] (Section 4.1.10.1.14)
Duplicate reflex testing without initial phlebotomy appointment, 35.8% of ptxs w/LDL[3,4]				
Duplicate reflex testing with initial phlebotomy appointment, 35.8% of ptxs w/LDL[4]				
Singlet reflex testing without initial phlebotomy appointment, 35.8% of ptxs w/LDL[3,4]				
Singlet reflex testing with initial appointment, 35.8% of ptxs w/LDL[4]				
Duplicate concurrent testing without initial phlebotomy appointment[3]				

Sensitivity analysis	Assumptions	ICER		Source (relevant sections)
		Ucar and colleagues (2017)	Arango and colleagues (2017)	
Duplicate reflex testing without initial phlebotomy appointment, 4.7% of ptxs w/LDL[3,5]				
Duplicate reflex testing with initial phlebotomy appointment, 4.7% of ptxs w/LDL[3,5]				
Singlet concurrent testing without initial phlebotomy appointment[3]				
Singlet reflex testing without initial phlebotomy appointment, 4.7% of ptxs w/LDL[3,5]				
Singlet reflex testing with initial appointment, 4.7% of ptxs w/LDL[5]				

Key: AE: adverse events; HAQ: health assessment questionnaire; ICER, incremental cost-effectiveness ratio; LDA: low disease activity; LDL: low drug level

Notes:

All costs are reported in 2017-18 prices.

1 Based on the average cost of joint replacement surgery in rheumatoid arthritis patients from the Royal Devon & Exeter NHS Foundation Trust (Appendix 8).

2 Utilities for the mixed disease population (as in the INGEBIO study) were assumed to be the same as those for people with RA

3 The cost of testing does not include the cost of an additional phlebotomy appointment which might not be required if people will be receiving regular hematological analysis as part of on-going treatement.

4 Assuming 35.8% of people have low drug level (Laine and colleagues 2016)[2]

5 Assuming 4.7% of people have low drug level (Chen and colleagues 2015)[65]

Of the sensitivity analyses conducted, only one impacted on the results: the assumption that the rate of flares alone changes as a consequence of monitoring (i.e. health states and AEs were not included) The ICERs in the analyses by Ucar and colleagues (2017)[11] and Arango and colleagues (2017)[15] were £72,645/QALY and £8,804/QALY, respectively (Table 73).

4.2.1.4 Deterministic sensitivity analysis

One-way sensitivity analyses for some of the parameters used to estimate the ICERs based on data from Arango and colleagues (2017)[15] were also conducted (Table 74). The deterministic sensitivity analyses used estimates from Arango and colleagues (2017)[15] only, as it was expected that results based on data from Ucar and colleagues (2017)[11] would be similar to those in the major analysis, i.e. the intervention would dominate standard care. Changing these parameters had no impact on the findings, standard care was estimated to dominate the intervention in all analyses.

Table 74: One-way deterministic sensitivity analyses based on data from Arango and colleagues (2017)

Parameter	Assumption	ICER	Source
Percentage of people in whom the biologic was tapered	+20% in the intervention arm and -20% in the control arm	ICER not relevant - Standard care dominates Intervention	Arango and colleagues (2017)
Flare rate	-20% in the intervention arm, +20% in the control arm	ICER not relevant - Standard care dominates Intervention	Arango and colleagues (2017)
Differential time in remission	+10% in the intervention arm, -10% in the control arm of the differential time in remission	ICER not relevant - Standard care dominates Intervention	Arango and colleagues (2017)
Costs of managing health states	- 20%	ICER not relevant - Standard care dominates Intervention	Arango and colleagues (2017), Radner and colleagues (2014), Barbieri and colleagues (2005)

Key: ICER: incremental cost-effectiveness ratio

4.2.1.5 Probabilistic sensitivity analysis

Probabilistic sensitivity analysis was deemed inappropriate because of a very substantial variation in clinical practice with respect to disease management in people with RA in England.

4.2.2 Etanercept and Infliximab and Promonitor

The cost-effectiveness of TNF testing in people treated with Etanercept (originator and biosimilar) and Infliximab (biosimilar) using the Promonitor test kit was explored in scenario analyses. In those analyses, it was assumed, based on clinical advice (and a lack of evidence to the contrary), that the clinical effectiveness of the different TNF inhibitors is likely to be the same, and the clinical effectiveness estimates from Ucar and colleagues (2017)[11] were adopted, with all assumptions, except acquisition and administration costs, as in *Table 42*. The information on the actual costs to the NHS of the TNF inhibitors was not available to the AG at the time of writing, and therefore the list prices of the biologics were assumed. The results are presented in *Table 75*.

Table 75: **Cost-effectiveness results for the other tests and TNF inhibitors: people in remission/LDA**

Treatment		ICER	
	Cost per year (£)	**Ucar and colleagues (2017)**	**Arango and colleagues (2017)**
ETN			
*Enbrel®**	9,327	ICER not relevant - Intervention dominates standard care (ICER -£38,247 (total costs -£261; total QALYs 0.007])	ICER not relevant - Standard care dominates Intervention (ICER £53,203 (total costs £360; total QALYs -0.007])
Erelzi	8,394	ICER not relevant - Intervention dominates standard care (ICER -£37,597 [total costs -£256; total QALYs 0.007)	ICER not relevant - Standard care dominates Intervention (ICER -£54,351 [total costs £368; total QALYs -0.007)
IFX[1]			
Flixabi/ Renflexis (no wastage)	5164	ICER not relevant - Intervention dominates standard care (ICER -£36,580 [total costs -£249; total QALYs 0.007)	ICER not relevant - Standard care dominates Intervention (ICER -£56,144 [total costs £380; total QALYs -0.007)

Key: ETN: etanercept; ICER: incremental cost-effectiveness ratio; IFX: infliximab

Notes:

It was assumed that blood samples would be sent for testing to UK laboratories, and the postage of £4 (per small parcel) was applied[17].

* The originator (or reference) product

[1] IFX administration cost was assumed to be 283 per injection (Section 4.1.10.1.7).

Source: Ucar and colleagues (2017)[11] and Arango and colleagues (2017)[15]

Other scenario analyses considered but not conducted due to no or low quality clinical data were: analysis of testing in the context of primary or secondary non-response; analysis for

non-responders who did not adhere to treatment with biologic therapies, including switching to intravenously administered IFX.

4.3 Discussion

Despite substantial weaknesses in the clinical effectiveness evidence base (Section 2), a simple model was developed to estimate the cost-utility of ELISA test-based monitoring for people with RA taking bDMARDS.

The analyses conducted are inconclusive and suggest considerable uncertainty in the cost-effetciveness of therapeutic monitoring of TNF-alpha inhibitors in RA. Data from 2 reports of the same study produced very different conclusions on the cost-effectiveness of Promonitor testing in people receiving ADL who are in remission/LDA. The results based on the longer follow-up (Arango and colleagues 2017[15]) suggested that monitoring is more costly and produces fewer QALYs than standard care.

Of the sensitivity analyses conducted, only the assumption that the rate of flares alone changes as a consequence of monitoring, impacted on the results. The ICERs in the analyses by Ucar and colleagues (2017)[11] and Arango and colleagues (2017)[15] were £72,645/QALY and £8,804/QALY, respectively (Table 70).

Exploratory analyses of using Promonitor to monitor patients in remission/LDA receiving ETN or INF were undertaken, and showed the same results as that for ADL: using the longer follow-up (Arango and colleagues 2017[15]) monitoring is more costly and produces fewer QALYs than standard care.

The main effectiveness evidence in the model was from the poorly reported INGEBIO study (a non-randomised controlled trial from Spain, where <40% of participants had RA), heavily supplemented by input parameters from other studies and expert advice. The results of the economic analysis should therefore be viewed as exploratory and highly speculative. For example, although the INGEBIO study only evaluated testing using Promonitor ELISA kits, for those in remission/LDA treated with Humira® (ADL), with further assumptions these results have been used to estimate the threshold testing costs at which TDM would become cost-effective with people taking other TNF inhibitors (and taking either originator products or biosimilars.

In summary, there is much uncertainty in relation to key potential drivers of the effectiveness and cost-effectiveness of using ELISA based testing to monitoring treatment with bDMARDs in people with RA, that no firm conclusions can be drawn.

SECTION 5: PAGE 210 to 215

5. Discussion

5.1 Statement of principal findings

5.1.1 Clinical effectiveness

Two studies (reported in four publications)[11,15,59,64] were included in the systematic review of the evaluation of using ELISA tests for therapeutic drug monitoring (TDM) on clinical outcomes in people with RA who had achieved remission or low disease activity [LDA], or in those who had experienced a primary non-response or a secondary non-response. Three articles[11,15,64] reported the same non-randomised controlled trial (the INGEBIO study). The remaining study was an observational study evaluating the impact of TDM. The non-randomised controlled study[11,15,64] was judged to be at serious risk of bias. The observational study[59] had a historical control.and was judged to be at moderate risk of bias. However, the study design should be taken into consideration in interpreting the risk of bias assessment (non-randomised controlled study vs. observational study).

The INGEBIO study used Promonitor ELISA kits for monitoring drug levels and/or anti-drug antibody levels, and the study by Pascual-Salcedo (2013) used Sanquin ELISA kits to measure drug levels and/or anti-drug antibody levels. The included studies measured drug levels and/or anti-drug antibody levels in patients who were being treated with adalimumab (ADL), etanercept (ETN) and/or infliximab (IFX). There were no studies identified for people who were being treated with certolizumab pegol (CTZ) and golimumab (GLM). No studies were identified evaluating eligible ELISA kits including IDKmonitor ELISA kits, LISA-TRACKER ELISA kits, RIDASCREEN ELISA kits and MabTrack ELISA kits. Both studies included individuals in remission, with the INGEBIO study also including individuals with low disease activity (at baseline).

5.1.1.1 Comparative controlled evidence

Three articles[11,15,64] reported the same non-randomised controlled trial (the INGEBIO study), which focused on the population who had achieved treatment target (remission or low disease activity [LDA]). In this trial, ADL and anti-ADL antibody levels were measured using Promonitor-ADL and Promonitor-ANTI-ADL (Grifols-Progenika). Monitoring testing results were revealed to physicians in the intervention arm. Such monitoring test results were not

revealed to physicians in the control arm. This reflected standard care in Spain where treatment decisions were based on clinical judgements without the knowledge of drug levels and anti-drug antibodies of patients. This INGEBIO study recruited a mixed population of 169 including a cohort of 63 people with RA. The results of the total mixed population were reported in the review as the authors were not able to provide the results for the cohort of people with RA. The three cohorts with different conditions (rheumatoid arthritis [RA], psoriatic arthritis [PsA] and ankylosing spondylitis [AS]), may have different treatment responses to TNF-α inhibitor therapies. Therefore, there was limited generalisability of findings from this mixed population to the target RA population.

The findings from this trial (Ucar and colleagues 2017)[11] showed that, at 18-month follow-up, the rate of flares per patient-year was 0.463 for the intervention group and 0.639 for the control group, with rate difference of -0.176 (95% CI -0.379 to 0.0289). There was a non-significant reduction in risk of flare in the intervention group compared with the control group (incidence rate ratio (IRR) 0.7252, 95% CI 0.4997 to 1.0578). Median time to first flare was 145 days for participants in the intervention group and 136.5 days for participants in the control group. This trial (Ucar and colleagues, 2017[11]) further presented the results of health-related quality of life (HRQoL) outcomes. Results showed that HRQoL (EQ-5D-5L) measures were higher in the intervention group at all visits compared with the control group. However, the statistically significant results were only observed at Visit 2 (p=0.001) and Visit 3 (p=0.035). Further details of results for this outcome were not reported.

Overall, the findings from this non-randomised controlled trial (the INGEBIO study) showed that there was a non-significant reduction in risk of flare in the intervention group (where treatment decisions were made on the basis of the results of therapeutic drug monitoring) compared with the control group (i.e. standard care where treatment decisions were based on clinical judgements without the knowledge of drug levels and anti-drug antibodies of patients). HRQoL outcomes were higher in the intervention group at all visits compared with the control group, with statistically significant results being observed at two visits. However, the quality of this trial was judged to be at serious risk of bias due to potential attrition bias and baseline imbalance in disease severity between the two groups. Therefore, the results should be interpreted with caution.

5.1.1.2 Evidence from observational studies

One observational study was identified that evaluated the effect of TDM on clinical outcomes in people with RA who had achieved remission or LDA, or in those people who had experienced a primary non-response or a secondary non-response.

5.1.1.2.2 Change in disease activity

The observational study (Pascual-Salcedo and colleagues 2013[59] evaluated the effect of TDM on change in disease activity at duration of follow-up of seven years, with a sample size of 43 individuals. The study focused on participants who had achieved remission or LDA and examined two different time periods, pre- and during-TDM practice. The study showed a non-significant reduction in the mean DAS28 score following the implementation of TDM at seven-year follow-up (pre-TDM: mean 2.51 [SD 0.85] vs.during-TDM: 2.31, [SD 0.52]; p=0.061).

Overall, the finding from this historically controlled study[59] showed that therapeutic drug monitoring was associated with a non-significant reduction in mean DAS28 scores at seven-year follow-up compared with the historical control period. It should be noted that the quality of data was judged to be at moderate risk of bias, which compromises the reliability of the findings.

5.1.1.2.3 Change in direction and magnitude of therapeutic dose

The observational study (Pascual-Salcedo 2013)[59] evaluated the outcome of changes in direction and magnitude of therapeutic dose in people with RA who had achieved remission or LDA.

The findings from the study demonstrated that, compared with the historical control period without TDM, there were statistically significant reductions in weekly mean dose per patient by each anti-TNF (AFX, ADL, ETN) during the 2nd period where TDM was introduced. The findings from this study further showed that, compared with the historical control, there were statistically significant increases in the mean interval of administration for each anti-TNF during the 2nd period when TDM was implemented.

Overall, the limited data from this observational study showed that TDM for optimising anti-TNF therapies was associated with reductions in therapeutic dose of anti-TNFs in people with RA who had achieved remission or LDA. This would be expected to lead to cost saving associated with TDM. However, the reliability of the findings may be compromised by the poor quality of the study.

5.1.2 Cost effectiveness

The analyses conducted are inconclusive and suggest considerable uncertainty in the cost-effetciveness of therapeutic monitoring of TNF-alpha inhibitors in RA. Data from 2 reports of the same study produced very different conclusions on the cost-effectiveness of Promonitor testing in people receiving ADL who are in remission/LDA. The results based on the longer

follow-up (Arango and colleagues 2017[15]) suggested that monitoring is more costly and produces fewer QALYs than standard care.

Of the sensitivity analyses conducted, only one impacted on the results: the assumption that the rate of flares alone changes as a consequence of monitoring. The ICERs in the analyses by Ucar and colleagues (2017)[11] and Arango and colleagues (2017)[15] were £72,645/QALY and £8,804/QALY, respectively.

Exploratory analyses of using Promonitor to monitor patients in remission/LDA receiving ETN or INF were undertaken, and showed the same results as that for ADL: using the longer follow-up (Arango and colleagues 2017[15]) monitoring is more costly and produces fewer QALYs than standard care.

5.2 Strengths and limitations of the assessment

5.2.1 Clinical effectiveness

Extensive literature searches were conducted with an attempt to maximize the retrieval of potentially relevant studies for the systematic review of clinical effectiveness. These included electronic searches of a variety of bibliographic databases as well as screening of clinical trial registers and conference proceedings to identify unpublished studies. The search strategy did not restrict by study design. The review process followed recommended methods to minimise the potential for error and/or bias. The quality of included studies was assessed and accounted for when interpreting the review results. Appropriate synthesis methods were employed by taking into account the heterogeneity of study characteristics.

In terms of limitations, only studies in English were included, therefore some potentially relevant non-English language studies may have been missed. There was scarce evidence relating to clinical effectiveness of TDM on clinical outcomes in people RA who had experienced a primary non-response or a secondary non-response. No studies were identified assessing ELISA kits including IDKmonitor ELISA kits, LISA-TRACKER ELISA kits, RIDASCREEN ELISA kits and MabTrack ELISA kits. There was considerable clinical heterogeneity associated with interventions, outcomes and *length of follow-up between included studies. It was not possible to* investigate publication bias, because quantitative synthesis was not possible in this systematic review due to considerable clinical heterogeneity.

5.2.2 Cost effectiveness

A systematic review of published economic evaluations of using ELISA tests relative to the alternatives and standard care was undertaken to help inform the type and structure of the

SECTION 5 PAGE 218 to 219

HRQoL from the INGEBO study was limited. It is recognised, however, that the HAQ is a functional measure, and does not capture the full impact of RA on quality of life.

- Utility values estimated from HRQoL data for people with RA were applied based on clinical outputs from the INGEBIO study which had a *mixed* population of people with RA, PsA, and AS. Since people with RA are usually older and more likely to be female when compared with people with PsA or AS, the utility values for people with RA, used in the economic analysis, are likely to be lower than those for the mixed population (since men tend to value health states higher than women, and the same applies to younger versus older people).[18] This may have overestimated the incremental cost-effectiveness ratios (ICERs).
- Since the rates of AE were not reported in the INGEBIO study, the impact of AEs was modelled using evidence from another study, which is a limitation of this analysis. However, based on clinical advice and published literature on adverse events in people with RA treated with TNF inhibitors, those AEs which carry a significant cost and disutility burden are relatively rare.
- Finally, limited evidence on utilities, based on EQ-5D scores, directly relevant to people with flares, people experiencing serious adverse events as well as people with remission/LDA or active disease health status in the UK settings was identified in this study. Therefore, utilities were derived from HAQ scores which were estimated in studies conducted in people with RA in non-UK settings. It should be noted however that utilities were estimated by mapping to EQ-5D outcomes from UK tariffs.

5.3 Uncertainties

5.3.1 Clinical effectiveness

In this assessment limited data were identified that evaluated clinical effectiveness of using ELISA tests for monitoring response to TNF-α inhibitors in people with RA who had achieved remission or LDA, and no data were identified for people who had experienced a primary non-response or a secondary non-response. In particular, no RCTs were identified that evaluated patient-related outcomes and disease activities associated with using ELISA tests for TDM in the target populations.

The non-randomised controlled study[11,15,64] was judged to be at serious risk of bias. The historical controlled observational study was judged to be at moderate risk of bias. In the non-randomised controlled trial (the INGEBIO study), there was baseline imbalance in disease severity between the intervention and control groups. Furthermore, there was a lack

of adjusting for this variable in the analysis of clinical outcomes. There were high attrition rates for some outcomes, which could lead to attrition bias. The historically controlled study by Pascual-Salcedo and colleagues (2013)[59] was associated with non-contemporaneous control bias due to the use of a historical control. Given the poor quality of included studies, the potential role of ELISA testing in terms of its clinical impact on monitoring response to TNF-α inhibitors in the target populations remains unclear.

5.3.2 Cost effectiveness

Since there is neither gold standards nor guidelines available to monitor the TNF inhibitors considered in this assessment, economic analyses of test-based treatment strategies with biologics represent a substantial challenge.

Due to data limitations and the lack of clarity with regard to test-based treatment strategies, the AG deliberately refrained from data-intensive modelling approaches, which would be impossible to implement without making strong assumptions not supported by evidence.

The studies identified in the clinical-effectiveness systematic review and used to inform the model structure and parameters are limited by study design (e.g. none of the studies were randomised, one study was observational). Furthermore, those studies are characterised by relatively small sample sizes. In one study (INGEBIO), treatment of RA patients based on the results of TNF testing was compared against usual care, and this was in a mixed disease population with only 37% of RA patients. The studies, including the INGEBIO study, were reported as abstracts only. The study sponsors were not categorically stated but some of the authors in the INGEBIO study worked for pharmaceutical companies.

The AG is aware of several test-based treatment algorithms used by physicians in England. However, in the only study comparing test versus no-test treatment strategies, the INGEBIO study (which was utilised in our economic analysis), physicians were not required to follow any therapeutic algorithm based on TDM results but could use tests to alter doses based on their clinical judgement. It is unclear, however, whether there are variations in clinical practice between England and Spain, which could have impacted the results presented here.

In the INGEBIO study (identified in the systematic review), a disease flare was defined as an increase in DAS28 >1.2, or an increase in DAS28 >0.6 if the current DAS28 ≥3.2. However, the AG is aware of several RA flare criteria, which have been used in clinical research.[9,67] Our clinical advisors confirmed that such a variation also exists in clinical practice. To address this uncertainty, the effect of the variation in the definition of flare was

SECTION 6: PAGES 222 to 223

6. Conclusions

6.1 Implications for service provision

The findings from this assessment demonstrate very limited evidence on the effect of TDM based on ELISA tests for optimising anti-TNF therapies in people with RA, either in those who had achieved remission or LDA, or in those who had experienced a primary non-response or a secondary non-response.

In relation to clinical effectiveness, limited data were identified evaluating TDM in the target populations. One non-randomised trial compared TDM with standard care (the INGEBIO study) had serious limitations in relation to the NICE scope: only one-third of the participants had RA, many of the analyses were not by intention-to-treat, follow-up was for only 18 months, there was no explicit algorithm for guiding clinicians in how the results of testing should change treatment (e.g. tapering), and the study was only reported in three abstracts. In addition, one observational study was also identified but was of minimal value in informing whether ELISA test-based monitoring is clinically effective or not.

Despite these substantial weaknesses in the clinical effectiveness evidence base, a simple model was developed to estimate the cost-utility of ELISA test-based monitoring for people with RA taking bDMARDS. The main effectiveness evidence in the model was also from the poorly reported INGEBIO study, heavily supplemented by input parameters from other studies and expert advice. The results of the economic analysis should therefore be viewed as exploratory and highly speculative. For example, although the INGEBIO study only evaluated testing using Promonitor ELISA kits, for those in remission/LDA treated with Humira® (ADL), with further assumptions these results have been used to estimate the threshold testing costs at which TDM would become cost-effective with people taking other TNF inhibitors (and taking either originator products or biosimilars.

In summary, there is limited valid and applicable research evidence, and much uncertainty in relation to key potential drivers of the effectiveness and cost-effectiveness of using ELISA based testing to monitoring treatment with bDMARDs in people with RA, that no firm conclusions regarding the implications for service provision can be drawn

6.2 Suggested research priorities

One ongoing Norwegian multicentre randomised controlled trial (the NOR-DRUM Study)[66] that evaluates the effect of TDM in people with RA in remission compared with standard

care. This ongoing trial will provide further useful data on the impact of TDM in the target population.

Further controlled trials with a large sample size (especially randomised controlled trials [RCTs]) are required to assess the impact of using Promonitor ELISA tests for monitoring anti-TNF therapies in people with RA who had achieved remission or LDA.

No studies were identified evaluating other eligible ELISA kits including IDKmonitor ELISA kits, LISA-TRACKER ELISA kits, RIDASCREEN ELISA kits and MabTrack ELISA kits. Therefore, future large RCTs are required to assess the impact of using those ELISA tests for monitoring anti-TNF therapies in people with RA who had achieved remission or LDA. More robust evidence is also needed to evaluate the impact of using Sanquin tests for monitoring anti-TNF therapies in this population.

There were no studies identified for people with RA treated with CTZ or GLM. Future RCTs are required to assess the clinical effectiveness of using ELISA tests for monitoring such anti-TNF therapies in the target populations.

There was no relevant data identified for the population of people with RA who had experienced a primary non-response or a secondary non-response. Future RCTs are warranted to assess the clinical effectiveness of using ELISA tests for monitoring anti-TNF therapies in those who had developed clinical inefficacy.

Limited evidence on healthcare resource use and utilities, based on EQ-5D scores, directly relevant to the population considered in this assessment was identified in this study. This warrants further research on medium/long term cost and health outcomes in people with RA treated with TNF inhibitors.

Appendix 2. Included and excluded studies

Table 76: **Studies included in the clinical-effectiveness systematic review**

	Source	Title	Article type	Contributed data
Non-randomised controlled studies				
C. G. Arango, M. L. G. Vivar, E. U. Angulo, I. Gorostiza, C. E. Perez, J. R. De Dios, B. Alvarez, A. R. Escribano, C. Stoye, M. Vasques, J. B. Otano, A. Escobar, Z. Trancho, A. R. Del Agua, L. Del Rio, C. Jorquera, A. Martinez and D. Nagore	Arthritis and Rheumatology. Conference: American College of Rheumatology/Association of Rheumatology Health Professionals Annual Scientific Meeting, ACR/ARHP	Prospective, intervention, multicenter, non-inferiority study of utility of therapeutic drug monitoring with respect to the efficacy and cost of adalimumab tapering in patients with rheumatic diseases	Conference abstract	Yes
I. Gorostiza, E. U. Angulo, C. G. Arango, C. E. Perez, J. R. De Dios, B. Alvarez, A. R. Escribano, C. Stoye, M. Vasques, J. B. Otano, A. Escobar, Z. Trancho, A. R. Del Agua, L. Del Rio, A. Martinez and D. Nagore	Arthritis and Rheumatology	Prospective, intervention, multicenter study of utility of biologic drug monitoring with respect to the efficacy and cost of adalimumab tapering in patients with rheumatic diseases (34-week descriptive data)	Conference abstract	Yes
Ucar, E., Gorostiza, I., Gomez, C., Perez, Ce, Dios, Jr, Alvarez, B., Ruibal, A., Stoye, C., Vasques, M., Belzunegui, J., Escobar, A., Trancho, Z.,Ruiz, Del Agua A., Martinez, A., Jorquera, C., Nagore, D	*Annals of the rheumatic diseases. Conference: annual european congress of rheumatology, EULAR 2017. Spain*	*Prospective, intervention, multicenter study of utility of biologic drug monitoring with respect to the efficacy and cost of adalimumab tapering in patients with rheumatic diseases: preliminary results of ingebio study*	*Conference abstract*	*Yes*
Observational study				
D. Pascual-Salcedo, C. Plasencia, L. Gonzalez Del Valle, T. Lopez Casla, F. Arribas, A. Villalba, G. Bonilla, E.	Annals of the Rheumatic Diseases. Conference: Annual European Congress of Rheumatology of the European League Against Rheumatism, EULAR	Therapeutic drug monitoring (TDM) in rheumatic day clinic enables to reduce pharmaceutical cost maintaining clinical efficacy	Conference	Yes

Source	Title	Article type	Contributed data
Lopez Granados, E. Martin Mola and A. Balsa			

APPENDIX 2: PAGE 329

Authors	Source	Title	Reasons for exclusion
Wilson, D. Plant, K. Watson, A. Barton, K. Hyrich		rheumatology biologics register for rheumatoid arthritis Ann Rheum	
Jani M, Dixon WG, Lunt M, De Cock D, Isaacs J, Morgan A, Watson K, Wilson AG, Barton A, Hyrich KL	Arthritis Rheumatol	The Association of Biologic Drug-Levels with Infection Risk: Results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis	Population
l'Ami MJ, Krieckaert CL, Nurmohamed MT, van Vollenhoven RF, Rispens T, Boers M, Wolbink GJ	Ann Rheum Dis	Successful reduction of overexposure in patients with rheumatoid arthritis with high serum adalimumab concentrations: an open-label, non-inferiority, randomised clinical trial	Population
D. Y. Chen, Y. M. Chen, T. Y. Hsieh, W. T. Hung, C. W. Hsieh, H. H. Chen, K. T. Tang and J. L. Lan	Rheumatology	Drug trough levels predict therapeutic responses to dose reduction of adalimumab for rheumatoid arthritis patients during 24 weeks of follow-up	Intervention
J. Inciarte-Mundo, M. Hernandez, V. Ruiz-Esquide, J. Ramirez, A. Cuervo, S. Cabrera-Villalba, M. Pascal, J. Yague, J. Canete and R. Sanmarti	Annals of the Rheumatic Diseases	Prediction of flare in rheumatoid arthritis and psoriatic arthritis patients with low disease activity receiving TNF inhibitors: Role of calprotectin and drug trough serum levels. A one-year prospective cohort study	Intervention
M. T. Lopez-Casla, D. Pascual-Salcedo, C. Plasencia, P. Alcozer, S. Garcia-Carazo, G. Bonilla, A. Villalba, D. Peiteado, F. Arribas, E. Martin-Mola and A. Balsa	Annals of the Rheumatic Diseases. Conference: Annual European Congress of Rheumatology of the European League Against Rheumatism, EULAR	The infliximab dose increase is not correlated with clinical improvement in RA patients	Intervention
B. Paredes, C. Plasencia, D. Pascual-Salcedo, I. Monjo, A. Pieren, E. Moral, C. Tornero, P. Bogas, G. Bonilla, L. Nuno, A.	Annals of the Rheumatic Diseases	Influence of tapering biological therapies in immunogenicity in a cohort of rheumatoid arthritis with low disease activity	Intervention

Authors	Source	Title	Reasons for exclusion
Villalba, D. Peiteado, S. Ramiro, T. Jurado, J. Diez, E. Martin-Mola and A. Balsa			
B. Paredes, C. Plasencia, D. Pascual-Salcedo, I. Monjo, A. Pieren, E. Moral, C. Tornero, G. Bonilla, L. Nuno, A. Villalba, D. Peiteado, S. Ramiro, T. Jurado, J. Diez, E. Martin-Mola and A. Balsa	Annals of the Rheumatic Diseases	Influence of optimization of biological therapies on immunogenicity in a cohort of rheumatoid arthritis with low disease activity	Intervention
J. Rosas, F. Llinares-Tello, J. Miguel Senabre, G. Santos-Soler, E. Salas-Heredia, X. Barber, A. Pons, C. Cano, M. Lorente and J. Molina	Arthritis and Rheumatology. Conference: American College of Rheumatology/Association of Rheumatology Health Professionals Annual Scientific Meeting, ACR/ARHP	Economic impact of decreasing adalimumab and etanercept doses and drug monitoring in patients with rheumatoid arthritis in clinical remission: Preliminary study from a local biologics unit	Intervention
J. M. Senabre Gallego, J. Rosas Gomez De Salazar, M. Marco Mingot, A. Naranjo, F. Llinares-Tello, A. Pons, X. Barber-Valles, G. Santos-Soler, E. Salas-Heredia, C. Cano, M. Lorente, J. A. Garcia Gomez and J. Molina	Annals of the Rheumatic Diseases	Clinical activity, ultrasound assessment and drug monitoring in rheumatoid arthritis patients receiving anti-TNF-alpha therapy with extended interval of administration	Intervention

APPENDIX 3: PAGES 331 to 332

Appendix 3. Quality Assessment

A3.1 (Part I): Quality assessment on the basis of specific outcomes

Table below presents the risk of bias assessment on the basis of specific outcomes: clinical disease activity (disease flare, remission, and change in disease activity), proportion of patients receiving dose tapering, health-related quality of life (HQoL) and treatment dose-related outcomes. For each specific outcome, the following bias domains were assessed: bias due to confounding, bias in selection of participants into the study, bias in measurement of interventions, bias due to departures from intended interventions, bias due to missing data, bias in taking measurements, and bias in selection of the reported result.

In terms of outcome-specific assessments, both the clinical activity outcome (disease flare, remission and change in disease activity) and the health related quality of life were judged to be at serious risk of bias, given that there was serious risk of bias in the domain of bias due to confounding. For both outcomes, there were low to moderate risks of biases for the remaining bias domains: bias in selection of participants into the study, bias in measurement of interventions, bias in taking measurements, and bias in selection of the reported results.

Regarding dose-related outcome and proportion of patients receiving dose tapering, both outcomes were judged to be at moderate risk of bias because there was moderate risk of bias for two bias domains (bias in taking measurements and bias due to confounding). For both outcomes, there was low risk of bias for the remaining bias domains: bias in selection of participants into the study, bias in measurement of interventions, and bias in selection of the reported results.

Risk of bias in outcome-specific assessments

Domain	Clinical activity (disease flare, remission, change in disease activity)	Proportion tapered	Health related quality of life	Dose-related outcomes
Bias due to confounding	Moderate - serious	Moderate	Serious	Moderate
Bias in selection of participants into the study	Low - moderate	Low	Low	Mmoderate
Bias in measurement of interventions	Low - moderate	Low	Low	Moderate
Bias due to departures from intended interventions	NI	NI	NI	NI
Bias due to missing data	Serious	NI	Serious	NI
Bias in taking measurements	Moderate	Moderate	Moderate	Moderate
Bias in selection of the reported result	Low	Low	Low	Low
Overall risk of bias	Moderate - serious	Moderate	Serious	Moderate

Key: Risk of bias judgement: low/moderate/serious/critical/NI. NI: no information

APPENDIX 3: PAGES 333 to

A3.2 (Part II): Quality assessment of individual studies

The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool
(version for cohort-type studies) **ARANGO 2017**
Version 19 September 2016

ROBINS-I tool (Stage I): At protocol stage

Specify the review question

Participants	Remission/primary non-responders/secondary non-responders
Experimental intervention	Therapeutic drug monitoring
Comparator	Standard care
Outcomes	13 outcomes; inconclusive results, time to results, dose changes, dose adjustment, treatment switch, discontinuation, changes in disease activity, rate of disease response, relapse and remission, hospitalization, rates of surgical intervention, adverse effects, health-related quality of life

List the confounding domains relevant to all or most studies

From protocol; time of testing, testing method (e.g. reflex vs. concurrent) Others (suggested); drug dose/levels, disease stage at enrollment, time of assessment for response/follow-up, type of drug manipulation (e.g. optimisation or tapering)

List co-interventions that could be different between intervention groups and that could impact on outcomes

Methotrexate, other DMARDs, combination or monotherapy

ROBINS-I tool (Stage II): For each study

Specify a target randomized trial specific to the study

Design	Individually randomized√ / Cluster randomized / Matched (e.g. cross-over)
Participants	Adult patient treated with Adalimumab (40mg sc) who remained clinically stable for at least 6 months
Experimental intervention	Adjustment of ADL frequency (tapering) plus therapeutic drug monitoring (TDM) data to revealed to physicians
Comparator	Adjustment of ADL frequency (tapering), physicians blinded to TDM data

Is your aim for this study…?

- □ to assess the effect of *assignment to* intervention
- √ to assess the effect of *starting and adhering to* intervention

Specify the outcome

Specify which outcome is being assessed for risk of bias (typically from among those earmarked for the Summary of Findings table). Specify whether this is a proposed benefit or harm of intervention.

Proportion of patients tapered (benefit), rate of flare (harm)

Specify the numerical result being assessed

In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, **figure** or paragraph) that uniquely defines the result being assessed.

Proportion tapered; 34.6% (CG), 35.7% (IG)

Preliminary consideration of confounders

Complete a row for each important confounding domain (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as potentially important.

"Important" confounding domains are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention. "Validity" refers to whether the confounding variable or variables fully measure the domain, while "reliability" refers to the precision of the measurement (more measurement error means less reliability).

(i) Confounding domains listed in the review protocol				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator?
			Yes / No / No information	Favour experimental / Favour comparator / No information
Disease stage (proportion in remission/LDA)		No	Yes	Expected to favour control group (28.6% IG had LDA vs. 17.3% of CG)
Time of assessment for response		No	No information	No information but likely to be unimportant. Measurement believed to be done at similar time points (at 8 scheduled visits over 18 months)
Serum Adalimumab levels		No	Yes	NA – serum ADL levels 5.76mg/L in the CG and 5.04mg/L in IG.
Serum anti-Adalimumab antibody levels		No	No information	No information

(ii) Additional confounding domains relevant to the setting of this particular study, or which the study authors identified as important				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator?
			Yes / No / No information	Favour experimental / Favour comparator / No information

* In the context of a particular study, variables can be demonstrated not to be confounders and so not included in the analysis: (a) if they are not predictive of the outcome; (b) if they are not predictive of intervention; or (c) because adjustment makes no or minimal difference to the estimated effect of the primary parameter. Note that "no statistically significant association" is not the same as "not predictive".

Preliminary consideration of co-interventions

Complete a row for each important co-intervention (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as important.

"Important" co-interventions are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention.

(i) Co-interventions listed in the review protocol		
Co-intervention	Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)?	Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information

(ii) Additional co-interventions relevant to the setting of this particular study, or which the study authors identified as important		
Co-intervention	Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)?	Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator
Methotrexate and other DMARDs	No	Favour experimental / Favour comparator / No information√
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information

Risk of bias assessment

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	Yes, differential baseline LDA rates and no information on co-intervention	Y / PY √ / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.		NA / Y / PY / PN / N √ / NI
1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)		NA / Y / PY / PN / N / NI
Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?		NA / Y / PY / PN / N √ / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?		NA√ / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?		NA√ / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?		NA / Y / PY / PN / N√ / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?		NA√ / Y / PY / PN / N / NI
Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?		Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4		Y / PY / PN / N√ / NI

2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?		NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?		Y / PY√ / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?		NA √ / Y / PY / PN / N / NI
Risk of bias judgement		Low √ / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?		Y √ / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?		Y√ / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?		Y / PY / PN / N√ / NI
Risk of bias judgement		Low√ / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions		
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?		Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?		NA√ / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?		Y / PY / PN / N / NI√
4.4. Was the intervention implemented successfully for most participants?		Y √ / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?		Y / PY√ / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?		NA√ / Y / PY / PN / N / NI
Risk of bias judgement		
Optional: What is the predicted direction of bias due to deviations from the intended interventions?		

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?		Y / PY / PN / N / NI√
5.2 Were participants excluded due to missing data on intervention status?		Y / PY / PN / N / NI√

5.3 Were participants excluded due to missing data on other variables needed for the analysis?		Y / PY / PN / N / NI√
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?		NA√ / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?		NA√ / Y / PY / PN / N / NI
Risk of bias judgement		Low / Moderate / Serious / Critical / NI√
Optional: What is the predicted direction of bias due to missing data?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?		Y / PY / PN √ / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?		Y √ / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?		Y / PY / PN / N / NI√
6.4 Were any systematic errors in measurement of the outcome related to intervention received?		Y / PY / PN / N / NI√
Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?		Y / PY / PN / N√ / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?		Y / PY / PN / N√ / NI
7.3 ... different *subgroups*?		Y / PY / PN / N√ / NI
Risk of bias judgement		Low√ / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias		
Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool
(version for cohort-type studies) **GOROSTIZA 2016**
Version 19 September 2016

ROBINS-I tool (Stage I): At protocol stage

Specify the review question

Participants	Remission/primary non-responders/secondary non-responders
Experimental intervention	Therapeutic drug monitoring
Comparator	Standard care
Outcomes	13 outcomes; inconclusive results, time to results, dose changes, dose adjustment, treatment switch, discontinuation, changes in disease activity, rate of disease response, relapse and remission, hospitalization, rates of surgical intervention, adverse effects, health-related quality of life

List the confounding domains relevant to all or most studies

From protocol; time of testing, testing method (e.g. reflex vs. concurrent)
Others (suggested); drug dose/levels, disease stage at enrollment, time of assessment for response/follow-up, type of drug manipulation (e.g. optimisation or tapering)

List co-interventions that could be different between intervention groups and that could impact on outcomes

Methotrexate, other DMARDs, combination or monotherapy

ROBINS-I tool (Stage II): For each study

Specify a target randomized trial specific to the study

Design	Individually randomized√ / Cluster randomized / Matched (e.g. cross-over)
Participants	Patient treated with adalimumab (40mg sc) who remained clinically stable for at least six months
Experimental intervention	Biological monitoring data (BMD) were released to physicians
Comparator	Physicians were blinded to BMD data

Is your aim for this study…?

- □ to assess the effect of *assignment to* intervention
- √ to assess the effect of *starting and adhering to* intervention

Specify the outcome

Specify which outcome is being assessed for risk of bias (typically from among those earmarked for the Summary of Findings table). Specify whether this is a proposed benefit or harm of intervention.

Proportion remaining in remission (benefit)

Specify the numerical result being assessed

In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.

Promotion remaining in remission = 69.6% (CG), 76.1% (IG)

Preliminary consideration of confounders

Complete a row for each important confounding domain (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as potentially important.

"Important" confounding domains are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention. "Validity" refers to whether the confounding variable or variables fully measure the domain, while "reliability" refers to the precision of the measurement (more measurement error means less reliability).

(i) Confounding domains listed in the review protocol				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator?
			Yes / No / No information	Favour experimental / Favour comparator / No information
Disease stage (proportion in remission/LDA)		No	Yes	Expected to favour control group (26.6% IG had LDA vs. 16.7% of CG)
Time of assessment for response		No	No information	No information but likely to be unimportant. Measurement believed to be done at similar time points (at 8 scheduled visits over 18 months)
Serum Adalimumab levels		No	Yes	NA – serum ADL levels 5.5mg/L in the CG and 5.3mg/L in IG.
Serum anti-Adalimumab antibody levels		No	No information	No information

(ii) Additional confounding domains relevant to the setting of this particular study, or which the study authors identified as important				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator?
			Yes / No / No information	Favour experimental / Favour comparator / No information

* In the context of a particular study, variables can be demonstrated not to be confounders and so not included in the analysis: (a) if they are not predictive of the outcome; (b) if they are not predictive of intervention; or (c) because adjustment makes no or minimal difference to the estimated effect of the primary parameter. Note that "no statistically significant association" is not the same as "not predictive".

Preliminary consideration of co-interventions

Complete a row for each important co-intervention (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as important.

"Important" co-interventions are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention.

(i) Co-interventions listed in the review protocol		
Co-intervention	Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)?	Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information

(ii) Additional co-interventions relevant to the setting of this particular study, or which the study authors identified as important		
Co-intervention	Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)?	Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator
Methotrexate and other DMARDs	No	Favour experimental / Favour comparator / No information√
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information

Risk of bias assessment

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	Yes, differential baseline LDA rates and no information on co-intervention	Y / PY √/ PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.		NA / Y / PY / PN / N√ / NI
1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)		NA / Y / PY / PN / N / NI
Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?		NA / Y / PY / PN / N √/ NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?		NA √/ Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?		NA√ / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?		NA / Y / PY / PN / N√ / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?		NA√ / Y / PY / PN / N / NI
Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?		Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4		Y / PY / PN / N√ / NI

2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?		NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?		Y / PY√ / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?		NA√ / Y / PY / PN / N / NI
Risk of bias judgement		Low √/ Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	BDM data were released only to IG	Y√ / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?		Y√ / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?		Y / PY / PN / N√ / NI
Risk of bias judgement		Low√ / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions		
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?		Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?		√NA / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?		Y / PY / PN / N / NI√
4.4. Was the intervention implemented successfully for most participants?		Y √/ PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?		Y / PY√ / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?		NA√ / Y / PY / PN / N / NI
Risk of bias judgement		
Optional: What is the predicted direction of bias due to deviations from the intended interventions?		

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?		Y / PY / PN / N / NI√
5.2 Were participants excluded due to missing data on intervention status?		Y / PY / PN / N / NI√

5.3 Were participants excluded due to missing data on other variables needed for the analysis?		Y / PY / PN / N / NI√
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?		NA√ / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?		NA√ / Y / PY / PN / N / NI
Risk of bias judgement		Low / Moderate / Serious / Critical / NI√
Optional: What is the predicted direction of bias due to missing data?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?		Y / PY / PN √/ N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?		Y√ / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?		Y / PY / PN / N / NI√
6.4 Were any systematic errors in measurement of the outcome related to intervention received?		Y / PY / PN / N / NI√
Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?		Y / PY / PN / N√ / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?		Y / PY / PN / N√ / NI
7.3 ... different *subgroups*?		Y / PY / PN / N√ / NI
Risk of bias judgement		Low√ / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias		
Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool
(version for cohort-type studies) **PASCUAL-SALCEDO 2013**
Version 19 September 2016

ROBINS-I tool (Stage I): At protocol stage

Specify the review question

Participants	Remission/primary non-responders/secondary non-responders
Experimental intervention	Therapeutic drug monitoring
Comparator	Standard care
Outcomes	13 outcomes; inconclusive results, time to results, dose changes, dose adjustment, treatment switch, discontinuation, changes in disease activity, rate of disease response, relapse and remission, hospitalization, rates of surgical intervention, adverse effects, health-related quality of life

List the confounding domains relevant to all or most studies

From protocol; time of testing, testing method (e.g. reflex vs. concurrent)
Others (suggested); drug dose/levels, disease stage at enrollment, time of assessment for response/follow-up, type of drug manipulation (e.g. optimisation or tapering)

List co-interventions that could be different between intervention groups and that could impact on outcomes

Methotrexate, other DMARDs, combination or monotherapy

ROBINS-I tool (Stage II): For each study

Specify a target randomized trial specific to the study

Design	Individually randomized / Cluster randomized / Matched (e.g. cross-over)
Participants	RA Patients in remission or LDA
Experimental intervention	Down-titration or cessation of Infliximab, adalimumab, etanercept plus therapeutic monitoring period
Comparator	Down-titration or cessation of Infliximab, adalimumab, etanercept, prior to therapeutic monitoring period

Is your aim for this study…?

- □ to assess the effect of *assignment to* intervention
- √ to assess the effect of *starting and adhering to* intervention

Specify the outcome

Specify which outcome is being assessed for risk of bias (typically from among those earmarked for the Summary of Findings table). Specify whether this is a proposed benefit or harm of intervention.

Mean DAS28 score (harmful), weekly mean dose (lower better), interval of administration (higher better)

Specify the numerical result being assessed

In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.

Mean DAS28 score; 1st period: 2.51±0.85 vs 2nd period: 2.31±0.52

Preliminary consideration of confounders

Complete a row for each important confounding domain (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as potentially important.

"Important" confounding domains are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention. "Validity" refers to whether the confounding variable or variables fully measure the domain, while "reliability" refers to the precision of the measurement (more measurement error means less reliability).

(i) Confounding domains listed in the review protocol				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator?
			Yes / No / No information	Favour experimental / Favour comparator / No information
Disease stage (proportion in remission/LDA)		No	No information	No information
Time of assessment for response		No	No information	No information
Serum Adalimumab levels		No	No information	No information
Serum anti-Adalimumab antibody levels		No	No information	No information

(ii) Additional confounding domains relevant to the setting of this particular study, or which the study authors identified as important				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator?
			Yes / No / No information	Favour experimental / Favour comparator / No information

* In the context of a particular study, variables can be demonstrated not to be confounders and so not included in the analysis: (a) if they are not predictive of the outcome; (b) if they are not predictive of intervention; or (c) because adjustment makes no or minimal difference to the estimated effect of the primary parameter. Note that "no statistically significant association" is not the same as "not predictive".

Preliminary consideration of co-interventions

Complete a row for each important co-intervention (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as important.

"Important" co-interventions are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention.

(i) Co-interventions listed in the review protocol		
Co-intervention	Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)?	Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information

(ii) Additional co-interventions relevant to the setting of this particular study, or which the study authors identified as important		
Co-intervention	Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)?	Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator
Methotrexate, other DMARDs, combination or monotherapy	Not done/no information	Favour experimental / Favour comparator / No information√
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information

Risk of bias assessment

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered		Y / PY√ / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.		NA / Y / PY / PN / N√ / NI
1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)		NA / Y / PY / PN / N / NI
Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?		NA / Y / PY / PN√ / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?		√NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?		NA / Y / PY / PN / N√ / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?		NA / Y / PY / PN√ / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?		√NA / Y / PY / PN / N / NI
Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?		Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	'…….a total of 88 patients (43 RA and 45 SpA), treated with three TNF inhibitors ……were included…'	Y / PY / PN / N√ / NI

2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?		√NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		√NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?		Y / PY / PN / N / NI√
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?		√NA / Y / PY / PN / N / NI
Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?		Y√ / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?		Y / PY / PN / N√ / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?		Y / PY / PN / N√ / NI
Risk of bias judgement		Low / Moderate √/ Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions		
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		Not applicable
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?		Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?		NA / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?		Y / PY / PN / N / NI√
4.4. Was the intervention implemented successfully for most participants?		Y / PY√ / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?		Y / PY / PN / N / NI√
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?		√NA / Y / PY / PN / N / NI
Risk of bias judgement		NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?		

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	Results reported were basically means (SD); difficult to determine	Y / PY / PN / N / NI√
5.2 Were participants excluded due to missing data on intervention status?		Y / PY / PN / N / NI√

5.3 Were participants excluded due to missing data on other variables needed for the analysis?		Y / PY / PN / N / NI√
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?		√NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?		√NA / Y / PY / PN / N / NI
Risk of bias judgement		Low / Moderate / Serious / Critical / NI√
Optional: What is the predicted direction of bias due to missing data?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?		Y / PY / PN√ / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?		Y √/ PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?		Y / PY√ / PN / N / NI
6.4 Were any systematic errors in measurement of the outcome related to intervention received?		Y / PY / PN√ / N / NI
Risk of bias judgement		Low / Moderate √/ Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?		Y / PY / PN / N√ / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?		Y / PY / PN / N √/ NI
7.3 ... different *subgroups*?		Y / PY / PN / N√ / NI
Risk of bias judgement		Low√ / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias		
Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool
(version for cohort-type studies) **UCAR 2017**
Version 19 September 2016

ROBINS-I tool (Stage I): At protocol stage

Specify the review question

Participants	Remission/primary non-responders/secondary non-responders
Experimental intervention	Therapeutic drug monitoring
Comparator	Standard care
Outcomes	13 outcomes; inconclusive results, time to results, dose changes, dose adjustment, treatment switch, discontinuation, changes in disease activity, rate of disease response, relapse and remission, hospitalization, rates of surgical intervention, adverse effects, health-related quality of life

List the confounding domains relevant to all or most studies

From protocol; time of testing, testing method (e.g. reflex vs. concurrent)
Others (suggested); drug dose/levels, disease stage at enrollment, time of assessment for response/follow-up, type of drug manipulation (e.g. optimisation or tapering)

List co-interventions that could be different between intervention groups and that could impact on outcomes

Methotrexate, other DMARDs, combination or monotherapy

ROBINS-I tool (Stage II): For each study

Specify a target randomized trial specific to the study

Design	Individually randomized √/ Cluster randomized / Matched (e.g. cross-over)
Participants	Patient treated with adalimumab (40mg sc) who remained clinically stable for at least six months
Experimental intervention	Biological monitoring data (BMD) were released to physicians
Comparator	Physicians were blinded to BMD data

Is your aim for this study…?

□ to assess the effect of *assignment to* intervention
√ to assess the effect of *starting and adhering to* intervention

Specify the outcome

Specify which outcome is being assessed for risk of bias (typically from among those earmarked for the Summary of Findings table). Specify whether this is a proposed benefit or harm of intervention.

Disease flare (harm)

Specify the numerical result being assessed

In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.

IRR = 0.7252 (95% CI = 0.49997 to 1.0578)

Preliminary consideration of confounders

Complete a row for each important confounding domain (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as potentially important.

"Important" confounding domains are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention. "Validity" refers to whether the confounding variable or variables fully measure the domain, while "reliability" refers to the precision of the measurement (more measurement error means less reliability).

(i) Confounding domains listed in the review protocol				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator?
			Yes / No / No information	Favour experimental / Favour comparator / No information
Disease stage (proportion in remission/LDA)		No	Yes	Expected to favour control group (26.6% IG had LDA vs. 16.7% of CG)
Time of assessment for response		No	No information	No information but likely to be unimportant. Measurement believed to be done at similar time points (at scheduled visits)
Serum Adalimumab levels		No	Yes	NA – serum ADL levels 5.5mg/L in the CG and 5.3mg/L in IG.
Serum anti-Adalimumab antibody levels		No	No information	No information

(ii) Additional confounding domains relevant to the setting of this particular study, or which the study authors identified as important				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator?
			Yes / No / No information	Favour experimental / Favour comparator / No information

* In the context of a particular study, variables can be demonstrated not to be confounders and so not included in the analysis: (a) if they are not predictive of the outcome; (b) if they are not predictive of intervention; or (c) because adjustment makes no or minimal difference to the estimated effect of the primary parameter. Note that "no statistically significant association" is not the same as "not predictive".

Preliminary consideration of co-interventions

Complete a row for each important co-intervention (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as important.

"Important" co-interventions are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention.

(i) Co-interventions listed in the review protocol		
Co-intervention	Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)?	Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information

(ii) Additional co-interventions relevant to the setting of this particular study, or which the study authors identified as important		
Co-intervention	Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)?	Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator
Methotrexate and other DMARDs	No	Favour experimental / Favour comparator / No information √
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information
		Favour experimental / Favour comparator / No information

Risk of bias assessment

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	Yes, differential baseline LDA rates and no information on co-intervention	Y / PY√ / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.		NA / Y / PY / PN / N√ / NI
1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)		NA / Y / PY / PN / N / NI
Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?		NA / Y / PY / PN / N √/ NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?		NA√ / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?		NA√ / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?		NA / Y / PY / PN / N √/ NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?		NA√ / Y / PY / PN / N / NI
Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?		Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study

2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4 2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention? 2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		Y / PY / PN / N√ / NI NA / Y / PY / PN / N / NI NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?		Y / PY√ / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?		NA√ / Y / PY / PN / N / NI
Risk of bias judgement		Low√ / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	BDM data were released only to IG	Y √/ PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?		Y√ / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?		Y / PY / PN / N √/ NI
Risk of bias judgement		Low √/ Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions		
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?		Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?		NA√ / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?		Y / PY / PN / N / NI√
4.4. Was the intervention implemented successfully for most participants?		Y √/ PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?		Y / PY√ / PN / N / NI

4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?		NA√ / Y / PY / PN / N / NI
Risk of bias judgement		NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?		

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?		Y / PY / PN / N / NI√
5.2 Were participants excluded due to missing data on intervention status?		Y / PY / PN / N / NI√
5.3 Were participants excluded due to missing data on other variables needed for the analysis?		Y / PY / PN / N / NI√
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?		NA√ / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?		NA √/ Y / PY / PN / N / NI
Risk of bias judgement		Low / Moderate / Serious / Critical / NI√
Optional: What is the predicted direction of bias due to missing data?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?		Y / PY / PN√ / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?		Y √/ PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?		Y / PY / PN / N / NI√
6.4 Were any systematic errors in measurement of the outcome related to intervention received?		Y / PY / PN / N / NI√
Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?		Y / PY / PN / N√ / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?		Y / PY / PN / N√ / NI
7.3 ... different *subgroups*?		Y / PY / PN / N√ / NI

	Risk of bias judgement		Low√ / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to selection of the reported result?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias			
	Risk of bias judgement		Low / Moderate√ / Serious / Critical / NI
	Optional: What is the overall predicted direction of bias for this outcome?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable