A Description of the IVI-RA Model v1.0[[1]](#footnote-21)[[2]](#footnote-22)

Devin Incerti[[3]](#footnote-23)

Jeroen P. Jansen

# Executive summary

This document describes version 1.0 of the [Innovation and Value Initiative’s (IVI’s)](http://www.thevalueinitiative.org/) individual patient simulation model for rheumatoid arthritis (RA) (the IVI-RA model). The model simulates the costs, health outcomes, and risks associated with disease-modifying anti-rheumatic drugs (DMARDs) including conventional DMARDs (cDMARDs), biologic DMARDs (bDMARDs), and Janus kinase/signal transducers and activators of transcription (JAK/STAT) inhibitors for patients with moderate to severe rheumatoid arthritis (RA) who have previously failed treatment with cDMARDs. The model is intended to help decision-makers assess the value of treatments for a population of patients with RA.

## Open-Source Value Project

The IVI-RA model is part of IVI’s Open Source Value Project (OSVP), which is building an open, collaborative, and consensus-based process for the development of tools for value assessment. Models developed by the OSVP process are iterative, evolving as the science of value assessment advances and as new evidence becomes available.

OSVP models are released and updated using a four step process:

1. Public release of the model.
2. Invite feedback and suggested improvements to the model in a public comment period.
3. A panel of experts determines which of the evidence-based suggestions for improvement suggested in Step 2 should be implemented by means of peer-review and a formal voting process.
4. Revise the model based on the feedback from the technical expert panel in Step 3.

To provide a starting point for debate, the initial release of each OSVP model (i.e., version 1.0) must be flexible and allow users to choose from a large number of plausible model structures and approaches based on clinical practice and previous modeling efforts. The four-step process is designed to be repeated many times so that the scientific approach and evidence considered can be refined over time. Over time, the number of model structures may shrink as the OSVP process moves toward scientific consensus. To be sure, the OSVP process will not eliminate all the variation in results of value assessment since perspectives on value will vary and disagreements about relevant clinical evidence may persist. But the consensus-based approach will allows users to better understand legitimate and intrinsic reasons why value estimates vary.

## Contents of the IVI-RA model

Version 1.0 is IVI’s first release of the IVI-RA model, and is designed to provide a starting point for open debate. As such, the model is very flexible and allows users to choose from a large number of the plausible model structures supported by clinical practice and prior decision-analytic modeling research in RA. The IVI-RA model is a collaborative multistakeholder effort that produces tools to help decision-makers evaluate the value of pharmaceutical treatments for RA. To facilitate transparency, understanding, and debate among diverse stakeholders, the IVI-RA model consists of the following components:

* **Source code**: and C++ code for the model is available in our IVI GitHub [repository](https://github.com/InnovationValueInitiative/IVI-RA). Modelers and programmers may adapt the source code for their own purposes or collaborate with IVI to improve the code.
* **package**: The IVI-RA model is released as an package with documentation available [online](https://innovationvalueinitiative.github.io/IVI-RA/index.html). Researchers can use the package to run the IVI-RA model for custom analyses. Use of the package is recommended when peforming analyses for academic publication.
* **Model documentation**: This document provides provides technical details on the model structure, statistical methods for parameter estimation, and source data.
* **IVI-RA Model Interface**: For users not be well-versed in the programming language, we provide a web application for running the model online. The web application is designed for custom analyses and allows users full control over the treatments, patient population, model structures, parameter values, and simulation settings.
* **The IVI-RA Value Tool**: An important aim of the OSVP project is to obtain feedback from as many relevant stakeholders as possible. The IVI-RA Value Tool is a general audience web-application allowing those who are not experts in modeling, health economics, or RA to interact with the IVI-RA model.

## Intended use of the IVI-RA model

The IVI-RA model is not a value assessment framework but a model that simulates the costs, health outcomes, and risks associated with treatments for RA. It can therefore be used with any value framework preferred by the user. Currently, our online tools support both cost-effectiveness analysis (CEA) and multi-criteria decision-analysis (MCDA). IVI has also developed an R package, hesim, for health-economic simulation modeling and decision analysis that can be used to perform individualized CEA (Basu and Meltzer 2007; Ioannidis and Garber 2011; Espinoza et al. 2014) on simulation output from the IVI-RA model.

## About the IVI-RA model

### Overview

The IVI-RA model is a discrete-time individual patient simulation that simulates outcomes for individual patients. Model cycles are 6-months long, which is consistent with clinical trial evidence. The model simulates the progression of the health assessment questionnaire disability index (HAQ), a measure of functional status in RA.

Serious infection rates and changes in HAQ score during the first 6 months from baseline are based on clinical trial evidence. The change in HAQ can be modeled indirectly as a function of the American College of Rheumatology (ACR) response to treatment, the European League Against Rheumatism (EULAR) response to treatment, or directly as a function of the treatment. Patients switch treatment during the initial 6 months if they have a serious infection. Additionally, the user can chose whether treatment switching should be based on disease activity level or treatment response.

After the first 6 months on a new treatment, the HAQ score progresses over time at a rate based on observational data. Progression can either be assumed to be linear (F. Wolfe and Michaud 2010; Michaud, Wallenstein, and Wolfe 2011) or modeled using a non-linear mixture model (Norton et al. 2014).

Patients remain on treatment until treatment discontinuation or death. Time to treatment discontinuation is based on parametric survival analyses of real-world data. Seven possible distributions (exponential, Weibull, Gompertz, log-logistic, lognormal, and generalized gamma) can be chosen by the user. Male and female mortality is based on US lifetables and increases with the HAQ score at baseline and the change in the HAQ score from baseline.

Health care sector costs consist of drug acquisition and administration costs, hospital costs (which increase with the HAQ score), general management costs, and costs caused by serious infections. Non-health care sector costs are those due to lost wages.

Users wishing to calculate utility for CEA can map HAQ and individual characteristics to utility using the logistic regression algorithm of A. Wailoo et al. (2006) or the Hernández-Alava et al. (2013) mixture model. With both the A. Wailoo et al. (2006) and A. Wailoo et al. (2006) mappings, utility is calculated as a function of the HAQ and individual patient characteristic mapping, serious infections, and preferences for treatment attributes unrelated to safety and efficacy. QALYs combine life expectancy with per cycle utility.

### Patient preferences and heterogeneity

The IVI-RA model is desgned to capture differences in individual characteristics, preferences, circumstances, and response to treatment. First, progression of disease, mortality, and preferences for treatment vary according to individual characteristics. Second, although current evidence is scarce, users can adapt the model so that treatment effects vary across patients (e.g., as a function of patient characteristics or prognostic factors). Third, the IVI-RA model incorporates preferences for treatment attributes unrelated to safety and efficacy—such as mode of administration and the time a medication has been on the market—that are not typically included in decision-analytic models for value assessment.

### Uncertainty analysis

Since there will always be gaps in the available evidence and the appropriate scientific assumptions, it is important to quantify uncertainty. The IVI-RA model consequently contains 384 possible model structures, which can be used to quantify structural uncertainty or to evaluate the implications of different modeling assumptions. Parameter uncertainty is quantified using probabilistic sensitivity analysis (PSA).

We have found that model outcomes are especially sensitive to certain parameters and model structures, which highlights the importance of a flexible and consensus-based model. Primary sources of uncertainty include:

* The effect of treatment on the change in HAQ from baseline during the first 6 months of treatment
* The long-term progression of HAQ
* The reduction in treatment response after previous treatment failures
* The extent to which the HAQ score “rebounds” to its initial level after failing treatment
* Time on biologic treatment
* The relationship between HAQ and quality of life

### Real-world evidence

To ensure that simulated clinical and economic outcomes reflect outcomes in routine practice, we model “baseline event rates” (i.e., disease progression, mortality, time on treatment), patient preferences, and costs using real-world data. To minimize bias, relative treatment effects (i.e., differences in safety and efficacy across treatments) are, when possible, based on randomized clinical trials (RCTs), and then applied to the baseline event rates.

### Perspective of the decision-maker

Models should be flexible enough to meet the specific needs (e.g., a specific patient population) and perspectives (e.g., relevant sources of value) of different decision-makers. The current model is suitable for decision-makers making decisions for specific populations or subpopulations (e.g., policymakers, insurers, provider groups) but is not suitable for making predictions at the indiviudal level. Future iterations of the model may expand its use so that that it can be used for patients making resource allocation decisions (e.g., individualized cost-effectiveness analysis).

Cost components included in the model are based on the framework suggested by the Second Panel on Cost-Effectiveness in Health and Medicine (Sanders et al. 2016). Analyses based on a health care sector perspective can be performed by only incorporating health care sector costs. Analyses based on a (limited) societal perspective would include lost wages in addition to health care sector costs.

### Value to the healthy

Conventional value assessments focus on value to the sick, but recent research provides a framework for valuing technology for the healthy (i.e., “insurance value”) as well D. Lakdawalla, Malani, and Reif (2017). The IVI-RA model allows users to optionally incorporate insurance value, but we note that it is less well established than conventional approaches.

# Open-source consensus-based models for value assessment

The continuing increase in US health care costs has stimulated the introduction of initiatives to promote the use of high-value care. Decision-analytic models can be used to inform efficient use of health care resources, but are only relevant when deemed credible by different stakeholders, are representative of the local context and patient population, and can be easily updated without duplication of effort.

The nature of simulation modeling often leads to scientific disagreements and mistrust among decision-makers. Models are typically complex and difficult to understand. Even modeling experts may not be able to fully understand a model without public source code and detailed model documentation. Furthermore, efforts to make models accessible to non-experts are lacking. Models also become quickly outdated as new evidence arises or new scientific approaches are developed, which means that previous finding quickly become irrelevant to decision-makers.

The OSVP aims to increase understanding and relevance to diverse stakeholders by developing open-source consensus-based models. The hope is that these efforts can increase confidence in efforts to base reimbursement and policy decisions on value.

OSVP models are released and updated using a four step process:

1. Public release of the model.
2. Invite feedback and suggested improvements to the model in a public comment period.
3. A panel of experts determines which of the evidence-based suggestions for improvement suggested in Step 2 should be implemented by means of peer-review and a formal voting process.
4. Revise the model based on the feedback from the technical expert panel in Step 3.

To provide a starting point for debate, the initial release of each OSVP model (i.e., version 1.0) must be flexible and allow users to choose from a large number of plausible model structures and approaches based on clinical practice and previous modeling efforts. The four-step process is designed to be repeated many times so that the scientific approach and evidence considered can be refined over time.

# Overview of the IVI-RA model

## Why IVI is modeling rheumatoid arthritis

Treatment for rheumatoid arthritis (RA) is well suited for the OSVP approach for three reasons. First, modeling methods and assumptions vary considerably across existing simulation models (Brennan et al. 2003; Wailoo et al. 2008; Tosh et al. 2011; Carlson et al. 2015; Stephens et al. 2015; Athanasakis et al. 2015; M. Stevenson et al. 2016; Institute for Clinical and Economic Review 2017; M. D. Stevenson et al. 2017). Predicting disease progression is complex and there are a number of different measures of treatment response and morbidity (Madan et al. 2015). Analyses have, not surprisingly, been performed using different modeling approaches and have reached different conclusions about the cost-effectiveness of treatments for RA.

Second, RA is an area of significant innovation. There have been important advancements in the treatment of RA over the past decade, which suggests that there is an increasing need for tools to assess the cost-effectiveness of these treatments.

Third, not only have new treatments come to market recently, but evidence on existing RA treatments is growing rapidly. Thus, there is a strong need for models that can be updated in a straightforward manner as the evidence base evolves.

## Contents

To facilitate transparency, understanding, and debate among diverse stakeholders, the IVI-RA model consists of the following components:

* **Source code**: and C++ code for the model is available in our IVI GitHub [repository](https://github.com/InnovationValueInitiative/IVI-RA). Modelers and programmers may adapt the source code for their own purposes or collaborate with IVI to improve the code.
* **package**: The IVI-RA model is released as an package with documentation available [online](https://innovationvalueinitiative.github.io/IVI-RA/index.html). Researchers can use the package to run the IVI-RA model for custom analyses. Use of the package is recommended when peforming analyses for academic publication.
* **Model documentation**: This document provides provides technical details on the model structure, statistical methods for parameter estimation, and source data.
* **IVI-RA Model Interface**: For users not be well-versed in the programming language, we provide a web application for running the model online. The web application is designed for custom analyses and allows users full control over the treatments, patient population, model structures, parameter values, and simulation settings.
* **The IVI-RA Value Tool**: An important aim of the OSVP project is to obtain feedback from as many relevant stakeholders as possible. The IVI-RA Value Tool is a general audience web-application allowing those who are not experts in modeling, health economics, or RA to interact with the IVI-RA model.

These components along with the OSVP process are designed to encourage collaboration among stakeholders. Stakeholders may collaborate with IVI in at least two ways. First, they can provide feedback on any of the components during the public comment period. Second, programmers can make direct changes to the source code by making a “pull request” on GitHub. IVI will review the proposed changes. Code modifications that affect the scientific approach or evidence considered will only be incorporated after a review by the technical panel but other changes such as bug fixes or performance improvements may be immediately accepted.

## About

The IVI-RA model is a discrete-time individual patient simulation (IPS) with 6 month cycles that simulates patients one at a time. The model accounts for both parameter and structural uncertainty. Since the range of defensible scientific approaches is large, the IVI-RA model consists of 384 possible model structures. Structural uncertainty can be quantified by estimating cost-effectiveness across these different model structures and parameter uncertainty is quantified using probabilistic sensitivity analysis (PSA). (Note that the simulation was primarily written in C++ so that PSAs and analyses of structural uncertainty can be run in a reasonable amount of time.)

To ensure that simulated outcomes reflect outcomes in routine practice, we model “baseline event rates” (i.e., disease progression, mortality, time on treatment), patient preferences, and costs using real-world data. To minimize bias, relative treatment effects (i.e., differences in safety and efficacy across treatments) are, when possible, based on randomized clinical trials (RCTs), and then applied to the baseline event rates.

The IPS approach allows us to take an “individualized” modeling approach that captures both observable and unobservable patient heterogeneity. Disease progression, mortality, and preferences all vary across patients. In addition, although the evidence base is limited, users of the package can model treatment effects as a function of any combination of patient characteristics (e.g., demographics, prognostic factors). Finally, the model incorporates preferences for treatment attributes unrelated to safety and efficacy.

As recommended by the Second Panel on Cost-Effectiveness in Health and Medicine (Sanders et al. 2016), costs are simulated from both a health care sector perspective and a societal perspective. Productivity losses from lost earnings are included in the societal perspective but not the health care sector perspective. As discussed below (), our individualized approach implies that future iterations of the model could be tailored to fit the perspective of a patient or provider.

## Intended use

The model simulates the costs, health outcomes and risks associated with treatments for RA for each individual in a given population (see ). As described in users can model any sequence of biologic treatments and conventional disease-modifying antirheumatic drugs (cDMARDs).

The model can therefore be used for a number of purposes, conditional on the population of interest and the perspective of the decision maker. Here we describe a few possibilities.

The first and most obvious use of the model is for value assessment. Two approaches, cost-effectiveness analysis (CEA) and multi-criteria decision analysis (MCDA), are discussed in more detail in . Within the CEA approach, cost-effectiveness can be evaluated from the conventional perspective of a sick individual or from the perspective of a healthy individual using the “insurance value” framework developed by D. Lakdawalla, Malani, and Reif (2017).

Second, the model can be used to evaluate the consequences of clinical guidelines such as the current treat-to-target guidelines in the US (J. A. Singh et al. 2016) or guidelines based on treatment response like in the UK (Deighton et al. 2010). Unlike most previous models, our flexible framework allows treatment switching decisions to depend on disease activity level or treatment response, so outcomes under different decision rules can be simulated.

Third, although the model is currently desnged for population level decision-making, it could, in principle, be used to predict long-term health and economic consequences for patients. The predicted outcomes could, for example, be used to inform patient and providers decision making. For instance, Ioannidis and Garber (2011) argue that cost-effectiveness has relevance to patients spending their own money on health care services, particularly as out-of-pocket costs grow. Likewise, providers have a growing interest in cost-effectiveness models to demonstrate the value of their care whether through participation in Accountable Care Organizations (ACOs), to ensure coverage of medical interventions for their patients, or to reduce unwanted variability in management.

# Value assessment

The IVI-RA model simulates clinical and economic outcomes for each individual in a given population of interest. Outcomes can be simulated over a particular time horizon or over a lifetime.

Although simulation output can be used with any value assessment framework, IVI tools currently support two methodologies for decision analysis: CEA and MCDA. Cost-effectiveness results and MCDA value scores are automatically generated when users run IVI’s web-based user interfaces. In addition, IVI has developed an R package, hesim, for health-economic simulation modeling and decision analysis that can be used to perform individualized CEA (Basu and Meltzer 2007; Ioannidis and Garber 2011; Espinoza et al. 2014).

## Cost-effectiveness analysis

CEA is a well-established approach for value assessment grounded in economic theory and widely used in the scientific literature (A. H. Briggs, Claxton, and Sculpher 2006; Meltzer, Smith, and others 2011; Drummond et al. 2015). In general, CEA can be thought of as a methodology for maximizing health or well being subject to a resource constraint (Garber and Phelps 1997). The total value of a new health technology relative to a comparator is typically assessed using the incremental net monetary benefit (INMB),

where is a measure of the incremental health benefits from the new technology relative to the comparator, is a measure of the incremental cost of the new technology, and is the willingness to pay for a one-unit health gain. The new technology can be deemed cost-effective if the , or equivalently, in terms of the incremental cost-effectiveness ratio (ICER), if,

Incremental health benefits are typially measured in terms of health gains or patient well-being. Since treatments can affect both morbidity and mortality, CEAs typically use the quality-adjusted life-year (QALY). Since costs and benefits vary across patients, some researchers have argued for individualized CEA (Basu and Meltzer 2007; Ioannidis and Garber 2011; Espinoza et al. 2014) so that INMBs and ICERs are calculated separately for different subpopulations. It can be shown that if treatment response varies across the population, then making separate decisions in different populations will increase social welfare (Basu and Meltzer 2007).

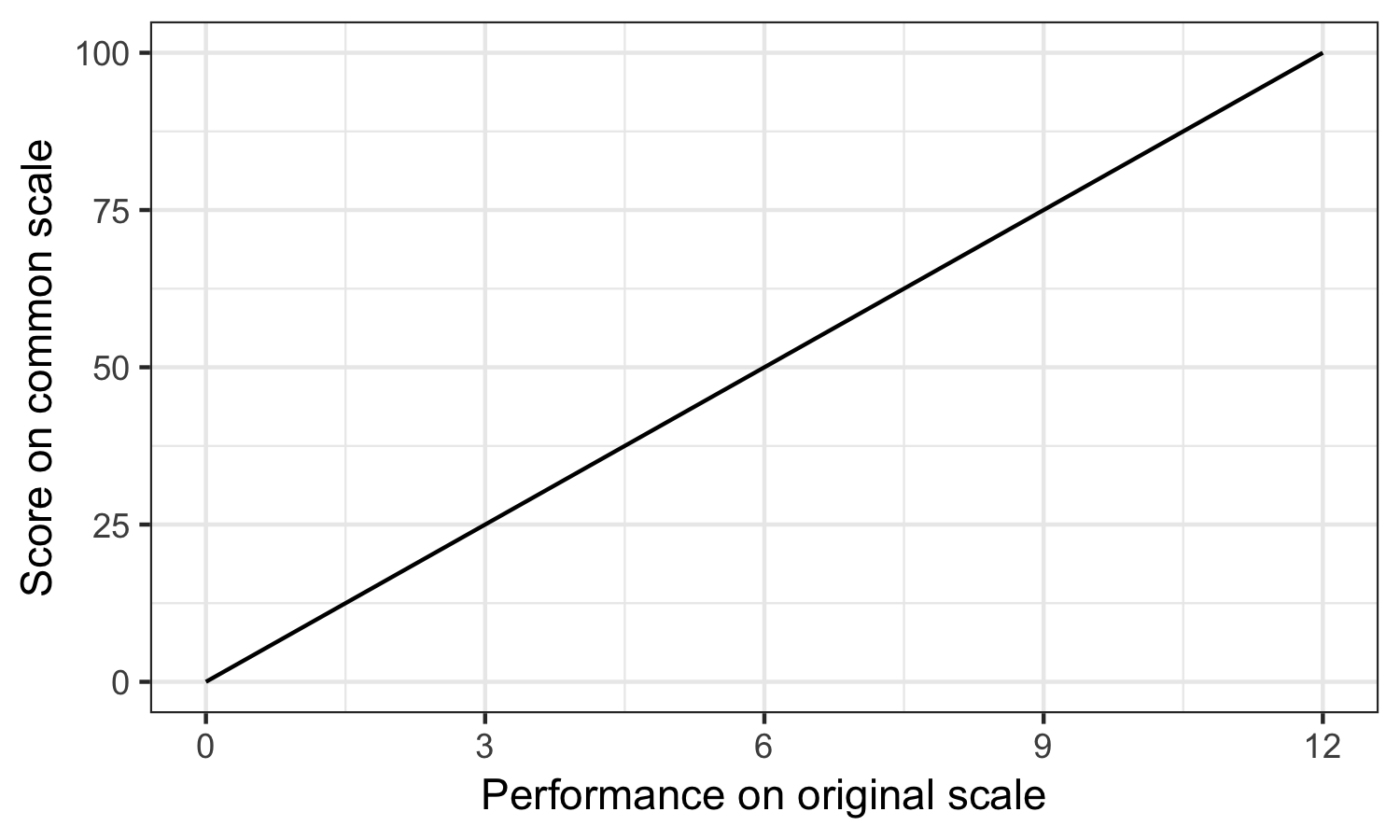
In practice, costs and health benefits are subject to statistical uncertainty. We quantify this uncertainty using probabilistic sensitivity analysis (PSA) and structural uncertainty analysis, which is described in more detail in . This approach allows us to generate standard measures of uncertainty in CEA including cost-effectiveness planes (Black 1990), cost-effectiveness acceptability curves (CEACs) (Van Hout et al. 1994; A. H. Briggs and others 1999; Fenwick, Claxton, and Sculpher 2001), and estimates of the expected value of perfect information (EVPI) (Fenwick, Claxton, and Sculpher 2001).

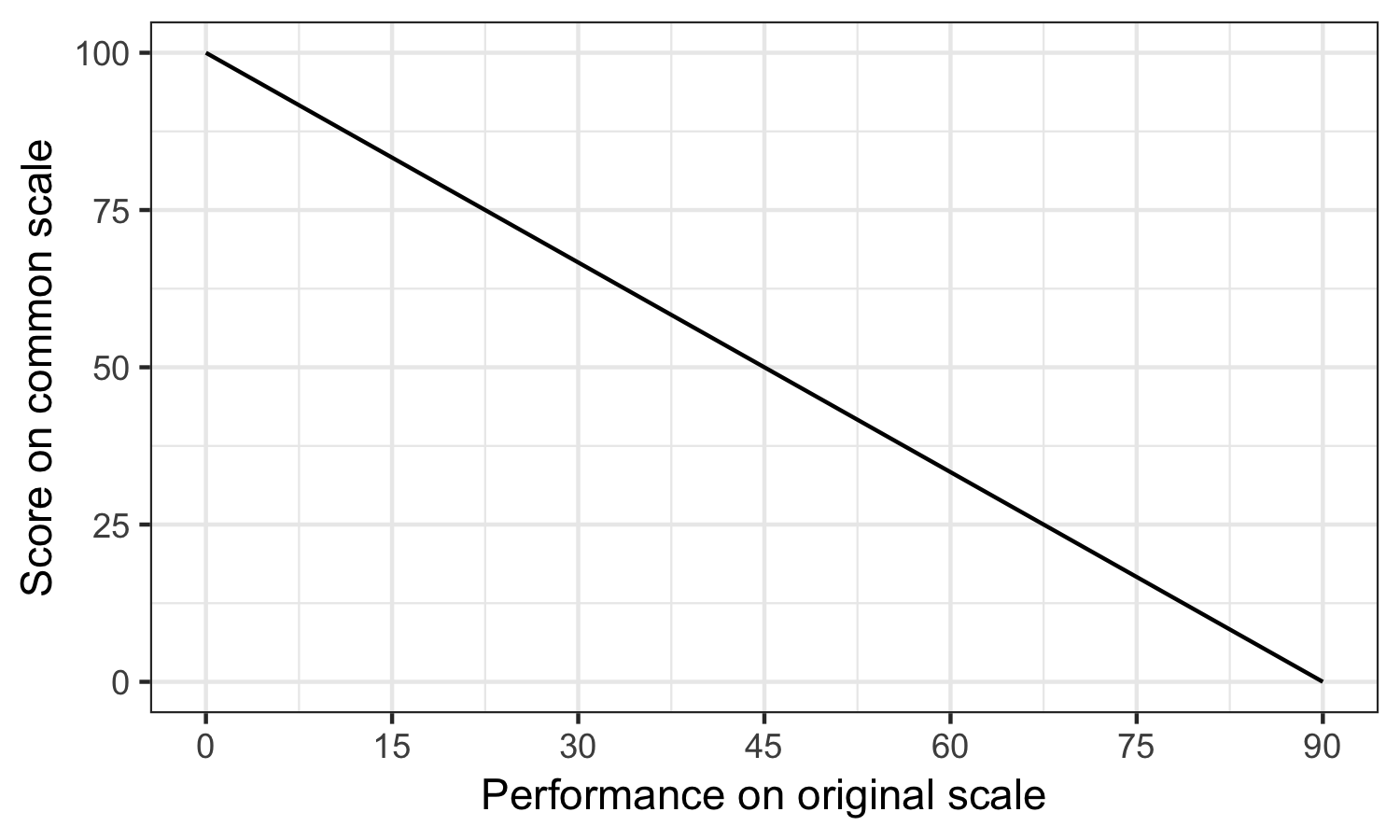
## Multi-criteria decision-analysis

An alternative approach to CEA is MCDA. Keeney and Raiffa (1993) define MCDA as “an extension of decision theory that covers any decision with multiple objectives. A methodology for appraising alternatives on individual, often conflicting criteria, and combining them into one overall appraisal...” We use a similar approach, which implies that separate criteria are aggregated into a single measure of value.

There are many approaches to MCDA; here, we discuss the approach used by IVI in the web-based user interface, which is based on the discussion in Thokala et al. (2016). First, decision-makers must select the relevent criteria for the analysis. These criteria are based on the costs, health outcomes, and risks simulated from the underlying health-economic model. We discuss the criteria relevant to the IVI-RA model in .

Since different criteria may be measured using different units, performance on each criterion is converted into a common scale, for instance, ranging from 0 to 100. There are a number of techniques for creating a common scale; we use a simple linear partial value function to translate scores, which assumes a linear relationship between performance on the original scale of a given criterion and the common scale. To illustrate, demonstrates two mappings between the original scale and the common scale.

.5 

.5 

Performance on the first criterion, shown in , ranges from 0 to 12 on the original scale, with higher scores denoting better performance. In contrast, performance on the second criterion, shown in , ranges from 0 to 90, with lower scores denoting better performance. The relationship between performance on the original scale and the score on the common scale is therefore positive for the first criterion and negative for the second criterion. In both cases, the relationship follows a straight line because we assume a linear relationship.

Each criterion is assigned points, say ranging from 0 to 10, by the decision maker, and weighted by dividing each criterion’s points by the sum of points across all criteria. For example, if there were 3 criteria and each criterion was given a score of 5, then each criterion would receive a weight of 1/3. If, on the other hand, the three criteria were given scores of 2.5, 5, and 7.5, then they would be given weights of .167, .33, and .5, respectively.

To aggregate results, we assume an additive model. In other words, the total score for a given treatment sequence is calculated by multiplying each criterion by the simulated standardized score and summing across criteria.

As with CEA, MCDA results are subject to statistical uncertainty. In our web applications, users choose a single model structure at a time, so uncertainty in MCDA outcomes is quantified using PSA. This produces a probability distribution around the simulated total score for each treatment sequence, which can be used to derive quantities of interest such as Bayesian credible intervals around the total score or the probability that each treatment sequence obtains a particular ranking among relevant treatment sequences.

# Broader concepts of value

Garrison, Kamal-Bahl, and Towse (2017) suggest five concepts of value that researchers should consider adding to the standard cost per QALY based CEA: (1) a reduction in uncertainty from a diagnostic test; (2) insurance value for healthy patients due to reduction against physical risk; (3) the value of hope for individuals who become risk-loving and would rather pay for a therapy with a long right survival tail than a therapy with a shorter right survival tail but an equivalent (or shorter) expected life-expectancy; (4) real option value when a therapy allows an individual to benefit from future medical innovations; and (5) scientific spillovers when the benefits of an innovation cannot be entirely appropriated by the innovator.

The concept that is arguably most salient to RA is insurance value, which focuses on valuing morbidity-reducing innovations and has the largest effects relative to conventional CEA on treatments for severe diseases where the burden of illness is the greatest. The IVI-RA model allows users to incorporate insurance value into their analyses, while noting that the approach is less well-established than conventional CEA.

Other concepts of value may be incorporated in the future, but likely in future disease areas. For example, real option value is most relevant for innovations that increase longevity and might be particularly well suited to analyses of treatments in oncology. Likewise, survey evidence for the value of hope is based on technologies that increase survival D. N. Lakdawalla et al. (2012) rather than those that affect morbidity. Reductions in uncertainty from diagnostic tests are clearly most relevant to diagnostics and scientific spillovers are most relevant to diseases with large externalities such infectious diseases.

D. Lakdawalla, Malani, and Reif (2017) provide a general mathematical framework for incorporating the effects of medical innovation on physical and financial risk. Conceptually, innovation can lower physical risk to healthy patients who might get sick in the future. New medical technologies act like “insurance policies” that protect a healthy person from all or part of the costs of falling ill. And while innovation certainly increases financial risk, this increase in financial risk can be mitigated by health care insurance.

The insurance value framework is an extension of the conventional CEA approach from the perspective of a healthy individual deriving utility from non-health consumption, and health, , according to . The individual is sick with probability and well with probability . Health when well is and health when sick is . Income is when well and when sick. The marginal utility of good in state is denoted by .

The value of a technology to a healthy consumer (with no health insurance), is derived implicitly by,

The marginal value of the technology, , can be shown to be,

where is the marginal value of a one unit health gain in dollar terms, is the marginal health gain from the technology, and is the marginal cost of the technology. The term is equivalent to the INMB in conventional CEA. The insurance value framework can therefore be implemented with knowledge of only two additional parameters beyond those in conventional CEA: the probability of illness, , and the marginal rate of substitution between the sick and the well states, .

The probability of illness can be estimated using incidence of disease in the population of interest (e.g., in the RA population). The second term, , is harder to estimate, but we allows users to specify it directly in our model and web-based user interfaces. Intuitively, this term reflects the amount of money the consumer would give up when healthy in exchange for gaining an additional dollar when sick. It rises when the consumer faces greater risks from illness.

It is worth emphasizing that insurance value is only larger than conventional value if the consumer is willing to give up more than $1 in the well state in exchange for an additional $1 in the sick state (i.e., ). This is likely to be true, because if the demand for health care insurance is positive, then .

The difference between the insurance value of a technology and its conventional value is even larger when individuals can purchase health insurance. For example, consider an actuarially fair insurance contract that pays the consumer I(p) when she falls sick. In this case, the insurance value of a health technology can be shown to be:

The term is the marginal payment made to the insuree per 1 dollar spent on health care. In the extreme case where there is no cost-sharing so that and . Here, health insurance completely eliminates spending risk the value of a technology is equal to its conventional value plus the value of physical risk reduction. More generally, and the value of a health technology with health insurance is equal to the sum of its conventional value, the insurance value absent health insurance, and the value of health insurance made possible by the technology.

# Populations

To run the IPS, a patient population must be specified. The model is designed for patients who are cDMARD experienced. The patient characteristics that must be included in the analysis are age, HAQ, gender, weight, the number of previous DMARDs, and disease activity. These variables are measured at the start of the simulation (i.e., model cycle 0).

Two default options for the patient population are available. First, a homogeneous cohort of men and women with gender-specific weights but otherwise identical characteristics can be used. Second, a heterogeneous cohort of patients with gender-specific weights but varying across all other characteristics can be specified. Other populations (i.e., for certain subgroups or based on registry data) can be used as well but are not prespecified in our package.

Our default population consists of individuals that, on average, have high disease activity. The proportion that is female, age, the number of previous DMARDs, baseline HAQ, and DAS28 are based on the values reported in Curtis et al. (2010). Mean values for the SDAI and CDAI are from the US301 clinical trial—which had a DAS28 score similar to the value from Curtis et al. (2010)—summarized in J. Smolen et al. (2003). Summaries of each variable are reported in . Details on the algorithm for simulating heterogeneous patients are described in .

@lcccc & & & &  
@inputtables/pat-inputs.txt

# Treatment strategies

Since patients typically use multiple treatments over a lifetime, the model is capable of simulating a treatment sequence of any arbitrary length. Treatments that can be included in a sequence include conventional disease-modifying anti-rheumatic drugs (cDMARDs), biologic DMARDs (bDMARDs), and Janus kinase/STAT (JAK/STAT) pathway inhibitors. The bDMARDs and JAK/STAT inhibitors, which we refer to collectively as targeted DMARDs (tDMARDs), included in the current version of the model are:

* **Tumor necrosis factor (TNF) inhibitors**: etanercept, adalimumab, infliximab, certolizumab, golimumab
* **non-TNF inhibitors**: abatecept, tocilizumab, rituximab
* **Janus kinase/signal transducers and activators of transcription (JAK/STAT) inhibitors**: tofacitinib

At the end of a sequence, patient switch to non-biologic therapy (NBT), which encompasses a range of therapies that clinicians may feel is appropriate for all patients such as methotrexate and sulfasalazine (M. Stevenson et al. 2016; M. D. Stevenson et al. 2017).

# Competing model structures

The IVI-RA model is a discrete-time IPS with 6 month cycles that can be run using a number of different model structures. Like most decision-analytic models in RA, version 1 of the model measures changes in disease severity using the Health Assessment Questionnaire (HAQ) Disability Index score (Brennan et al. 2003; Wailoo et al. 2008; Tosh et al. 2011; Carlson et al. 2015; Stephens et al. 2015; Athanasakis et al. 2015; M. Stevenson et al. 2016; Institute for Clinical and Economic Review 2017; M. D. Stevenson et al. 2017). At the start of the simulation, each patient is assigned a baseline HAQ score. Subsequently, the impact of the disease measured by the HAQ trajectory over time is modeled as a function of a sequence of treatments (). In the absence of treatment, HAQ deteriorates at a certain rate as depicted by the dashed line in the figure. For each treatment in a treatment sequence, treatment is separated into two distinct phases: an initial phase of up to 6 months, consistent with data reported from randomized controlled trials (RCTs), and a maintenance phase thereafter until discontinuation.

## Initial treatment phase

During the initial treatment phase HAQ is modeled as a change from baseline.

* **H1**: Treatment ACR HAQ
* **H2**: Treatment ACR EULAR HAQ
* **H3**: Treatment HAQ

In **H1**, treatment influences HAQ through its effect on the American College of Rheumatology (ACR) response criteria, which is similar to the structure used in US based cost-effectiveness models (e.g. Carlson et al. 2015; Institute for Clinical and Economic Review 2017). ACR 20/50/70 response is defined as at least a 20/50/70% improvement. In the simulation, we convert these overlapping ACR categories to four mutually exclusive categories: no response (defined as less than 20% improvement), ACR 20% to <50% improvement, ACR 50% to <70% improvement, and ACR 70% improvement or greater. The rationale for using ACR response rather than HAQ directly is that the evidence base relating treatment to ACR response is larger than the evidence based relating treatment to HAQ. **H2** follows the National Institute for Health and Care Excellence (NICE) cost-effectiveness model (M. Stevenson et al. 2016; M. D. Stevenson et al. 2017) and models the effect of treatment on HAQ indirectly through its effect on ACR response and, in turn, the three categories of the European League Against Rheumatism (EULAR) response (no response, moderate response, or good response). Finally, since modeling the effect of treatment on HAQ through intermediary variables may mediate treatment response, in **H3**, treatment impacts HAQ directly.

Treatment switching during the initial treatment phase is modeled using 6 different pathways **S1-S6**.

* **S1:** Treatment ACR Switch
* **S2:** Treatment ACR DAS28 DAS28 Switch
* **S3:** Treatment ACR SDAI SDAI Switch
* **S4:** Treatment ACR CDAI CDAI Switch
* **S5:** Treatment DAS28 DAS28 Switch
* **S6:** Treatment ACR EULAR Switch

**S1** follows a common approach where ACR non-responders discontinue treatment (e.g. Carlson et al. 2015; Institute for Clinical and Economic Review 2017). One drawback of this approach is that it is not consistent with current treat-to-target guidelines in the United States (J. A. Singh et al. 2016). In **S2-S5**, treatment switching consequently depends on disease activity (remission, low, moderate, high) (Anderson et al. 2012). In **S2-S4**, ACR response predicts the change in disease activity from baseline, which along with baseline disease activity, predicts absolute disease activity. Patients with moderate or high disease switch treatment while patients with low disease activity or in remission continue treatment. Disease activity is measured using either the Disease Activity Score with 28-joint counts (DAS28) (Prevoo et al. 1995), Simplified Disease Activity Index (SDAI) (J. Smolen et al. 2003; D Aletaha and Smolen 2005), or the Clinical Disease Activity Index (CDAI) (Daniel Aletaha et al. 2005).

**S5** is similar to **S2-S4**, but models the effect of treatment on changes in DAS28 directly, rather than indirectly through ACR response. We also aimed to model the direct effect of treatment on SDAI and CDAI, but sufficient clinical trial data are not available. Finally, since in the UK, the British Society for Rheumatology and the British Health Professionals in Rheumatology recommends using the EULAR response (Deighton et al. 2010), treatment switching in **S6** depends on EULAR response. In particular, following the NICE model, we assume that EULAR non-responders discontinue treatment while moderate and good responders continue treatment (M. Stevenson et al. 2016). The reasoning is that rules stipulated by NICE require a DAS28 improvement of more than 1.2 to continue treatment which is associated with moderate or good EULAR response.

Not all pathways **S1-S6** can be used with each of **H1-H3**. If **H1** is used, then **S1-S5** are available, but **S6** is not because EULAR response is not simulated. In **H2**, **S1-S6** are all available while in **H3** only **S5** can be used since ACR response is not simulated. The 12 possible combinations are outlined in .

@lcccccc & & & & & &  
H1 & 1 & 2 & 3 & 4 & 5 & -  
H2 & 6 & 7 & 8 & 9 & 10 & 11  
H3 & - & - & - & - & 12 & -

Notes: Rows denote the pathway used to relate treatment to HAQ and columns denote the pathway used to determine treatment switching. Each number denotes a unique combination of pathways (i.e., 1 corresponds to H1 and S1, and 8 corresponds to H2 and S3) and the “-” denotes a combination of pathways that is not possible. There are 12 possible model structures for the initial treatment phase.

## Maintenance phase

In the maintenance phase, the long-term progression of HAQ can be modeled in two ways. First, as is common in cost-effectiveness analyses (CEAs) of therapies for RA, HAQ is assumed to progress at a constant linear rate over time (see Tosh et al. 2011; Wailoo et al. 2008). However, since emerging evidence suggests that the rate of HAQ progression is non-linear and varies across patients (L. Gibson, Alava, and Wailoo 2015), our second scenario simulates HAQ progression using a latent class growth model (LCGM) (Norton et al. 2014) with 4 distinct HAQ trajectories and a rate of HAQ progression that decreases over time within each trajectory. Upon discontinuation of treatment, the HAQ score rebounds by a proportion of the improvement experienced at the end of the initial 6-month period with that treatment.

The duration of the maintenance phase (i.e., time to discontinuation of maintenance treatment) is simulated using parametric time-to-event distributions. When **S1** is used, time to treatment discontinuation is simulated using a single time-to-event curve because we have been unable to obtain curves stratified by ACR response categories. In contrast, when **S2-S5** are selected, the time-to-event curves are a function of disease activity level so patients with lower disease activity at the end of the initial treatment phase stay on treatment longer, on average. Likewise, when structure **S6** is used, the time-to-event distributions are stratified by EULAR response category and patients with good response at the end of the initial treatment phase tend to stay on treatment longer than patients with a moderate response. In each case, time to discontinuation can be simulated using one of 7 possible distributions (exponential, Weibull, Gompertz, gamma, log-logistic, lognormal, generalized gamma).

## Adverse events

In line with M. Stevenson et al. (2016) the adverse events included in the model are limited to serious infections; we assume that only serious infections have a significant cost impact and increased risk over background rates to be meaningful to include (Ramiro et al. 2017). During the initial treatment phase, a patient immediately stops treatment if a serious infection occurs; during the maintenance phase, time on treatment depends on the sampled time to treatment discontinuation and a patient experiences a serious infection if the individual’s sampled time to the adverse event is shorter than the sampled time to treatment discontinuation.

## Mortality

Baseline HAQ scores (and changes in HAQ scores from baseline) are used to determine mortality relative to age/sex specific rates for the US general population (assumed to have a HAQ score of 0). Treatment, therefore, has an indirect effect on mortality through its effect on HAQ.

## Utility

Individual HAQ scores at a particular point in time were also used to simulate EuroQol five dimensions questionnaire (EQ-5D) utility scores (0-1 range), which, in turn, are used to simulate quality-adjusted life-years (QALYs). However, since a number of different methods have been used to convert HAQ into utility, our model contains two different possible mapping algorithms. Our preferred algorithm is the Hernández-Alava et al. (2013) mixture model, which uses a much larger sample size than other statistical models and has been shown to have better predictive accuracy. Other algorithms are typically estimated using clinical trial data (e.g. Carlson et al. 2015; Stephens et al. 2015) and consequently have limited generalizability. The second utility algorithm available within our model is based on a linear regression analysis of real-world data by A. Wailoo et al. (2006) that has been used in a few previous CEAs (e.g. Wailoo et al. 2008; Institute for Clinical and Economic Review 2017).

## Costs

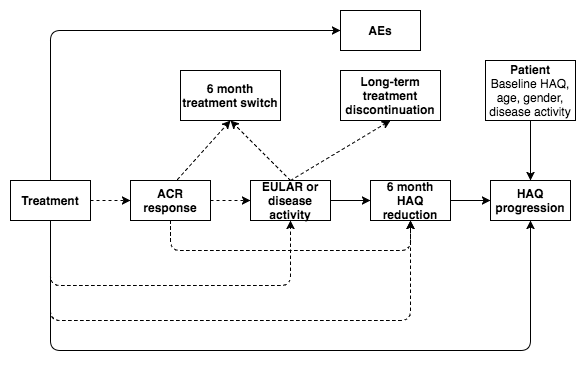
Annual hospitalization days and productivity losses are simulated as a function of HAQ. Health sector costs considered in the models are related to drug acquisition and administration, adverse events, general management of RA, and hospitalization. Non-health sector costs are limited to work-related productivity loss.

## Summary of simulation

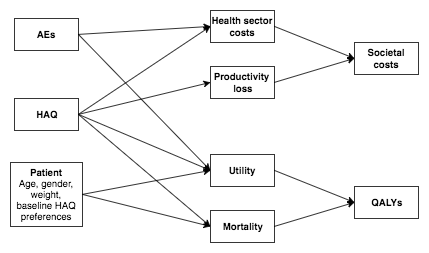
The flow diagram in describes the flow of a single patient through the simulation. The simulation runs for a patient’s entire lifespan beginning with treatment initiation and ending in death. The rectangles in the figure represent “processes” determining the effect of treatment on disease progression and the diamonds represent “decisions” that determine whether a patient will switch to a new treatment.

Notes: Rectangles represent “processes” determining the effect of treatment on disease progression, Diamonds represent “decisions” that determine whether a patient will switch to a new treatment. Dotted lines denote start of a new treatment or the end of the simulation.

The influence diagram in summarizes the assumed relationships among different variables in the model. Each arrow represents the direct effect of one parameter on another. Dashed lines represent relationships that depend on the structural assumptions used. focuses on the effect of treatment on disease progression and adverse events while looks at the relationships between the health and cost outcome variables.



Influence diagram outlining structural relationships



Influence diagram outlining structural relationships

Notes: ACR: American College of Rheumatology; EULAR: European League Against Rheumatism; HAQ: Health Assessment Questionnaire: AEs: adverse events; QALYs: quality-adjusted life-years; WTP: willingness to pay. Disease activity refers to the Disease Activity Score with 28-joint counts (DAS28), the Simplified Disease Activity Index (SDAI), or the Clinical Disease Activity Index (CDAI).

The model accounts for patient heterogeneity in two ways. First, baseline event rates vary across patients by both observable and unobservable factors. For example, long-term HAQ progression, mortality, and utility depend on patient specific variables including age, gender, and baseline disease level. Moreover, unobserved differences in long-term HAQ progression and utility across patients are modeled using mixture models. Second, relative treatment effects for ACR response, the change in HAQ at 6 months, and the change in DAS28 at 6 months, can be modeled as a function of explanatory variables in the package.

## Model outcomes

### Benefits, costs, and risks

The model simulates the health outcomes, costs, and risks associated with treatment. Depending on the model structure, model outcomes include the following:

* **Clinical outcomes during initial treatment phase**: ACR response, EULAR response, DAS28, SDAI, CDAI
* **Long-term clinical outcomes**: HAQ, QALYs
* **Adverse events**: number of serious infections
* **Health care sector costs**: drug acquisition and administration costs, general management and monitoring costs, adverse event costs, hospitalization costs
* **Non-health care sector costs**: productivity losses

### Outcomes for value assessment

If CEA is used for value assessment, then the value of treatment is estimated using the NMB, as described in . CEA from a societal perspective would include productivity losses while analyses from a health care sector perspective would not.

Any combination of simulated model outcomes can be used for MCDA. In IVI’s web interfaces, the MCDA is currently based on the following criteria: (i) QALYs, (ii) total health care sector costs, (iii) productivity losses, (iv) number of serious infections, (v) route of administration (oral/injection/infusion) and (vi) time the medication has been on the market. We measure performance for each route of administration by calculating the percentage of total life-years that were spent using that particular route of administration. If a combination therapy is used during the treatment sequence, we allocate time equally among all routes of administration within the combination therapy (i.e., during a time period in which tofacitinib citrate is used with methotrexate, we allocate half of the time to oral admnistration and half to admnistration by injection). Performance on the time since the medication has been on the market criterion is a weighted averaged of time since FDA approval for each treatment in a treatment sequence, where weights are equal to the number of life-years spent using a particular treatment within the sequence. In the web interfaces users can input their own weights for each of the criteria, but it is important to note that we have not conducted the surveys required to elicit weights in a representative sample of patients.

When analyzing value to healthy individuals—rather than sick patients—we use the framework described in . Following D. Lakdawalla, Malani, and Reif (2015) we calculate annual value for patients (e.g., benefits to an insurance enrollee during a plan year) by annualizing lifetime health gains (i.e., QALYs) and costs (see for more details). To calculate the conventional value of a treatment to a healthy individual (i.e., from ), we estimate using annualized incremental QALYs, using annualized incremental costs, using willingness to pay thresholds, and as the probability of obtaining RA within the next year.

# Source data and parameter estimation

## Treatment effects at 6 months

The effect of treatment on ACR response, DAS28, and HAQ at 6 months for tDMARD naive patients are estimated using Bayesian network meta-analyses (NMA) of published randomized controlled trials (RCTs). Primary outcomes were ACR response, change in DAS28 from baseline at 6 months, and the change in HAQ from baseline at 6 months. Results from the NMA are shown in . Details of the systematic literature review and the statistical methodology are provided in the Appendix ().

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| (lr)2-4 |  |  |  |  |  |
| @inputtables/nma-acr-das28-haq.txt |  |  |  |  |  |

Notes: ACR20/50/70 categories are the probability of at least a 20/50/70% improvement. 95% credible intervals are in parentheses. Estimates are based on 1,000 random draws of the NMA parameters. DAS28 and HAQ are changes in the DAS28 and HAQ score from their baseline scores respectively; negative numbers denote reductions in baseline values. cDMARDs = conventional disease-modifying antirheumatic drugs; MTX = methotrexate; ABT IV = abatacept intravenous; ADA = adalimumab; ETN = etanercept; GOL = golimumab; IFX = infliximab; TCZ = tocilizumab; CZP = certolizumab pegol; ABT SC = abatacept subcutaneous; RTX = rituximab; TOF = tofacitinib. ACR = American College of Rheumatology.

Its important to note that treatment effects for each tDMARD were estimated relative to cDMARDs and then applied to the average response for patients using cDMARDs. A limitation of our current approach is that that the average response for patients using cDMARDs is estimated using data from the clinical trials include in the NMA, and may not reflect outcomes seen in routine practice. Future versions of the model could consider using real-world data instead of clinicial trial evidence to estimate this average response.

Given that there is limited evidence that treatment effects vary across patients in the published literature, treatment response at 6 months for a given treatment does not vary according to patient characteristics. Nonetheless, in our package, treatment effects for each simulated patient can be modeled as a function of any variables chosen by the user. Our approach to modeling treatment effect heterogeneity is described in .

Treatment effects for tDMARD experienced patients are reduced by multiplying treatment effects for tDMARD naive patients by a constant . Based on evidence reported in Carlson et al. (2015), we assume that is uniformly distributed and ranges between .75 and .92, implying that (rounding up) the average value of is .84. In other words, reductions in DAS28 and HAQ scores for tDMARD experienced patients are, on average, 84% of the reduction in DAS28 and HAQ scores for tDMARD naive patients, and an ACR response of 60/40/20 for tDMARD naive patients would, on average, be reduced to 50/33/16 for tDMARD experienced patients.

In the simulation, treatment response depends on the line of therapy and whether a patient is tDMARD naive or tDMARD experienced at baseline. For tDMARD naive patients, first line treatment response is based on the NMA results for tDMARD naive patients while response for all other treatments in a treatment sequence is reduced using the constant . For tDMARD experienced patients, treatment response is reduced using at each line of therapy including the first line. One limitation of this approach is that we are unable to model the relationship between line of therapy and ; that is, treatment response for a patient who has failed at least one biologic is assumed to be reduced by, on average, .84, regardless of line of therapy.

## Treatment switching at 6 months

The data required to determine treatment switching at 6 months depends on the selected model structure. If **S1** is selected, then treatment switching depends on the simulated ACR response; likewise, if **S5** is selected, then treatment switching depends on the simulated level of DAS28 at 6 months. When **S2-S4** are used, treatment switching is determined by the relationship between ACR response and the change in disease activity, and in **S6**, switching is based on the relationship between ACR response and EULAR response. Details of the mapping between ACR response and change in disease activity and between ACR response and EULAR response are provided below.

### ACR response and change in disease activity

There are currently no established mappings between mutually exclusive ACR response categories and DAS28, SDAI, or CDAI (Madan et al. 2015). However, D Aletaha and Smolen (2005) provides evidence on the relationship between overlapping ACR response categories (ACR 20/50/70) and mean changes in each of the three disease activity measures. Results are reported for three cohorts—the Leflunomide datasets, the inception cohort, and the routine cohort—with 1,839, 91, and 279 patients, respectively. We transformed mean changes by overlapping ACR response categories to mean changes by mutually exclusive ACR response categories by using the number of patients in each mutually exclusive ACR response category as described in . J. Smolen et al. (2003) provided the number of patients in each ACR response category in the Leflunomide dataset and Daniel Aletaha et al. (2005) provided the number of patients in the inception cohort. Mean changes in disease activity in each mutually exclusive ACR response category are shown in . However, note that the referenced publications did not report mean outcomes, so we were unable to generate standard errors for the estimates. We consequently assume allow the estimates to vary by 20% in either direction.

@lcccc &  
& &  
(lr)2-2 (lr)3-5 & & & &  
@inputtables/acr2da.txt

We did not include estimates from the routine cohort for two reasons. First, we were unable to find information on the number of patients in each ACR response category. Second, patients in the routine cohort had considerably lower disease activity levels (D Aletaha and Smolen 2005; Daniel Aletaha et al. 2005) and our default population (see ) consists of patients with high disease activity at baseline. Mean DAS28 in the inception cohort and routine cohort were 5.62 and 4.09, respectively, while the mean DAS 28 ranged from 6.3 to 7 across the clinical trials making up the Leflunomide dataset.

### ACR response and change in EULAR response

ACR responses were translated into EULAR response probabilities based on evidence of their relationship reported in M. Stevenson et al. (2016) and obtained from the US Veterans Affairs Rheumatoid Arthritis (VARA) registry ().

@lcccc &  
& & &  
@inputtables/acr2eular.txt

Notes: The VARA registry is a multicentre, US database of veterans age 19 and older. Each cell represents the number of patients in the database in a given category.

## Change in HAQ at 6 months

In model structures including **H1**, the impact of treatment on changes in HAQ at 6 months is modeled by first estimating the effect of treatment on ACR response and then mapping ACR response to a change in HAQ. As in Institute for Clinical and Economic Review (2017), ACR responses from the NMA were translated into HAQ scores based on evidence from the adalimumab monotherapy for treatment of rheumatoid arthritis (ADACTA) trial reported in Carlson et al. (2015) ().

@lcc &  
& &  
@inputtables/acr2haq.txt

Source: Carlson et al. (2015)

The relationship between EULAR response and HAQ is based on analyses conducted by M. Stevenson et al. (2016) using the BSRBR database. Their analysis is based on predictions from a mixture model with covariates set to sample means. Moderate and good EULAR responses are associated with -0.317 (SE = 0.048) and -0.672 (SE = 0.112) changes in HAQ scores respectively ().

@lcc & &  
@inputtables/eular2haq.txt

compares the impact of treatment on HAQ when using **H1-H3**. Results were estimated by simulating patients for 6 months and randomly sampling parameter sets. For each randomly sampled parameter set, we calculated the average decrease in HAQ at 6 months across the patients. Estimates reported in the table are the mean and 95% credible interval of the mean decrease in HAQ at 6 months. To maintain consistency across **H1-H3**, we did not allow HAQ scores for patients who might have otherwise switched treatments accoring to **S1-S6** to rebound back to their baseline levels (i.e., levels at the start of the simulation) at the end of the 6 month period.

@lccc & & &  
@inputtables/sim-dhaq-6months.txt

Notes: **H1**, **H2**, and **H3** are the Treatment ACR HAQ, Treatment ACR EULAR HAQ, and Treatment HAQ pathways respectively. 95% credible intervals are in parentheses. Estimates are based on 6-month simulations of 1,000 patients and 1,000 parameters sets for each therapy. HAQ denotes a change in the HAQ score at 6 months from baseline; a negative value indicates a reduction in the HAQ score. Mean HAQ is calculated for each parameter set by averaging across the 1,000 patients. cDMARDs = conventional disease-modifying antirheumatic drugs; MTX = methotrexate; ABT IV = abatacept intravenous; ADA = adalimumab; ETN = etanercept; GOL = golimumab; IFX = infliximab; TCZ = tocilizumab; CZP = certolizumab pegol; ABT SC = abatacept subcutaneous; RTX = rituximab; TOF = tofacitinib. ACR = American College of Rheumatology.

Estimates for **H1** and **H3** are generally similar but treatment response is considerably smaller when using **H2**. This suggests that the additional mapping between ACR response and EULAR response attenuates treatment response. Given these varying estimates of the change in HAQ during the initial treatment phase and the impact of HAQ on other important outcomes within the model including utility and health care costs, the choice of **H1-H3** (and in particular **H2** vs. **H1/H3**) appears to have important consequences for value assessment.

## HAQ progression in the absence of tDMARD treatment

The natural course of HAQ progression in the absence of tDMARDs develops over time according to an estimated natural course for patients remaining on cDMARDs or following discontinuation of the last tDMARD of the sequence (i.e., on NBT). The natural course of HAQ can either be assumed to change at a constant linear rate or be modeled using a LCGM that accounts for non-linear progression and heterogeneity across patients.

### Constant linear rate of progression

The rate of progression in the linear case is based on the observational study by F. Wolfe and Michaud (2010). They assessed the development of HAQ over time at six month intervals for up to 11 years among 3,829 RA patients who switched from non-biologic treatment to biologic treatment and participated in the National Data Bank for Rheumatic Diseases (NDB) longitudinal study of RA outcomes. The annual HAQ progression rate prior to biologic therapy was 0.031 (95% confidence interval (95%CI): 0.026 to 0.036) and is assumed to reflect the course of progression of HAQ in the absence of tDMARDs.

Based on the same data, Michaud, Wallenstein, and Wolfe (2011) reported overall and age-specific specific HAQ progression rates. The differences between the overall and age specific rates are as follows: <40: -0.020 (95%CI: -0.0223 to -0.0177); 40-64: -0.008 (95%CI: -0.0101 to -0.0059); 65 0.017 (95%CI: 0.0136 to 0.0204). These estimates are applied to the overall progression rate of 0.031 to obtain age specific HAQ progression rates (see ).

@lrrrl & &  
& & & &  
Overall progression rate  
@inputtables/haq-lprog-cdmards.txt Change in overall progression rate by age  
@inputtables/haq-lprog-diff-byage.txt

Notes: 95% confidence intervals are calculated using a normal distribution. Confidence intervals for changes in HAQ progression rates by age assume no covariance between the overall progression rate and the age-specific rates reported by Michaud, Wallenstein, and Wolfe (2011).

### Latent class growth model

We also model the rate of HAQ progression in the absence of tDMARDs using a mixture model approach that has increasingly been used to model HAQ progression over time (M. Stevenson et al. 2016; Norton et al. 2013; Norton et al. 2014). These models suggest that different subgroups have distinct HAQ trajectories and that the rate of worsening of HAQ progression decreases over time. We use the LCGM estimated by Norton et al. (2014) and since we aim to model trajectories for cDMARDs and NBTs we chose the specification based on data from the Early Rheumatoid Arthritis Cohort Study (ERAS) cohort, which has a high percentage of patients receiving methotrexate and a very small percentage receiving biologics. Complete details of the LCGM are provided in .

The Norton et al. (2014) LCGM determined that there are four classes of patients and thus four distinct HAQ trajectories. The probability of class membership depends on 7 variables: age, gender, DAS28, disease duration, rheumatoid factor, the ACR 1987 criteria for RA, and a measure of socioeconomic status. Age, gender, and the DAS28 are relevant to the way the population is defined within our model (see ) and are therefore important determinants of the HAQ trajectory. Other variables (disease duration, rheumatoid factor, ACR criteria, and socioeconomic status) are not defined within our population. We consequently set disease duration (8.2 months), rheumatoid factor (0.73), and the socioeconomic status variable (0.49) equal to their mean values with the ERAS cohort. The ACR criteria was set to 1.

HAQ trajectories (in levels) by class are shown . The dotted lines plot observed mean values. There are clear distinguishable classes as both the level of the HAQ score and its slope vary between groups. Norton et al. (2014) refer to the groups as “low”, “moderate”, “high”, and “severe” groups, in order from the lowest to highest HAQ scores. The observed trends for the low, medium, and high groups follow a J-shaped pattern with a sharp drop following treatment initiation and an upward slope thereafter, while the severe group experiences persistently high HAQ scores. Since our model separates the initial treatment phase from the maintenance phase, we are only concerned with HAQ progression following the initial drop. As in M. Stevenson et al. (2016), we consequently only predict values from year 2 onward. The fitted values are the solid upward sloping lines in the plot.

Notes: The first three data points corresponds to years 0, 0.5, and 1, respectively; all other data points are spaced 1 year apart.

An important question for modeling disease progression in RA is how the rate of progression within each class in the LCGM compares to a constant linear trajectory. We examine this question in , which compares yearly rates of changes in HAQ using the LCGM and with constant annual rates of change (0.031 per year) based on the F. Wolfe and Michaud (2010) analysis. The LCGM was simulated over 30 years and differences between year and year were used to assess changes in HAQ score from one year to the next.

In the moderate, high, and severe groups the rate of HAQ progression is higher initially in the LCGM than in the F. Wolfe and Michaud (2010) analysis; however, the LCGM modeled rate of HAQ progression declines over time and eventually begins to approach zero. In the low group, HAQ increases at a rate less than 0.031 per year and the rate of increase declines over time.

## HAQ trajectory with tDMARD maintenance treatment

Based on the NDB longitudinal study, F. Wolfe and Michaud (2010) estimated the overall annual HAQ progression rate among RA patients who had switched to biologic treatment at -0.001 (95CI: -0.004 to 0.002). In a separate analysis, also based on NDB data, Michaud, Wallenstein, and Wolfe (2011) reported annual HAQ progression rates by treatment adjusted for baseline HAQ score, age, sex, education, smoking, BMI, comorbidity, and RA onset. The average HAQ rate among patients on a biologic was -0.001 as well, which instills confidence that the reported HAQ progression rates for different biologics as reported by Michaud, Wallenstein, and Wolfe (2011) can be directly compared with the overall annual HAQ progression rate of 0.031 reported by F. Wolfe and Michaud (2010). Accordingly, biologic specific HAQ progression rates by Michaud, Wallenstein, and Wolfe (2011) are used in the model. For tDMARD treatments evaluated in the model for which no HAQ progression rate was reported by Michaud, Wallenstein, and Wolfe (2011), the overall biologic rate of -0.001 is used.

## Duration of maintenance treatment

Time to treatment discontinuation in the maintenance phase depends on the pathway (**S1-S6**) used to model treatment switching. If **S1** is selected, a single treatment discontinuation curve based on an analysis from the CORRONA database is used for all patients. In **S2-S5**, time to treatment discontinuation is stratified by the level of disease activity, and in **S6** treatment duration depends on EULAR response.

### Treatment duration in the US

We based our estimates of treatment duration during the maintenance phase for patients in the US with analyses of the CORRONA database (Strand et al. 2013). The analysis sample consisted of 6,209 patients age 18 or older treated between 2002 and 2011 receiving either TNF inhibitors or other bDMARDs. The mean age was 57.6 years, 43% of patients were biologic naive, the mean CDAI was 16, and just over 26% of patients had high disease activity (CDAI 22).

7 parametric survival models (exponential, Weibull, Gompertz, gamma, log-logistic, lognormal, and generalized gamma) were estimated on individual patient data reconstructed from a Kaplan-Meier curve from the CORRONA analysis using the algorithm developed in Guyot et al. (2012). We compared fit using the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). The generalized gamma had the lowest AIC and BIC, so we consider it to be the preferred model. A plot of of the generalized gamma distribution against the Kaplan-Meier curve is shown in . As can be seen in the plot, the shape of the survival curve estimated using a generalized gamma distribution tracks the Kaplan-Meier curve closely.

@lcccc & &  
@inputtables/ttd-corrona-ipdsurv-ic.txt

We considered estimating separate time to discontinuation curves for each treatment, but did not for a number of the reasons cited in M. Stevenson et al. (2016). The majority of the literature focuses on anti-TNFs (e.g., infliximab, etanercept, and adalimumab) (e.g. Gomez-Reino and Carmona 2006; Yazici et al. 2009; Pan et al. 2009), which makes it difficult to estimate discontinuation curves for the other treatments. Furthermore, studies comparing rates of discontinuation across treatments tend to be observational because clinical trials are of short duration and do not reflect real-world patient populations. However, although observational studies provide accurate predictions on time to discontinuation, it is difficult to avoid bias from confounding when estimating differences across treatments because patients are not randomized into treatment and control groups (Souto, Maneiro, and Gómez-Reino 2015) .

We also lack data on treatment duration for patients on cDMARDs. Following M. Stevenson et al. (2016), we assume that, conditional on continuing treatment at 6 months, treatment duration for tDMARDs is applicable to treatment duration for cDMARDs. This is, in turn, based on the assumption that cDMARDs are not likely to be more toxic than biologics used in combination with cDMARDs.

### Treatment duration by disease activity level

When **S2-S5** are selected, treatment duration is stratified by the level of disease activity. Since patients in the CORRONA database tended to have moderate disease activity (mean CDAI = 16), we use the CORRONA survival curve to model treatment duration for patients with moderate disease activity. We adjust this curve for patients in remission or low disease activity using the odds ratios reported in Zhang et al. (2011), which imply that patients in remission or with low disease activity have .52 times the odds of stopping treatment as patients with moderate disease activity. In particular, we adjust the probability of treatment failure at each point in time using the methodology described in . As with the analysis described in , we then fit 7 parametric survival models to individual patient data reconstructed from the adjusted survival curve using the Guyot et al. (2012) algorithm. Generalized gamma time to treatment discontinuation curves stratified by disease activity level are shown in . Survival curves for patients with severe disease activity are not displayed because patients with severe disease activity are assumed to switch treatments after the first 6 months of treatment.

Notes: The shaded region denotes the simulation based 95% confidence interval (Mandel 2013).

### Treatment duration by EULAR response

In **S6**, we stratify time to treatment discontinuation by EULAR response based on analyses of the British Society for Rheumatology Biologics Registers (BSRBR) database (M. Stevenson et al. 2016). We again fit 7 parametric survival models using reconstructed individual patient data. The survival curves reported in M. Stevenson et al. (2016) were used to create the patient data. The AIC and BIC of each model by EULAR response category are shown in .

@lcccc & &  
& & & &  
@inputtables/ttd-bsrbr-ipdsurv-eular-ic.txt

One concern is that the BSRBR is representative of the UK but not the US. As a result, we also estimate “adjusted” survival models appropriate for US based analyses. The adjustment is made in six steps using the analyses from the CORRONA database described in .

1. Calculate a hazard function based on a survival curve from an analysis of the CORRONA database. In particular, reconstruct individual patient data from the survival curve (Guyot et al. 2012) and fit a spline-based survival model. Then use the spline-based model to estimate the hazard function .
2. Calculate a hazard function based on the BSRBR. To do so, first calculate hazard functions for both moderate and good EULAR responders using the same method described in step 1. Then calculate an overall hazard function with the proportion of moderate and good responders in the BSRBR analysis. Given that the number of moderate responders is and the number of good responders is the overall hazard function is .
3. At each point in time, calculate the ratio of the CORRONA and BSRBR hazard functions: .
4. Apply the hazard ratio in step 3 to the BSRBR hazard functions for each EULAR response category. That is and .
5. Generate survival curves using the hazard functions from step 4. Specifically, given a general hazard function , calculate the cumulative hazard function, , convert this to a survival function using , and reconstruct individual patient data using the survival curve.
6. Fit parametric survival models to the individual patient data generated in step 5.

Both adjusted and unadjusted survival curves by EULAR response fit using a generalized gamma distribution are shown in . AIC and BIC for the parametric models fit in step 6 to the adjusted individual patient data are shown in .

Notes: The shaded region denotes the simulation based 95% confidence interval (Mandel 2013).

@lcccc & &  
& & & &  
@inputtables/ttd-bsrbr-ipdsurv-eular-adjusted-ic.txt

## Rebound post treatment

Since no data exists on the size of the HAQ rebound post treatment, we vary its size as a proportion of the initial 6-month HAQ decline. 1 is used as an upper bound, which implies that the HAQ rebound is equal to the improvement experienced at the end of the initial 6-month period with that treatment. 0.7 is currently used as a lower bound.

## Serious infections

Based on the NMA by J. A. Singh et al. (2011) and in accordance with M. Stevenson et al. (2016), we assume a rate of 0.035 (95% CI: 0.027 to 0.046) infections per person-year with all tDMARDs and a rate of 0.026 (no CI reported) infections per person-year with cDMARDs. The rate of infection is assumed to be equal across tDMARDs because the published results for specific tDMARDs are estimated with very little precision. The standard error on the infection rate for tDMARDs is assumed to be the same as the standard error for cDMARDs since no standard error was reported for tDMARDs in J. A. Singh et al. (2011).

A patient in the IPS has a serious infection if the simulated time to serious infection occurs before the simulated time of treatment discontinuation. shows the probability of this occurring when treatment duration is modeled using a generalized gamma distribution. The probability of a serious infection is relatively rare as only 3.53% of patients using cDMARDs and 8.13% of patients using tDMARDs have serious infections. However, differences between cDMARDs and tDMARDs are not insignificant as the probability of a serious infection is almost 5 percentage points higher with tDMARDs than with cDMARDs.

@lrrr &  
&  
& & &  
@inputtables/si-prop-gengamma.txt

Notes: Probabilities are estimated by simulating 1,000 patients and 1,000 parameter sets. Treatment duration is simulated using a generalized gamma distribution.

An important question related to the sensitivity of cost-effectiveness to the model specification is whether the probability of serious infections depends on the distribution used to model time to treatment discontinuation. We consequently simulated time to treatment discontinuation using each of the 7 possible probability distributions. We used the pathway **S1** to model treatment switching, so survival is based on the discussion in . Results from the simulation are reported in . There are very small differences across distributions, suggesting that the treatment duration distribution has almost no impact on the probability of serious infections.

.5l@r &  
@inputtables/si-prop-mtx-bydist.txt

Notes: Probabilities are estimated by simulating 1,000 patients and 1,000 parameter sets.

## Utility

Two algorithms can be used to map HAQ to an EQ-5D utility score. Each is used to simulate utility for every patient in the model to obtain a distribution of utility over time. Our preferred algorithm is the mixture model developed by Hernández-Alava et al. (2013), which is described in detail in . The second algorithm uses the logistic regression equation reported in A. Wailoo et al. (2006). Regression coefficients are reported in .

compares results from the two algorithms. Mean utility scores from the Hernández-Alava et al. (2013) mixture model lie above those from the A. Wailoo et al. (2006) equation for all values of HAQ. Moreover, the slope of utility curve produced from the mixture model is steeper (although less so for the commonly observed HAQ scores between 1 and 1.5), implying that changes in HAQ from the mixture model predict larger changes in utility. Given that the mixture models have been shown to predict utility more accurately (Hernández-Alava, Wailoo, and Ara 2012; Hernández-Alava et al. 2013; Hernández-Alava et al. 2014), this suggests that standard models underestimate the quality-adjusted life-year benefits, and hence, the cost-effectiveness of treatments.

The utility score depends on serious infections in addition to HAQ. In particular, disutility due to serious infections is assumed to be 0.156 for the duration of the month of infection based on prior studies (M. Stevenson et al. 2016; Oppong et al. 2013). However, given the weak evidence for this estimate, the disutility of an infection is allowed to vary by 20% in either direction.

Finally, in the package, we also allow users to incorporate treatment attributes unrelated to safety or efficacy that might impact utility. In particular, users can specify a vector of variables and a vector of corresponding coefficients. Each coefficient is the impact of the corresponding variable on utility in a given 6 month period. By default, we include variables related to mode of administration (infusion, injection, oral) and years since FDA approval; however, since we have no evidence on the impact of each variable on utility, the coefficients are set to 0 in our default settings.

## Mortality

The probability of death is simulated as a function of age/sex specific mortality from U.S. lifetables (Arias 2015), baseline HAQ, and changes in HAQ from baseline. F. Wolfe et al. (2003) estimate an odds ratio for the effect of HAQ on mortality of 2.22, which is applied to the absolute mortality rates of the general population (HAQ score of 0). To capture the effect of treatment on mortality, we assume that, for every 0.25-unit increase in HAQ score, subsequent 6-month mortality increases according to the hazard ratios reported in Michaud, Vera-Llonch, and Oster (2012). Parameter estimates are shown in .

@lcccc & &  
& & & &  
Impact of baseline HAQ on mortality  
@inputtables/mort-or.txt Impact of 0.25-unit change in HAQ from baseline on mortality  
@inputtables/mort-hr-haqdif.txt

Notes: 95% confidence intervals are calculated using normal distributions on the log odds and log hazard ratio scales.

plots survival curves by gender for patients with a baseline age of 55. Survival was simulated by setting the log odds ratios and log hazard ratios from equal to their point estimates. Three scenarios are considerd. In scenario one, patients do not have RA (i.e., HAQ score of 0). In the second scenario, patients have baseline HAQ score of 1 but it does not increase over time. In the third scenario, patients still have a baseline HAQ score of 1, but it increases by 0.03 per year. The third scenario, therefore, utilizes the relationship between changes and HAQ and mortality from Michaud, Vera-Llonch, and Oster (2012) while the second scenario does not.

Mean survival for females without RA was 82.5 years and declined to 77.0 for females with a constant baseline HAQ of 1 and to 72.4 when HAQ increased by 0.03 per year. Mean survival for males in the first, second, and third scenario were 79.4, 73.2, and 70.1 years respectively. Overall, the figure suggests that RA increases mortality and that larger increases in HAQ over time increase mortality by even more.

Notes: Baseline HAQ is 1 for the “Constant HAQ” and “HAQ increase of 0.03 per year” scenarios; baseline HAQ is 0 for the “No RA” scenario.

## Costs

An overview of drug acquisition and administration costs is presented in . Costs are a function of dose and frequency of administration, strength and dosage form, price, and infusion costs. Since infliximab dosing depend on patient weight, the costs for infliximab reported in the table average over a patient population that is 21% male. The prices in the table are based on the wholesale acquisition cost (WAC) and do not include discounts or rebates so they may be higher than actual drug costs. In the simulation, a unique discount can be used for each drug; currently the discount is assumed to range from 20% to 30%. The methodology used to calculate drug acquisition and administration costs is described in more detail in .

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| @inputtables/treat-cost.txt |  |  |  |  |  |  |  |  |

Notes: Costs in the table do not include rebates or discounts, but rebates and discounts are used in the simulation. Cost for infliximab are calculated by assuming that 21% of patients are male and that the weight of men and women are 89 kg and 75 kg respectively. Tocilizumab is dosed weekly if weight is greater than 100 kg; costs for tocilizumab reported in the table are for patients weighing less than 100 kg. IV = intravenous; SC = subcutaneous.

Parameters associated with resource use are show in in . Costs related to physician visits, chest X-rays, tuberculosis tests, and outpatient follow-up are based on L. Claxton et al. (2016). The cost per hospital day and the relationship between the HAQ score and the annual number of hospital days are from Carlson et al. (2015). Cost of any serious infection are assumed to be equal to the cost of pneumonia hospitalization at $5,873, based on Medicare reimbursement rates. F. Wolfe et al. (2005) provide an estimate of annual income loss in relation to HAQ scores: $4,372 (95% CI: 2,078 to 6,607; 2002 dollars) change per unit HAQ change, which are inflated to 2016 dollars.

@lcccc & &  
& & & &  
Days in hospital per year  
@inputtables/hosp-days-haq.txt @inputtables/hosp-cost-perday.txt General management cost  
@inputtables/mgmt-cost.txt Productivity loss  
@inputtables/prod-loss.txt

Notes: 95% confidence intervals for hospital days per year by HAQ score and hospital cost per day are calculated by using the methods of moments to generate the parameters of the gamma distribution given a mean and standard error. The 95% confidence intervals for general management costs are based on normal distributions as assumed in L. Claxton et al. (2016). 95% confidence interval for productivity loss are calculated using a normal distribution and inflated to 2016 dollars.

### Insurance value

In the IVI-RA Model interface, users have complete control over the probability of illness parameter and the marginal rate of substitution between the sick and the well states. In the IVI-RA Value Tool, we set the probability of obtaining RA in the next year equal to , based on the annual incidence rate reported for individuals age 55 to 64 in Myasoedova et al. (2010). Furthermore, we set the the marginal rate of substitution between the sick and well states equal to 1.5 given that positive demand for health insurance suggests that it is positive, but note that there is considerable uncertainty around this estimate and that more research is required.

# Simulation and uncertainty analysis

## Individual patient simulation

The IPS is a discrete-time simulation that simulates individual patients one at a time. Model cycles, denoted by , were chosen to be 6-months long to be consistent with most RCT and real-world data evidence. describes the main components of the IPS for a single patient and a single treatment. The full simulation cycles through each treatment in a treatment sequence and through each simulated patient.

1. **First 6 months ()**
   1. Simulate treatment switching using **S1-S6**, time to serious infection , and death ().
      1. **If** **S1-S6** leads to a treatment switch or if the sampled time to serious infection occurs during cycle 0 (i.e., ), **then** stop treatment. It is assumed that HAQ does not change. **Else**, continue treatment. Simulate change in HAQ using **H1-H3** and time to treatment discontinuation .
      2. **If** patient died, **then** move to next patient.
2. **Maintenance phase** (for )
   1. Simulate death and change in HAQ.
   2. **If** patient died, **then** move to next patient.
   3. **If** , **then** switch treatment. Treatment switch caused by a serious infection if time to serious infection occurred during or before cycle T (i.e., ).

## Parameter uncertainty

Parameter uncertainty is quantified using PSA, which propagates uncertainty in the model input parameters throughout the model by randomly sampling the input parameters from their joint probability distribution (Baio and Dawid 2015; Claxton et al. 2005). Probability distributions are determined according to the distributional properties of the statistical estimates, which, in turn, depend on the statistical techniques used and the distributions of the underlying data. We use normal distributions for sample means, gamma distributions for right-skewed data (e.g., hospital costs), and Dirichlet distributions for multinomial data. The multivariate normal distribution is used for regression parameters estimated using frequentist techniques, provided that the variance-covariance from the statistical analysis is available. For parameters estimated using a Bayesian NMA, we fit multivariate normal distributions to the posterior distribution of the parameters generated from the Markov-Chain Monte-Carlo (MCMC) algorithm using sample means and the sample covariance matrix. When we lack evidence on a parameter, we typically assume a uniform distribution with lower and upper limits that reflect the degree of uncertainty in the parameter. The PSA parameter distributions are summarized in .

@p.65 p.35 &  
Rebound factor & Uniform  
NMA parameters - ACR response & Multivariate normal  
NMA parameters - DAS28 & Multivariate normal  
NMA parameters - HAQ & Multivariate normal  
Drug acquisition and administration cost & Fixed  
Survival model parameters for treatment duration during maintenance phase & Multivariate normal  
US lifetable mortality rates & Fixed  
Mortality probability odds ratio - baseline HAQ & Normal  
Mortality probability hazard ratio - change in HAQ from baseline & Normal  
ACR response to EULAR response mapping & Dirichlet  
ACR response to SDAI mapping & Uniform  
ACR response to CDAI mapping & Uniform  
ACR response to HAQ mapping & Normal  
EULAR response to HAQ mapping & Normal  
Linear HAQ progression - by therapy & Normal  
Linear HAQ progression - by age & Normal  
Latent class growth model for HAQ progression & Normal  
Utility model - (Hernández-Alava et al. 2013) mixture model & Multivariate normal  
Utility model - A. Wailoo et al. (2006) & Normal  
Hospital costs - hospital days by HAQ & Gamma  
Hospital costs - hospital costs per day & Gamma  
General management cost & Gamma  
Serious infection - survival parameters & Normal  
Serious infection - cost per infection & Uniform  
Serious infection - utility loss & Uniform

## Structural uncertainty

We consider structural uncertainty due to two factors:

* The relationship between health states within the model.
* The statistical model used to estimate parameters.

summarizes the competing model structures, which are conditional on the perspective of the decision maker. In total, there are 12 x 2 x 8 x 2 = 384 possible model structures. The choice of model structure for the initial treatment phase (**H1-H3** and **S1-S6**) depends on the preferred measures of disease activity included in the model as well as whether statistical relationships should be modeled directly or indirectly. Likewise, model structures related to HAQ progression, treatment duration, and converting HAQ to utility all reflect uncertainty in the appropriate statistical model.

Competing model structures

|  |  |
| --- | --- |
| Initial effect of treatment on HAQ (**H1-H3**) and switching (**S1-S6**) | 12 |
| HAQ trajectory | 2 |
| Cause and probability distribution used to model treatment discontinuation | 8 |
| Utility algorithm | 2 |

## Implementation

We begin by describing the simulation procedure conditional on model structure, which uses PSA to capture uncertainty within but not between models. The procedure proceeds in two steps: first, model parameters are sampled from their joint probability distribution (), and second, for each parameter set, model outcomes are simulated one at a time for individual patients in the specified population ().

Analysts who wish to expand the analysis to capture uncertainty between models can follow the approach described in Bojke et al. (2009). In particular, for each randomly sampled parameter set, each model structure (or a subset of plausible model structures) can be simulated. The distribution of simulated outcomes across parameters and models will then reflect uncertainty both within and between models.

It’s important to note that simulation output for an individual patient captures differences in outcomes across patients due to random variation (often referred to as first order uncertainty). This information might be useful to patients since it is needed to predict the distribution of their future outcomes conditional on their characteristics, but less useful to a decision maker concerned with making treatment decisions for a population or subset of a population. Analysts wishing to use the model for CEA or MCDA should therefore estimate mean outcomes by averaging over the simulated patients for each parameter set and model structure. The number of simulated patients should be sufficiently large so that mean outcomes are stable across model runs (i.e., so that first order uncertainty is eliminated).

Although CEA and MCDA is concerned with mean outcomes, that does not imply that it does not account for heterogeneity. Instead, since outcomes depend on the characteristics of each patient, model averages are a function of the population analyzed. Subgroup analyses can be used to examine differences in cost-effectiveness across subgroups by simulating patients with certain shared characteristics.

Parameter and structural uncertainty imply decision uncertainty, or the degree to which decisions are made based on imperfect knowledge. Indeed, in CEA, with the aim to maximize health outcomes for a given budget, the optimal decision with current information is to choose the policy that maximizes the expected NMB; however, due to uncertainty, the incorrect policy may be considered the most cost-effective. To characterize uncertainty within a CEA framework, standard summary measures including 95% credible intervals for NMBs and other model outcomes, cost-effectiveness planes, and cost-effectiveness acceptability curves, and the expected value of perfect information can be calculated from the simulated output. Since the expected value of partial perfect information is computationally costly, it can be approximated using meta-modeling techniques (Jalal et al. 2013; Jalal, Goldhaber-Fiebert, and Kuntz 2015; Heath, Manolopoulou, and Baio 2016).

# Validation

We aim to validate the model using the five types of validation described by Eddy et al. (2012). Currently, we are able to use the first three validation types. First, we have checked the model for face validity by ensuring that simulated outcomes are consistent with current science and evidence. Second, we performed unit tests to verify that the individual units of code that are used to simulate the model return expected results. Third, we compared simulated results for key outcomes such as mortality, HAQ over time, and time to treatment discontinuation with real-world data and our underlying parameter values. In particular, we ran the model online under various scenarios using our R Shiny web application and checked the simulated outcomes.

In the future, we plan to use both external validation and predictive validation to help fine tune our model. External validation will be performed by comparing outcomes simulated using our model to real-world outcomes and predictive validity will involve using our model to forecast future events and comparing our forecasted outcomes to the observed outcomes.

# Limitations and areas for improvement

The IVI-RA model is an open-source model that is part of the OSVP process and therefore designed to be updated and improved over time. We believe that there are number of potential areas for improvement.

* **Adverse events other than serious infections:** The current model does not account for side effects other than serious infections even though these are important to patients and can result in treatment switching.
* **Adverse events that vary across biologics:** The model allows the serious infection rate to differ between cDMARDs and tDMARDs but assumes that the infection rate is equal among tDMARDs. Future model versions might want to reconsider the evidence underlying this assumption.
* **Time to treatment discontinuation:** Our time to treatment discontinuation curves are based on scanned data and combine information from multiple sources. Direct analyses of databases like the CORRONA database or the National Data Bank for Rheumatic Diseases (NDB) could generate more accurate estimates of treatment duration as well the effect of treatment response or disease activity level on discontinuation rates.
* **Patient preferences:** In the current model, patient utility is a function of the HAQ score and varies according to age, gender, and unobserved patient-specific factors. In other words, utility depends on treatment (through the effect of treatment on HAQ) and the characteristics of the patient. Future iterations of the model should consider other ways that treatment influences utility and that utility varies across patients. For example, disease activity level or the number of previous therapies might help predict utility conditional on HAQ. Furthermore, surveys could be used to estimate the effect of treatment attributes such as route of administration or frequency of administration on utility. Finally, since unobserved patient-specific factors are very important predictors of utility, the model could be run for specific classes of patients within the mixture model (e.g., subgroups where HAQ has the largest effect on utility), although it might be difficult to identify these patient subgroups in a real-world setting.
* **Treatment effect modifiers:** There is currently little evidence (that we are aware of) suggesting that treatment effects vary across patients. When there is sufficient evidence in the literature related to treatment response heterogeneity, we will allow treatment response at 6 months to depend on the characteristics of the patient.
* **Treatment effects after treatment failure:** There are two main limitations in the model related to reductions in treatment response after failing a biologic; first, there are not enough RCTs to reliably estimate tDMARD-specific treatment effects for tDMARD experienced patients using a NMA, and second, treatment response likely does not only depend on whether a patient is tDMARD naive or experienced, but on the number of previous failures as well. Our current approach is to assume that treatment response is reduced for tDMARD experience patients based on evidence from Carlson et al. (2015). One possible extension is to use a Bayesian NMA approach in which the Carlson et al. (2015) results are used to generate priors for the tDMARD experienced group. As new RCTs become available, the posterior distributions from the Bayesian analysis would move further from the prior and closer to estimates from the trials. The estimates could be further improved by combing NMA results with real-world data and modeling reductions in treatment response as a flexible function of the number of failed biologics.
* **A LCGM for the progression of tDMARDs over time:** The LCGM can be used to model HAQ progression for patients using cDMARDs or on NBT; however, we only have estimates of constant linear progression of HAQ for patients on biologics. Future studies that use non-linear mixture models to model the long-term progression of disease for patients using tDMARDs are needed.
* **Long-term trends in disease activity:** The current model uses results from RCTs to model changes in disease activity during the first 6 months of treatment. But there is, to our knowledge, no evidence on the progression of disease activity over time. New studies are needed to model both the long-term impact of treatment on disease activity and the correlation between changes in disease activity and changes in HAQ.
* **The patient population:** Our population characteristics are based on summary data reported in the published literature. As a result, the sampled patient populations within the model do not account for correlations across all of the variables. Distributions estimated from patient databases like the CORRONA database or the NDB would yield more realistic patient populations.
* **Estimating the rebound effect:** One of the most important predictors of cost-effectiveness is the degree to which the HAQ score increases following treatment failure. Most models currently assume that the HAQ score increases by the same amount as the initial 6 month decline in the HAQ score, but there is little evidence to support this. Studies that attempt to quantify the rebound effect are critical.
* **Insurance value:** Additional research is needed on insurance value and its use in value assessment. First, the framework presented here assumes that there is a single probability of illness and that treatment benefits and costs can be attributed to that sick state. However, in practice, the probability of illness depends on age and illness (e.g., RA) worsens over time. Moreover, in RA, treatment benefits and costs depend on disease severity. Future research should consider insurance value in a dynamic context, in which the value to a healthy individual today depends on the probability of all future health states. Second, new research is needed on the marginal rate of substitution between the sick and well states. D. Lakdawalla, Malani, and Reif (2017) suggest a few promising approaches.

# Rates, probabilities, and standard errors

## Using odds ratios to adjust probabilities

Let be a baseline probability, be a vector of log odds ratios, and be a vector of regressors. We apply the log odds ratios to to generate a new probability with the logistic equation,

where,

## Converting rates and probabilities

Given a *constant* rate during a given time period, we estimate the probability of an event occurring before time using the exponential distribution,

Given a probability , the rate parameter is recovered by applying the log transformation,

## Calculating standard errors from confidence intervals

Journal articles often report confidence intervals rather than standard errors. However, given that regression coefficients are asymptotically normally distributed, standard errors can be calculated from a confidence interval using the normal distribution. In particular, given a coefficient estimate (e.g., a log hazard ratio, log odds ratio, or linear regression coefficient) and an upper bound and lower bound of a two-sided 95% confidence interval, we calculate the standard error as,

where is the quantile function of the normal distribution.

# Heterogeneous populations

When generating heterogeneous patient populations, we sample binary variables from binomial distributions, continuous uncorrelated variables from normal distributions, and continuous correlated variables from multivariate normal distributions. Truncated distributions are used when variables are restricted to lie within certain intervals.

In particular, the proportion of the female population is drawn from a binomial distribution while age, disease duration and the number of previous DMARDs are drawn from truncated normal distributions. Each sampled value of the number of previous DMARDs is rounded to the nearest integer. Baseline HAQ and three disease activity measures (DAS28, SDAI, and CDAI) are drawn from truncated multivariate normal distributions. The covariance matrix is calculated using the correlations reported in Daniel Aletaha et al. (2005) ().

We used the correlations from the routine cohort (during visit 1) rather than correlations in the inception cohort (at baseline) since the correlation between HAQ and the disease activity measures were more similar to those from the Leflunomide database (J. Smolen et al. 2003). That said, correlations between the three disease activity measures were nearly identical in each cohort. The one exception was that the correlation between SDAI and CDAI of 1 in the routine cohort seemed unreasonably high so we used the value of 0.94 from the inception cohort.

We used this sampling procedure to simulate 1,000 patients. Summary statistics from a simulated patient cohort of size 1,000 are shown in .

@lccc &  
& & &  
@inputtables/pats-samp.txt

# Mapping ACR response to changes in disease activity

Let denote disease activity, the number of patients with ACR 20 to <50 response, the number of patients with ACR 50 to <70 response, the number of patients with ACR 70 response, and the number of patients with an ACR response greater than or equal to 20%. Mean changes in SDAI, CDAI, and DAS28 by overlapping ACR response categories are converted to mean changes by mutually exclusive ACR response categories as follows:

* **ACR 70**: Mean changes by ACR 70 were reported directly in D Aletaha and Smolen (2005).
* **ACR 50 to <70**: Mean change in disease activity given ACR 50 to <70 response is calculated by solving for :
* Mean change in disease activity given ACR 20 to <50 response is calculated by solving for

# HAQ progression

## Effect of age on linear HAQ progression

Michaud, Wallenstein, and Wolfe (2011) report an overall rate of linear HAQ progression and rates for three age groups (<40, 40-64, 65). Let be the overall rate of progression and be the rate of progression for age group . To estimate the effect of age on the progression rate, we calculated the difference between the overall progression rate and the age specific rate, . We estimated the standard error of this quantity assuming no covariance between and ,

## HAQ trajectory with a latent class growth model

Norton et al. (2014) model HAQ progression using a LCGM. The probability that individual is a member of class at time is modeled using a multinomial logistic regression,

where is the vector of regression coefficients associated with class and is the corresponding vector of regressors. The variables included in are age, gender, baseline DAS28, symptom duration, rheumatoid factor, ACR criteria, and socioeconomic status. Regression coefficients for classes 2-4 relative to class 1 are shown in . Older age and female gender are especially important predictors of membership in higher risk classes; a worse DAS28 score, rheumatoid factor positivity, fulfillment of the 1987 ACR criteria, lower socioeconomic status, and longer disease duration are also predictors of membership in classes with worse HAQ progression.

@lccc &  
& & &  
  
@inputtables/lcgm-coef-class2.txt   
  
@inputtables/lcgm-coef-class3.txt   
  
@inputtables/lcgm-coef-class4.txt

Notes: Class 1, or the “low” group, is the reference category.

The HAQ trajectory for a given class can be written as,

where is the HAQ score, is a variable that is a function of time, the are polynomial regression coefficients for members of class c, and is an error term.

Sam Norton generously provided us with statistical estimates of the 4 class LCGM used in Norton et al. (2014) from [MPlus](https://www.statmodel.com). Like M. Stevenson et al. (2016), we noted that the coefficient estimates the MPlus resulted in large fluctuations in the predicted HAQ scores, likely because three decimal places was not precise enough for the cubic term in . We consequently used the coefficient estimates to predict the probability of class membership—which are less likely to be influenced by the number of reported decimal places—but estimated using the observed HAQ values reported in Figure 2 in Norton et al. (2014). However, since standard errors were artificially high using grouped data, we standard errors in were based on those reported in the original paper. Moreover, since we are only interested in the HAQ trajectory following the HAQ decline during the initial treatment phase, we limited our analysis to HAQ values from year 2 and onwards. Using the post year 2 data, we estimated using separate linear regressions with cubic polynomials for each class (). Like Norton et al. (2014), we set equal to a reciprocal transformation of time,

@lccc & &  
  
@inputtables/lcgm-coef-growth-class1.txt   
  
@inputtables/lcgm-coef-growth-class2.txt   
  
@inputtables/lcgm-coef-growth-class3.txt   
  
@inputtables/lcgm-coef-growth-class4.txt

Notes: Class 1, or the “low” group, is the reference category.

In the simulation model, we simulate the HAQ score at 6 months as a function of the baseline HAQ score and the change in HAQ during the initial treatment phase. Since the Norton et al. (2014) model is not conditional on the HAQ score in the previous period, we use it to predict changes in HAQ rather than the level of the HAQ score. More precisely, for a patient in a given class, we model the change in HAQ as,

Since was estimated on aggregated data, we did not have reliable estimates of . We consequently set equal to 0, which implies that we are generating a mean response rather than a predicted response. In other words, we are not simulating the random variation associated with each individual, but are still accurately simulating mean outcomes across populations or subpopulations.

# Simulating mortality

Death is simulated for each patient during each model cycle based on age, gender, baseline HAQ, and change in HAQ from baseline. A 0/1 death indicator is randomly drawn using the following procedure:

1. Find , the probability that a patient of gender and age will die before age , from lifetables.
2. As described in , adjust using the effect of a change in baseline HAQ on the odds of mortality, ,
3. Following , convert the mortality probability, , into a mortality rate, .
4. Adjust the mortality rate, , using the estimated log hazard ratio of mortality, , of a change in HAQ from baseline, HAQ.
5. Following , convert the mortality rate into a probability given a 6-month cycle length,
6. Randomly draw a 0/1 death indicator, , given the probability of death, ,

# Simulate utility

## Mixture model

The mixture model estimated by Hernández-Alava et al. (2013) simulates utility in two stages. In the first stage, we sampled pain for a given individual in a particular model cycle based on the HAQ score. In the second stage, we simulated utility as a function of HAQ, pain and age/sex.

### Simulating pain

To simulate pain from HAQ, we used the summary statistics for pain and HAQ reported in Sarzi-Puttini et al. (2002). Pain was measured with the visual analog scale (VAS) with mean 61.65 and standard deviation 19.10, while HAQ was reported to have mean 1.39 and standard deviation 0.59.

We then estimated the correlation between pain and HAQ by digitally scanning the curve depicting the (linear) relationship between pain and HAQ (Figure 114) shown in M. Stevenson et al. (2016). Using the scanned data, we regressed pain on HAQ using simple ordinary least squares (OLS). The correlation between pain and HAQ, estimated as 0.52, was calculated by rearranging the OLS estimate for the slope, , of the regression model,

Pain was simulated using these parameters by assuming that pain was normally distributed conditional on HAQ,

However, since the VAS is constrained to lie between 0 and 100, pain was drawn from a truncated normal distribution with a lower limit of 0 and an upper limit of 100.

### Simulating utility

After simulating pain, we simulated utility with a mixture model. Within each class , the HAQ score for patient in period was modeled as,

where is a random error term and is a vector of regression coefficients corresponding to the vector of variables . is a random intercept for individual and class that is predicted by a class-specific intercept, , a vector of individual-specific variables , a coefficient vector , and an error term, . Variables included in are , , , , and ; contains a single indicator variable, , equal to 1 if the patient is male and 0 if female.

The probability of class membership was modeled using a multinomial logit model,

where there are four possible classes and is a vector of coefficients corresponding to the vector of variables, (which includes an intercept). Variables included in other than the intercept are , , and .

We sampled from the mixture model as follows.

1. For each individual , sample the error term, .
2. For each individual and time-period :
   1. Sample class membership conditional on ; that is, sample where is the probability of being in class .
   2. Predict the intercept .
   3. Sample the error term, .
   4. Predict the HAQ score, .

## Logistic regression model

A. Wailoo et al. (2006) use a logistic regression equation to predict utility as a function of patient demographics, disease history, and current disease status. The regression coefficients from the model are shown in and used to predict utility with the inverse logit function. Specifically, if the vector of coefficients is denoted by and the corresponding vector of explanatory variables is denoted by the vector , then predicted utility is given by .

@lcc & &  
@inputtables/util-wailoo-pars.txt

Notes: Coefficients are from the logistic regression reported in A. Wailoo et al. (2006).

# Drug acquisition and administration costs

Drug acquisition and administration costs are calculated separately during the initial treatment phase and the maintenance phase since dosing typically differs. Costs are separated into acquisition costs and infusion costs. Infusion costs are calculated by multiplying the number of doses in a 6 month period by the cost of an infusion and acquisition costs are calculated as,

where is the ceiling function and implies that products cannot be reused after opening, is the recommended dose of the drug, is the strength of the drug, is the number of doses in a 6 month period, and is the price per unit of the treatment after discounts and rebates. For example, as shown in , both the strength and the dose of adalimumab are 50 mg, so costs (before discounts and rebates) for the initial 6 month period are calculated by multiplying the number of doses (13) by the WAC ($2,220.62).

When dosing depends on weight, costs are calculated separately for each patient in the simulation. In particular, costs are calculated as,

where is patient weight, is the dose per weight, and , , and are defined in the same way as in the non-weight based scenario. To illustrate, the acquisition cost (before discounts and rebates) for infliximab after the first 6 months is calculated by multiplying each patient’s weight by the dose (6 mg/kg) and dividing by the size of a vial (100 mg), and then multiplying by the number of doses (8.67) and the price per unit ($1,113.27).

# Annualized costs and benefits

Annualized QALY gains and costs, which are used to estimate the annual insurance value of treatment, are calculated using the formula for a geometric series and assuming a constant flow rate each model cycle. In particular, if indexes time (in years), and denote lifetime costs and QALYs simulated by the model, respectively, then annualized costs, and QALYs, , can be derived from,

where is the discount factor for costs and is the discount factor for QALYs. The discount factor is a function of the annual discount rate (typically assumed to be 0.03); that is, where is the discount rate for . The time horizon is set to equal the maximum number of years that a patient could survive within the model—currently the maximum age that a patient could survive to is 100 given the default lifetables used within the model, so is equal to 100 minus a patient’s age at the start of the simulation.

# Network Meta-Analysis

Treatment effects with tDMARDs relative to cDMARDs for moderate to severe RA patients who failed treatment with a conventional DMARD were estimated based on currently available evidence as reported in the literature. The two populations of interest were: 1) RA patients without a history of prior tDMARD treatment (tDMARD naive population); and 2) RA patient with a history of tDMARD treatment (tDMARD experienced population). Relevant randomized controlled trials were identified with a systematic literature review and treatment effects were estimated by means of a network meta-analysis.

## Systematic literature review to identify relevant studies

### Eligibility criteria

Study eligibility criteria were defined in terms of the population, interventions, comparisons, outcomes, and study design (PICOS).

Study eligibility criteria

|  |  |
| --- | --- |
| Population | Adult (>18) patients with moderate to severe active rheumatoid arthritis who failed treatment with a conventional DMARD and were either targeted DMARD naive or experienced. |
|  |  |
| Interventions | Approved dosing regimens (or equivalent) of the following tDMARDs as monotherapy or in combination with a cDMARD: abatacept, adalimumab, anakinra, certolizumab, etanercept, golimumab, infliximab, rituximab, sarilumab, tocilizumab, tofacitinib |
|  |  |
| Comparators | cDMARDs (methotrexate, leflunomide, hydroxychloroquinine, sulfasalazine); placebo; any of the interventions of interest; any other intervention (or dosing regimen) that facilitates an indirect comparison between the interventions of interest |
|  |  |
| Outcomes | At least one of the following outcomes at 6 months of follow-up: ACR 20/50/70, HAQ-DI, DAS28 |
|  |  |
| Study design | Randomized clinical trials |
|  |  |
| Other | Only full text reported in Enlgish were included. Studies only available as conference abstractsor presentations were excluded. |

### Literature search

Relevant studies were identified by searching the following databases: Medline, EMBASE, and Cochrane Central Register of Controlled Trials. The searches were executed June 2017 with the following predefined search strategies and corresponding results.

@rp0.70r

& &

& &  
  
@inputtables/medline-search.txt

@rp0.70r

& &

& &  
  
@inputtables/embase-search.txt

@rp0.70r

& &

& &  
  
@inputtables/cochrane-search.txt

### Study selection

Two investigators working independently scanned all abstracts identified in the literature search. The same two investigators independently reviewed relevant abstracts in full-text. Discrepancies occurring between the studies selected by the two investigators were resolved by involving a third investigator and reaching consensus.

### Data extraction

Two investigators working independently extracted relevant data on study characteristics, interventions, patient characteristics, and outcomes from the final list of selected eligible studies. Discrepancies observed between the data extracted by the two investigators were resolved by involving a third investigator and reaching consensus.

## Analyses

In order to perform a network meta-analysis where the risk of biased relative treatment effect estimates is limited we need to have 1) a single evidence network where each randomized controlled trial has at least one intervention in common with another trial; and 2) no differences in study designs and the distribution of patient characteristics that act as relative treatment effect-modifiers across the studies in the network (ref). For both the tDMARD naive population and the tDMARD experienced population a connected evidence network could be defined, however for the latter population the treatment history across studies was considered too different to obtain valid results from a network meta-analysis, and the limited number of studies would not allow adjusting for these differences with statistical techniques. Hence, network meta-analyses were only performed for the tDMARD naive population. Studies that reported results for a mixed population regarding prior tDMARD use were excluded from the analyses; only studies with results reported for a tDMARD naive population were included.

Treatment effects at 6 months with each of the interventions in the network relative to cDMARDs were estimated in terms of ACR response, change from baseline in HAQ, and change from baseline in DAS28 with Bayesian random effects network meta-analysis models as presented by Dias et al. 2013. To avoid influencing the observed results by prior belief, uninformative prior distributions were used for the estimated treatment effect and between-study heterogeneity parameters. Posterior distributions for the model parameters are derived with the Markov chain Monte Carlo method using the JAGS software package (<http://mcmcjags.sourceforge.net/>). The studies included in each analysis for the tDMARD naive population were considered sufficiently similar regarding study design and distribution of patient characteristics that may act as relative treatment effect-modifiers. Accordingly, no meta-regression was performed to adjust for between-trial differences and the obtained estimated for each of the interventions were considered reflective of this target population of interest.

### ACR response at 6 months

The four mutually exclusive ACR response categories were estimated from the overlapping ACR categories using a ordered probit model appropriate for ordered categorical data (Dias et al. 2013). The model assumes that there is an underlying continuous variable (ACR20/50/70) categorized by specifying different cutoffs corresponding to the point at which an individual moves from one category to the next in each trial. The advantage of this approach over an analysis that considers ACR categories separately is that all possible outcomes are analyzed simultaneously based on the same randomized controlled trials, allowing for consistent estimates by category.

More specifically, let be the number of patients in trial for treatment in the mutually exclusive category . The model can be written as,

where is the probability that a patient from trial and treatment is in category and there are patients in trial and treatment . Probabilities are modeled using a probit function,

where is a trial specific intercept, is a cutpoint for trial and category , and is the effect of treatment relative to treatment . The cutpoint for category , , is modeled as random,

The study-specific relative treatment effects are also drawn from a common population distribution with mean and variance ,

where . To generate treatment responses, we estimate the response for treatment by averaging across trials containing treatment . In particular, letting be the set of all trials containing treatment , we estimate,

We calculate the probability of ACR < 20% improvement, ACR < 50% improvement, and ACR < 70% improvement with treatment as,

The probabilities of overlapping ACR response (i.e., ACR 20/50/70) are then,

where is the reduction in treatment response at a given line of therapy. is a patient is bDMARD naive and on average, equal to .84 after failing a biologic. The mutually exclusive ACR response categories are easily derived from the overlapping categories.

To avoid influencing the observed results by prior belief, uninformative prior distributions were used for the estimated model parameters. Posterior distributions for the model parameters are derived with the Markov chain Monte Carlo method.

### Change in HAQ and DAS28 at 6 months

The models of changes in HAQ and DAS28 at 6 months use a normal likelihood (since the sample mean is approximately normally distributed by the central limit theorem if the sample size is reasonably large) and an identity link.

More specifically, let be the outcome of interest in trial and treatment , and consider the model,

where,

is modeled using a random effect with ,

where . As with the ACR response model, we allow treatment response to depend on patient characteristics by modeling as a function of covariates for each individual patient ,

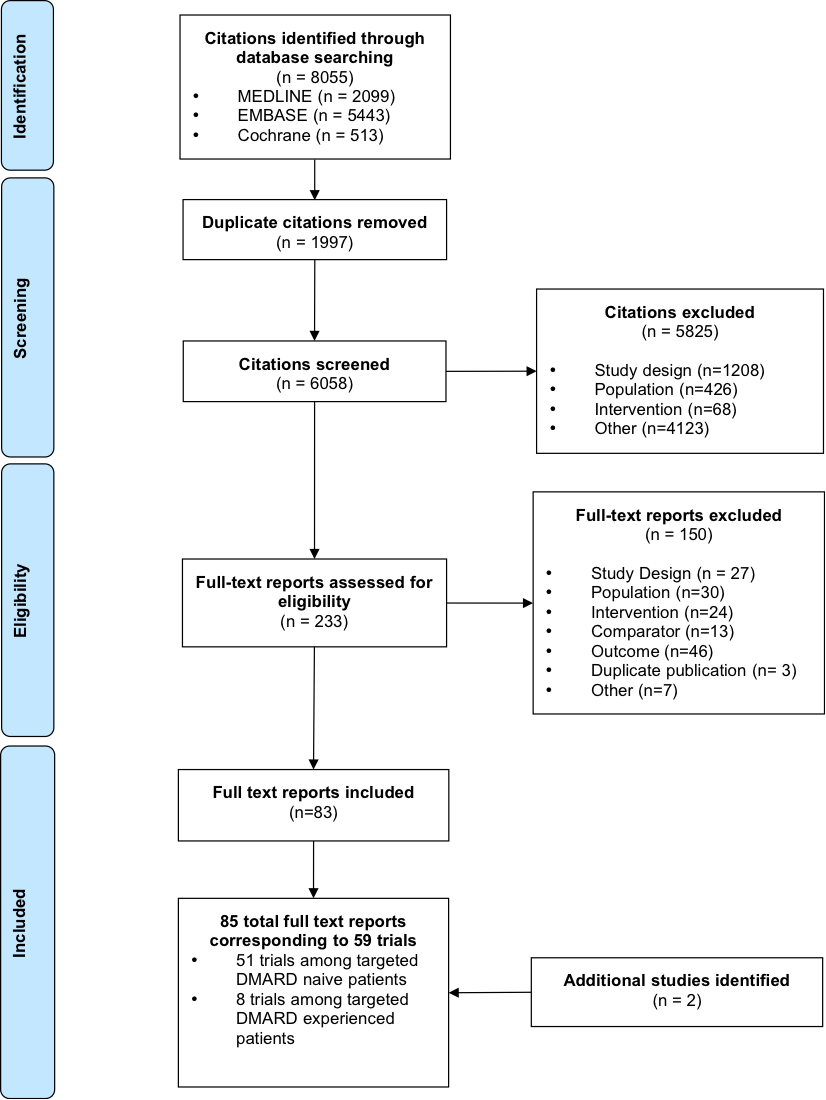
In the simulation, we allow for treatment effect modifiers by modeling as a function of covariates for each individual patient ,

The absolute treatment effect is estimated by calculating as in . The absolute treatment effect for treatment is then,

where is defined as in . Uninformative priors were used to derive the posterior distributions.

## Evidence base

### Study identification and selection



Study identification and selection

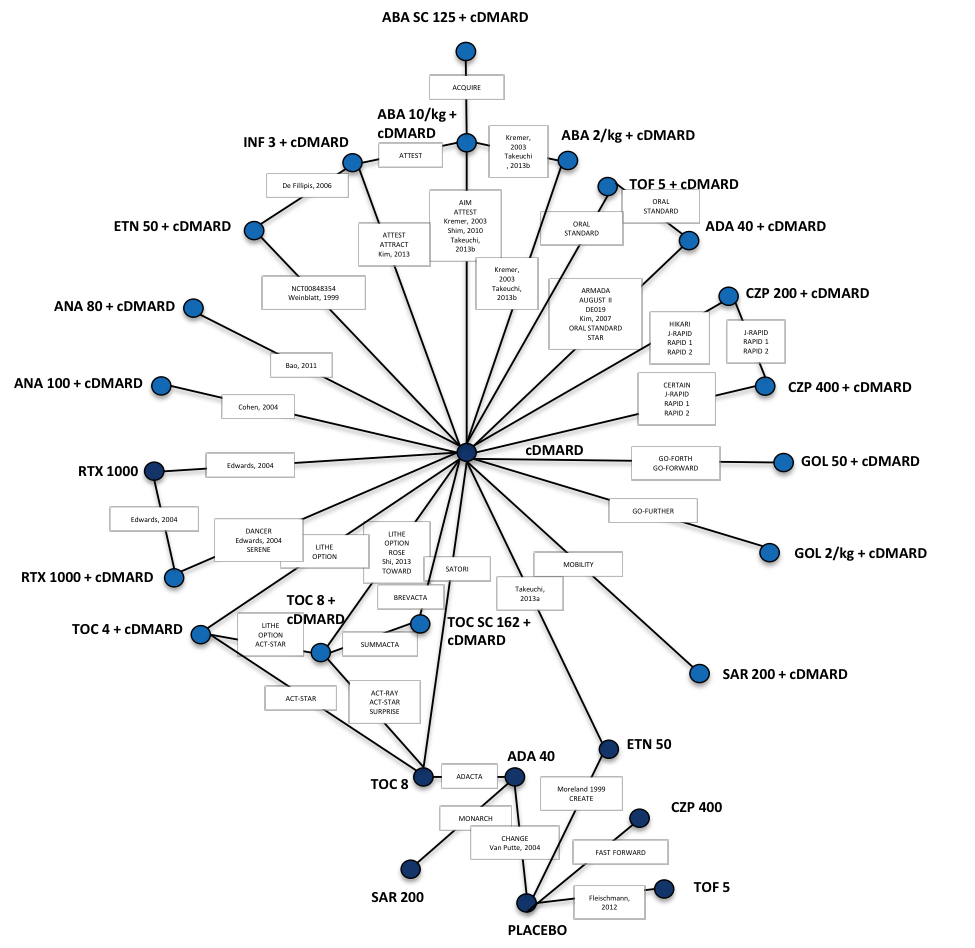
### Included studies

@lll

& &

& &  
  
@inputtables/included-studies.txt

### Evidence network, tDMARD naive population



Evidence network, tDMARD naive population

### Study specific data, tDMARD naive population

@lp.10rcrrrrr

& & & & & & & &  
& & & & & & & &

& & & & & & & &  
& & & & & & & &  
  
  
@inputtables/trial-data.txt

## Comparing the IVI NMA to the NICE NMA

To help ensure that differences in cost-effectiveness estimates from our model relative to others are not driven by the NMA results, we compared our NMA estimates to estimates reported by NICE in M. Stevenson et al. (2016). We focus on ACR response, since the NICE report and other models use treatment pathways similar to **H1** and **H2** and rarely use DAS28 to inform treatment duration. As shown in , our results are similar and the NICE point estimates are generally within the 95% credible intervals surrounding our point estimates.

@lrrrrrr & &  
(lr)2-4 (lr)5-7 & & & & & &  
@inputtables/nma-acr-nice-vs-ivi.txt

Notes: ACR20/50/70 categories are the probability of at least a 20/50/70% improvement. 95% credible intervals are in parentheses. IVI estimates are based on 6-month simulations of 1,000 patients and 1,000 parameters sets for each therapy. NICE estimates are from Table 37 in M. D. Stevenson et al. (2017). cDMARDs = conventional disease-modifying antirheumatic drugs; MTX = methotrexate; ABT IV = abatacept intravenous; ADA = adalimumab; ETN = etanercept; GOL = golimumab; IFX = infliximab; TCZ = tocilizumab; CZP = certolizumab pegol; ABT SC = abatacept subcutaneous; RTX = rituximab; TOF = tofacitinib. ACR = American College of Rheumatology.

Aletaha, D, and J Smolen. 2005. “The Simplified Disease Activity Index (Sdai) and the Clinical Disease Activity Index (Cdai): A Review of Their Usefulness and Validity in Rheumatoid Arthritis.” *Clinical and Experimental Rheumatology* 23 (5). CLINICAL AND EXPERIMENTAL RHEUMATOLOGY SAS: S100.

Aletaha, Daniel, Valerie PK Nell, Tanja Stamm, Martin Uffmann, Stephan Pflugbeil, Klaus Machold, and Josef S Smolen. 2005. “Acute Phase Reactants Add Little to Composite Disease Activity Indices for Rheumatoid Arthritis: Validation of a Clinical Activity Score.” *Arthritis Research & Therapy* 7 (4). BioMed Central: R796.

Anderson, Jaclyn, Liron Caplan, Jinoos Yazdany, Mark L Robbins, Tuhina Neogi, Kaleb Michaud, Kenneth G Saag, James R O’dell, and Salahuddin Kazi. 2012. “Rheumatoid Arthritis Disease Activity Measures: American College of Rheumatology Recommendations for Use in Clinical Practice.” *Arthritis Care & Research* 64 (5). Wiley Online Library: 640–47.

Arias, Elizabeth. 2015. “United States Life Tables, 2011.” *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System* 64 (11): 1–62.

Athanasakis, Kostas, Filippos Tarantilis, Konstantina Tsalapati, Thomais Konstantopoulou, Eleni Vritzali, and John Kyriopoulos. 2015. “Cost-Utility Analysis of Tocilizumab Monotherapy in First Line Versus Standard of Care for the Treatment of Rheumatoid Arthritis in Greece.” *Rheumatology International* 35 (9). Springer: 1489–95.

Baio, Gianluca, and A Philip Dawid. 2015. “Probabilistic Sensitivity Analysis in Health Economics.” *Statistical Methods in Medical Research* 24 (6). SAGE Publications Sage UK: London, England: 615–34.

Basu, Anirban, and David Meltzer. 2007. “Value of Information on Preference Heterogeneity and Individualized Care.” *Medical Decision Making* 27 (2). Sage Publications Sage CA: Los Angeles, CA: 112–27.

Black, William C. 1990. “The Ce Plane: A Graphic Representation of Cost-Effectiveness.” *Medical Decision Making* 10 (3). Sage Publications Sage CA: Thousand Oaks, CA: 212–14.

Bojke, Laura, Karl Claxton, Mark Sculpher, and Stephen Palmer. 2009. “Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods.” *Value in Health* 12 (5). Elsevier: 739–49.

Brennan, A, N Bansback, A Reynolds, and P Conway. 2003. “Modelling the Cost-Effectiveness of Etanercept in Adults with Rheumatoid Arthritis in the Uk.” *Rheumatology* 43 (1). Oxford University Press: 62–72.

Briggs, Andrew H, and others. 1999. “A Bayesian Approach to Stochastic Cost-Effectiveness Analysis.” *Health Economics* 8 (3): 257–61.

Briggs, Andrew H, Karl Claxton, and Mark J Sculpher. 2006. *Decision Modelling for Health Economic Evaluation*. Handbooks in Health Economic E.

Carlson, Josh J, Sarika Ogale, Fred Dejonckheere, and Sean D Sullivan. 2015. “Economic Evaluation of Tocilizumab Monotherapy Compared to Adalimumab Monotherapy in the Treatment of Severe Active Rheumatoid Arthritis.” *Value in Health* 18 (2). Elsevier: 173–79.

Claxton, Karl, Mark Sculpher, Chris McCabe, Andrew Briggs, Ron Akehurst, Martin Buxton, John Brazier, and Tony O’Hagan. 2005. “Probabilistic Sensitivity Analysis for Nice Technology Assessment: Not an Optional Extra.” *Health Economics* 14 (4). Wiley Online Library: 339–47.

Claxton, Lindsay, Michelle Jenks, Matthew Taylor, Gene Wallenstein, Alan M Mendelsohn, Jeffrey A Bourret, Amitabh Singh, Dermot Moynagh, and Robert A Gerber. 2016. “An Economic Evaluation of Tofacitinib Treatment in Rheumatoid Arthritis: Modeling the Cost of Treatment Strategies in the United States.” *Journal of Managed Care & Specialty Pharmacy* 22 (9). Academy of Managed Care Pharmacy: 1088–1102.

Curtis, Jeffrey R, Archana Jain, Johan Askling, S Louis Bridges, Loreto Carmona, William Dixon, Axel Finckh, et al. 2010. “A Comparison of Patient Characteristics and Outcomes in Selected European and Us Rheumatoid Arthritis Registries.” In *Seminars in Arthritis and Rheumatism*, 40:2–14. 1. Elsevier.

Deighton, Chris, Kimme Hyrich, Tina Ding, Jo Ledingham, Mark Lunt, Raashid Luqmani, Patrick Kiely, et al. 2010. “BSR and Bhpr Rheumatoid Arthritis Guidelines on Eligibility Criteria for the First Biological Therapy.” *Rheumatology* 49 (6). Oxford University Press: 1197–9.

Dias, Sofia, Alex J Sutton, AE Ades, and Nicky J Welton. 2013. “Evidence Synthesis for Decision Making 2: A Generalized Linear Modeling Framework for Pairwise and Network Meta-Analysis of Randomized Controlled Trials.” *Medical Decision Making* 33 (5). Sage Publications Sage CA: Los Angeles, CA: 607–17.

Drummond, Michael F, Mark J Sculpher, Karl Claxton, Greg L Stoddart, and George W Torrance. 2015. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford university press.

Eddy, David M, William Hollingworth, J Jaime Caro, Joel Tsevat, Kathryn M McDonald, and John B Wong. 2012. “Model Transparency and Validation: A Report of the Ispor-Smdm Modeling Good Research Practices Task Force–7.” *Medical Decision Making* 32 (5). Sage Publications Sage CA: Los Angeles, CA: 733–43.

Espinoza, Manuel A, Andrea Manca, Karl Claxton, and Mark J Sculpher. 2014. “The Value of Heterogeneity for Cost-Effectiveness Subgroup Analysis: Conceptual Framework and Application.” *Medical Decision Making* 34 (8). Sage Publications Sage CA: Los Angeles, CA: 951–64.

Fenwick, Elisabeth, Karl Claxton, and Mark Sculpher. 2001. “Representing Uncertainty: The Role of Cost-Effectiveness Acceptability Curves.” *Health Economics* 10 (8). Wiley Online Library: 779–87.

Garber, Alan M, and Charles E Phelps. 1997. “Economic Foundations of Cost-Effectiveness Analysis.” *Journal of Health Economics* 16 (1). Elsevier: 1–31.

Garrison, Louis P, Sachin Kamal-Bahl, and Adrian Towse. 2017. “Toward a Broader Concept of Value: Identifying and Defining Elements for an Expanded Cost-Effectiveness Analysis.” *Value in Health* 20 (2). Elsevier: 213–16.

Gibson, Laura, Monica Hernandez Alava, and Allan Wailoo. 2015. “Progression of Disease in People with Rheumatoid Arthritis Treated with Non Biologic Therapies.”

Gomez-Reino, Juan J, and Loreto Carmona. 2006. “Switching Tnf Antagonists in Patients with Chronic Arthritis: An Observational Study of 488 Patients over a Four-Year Period.” *Arthritis Research & Therapy* 8 (1). BioMed Central: R29.

Guyot, Patricia, AE Ades, Mario JNM Ouwens, and Nicky J Welton. 2012. “Enhanced Secondary Analysis of Survival Data: Reconstructing the Data from Published Kaplan-Meier Survival Curves.” *BMC Medical Research Methodology* 12 (1). BioMed Central: 9.

Heath, Anna, Ioanna Manolopoulou, and Gianluca Baio. 2016. “Estimating the Expected Value of Partial Perfect Information in Health Economic Evaluations Using Integrated Nested Laplace Approximation.” *Statistics in Medicine* 35 (23). Wiley Online Library: 4264–80.

Hernández-Alava, Mónica, Allan J Wailoo, and Roberta Ara. 2012. “Tails from the Peak District: Adjusted Limited Dependent Variable Mixture Models of Eq-5D Questionnaire Health State Utility Values.” *Value in Health* 15 (3). Elsevier: 550–61.

Hernández-Alava, Mónica, Allan Wailoo, Fred Wolfe, and Kaleb Michaud. 2013. “The Relationship Between Eq-5D, Haq and Pain in Patients with Rheumatoid Arthritis.” *Rheumatology* 52 (5). Br Soc Rheumatology: 944–50.

———. 2014. “A Comparison of Direct and Indirect Methods for the Estimation of Health Utilities from Clinical Outcomes.” *Medical Decision Making* 34 (7). Sage Publications Sage CA: Los Angeles, CA: 919–30.

Institute for Clinical and Economic Review. 2017. “Targeted Immune Modulators for Rheumatoid Arthritis: Effectiveness & Value.”

Ioannidis, John PA, and Alan M Garber. 2011. “Individualized Cost-Effectiveness Analysis.” *PLoS Medicine* 8 (7). Public Library of Science: e1001058.

Jalal, Hawre, Bryan Dowd, François Sainfort, and Karen M Kuntz. 2013. “Linear Regression Metamodeling as a Tool to Summarize and Present Simulation Model Results.” *Medical Decision Making* 33 (7). Sage Publications Sage CA: Los Angeles, CA: 880–90.

Jalal, Hawre, Jeremy D Goldhaber-Fiebert, and Karen M Kuntz. 2015. “Computing Expected Value of Partial Sample Information from Probabilistic Sensitivity Analysis Using Linear Regression Metamodeling.” *Medical Decision Making* 35 (5). SAGE Publications Sage CA: Los Angeles, CA: 584–95.

Keeney, Ralph L, and Howard Raiffa. 1993. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge university press.

Lakdawalla, Darius N, John A Romley, Yuri Sanchez, J Ross Maclean, John R Penrod, and Tomas Philipson. 2012. “How Cancer Patients Value Hope and the Implications for Cost-Effectiveness Assessments of High-Cost Cancer Therapies.” *Health Affairs* 31 (4). Health Affairs: 676–82.

Lakdawalla, Darius, Anup Malani, and Julian Reif. 2015. “The Insurance Value of Medical Innovation.” National Bureau of Economic Research.

———. 2017. “The Insurance Value of Medical Innovation.” *Journal of Public Economics* 145. Elsevier: 94–102.

Madan, Jason, Tony Ades, Pelham Barton, Laura Bojke, Ernest Choy, Philip Helliwell, Paresh Jobanputra, et al. 2015. “Consensus Decision Models for Biologics in Rheumatoid and Psoriatic Arthritis: Recommendations of a Multidisciplinary Working Party.” *Rheumatology and Therapy* 2 (2). Springer: 113–25.

Mandel, Micha. 2013. “Simulation-Based Confidence Intervals for Functions with Complicated Derivatives.” *The American Statistician* 67 (2). Taylor & Francis: 76–81.

Meltzer, David O, Peter C Smith, and others. 2011. “Theoretical Issues Relevant to the Economic Evaluation of Health Technologies.” *Handbook of Health Economics* 2. Elsevier New York: 433–69.

Michaud, Kaleb, Montserrat Vera-Llonch, and Gerry Oster. 2012. “Mortality Risk by Functional Status and Health-Related Quality of Life in Patients with Rheumatoid Arthritis.” *The Journal of Rheumatology* 39 (1). The Journal of Rheumatology: 54–59.

Michaud, Kaleb, Gene Wallenstein, and Frederick Wolfe. 2011. “Treatment and Nontreatment Predictors of Health Assessment Questionnaire Disability Progression in Rheumatoid Arthritis: A Longitudinal Study of 18,485 Patients.” *Arthritis Care & Research* 63 (3). Wiley Online Library: 366–72.

Myasoedova, Elena, Cynthia S Crowson, Hilal Maradit Kremers, Terry M Therneau, and Sherine E Gabriel. 2010. “Is the Incidence of Rheumatoid Arthritis Rising?: Results from Olmsted County, Minnesota, 1955–2007.” *Arthritis & Rheumatology* 62 (6). Wiley Online Library: 1576–82.

Norton, Sam, Bo Fu, David L Scott, Chris Deighton, Deborah PM Symmons, Allan J Wailoo, Jonathan Tosh, et al. 2014. “Health Assessment Questionnaire Disability Progression in Early Rheumatoid Arthritis: Systematic Review and Analysis of Two Inception Cohorts.” In *Seminars in Arthritis and Rheumatism*, 44:131–44. 2. Elsevier.

Norton, Sam, Amanda Sacker, Josh Dixey, John Done, Peter Williams, and Adam Young. 2013. “Trajectories of Functional Limitation in Early Rheumatoid Arthritis and Their Association with Mortality.” *Rheumatology*. Br Soc Rheumatology, ket253.

Oppong, Raymond, Billingsley Kaambwa, Jacqueline Nuttall, Kerenza Hood, Richard D Smith, and Joanna Coast. 2013. “The Impact of Using Different Tariffs to Value Eq-5D Health State Descriptions: An Example from a Study of Acute Cough/Lower Respiratory Tract Infections in Seven Countries.” *The European Journal of Health Economics* 14 (2). Springer: 197–209.

Pan, Sophie Martin Du, Silvia Dehler, Adrian Ciurea, Hans-Rudolf Ziswiler, Cem Gabay, and Axel Finckh. 2009. “Comparison of Drug Retention Rates and Causes of Drug Discontinuation Between Anti–tumor Necrosis Factor Agents in Rheumatoid Arthritis.” *Arthritis Care & Research* 61 (5). Wiley Online Library: 560–68.

Prevoo, MLL, MA Van’t Hof, HH Kuper, MA Van Leeuwen, LBA Van De Putte, and PLCM Van Riel. 1995. “Modified Disease Activity Scores That Include Twenty-Eight-Joint Counts Development and Validation in a Prospective Longitudinal Study of Patients with Rheumatoid Arthritis.” *Arthritis & Rheumatology* 38 (1). Wiley Online Library: 44–48.

Ramiro, Sofia, Alexandre Sepriano, Katerina Chatzidionysiou, Jackie L Nam, Josef S Smolen, Désirée van der Heijde, Maxime Dougados, et al. 2017. “Safety of Synthetic and Biological Dmards: A Systematic Literature Review Informing the 2016 Update of the Eular Recommendations for Management of Rheumatoid Arthritis.” *Annals of the Rheumatic Diseases*. BMJ Publishing Group Ltd, annrheumdis–2016.

Sanders, Gillian D, Peter J Neumann, Anirban Basu, Dan W Brock, David Feeny, Murray Krahn, Karen M Kuntz, et al. 2016. “Recommendations for Conduct, Methodological Practices, and Reporting of Cost-Effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine.” *Jama* 316 (10). American Medical Association: 1093–1103.

Sarzi-Puttini, Piercarlo, Tania Fiorini, Benedetta Panni, Maurizio Turiel, Marco Cazzola, and Fabiola Atzeni. 2002. “Correlation of the Score for Subjective Pain with Physical Disability, Clinical and Radiographic Scores in Recent Onset Rheumatoid Arthritis.” *BMC Musculoskeletal Disorders* 3 (1). BioMed Central: 18.

Singh, Jasvinder A, Kenneth G Saag, S Louis Bridges, Elie A Akl, Raveendhara R Bannuru, Matthew C Sullivan, Elizaveta Vaysbrot, et al. 2016. “2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis.” *Arthritis & Rheumatology* 68 (1). Wiley Online Library: 1–26.

Singh, Jasvinder A, George A Wells, Robin Christensen, Elizabeth Tanjong Ghogomu, Lara J Maxwell, John K MacDonald, Graziella Filippini, et al. 2011. “Adverse Effects of Biologics: A Network Meta-Analysis and Cochrane Overview.” *The Cochrane Library*. Wiley Online Library.

Smolen, JS, FC Breedveld, MH Schiff, JR Kalden, P Emery, G Eberl, PL Van Riel, and P Tugwell. 2003. “A Simplified Disease Activity Index for Rheumatoid Arthritis for Use in Clinical Practice.” *Rheumatology* 42 (2). Oxford University Press: 244–57.

Souto, Alejandro, José Ramón Maneiro, and Juan J Gómez-Reino. 2015. “Rate of Discontinuation and Drug Survival of Biologic Therapies in Rheumatoid Arthritis: A Systematic Review and Meta-Analysis of Drug Registries and Health Care Databases.” *Rheumatology* 55 (3). Oxford University Press: 523–34.

Stephens, Stephanie, Marc F Botteman, Mary A Cifaldi, and Ben A van Hout. 2015. “Modelling the Cost-Effectiveness of Combination Therapy for Early, Rapidly Progressing Rheumatoid Arthritis by Simulating the Reversible and Irreversible Effects of the Disease.” *BMJ Open* 5 (6). British Medical Journal Publishing Group: e006560.

Stevenson, Matt D, Allan J Wailoo, Jonathan C Tosh, Monica Hernandez-Alava, Laura A Gibson, John W Stevens, Rachel J Archer, et al. 2017. “The Cost-Effectiveness of Sequences of Biological Disease-Modifying Antirheumatic Drug Treatment in England for Patients with Rheumatoid Arthritis Who Can Tolerate Methotrexate.” *The Journal of Rheumatology*. The Journal of Rheumatology, jrheum–160941.

Stevenson, Matt, Rachel Archer, Jon Tosh, Emma Simpson, Emma Everson-Hock, John Stevens, Monica Hernandez-Alava, et al. 2016. “Adalimumab, Etanercept, Infliximab, Certolizumab Pegol, Golimumab, Tocilizumab and Abatacept for the Treatment of Rheumatoid Arthritis Not Previously Treated with Disease-Modifying Antirheumatic Drugs and After the Failure of Conventional Disease-Modifying Antirheumatic Drugs Only: Systematic Review and Economic Evaluation.” *Health Technology Assessment* 20 (35). NIHR Health Technology Assessment Programme: 1–610.

Strand, V, S Williams, PSJ Miller, K Saunders, S Grant, and JM Kremer. 2013. “OP0064 Discontinuation of Biologic Therapy in Rheumatoid Arthritis (Ra): Analysis from the Consortium of Rheumatology Researchers of North America (Corrona) Database.” *Annals of the Rheumatic Diseases* 72 (Suppl 3). BMJ Publishing Group Ltd: A71–A72.

Thokala, Praveen, Nancy Devlin, Kevin Marsh, Rob Baltussen, Meindert Boysen, Zoltan Kalo, Thomas Longrenn, et al. 2016. “Multiple Criteria Decision Analysis for Health Care Decision Making-an Introduction: Report 1 of the Ispor Mcda Emerging Good Practices Task Force.” *Value in Health* 19 (1). Elsevier: 1–13.

Tosh, Jonathan, Alan Brennan, Allan Wailoo, and Nick Bansback. 2011. “The Sheffield Rheumatoid Arthritis Health Economic Model.” *Rheumatology* 50 (suppl 4). Br Soc Rheumatology: iv26–iv31.

Van Hout, Ben A, Maiwenn J Al, Gilad S Gordon, and Frans FH Rutten. 1994. “Costs, Effects and c/E-Ratios Alongside a Clinical Trial.” *Health Economics* 3 (5). Wiley Online Library: 309–19.

Wailoo, A, A Brennan, N Bansback, R Nixon, F Wolfe, and K Michaud. 2006. “Modeling the Cost Effectiveness of Etanercept, Adalimumab and Anakinra Compared to Infliximab in the Treatment of Patients with Rheumatoid Arthritis in the Medicare Program.” *Rockville, MD: Agency for Healthcare Research and Quality*.

Wailoo, Allan J, Nick Bansback, Alan Brennan, Kaleb Michaud, Richard M Nixon, and Fred Wolfe. 2008. “Biologic Drugs for Rheumatoid Arthritis in the Medicare Program: A Cost-Effectiveness Analysis.” *Arthritis & Rheumatology* 58 (4). Wiley Online Library: 939–46.

Wolfe, Frederick, and Kaleb Michaud. 2010. “The Loss of Health Status in Rheumatoid Arthritis and the Effect of Biologic Therapy: A Longitudinal Observational Study.” *Arthritis Research & Therapy* 12 (2). BioMed Central: R35.

Wolfe, Frederick, Kaleb Michaud, Hyon K Choi, and Rhys Williams. 2005. “Household Income and Earnings Losses Among 6,396 Persons with Rheumatoid Arthritis.” *The Journal of Rheumatology* 32 (10). The Journal of Rheumatology: 1875–83.

Wolfe, Frederick, Kaleb Michaud, Olaf Gefeller, and Hyon K Choi. 2003. “Predicting Mortality in Patients with Rheumatoid Arthritis.” *Arthritis & Rheumatism* 48 (6). Wiley Online Library: 1530–42.

Yazici, Yusuf, Svetlana Krasnokutsky, Jaime P Barnes, Patricia L Hines, Jason Wang, and Lisa Rosenblatt. 2009. “Changing Patterns of Tumor Necrosis Factor Inhibitor Use in 9074 Patients with Rheumatoid Arthritis.” *The Journal of Rheumatology* 36 (5). The Journal of Rheumatology: 907–13.

Zhang, Jie, Ying Shan, George Reed, Joel Kremer, Jeffrey D Greenberg, Scott Baumgartner, and Jeffrey R Curtis. 2011. “Thresholds in Disease Activity for Switching Biologics in Rheumatoid Arthritis Patients: Experience from a Large Us Cohort.” *Arthritis Care & Research* 63 (12). Wiley Online Library: 1672–9.

1. We thank Darius Lakdawalla, Jason Shafrin, Mark Linthicum, Thanh Ton, Jeffrey Curtis, Carole Wiedmeyer, and members of the Innovation and Value Initiative’s scientific advisory panel for reviewing this document. In addition, we thank Sam Norton for help with the latent class growth model. Finally, we thank Maria Lorenzi, Andrew Frederickson, Jacky Zhang, and Mark Embrett for conducting the systematic literature review and running the network meta-analysis. [↑](#footnote-ref-21)
2. This report should be referenced as follows: Incerti, D, Jansen, JP. A Description of the IVI-RA Model. 2017; last updated October 2017. [↑](#footnote-ref-22)
3. [Innovation and Value Initiative](http://www.thevalueinitiative.org/) [↑](#footnote-ref-23)