Clinical effectiveness and cost-effectiveness of use of therapeutic monitoring of tumour necrosis factor alpha (TNF-α) inhibitors [LISA-TRACKER® enzyme-linked immunosorbent assay (ELISA) kits, TNF-α-Blocker ELISA kits and Promonitor® ELISA kits] versus standard care in patients with Crohn's disease: systematic reviews and economic modelling

Karoline Freeman,[1] Martin Connock,[1] Peter Auguste,[1] Sian Taylor-Phillips,[1] Hema Mistry,[1] Deepson Shyangdan,[1] Rachel Court,[1] Ramesh Arasaradnam,[2] Paul Sutcliffe[1]* and Aileen Clarke[1]

[1]Warwick Evidence, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK
[2]Clinical Sciences Research Institute, University of Warwick, University Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK

*Corresponding author

Declared competing interests of authors: Aileen Clarke is a member of the National Institute for Health Research, Health Technology Assessment and Efficacy and Mechanism Evaluation Editorial Boards. Aileen Clarke and Sian Taylor-Phillips are partly supported by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West Midlands at the University Hospitals Birmingham NHS Foundation Trust.

Published November 2016
DOI: 10.3310/hta20830

Scientific summary

Therapeutic monitoring of TNF-α inhibitors vs. standard care

Health Technology Assessment 2016; Vol. 20: No. 83
DOI: 10.3310/hta20830

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

Scientific summary

Introduction

Crohn's disease (CD) is a chronic fluctuating inflammatory condition of the digestive tract. It is currently estimated to affect approximately 115,000 patients in the UK, with 3000 new cases diagnosed each year. In severe active CD, biological therapies are used when other treatment options fail and before surgical removal of the affected bowel is considered. These more recent drugs are monoclonal antibodies that inactivate tumour necrosis factor alpha (anti-TNF-α). The two anti-TNF-α agents considered here are infliximab (IFX) (Remicade®, Merck Sharp & Dohme Ltd, Kenilworth, NJ, USA) and adalimumab (ADA) (Humira®, AbbVie Inc., North Chicago, IL, USA).

Response to anti-TNF-α agents is variable. Loss of response (LOR) is thought to be caused by subtherapeutic drug levels or the development of anti-drug antibodies that neutralise anti-TNF-α and hasten clearance from the circulation. This idea has led to the development of test kits able to measure circulatory levels of anti-TNF-α drugs and the antibodies directed against them, and to the use of test results in treatment algorithms to bring the anti-TNF-α levels into the therapeutic range and to prevent continuing use of ineffective agents.

Decision problem

The decision problem for this assessment is:

- Does testing of TNF-α inhibitor levels and antibodies to TNF-α inhibitors (IFX or ADA) represent a clinically effective and cost-effective use of NHS resources in patients with moderate or severe CD whose disease responds to treatment or who have lost response to treatment with TNF-α inhibitors?

The comparator for testing is standard care with an appropriate anti-TNF-α.

Three commercially available test kits for estimation of serum anti-TNF-α agents and anti-drug antibodies have been identified as interventions. These are LISA-TRACKER® enzyme-linked immunosorbent assay (ELISA) kits (Theradiag, Marne La Vallee, France, or Alpha Laboratories, Heriot, UK), TNF-α-Blocker ELISA kits (Immundiagnostik AG, Bensheim, Germany) and Promonitor® ELISA kits (Proteomika, Progenika Biopharma, Bizkaia, Spain).

Objectives

Objective A: review of comparative performance of tests

To review and critique studies:

- that compare two or more intervention tests, or an intervention test with another test method which can be used to perform a linked evidence assessment
- that report a test threshold analysis to determine the optimal drug cut-off level to predict or diagnose response.

Objective B: description of algorithms

To describe algorithms used in studies that include data on one or more intervention test or on a test which allows a linked evidence approach to be performed (i.e. algorithms used in studies identified in objective C1).

Objective C1: review of clinical effectiveness of test algorithm combinations

To systematically review the literature comparing the clinical effectiveness of an intervention or other assays for anti-TNF-α agents and/or for anti-drug antibodies used in conjunction with a treatment algorithm in patients with CD treated with IFX or ADA with the clinical effectiveness of standard care (no tests or test-informed algorithm used) in patients with CD treated with the same anti-TNF-α agent.

Objective C2: analysis of correlation between test results and clinical outcomes

To analyse correlation studies that investigate the relationship between test results for anti-TNF-α and anti-drug antibody levels, and clinical outcome measured as clinical response. This objective was added post protocol because of the paucity of studies which address the decision question.

Objective D: review of cost-effectiveness of test algorithm combinations versus standard care

To assess the cost-effectiveness of employing anti-TNF-α and anti-TNF-α antibody monitoring with LISA-TRACKER ELISA kits, TNF-α-Blocker ELISA kits and Promonitor ELISA kits compared with standard care in patients with CD.

In the absence of studies using the intervention tests, to use a linked evidence approach in which evidence of clinical effectiveness is taken from studies using alternative tests to the intervention tests.

Methods

Clinical effectiveness and cost-effectiveness systematic reviews

Multiple electronic databases were searched from inception up to the point of searching, during October to December 2014. Supplementary searches were used to identify additional published and unpublished studies. Reference lists and citation searches of included studies and review articles were undertaken. Further information was provided by the companies.

Two reviewers independently screened and assessed titles and abstracts of all records. Studies were included according to the following:

- population – adults and children with moderate to severe active CD treated with IFX or ADA
- intervention – monitoring of serum anti-TNF-α (IFX or ADA) and/or anti-drug antibody levels using intervention tests or other tests implemented using a test–treatment algorithm
- comparator – standard care (no anti-TNF-α or antibody monitoring)
- outcomes – any patient-related outcome, test agreement, cost-effectiveness estimates
- study design – any primary study design, systematic reviews with meta-analyses.

Study quality assessments were undertaken using an adapted Quality Assessment of Diagnostic Accuracy Studies-2 checklist, the Cochrane risk-of-bias tool, the Downs and Black checklist, Philips' checklist and the Consolidated Health Economic Evaluation Reporting Standards. Data were extracted by one reviewer and checked by a second reviewer. Disagreements were resolved by consensus or with a third reviewer. Evidence was synthesised using narrative review and statistical methods when appropriate. Individual patient data were reconstructed from available Kaplan–Meier plots using the method of Guyot *et al.* (Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan–Meier survival curves. *BMC Med Res Methodol* 2012;**12**:9). Meta-analyses were undertaken in Stata version 11 (StataCorp LP, College Station, TX, USA) or using 'MetaAnalyst' software (Tufts University, Medford, MA, USA) and RevMan version 5.3 (The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark). The Harbord and Whiting (Harbord R, Whiting P. metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata J* 2009;**9**:211–29) method of hierarchical meta-analysis was used for diagnostic studies.

Cost-effectiveness model

A de novo Markov model was built in TreeAge Pro 2013 (TreeAge Software, Inc., Williamstown, MA, USA) to evaluate the cost-effectiveness of a test algorithm strategy-based treatment versus standard care. Two populations were considered: (1) patients responding to treatment and (2) patients who had lost response to treatment. Two test strategies were assessed: (1) concurrent and reflex testing of drugs and (2) antibodies to the drugs (i.e. simultaneous or sequential drug and antibody testing). Concurrent testing yields four possible outcomes: drug+/antibody–, drug+/antibody+, drug–/antibody+ or drug–/antibody–. Reflex testing (antibody testing if drug tests are negative) yields three outcomes: drug+, drug–/antibody– or drug–/antibody+. The model structure was informed by the literature search and expert clinical advice. The model had a 4-week cycle, a 10-year time horizon and adopted a NHS and Personal Social Services perspective. Costs were adjusted to 2013/14 prices and discounted at 3.5% per annum. The starting point was a hypothetical cohort of patients aged 30 years. Outcomes are reported as incremental cost-effectiveness ratios (ICERs), expressed in terms of cost per quality-adjusted life-years (QALYs) gained. A linked evidence approach was adopted (evidence from studies using tests other than the designated intervention tests was employed as a proxy for intervention test evidence). A number of sensitivity analyses were undertaken, including a shortened 1-year time horizon with 4-week cycle lengths and different transition probabilities for LOR. Probabilistic sensitivity analysis was also undertaken (10,000 model runs).

Results

The searches identified 2434 studies of clinical effectiveness and 2466 studies of cost-effectiveness, of which 68 and four studies, respectively, were included.

Clinical effectiveness

Twenty-three studies comparing test methods were identified. Most studies did not investigate any of the three intervention tests. Evidence on concordance between the three intervention assays at a clinically relevant threshold was sparse and sometimes contradictory. Overall, there was insufficient evidence to reliably assess comparative performance of the three intervention assays or their performance relative to other assay methods or to any of the comparators with links to clinical outcomes [homogeneous mobility shift assay (HMSA), radioimmunoassay (RIA), PROMETHEUS® ELISA (Prometheus Laboratories Inc., San Diego, CA, USA) or Leuven in-house ELISA (Vande Casteele N, Ferrante M, Van Assche G, Ballet V, Compernolle G, Van Steen K, *et al.* Trough concentrations of infliximab guide dosing for patients with inflammatory bowel disease. *Gastroenterology* 2015;**148**:1320–9)].

Three studies – two randomised controlled trials (RCTs) and one retrospective observational study – provided comparative evidence on clinical outcomes following implementation of a test algorithm versus a non-algorithm strategy. None of these studies used the intervention tests; all investigated IFX. Neither of the RCTs found evidence of clinical benefit for a test–algorithm–treatment regimen. In the Trough level Adapted infliXImab Treatment trial, which investigated the effectiveness of drug monitoring following dose optimisation in patients with response to IFX treatment, 131 out of 178 (73.59%) patients with CD were in clinical remission before dose optimisation and 138 out of 173 (79.77%) after dose optimisation using a test–treatment algorithm; at 52 weeks post randomisation there was likewise no difference in clinical and biological remission between the intervention test–treatment group and the control group ($p = 0.353$). Both RCTs estimated cost savings in drug expenditure with a test–treatment algorithm compared with normal care. The retrospective observational study compared a proactive test–treatment algorithm with normal care and reported greater retention on IFX treatment for the intervention group. However, the algorithm was ill-defined. Much of the evidence comes from studies, including this retrospective study, that investigated mixed groups of patients with inflammatory bowel disease [CD and ulcerative colitis (UC)].

Thirty-one studies reported on the correlation between test results and subsequent clinical state (response/no response). The studies were meta-analysed to estimate test accuracy for predicting clinical status. Meta-analyses indicated moderate test accuracy; positive and negative predictive value estimates derived

from meta-analyses indicated that between 20% and 30% of positive and negative test results are likely to be inaccurate. This was confirmed by re-analysis of three meta-analyses of the ability to predict response/LOR using drug and/or anti-drug antibody levels.

Among these there were three studies that reported results from both drug and anti-drug antibody tests for individual patients (one for IFX-treated responders, one for IFX-treated patients with LOR and one for ADA-treated responders). However, the patients in these studies did not receive treatment according to a test–treatment algorithm, therefore no outcomes data from the studies were available, and outcomes data from the RCTs had to be used in the economic modelling.

Cost-effectiveness

The systematic review of cost-effectiveness studies identified four studies. All of these indicated that a testing strategy might be less costly than alternatives with variable small effects on effectiveness. Use of standard checklists suggested that all the studies are subject to some limitations. There was insufficient published information to model an ADA test-based treatment strategy. The model therefore addressed IFX therapy only.

In the base case, the de novo Markov model showed that for IFX reflex testing dominates concurrent testing (that means that it is less costly and produces more QALYs); however, no-testing is more costly and produces more QALYs than reflex testing, with an ICER of approximately £50,800.

However, the probabilistic sensitivity analysis indicated a 92% likelihood that the 'no-testing' strategy was cost-effective at a willingness to pay of £20,000 per QALY.

Discussion and conclusions

Main findings

The meta-analysis indicates that tests have only moderate predictive accuracy for clinical status. There was insufficient evidence to assess the performance of the intervention tests properly relative to one another or to tests using alternative methodology. The literature indicates a lack of clinical consensus about which are the best and most appropriate tests to employ in clinical practice.

The limited RCT evidence from short-term studies indicated that there is little or no benefit from a test algorithm strategy, although there may be some cost savings.

The base-case cost-effectiveness analysis indicated that standard care, the no-testing strategy, accumulates slightly greater QALYs, albeit at a higher cost. This strategy is 92% likely to be cost-effective at standard levels of willingness to pay.

Strengths and limitations

Strengths of the work include a robust and comprehensive systematic review (literature search, data extraction and analysis) strategy and the building of a de novo Markov model for the cost-effectiveness assessment.

The main limitation relates to the availability of relevant high-quality evidence. Although we undertook extensive systematic searches and screened more than 30,000 titles, the findings of the systematic review warrant a cautious interpretation. Definitions of severity of disease (including response and LOR) lack standardisation, which impact on the classification of patients in different studies. Consensus on treatment algorithms is missing, possibly impacting on clinicians' confidence in using them. The evidence on assay performance was sparse and sometimes conflicting, with lack of an agreed gold or reference standard for tests. There were very limited concordance data from studies comparing test performance of different assays. Evidence on ADA was lacking.

Populating the economic model with information from the literature was problematic because of the small size and short duration of the studies and their use of subjective methods for outcomes measurement. None of the studies used an appropriate standard care arm for economic modelling and many external sources of data and assumptions were required to populate the model. Inputs for the economic model needed to be drawn from disparate studies so that conclusions need to be tested with data from further research. Several studies sourced for model inputs included a proportion of patients with UC; the impact of this on model outputs is difficult to gauge. Variation in clinical practice in the management of patients with CD further complicated assumptions for model structure and inputs. We were unable to include adverse events and their treatment costs, and this may have underestimated the costs.

Implications

Our finding that testing for levels of IFX and its antibodies is not cost-effective should be viewed cautiously by clinicians and policy-makers, in view of the linked evidence approach required and the poor quality of the evidence available to us. Clinicians should be mindful of the potential variation in performance of the different testing methods and strategies in their day-to-day practice.

Research priorities

We found that there is uncertainty about underlying treatment pathways, about the relative effectiveness of assays in the absence of a gold standard or agreed reference test, about which assays to use under which circumstances and about which clinical algorithms to follow as a result of testing. There is very little research on ADA or the use of testing strategies and algorithms in children. The key questions for future research consideration are:

1. What is the relative performance of methods of measuring anti-TNF-α drug and their antibodies by ELISA kits compared with other methods, such as RIA and HMSA, and are any potential differences clinically significant? For example, is there a validated drug threshold that is a useful predictor of clinical outcome?
2. What are the best criteria for estimating response, non-response and LOR in CD?
3. At what time should assessments of drug and antibody take place?
4. What is the effectiveness of clinical algorithms for disease management in response to testing in the UK?
5. What is the clinical effectiveness and cost-effectiveness of monitoring patients with CD on ADA and for paediatric patients with CD?
6. What is the relevance of cotreatment with immunosuppressants in the monitoring of anti-TNF-α agents and their antibodies?
7. Is there a benefit of measuring total drug/antibodies compared with measurements of free drug/antibody alone?

Study registration

This study is registered as PROSPERO CRD42014015278.

Funding

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

Health Technology Assessment

NICE TAR and DAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.058

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

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

Editorial contact: nihredit@southampton.ac.uk

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

Criteria for inclusion in the *Health Technology Assessment* journal

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

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

HTA programme

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

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

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

This report

The research reported in this issue of the journal was commissioned and funded by the HTA programme on behalf of NICE as project number 14/69/03. The protocol was agreed in January 2015. The assessment report began editorial review in May 2015 and was accepted for publication in December 2015. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

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