A Description of the IVI-RA Model

Devin Incerti* Jeroen P. Jansen†

June 21, 2017

Contents

*Innovation and Value Initiative

†Innovation and Value Initiative

List of Figures

List of Tables

1 Overview

This document describes version 1.0 of IVI's "family" of rheumatoid arthritis cost-effectiveness models. The family of models consists of a set of separate models with different underlying structural assumptions. Although the underlying modeling assumptions vary, each model is an individual patient simulation (IPS), which allows us to model treatment and disease progression as realistically as possible.

The IVI026 family of models is available as an R package, but the majority of the code for the IPS is written in C++ so that the models can be simulated in a reasonable time frame. Documentation is available online at `https://www.google.com`.

This document is structured as follows. Section 2 outlines the competing model structures. Section 3 describes the statistical techniques used to estimate the model parameters and the data sources used. Section 4 describes the simulation techniques used to implement the RA family of models and quantify uncertainty. Finally, Section 5 validates the different components of the models and examines the implications of the alternative structural assumptions.

2 Competing model structures

The competing model structures are based on the Sheffield RA health economic model (Tosh et al. 2011) and the IPS model by Stevenson et al. (2016) recently used by NICE in the United Kingdom.

The models simulate treatment sequence of arbitrary length. bDMARDS that can be included in the sequence include tumor necrosis factor (TNF) inhibitors (etanercept, adalimumab, certolizumab, golimumab), non-TNF inhibitors (abatecept, tocilizumab, rituximab), and Janus kinase/signal transducers and activators of transcription (JAK/STAT) inhibitors (tofacitinib). Conventional DMARDS (cDMARDs) such as methotrexate can also be included in a treatment sequence.

At the start of the model each patient is assigned a baseline Health Assessment Questionnaire (HAQ) Disability Index score. Subsequently, the impact of the disease measured by the HAQ trajectory over time is modeled as a function of sequenced treatments (Figure 1). In the absence of treatment, HAQ deteriorates at a certain rate as depicted by the dashed line in the figure. 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. During the initial treatment phase HAQ is modeled as a change from baseline. In the maintenance

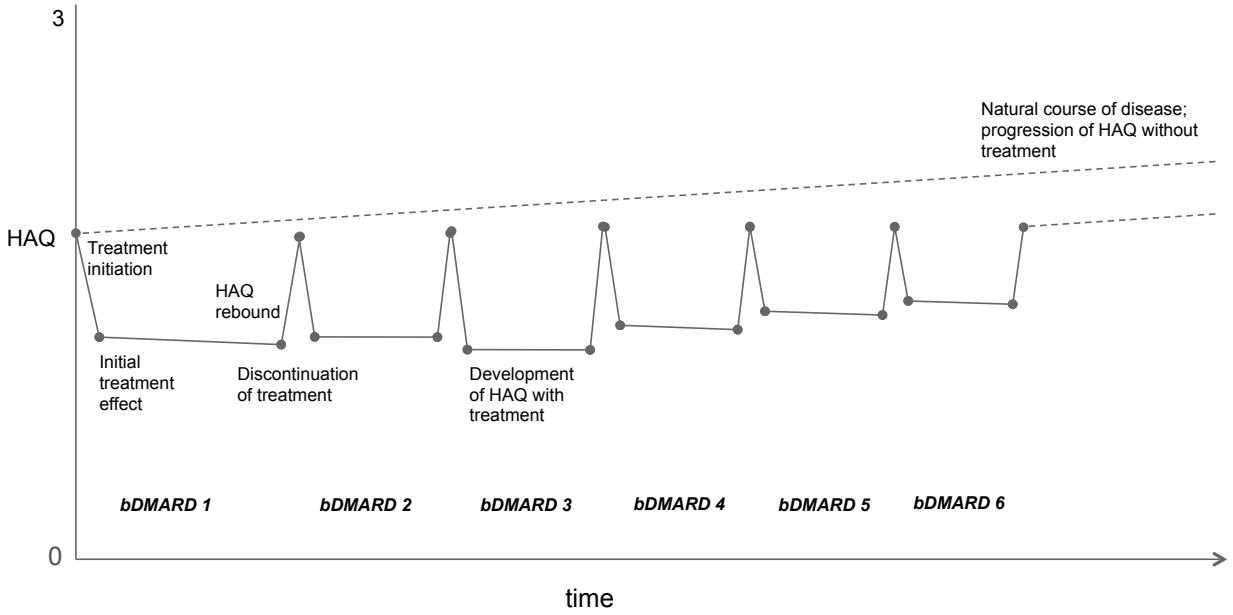

Figure 1: Model structure regarding development of HAQ with sequential biologic treatment

phase, statistical models are used to simulate the progression of HAQ over time. Two possibilities are considered: a scenario in which HAQ progression is linear and a second scenario where HAQ progression is simulated using a non-linear mixture model.

Patients with insufficient response at the end of the initial 6-month period switch to the subsequent bDMARD, whereas those with a successful response continue with the same treatment as maintenance regimen until efficacy is lost or treatment is discontinued for other reasons. Most clinical trials use the American College of Rheumatology (ACR) 20/50/70 response criteria to assess efficacy. However, in routine practice ACR response criteria are not used; decisions regarding treatment switching are more likely to be informed by response criteria based on SDAI in the US and EULAR response in the UK. To capture differences in clinical practice in different settings we will model separate scenarios where treatment duration depends on either EULAR response or SDAI (Tosh et al. 2011; Madan et al. 2015). In the SDAI scenario patients with a SDAI major or minor response continue with maintenance treatment and in the EULAR scenario patients with a good or moderate response continue with maintenance treatment. In both cases, patients without a response switch to the next bDMARD. Duration of the treatment maintenance phase is modeled from real-world data using time-to-event distributions conditional upon SDAI or EULAR responses at 6 months. Patients with a better treatment response according to SDAI or EULAR tend to stay on maintenance treatment longer.

The effect of treatment on HAQ during the initial treatment phase is modeled using two separate scenarios. In the first scenario, treatment has a direct effect on HAQ, and in the second scenario treatment has an indirect effect on HAQ through its effect on ACR response. In both scenarios, more effective treatments are associated with improvements (i.e. reductions) in HAQ relative to the score 6 months prior. In addition, these patients, who continue with the same bDMARD during the maintenance phase, experience an improvement in HAQ relative to patients without treatment. 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. Given the more favorable

HAQ progression rates with maintenance bDMARD than without treatment, the absolute HAQ score after rebound is at a more favorable level than what the HAQ would have been in the absence of treatment.

In line with 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). While on a treatment, 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.

Baseline HAQ scores and changes in HAQ over time determine mortality relative to age/sex specific rates for the US general population. Furthermore, the individual HAQ score in combination with matching pain levels and age/sex at a particular point in time is used to estimate the EQ-5D utility score (0-1 range). Utility is reduced for the duration of a month when a serious infection occurred. Accrued utility estimates among patients alive determine estimated QALYs. HAQ scores also determine annual hospitalization days and productivity loss. 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 considered are limited to work related productivity loss.

Patient preferences for treatment attributes have a direct effect on long-term treatment duration and utility. Patients with treatments that more closely match their preferences have longer treatment duration and higher utility. Treatment attributes that are incorporated into the models include route of administration and frequency of administration.

The flow diagram in Figure 2 describes the flow of a single patient through the simulation. Each patient begins the simulation by initiating treatment and ends the simulation with 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.

The influence diagram in Figure 3 summarizes the assumed structural relationships among the different parameters. Each arrow represents the direct effect of one parameter on another. Dashed lines represent relationships that depend on the structural modeling assumptions used. Different stakeholder perspectives can be considered based on whether productivity is included in the analysis: a societal perspective includes productivity and a health-sector perspective does not.

Table 1 summarizes the competing model structures, which are conditional on the perspective of the decision maker. First, the initial treatment response can either have a direct effect on HAQ or have an indirect effect on HAQ through its effect on ACR (and the association between ACR and the response criteria used). Second, the response criteria can either be based on SDAI or EULAR, which impacts treatment duration. Third, HAQ can either be linear or non-linear. Fourth, seven probability distributions (exponential, Weibull, Gompertz, gamma, log-logistic, lognormal, and generalized gamma) can be used to model treatment duration conditional on a positive treatment response at 6-months. All told, there are 2 x 2 x 2 x 7 = 56 possible models.

For a given model, the outcomes of interest are:

- HAQ over time
- Cumulative QALYs
- Total drug acquisition and administration costs

Figure 2: Flow diagram of the simulation for a single patient

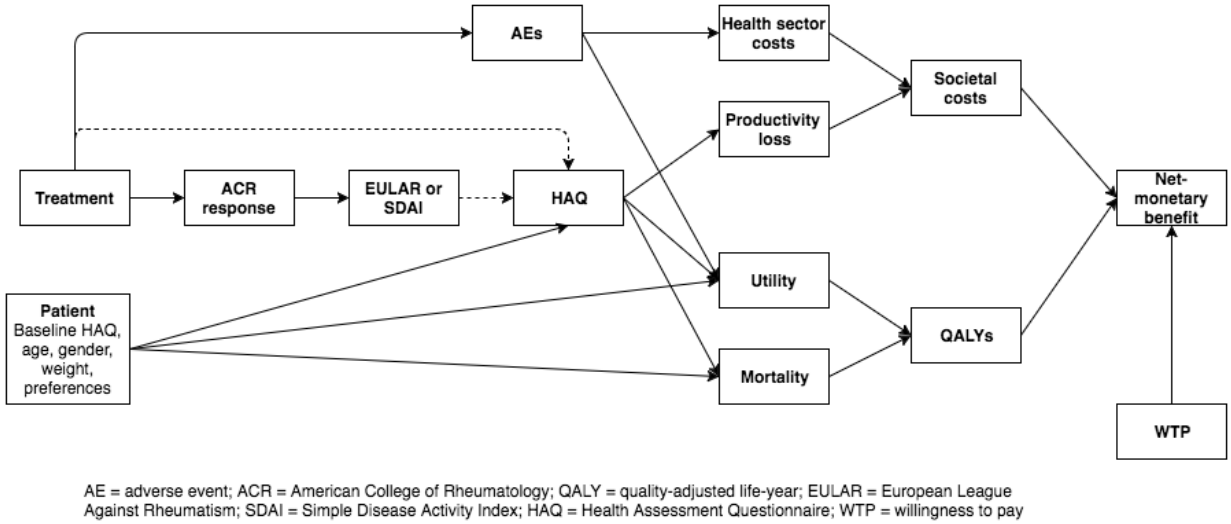

Figure 3: Influence diagram outlining structural relationships between model parameters

Table 1: Sets of models within the family of models

Component of model structure	Possible combinations
Initial treatment response based on indirect effect of treatment on HAQ through ACR or direct effect of treatment on HAQ	2
Treatment duration based on SDAI or EULAR	2
HAQ progression linear or non-linear	2
Probability distribution for treatment duration	7

- General management and monitoring costs
- Adverse event costs
- Hospitalization costs
- Total health sector costs
- Productivity loss
- Total costs
- Net-monetary benefit (NMB)

3 Source data and parameter estimation

3.1 HAQ progression in the absence of bDMARD treatment

The natural course of HAQ progression in the absence of bDMARDs is assumed to either be linear or non-linear. HAQ develops over time according to this estimated natural course for patients remaining on cDMARDs or following discontinuation of the last bDMARD of the sequence.

The rate of progression in the linear case is based on the observational study by 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 bDMARD.

Based on the same data, Michaud et al. (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); $\geq$ 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.

Table 2: Anuual linear progression of HAQ in the absense of bDMARDs beyond 6 months

		95% CI		
	Estimate	Lower	Upper	Reference
Overall progression rate				
MTX or non-biologic treatment	0.031	0.026	0.036	Wolfe and Michaud (2010)
Change in overall progression rate by age				
<40	-0.020	-0.028	-0.012	Michaud et al. (2011)
40-64	-0.008	-0.010	-0.006	Michaud et al. (2011)
65+	0.017	0.013	0.021	Michaud et al. (2011)

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 et al (2011).

The rate of progression in the non-linear case is based on a mixture model approach that has increasingly been used to model HAQ progression over time (Stevenson et al. 2016; Norton et al. 2013, 2014). These models suggest that different subgroups have distinct HAQ trajectories and that the rate of worsening of HAQ progression decreases over time. Parameter estimates are based on We use the statistical model estimated on the Early Rheumatoid Arthritis Cohort Study (ERAS) cohort, which has a high percentage of patients receiving methotrexate and a very small percentage receiving biologics. Following Stevenson et al. (2016), explanatory variables in the statistical model that are not used in the IPS will be set to their mean values in the ERAS cohort. Uncertainty in the parameters of the mixture model are based on standard errors since Norton et al. (2013) did not report the full covariance matrix needed to model the covariance between the parameters.

3.2 Initial treatment effect

Treatment effects for bDMARD *naive* patients at 6 months are estimated using a Bayesian network meta-analysis (NMA) of published randomized controlled trials (RCTs) (see Section B.1). Two primary endpoints were considered: ACR 20/50/70 response and HAQ. The rationale for using ACR response rather than HAQ directly is that the evidence base is larger than for HAQ. In one version of the model structure, EULAR response categories are used to determine treatment continuation and duration. ACR responses from the NMA were translated into EULAR response probabilities based on evidence of their relationship reported in Stevenson et al. (2016) and obtained from the US Veterans Affairs Rheumatoid Arthritis (VARA) registry (Table 3).

Table 3: Relationship between ACR response and EULAR response

	EULAR response		
ACR response	None	Moderate	Good
<20	755	136	57
20-50	4	27	26
50-70	2	2	10
70+	0	2	2

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

The model structure allows treatment to impact HAQ either directly or indirerectly through its effect on ACR response. In the indirect case, ACR response first affects EULAR response or SDAI using the evidence reported in Table 3, which, in turn, affects HAQ. The relationship between EULAR response and HAQ is based on analyses conducted by Stevenson et al. (2016) using the BSRBR databse. Their analysis is based on predictions from a mixture model with covariates set to sample means. Moderate and good EULAR response are associated with -0.317 (SE = 0.048) and -0.672 (SE = 0.112) changes in HAQ scores respectively. The impact of treatment on ACR response, EULAR response, and HAQ (given their ACR and EULAR responses) is shown in Table 4.

Table 4: Response at 6 months for 1st line treatment

	ACR response				EULAR response			Mean HAQ
Treatment	<20	20-50	50-70	70+	None	Moderate	Good	decrease
cDMARDs	0.70 (0.68-0.73)	0.17 (0.15-0.20)	0.08 (0.07-0.10)	0.04 (0.03-0.05)	0.58 (0.55-0.62)	0.21 (0.17-0.26)	0.20 (0.16-0.24)	0.20 (0.00-0.79)
ABT IV + MTX	0.40 (0.30-0.49)	0.24 (0.21-0.27)	0.19 (0.15-0.22)	0.17 (0.12-0.25)	0.36 (0.27-0.45)	0.29 (0.20-0.39)	0.35 (0.26-0.47)	0.33 (0.00-0.84)
ADA + MTX	0.41 (0.32-0.51)	0.24 (0.21-0.27)	0.18 (0.15-0.22)	0.16 (0.11-0.22)	0.37 (0.29-0.45)	0.28 (0.20-0.38)	0.35 (0.25-0.45)	0.32 (0.00-0.84)
ETN + MTX	0.37 (0.27-0.47)	0.24 (0.22-0.27)	0.19 (0.16-0.23)	0.19 (0.13-0.27)	0.34 (0.25-0.42)	0.29 (0.20-0.40)	0.37 (0.26-0.49)	0.34 (0.00-0.84)
GOL + MTX	0.39 (0.28-0.51)	0.24 (0.21-0.27)	0.19 (0.15-0.23)	0.18 (0.11-0.26)	0.36 (0.26-0.45)	0.29 (0.20-0.40)	0.36 (0.25-0.48)	0.33 (0.00-0.84)
IFX + MTX	0.41 (0.30-0.52)	0.24 (0.21-0.27)	0.18 (0.14-0.22)	0.16 (0.10-0.25)	0.37 (0.28-0.47)	0.28 (0.20-0.38)	0.35 (0.25-0.46)	0.32 (0.00-0.84)
TCZ + MTX	0.41 (0.23-0.61)	0.24 (0.20-0.27)	0.18 (0.11-0.24)	0.17 (0.07-0.32)	0.37 (0.22-0.52)	0.28 (0.19-0.40)	0.35 (0.22-0.50)	0.32 (0.00-0.83)
CZP + MTX	0.26 (0.18-0.34)	0.23 (0.20-0.26)	0.22 (0.19-0.25)	0.29 (0.21-0.37)	0.25 (0.18-0.34)	0.32 (0.20-0.47)	0.42 (0.29-0.57)	0.39 (0.00-0.85)
ABT SC + MTX	0.39 (0.30-0.49)	0.24 (0.21-0.27)	0.19 (0.15-0.23)	0.18 (0.11-0.25)	0.36 (0.27-0.44)	0.29 (0.20-0.39)	0.36 (0.26-0.47)	0.33 (0.00-0.84)
RTX + MTX	0.44 (0.32-0.57)	0.24 (0.20-0.27)	0.17 (0.12-0.22)	0.15 (0.08-0.23)	0.39 (0.29-0.50)	0.27 (0.20-0.37)	0.33 (0.24-0.45)	0.31 (0.00-0.83)
TOF + MTX	0.42 (0.31-0.52)	0.24 (0.21-0.27)	0.18 (0.14-0.22)	0.16 (0.10-0.24)	0.37 (0.28-0.47)	0.28 (0.20-0.38)	0.34 (0.25-0.46)	0.32 (0.00-0.83)

Notes: 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. cDMARDs = conventional disease-modifying antiheumatic 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; EULAR = European League Against Rheumatism; HAQ = Health Assessment Questionnaire.

The relationship between ACR response and SDAI...

In the direct case...

3.3 HAQ trajectory with bDMARD maintenance treatment

Based on the NDB longitudinal study, 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 et al. (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 bDMARDs as reported by Michaud et al. (2011) can be directly compared with the overall annual HAQ progression rate of 0.031 reported by Wolfe and Michaud (2010). Accordingly, bDMARD specific HAQ progression rates by Michaud et al. (2011) are used in the model. For bDMARD treatments evaluated in the model for which no HAQ progression rate was reported by Michaud et al. (2011), the overall biologic rate of -0.001 is used.

3.4 Duration of maintenance treatment

3.4.1 By EULAR response

Treatment duration as a function of EULAR response is estimated from survival curves based on analyses of the British Society for Rheumatology Biologics Registers (BSRBR) database (Stevenson et al. 2016). Seven parametric survival models (exponential, Weibull, Gompertz, gamma, log-logistic, lognormal, and generalized gamma) were estimated on individual patient data reconstructed from the BSRBR survival curves using the algorithm developed in Guyot et al. (2012). The Akaike Information Criteria (AIK) and Bayesian Information Criteria (BIC) of each model by EULAR response category (moderate, good) are shown in Table 5.

Table 5: AIC and BIC for parametric models of treatment duration by EULAR response

	Moderate EULAR response		Good EULAR response	
Distribution	AIC	BIC	AIC	BIC
Exponential	38,840	38,847	15,126	15,132
Weibull	38,478	38,492	15,090	15,101
Gompertz	38,099	38,112	15,066	15,077
Gamma	38,587	38,600	15,098	15,110
Log-logistic	38,142	38,155	15,062	15,073
Lognormal	37,988	38,001	15,047	15,059
Generalized gamma	37,869	37,889	15,048	15,065

One concern is that the BSRBR is representative of the UK but not the US. As a result, we also estimate "adjusted" survival models that are more representative of the US. The adjustment is made in six steps based on an analysis of the Consortium of Rheumatology Researchers of North America (Corrona) database.

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 $h(t)_{corrona}$.

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 $5,492$ and the number of good responders is $2,417$ the overall hazard function is $h(t)_{bsrbr} = \frac{5,492}{7,909} h(t)_{bsrbr,moderate} + \frac{2,417}{7,909} h(t)_{bbsrbr,good}$.

3. At each point in time, calculate the ratio of the Corrona and BSRBR hazard functions: $HR(t) = h(t)_{corrona}/h(t)_{bbsrbr}$.

4. Apply the hazard ratio in step 3 to the BSRBR hazard functions for each EULAR response category. That is $h(t)_{bsrbr,moderate,adj} = h(t)_{bsrbr,moderate} \cdot HR(t)$ and $h(t)_{bsrbr,good,adj} = h(t)_{bsrbr,good} \cdot HR(t)$.

5. Generate survival curves using the hazard functions from step 4. Specifically, given a general hazard function $h(t)$, calculate the cumulative hazard functions, $H(t) = \int_{z=0}^{t} h(z)dz$, convert this to a survival function using $S(t) = exp(-H(t))$, 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 Figure 4. AIC and BIC for the parametric models fit in step 6 do the adjusted individual patient data are shwon in Table 6.

Table 6: AIC and BIC for Corrona adjusted parametric models of treatment duration by EULAR response

	Moderate EULAR response		Good EULAR response	
Distribution	AIC	BIC	AIC	BIC
Exponential	42,304	42,310	18,098	18,103
Weibull	41,946	41,959	18,051	18,062
Gompertz	41,569	41,582	18,039	18,050
Gamma	42,098	42,111	18,063	18,074
Log-logistic	41,406	41,419	18,037	18,049
Lognormal	41,235	41,248	18,004	18,016
Generalized gamma	41,110	41,129	18,000	18,017

3.4.2 By SDAI

TBD

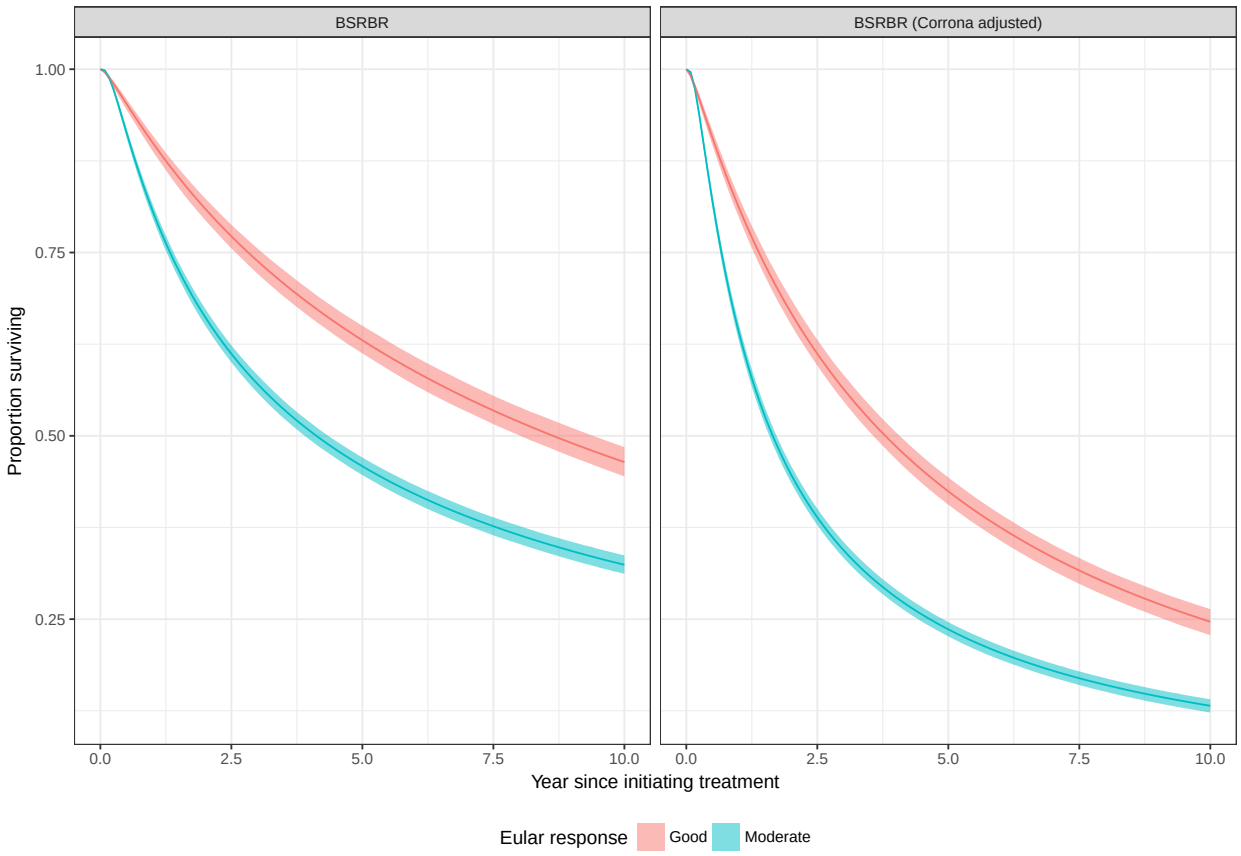

Figure 4: Generalized gamma survival curve of treatment duration using reconstructed individual patient data based on analyses from Stevenson et al. (2016) by EULAR response category

3.5 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. 0.5 is used as a lower bound based on expert opinion.

3.6 Serious infections

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

Table 7: Probability of serious infection

	Probability		
		95% CI	
	Mean	Lower	Upper
MTX or non-biologic treatment	0.0608	0.0410	0.0810
Biologic treatment	0.1148	0.0830	0.1520

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

Table 8: Probability of serious infection with methotrexate by distribution used to model treatment duration

Distribution	Mean probability
Exponential	0.0606
Weibull	0.0606
Gompertz	0.0608
Gamma	0.0611
Log-logistic	0.0612
Lognormal	0.0604
Generalized gamma	0.0608

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

3.7 Utility

Alava et al. (2013) developed a non-linear mixture model relating EQ-5D utility to HAQ, pain and age/sex. We simulate this mixture model for every patient in the model to obtain the distribution of utility over time. Since pain is not explicitly captured in our cost-effectiveness model, an individual's

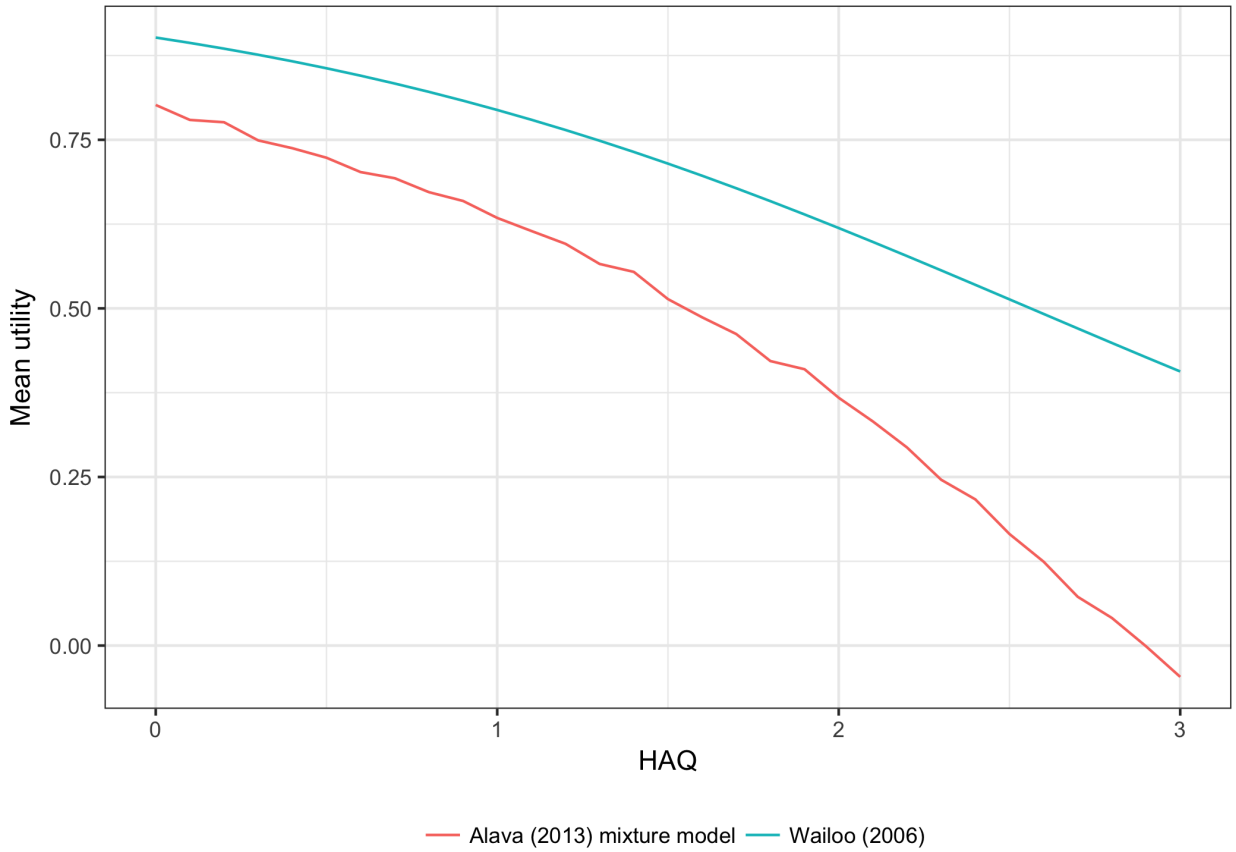

Figure 5: Simulated mean utility by current HAQ

pain score is first sampled given that individual's HAQ score and the stochastic relationship between pain and HAQ. Disutility due to serious infections is assumed to be 0.156 for the duration of the month of infection based on prior studies (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.

Figure 5

3.8 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. 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 et al. (2012).

Figure 6

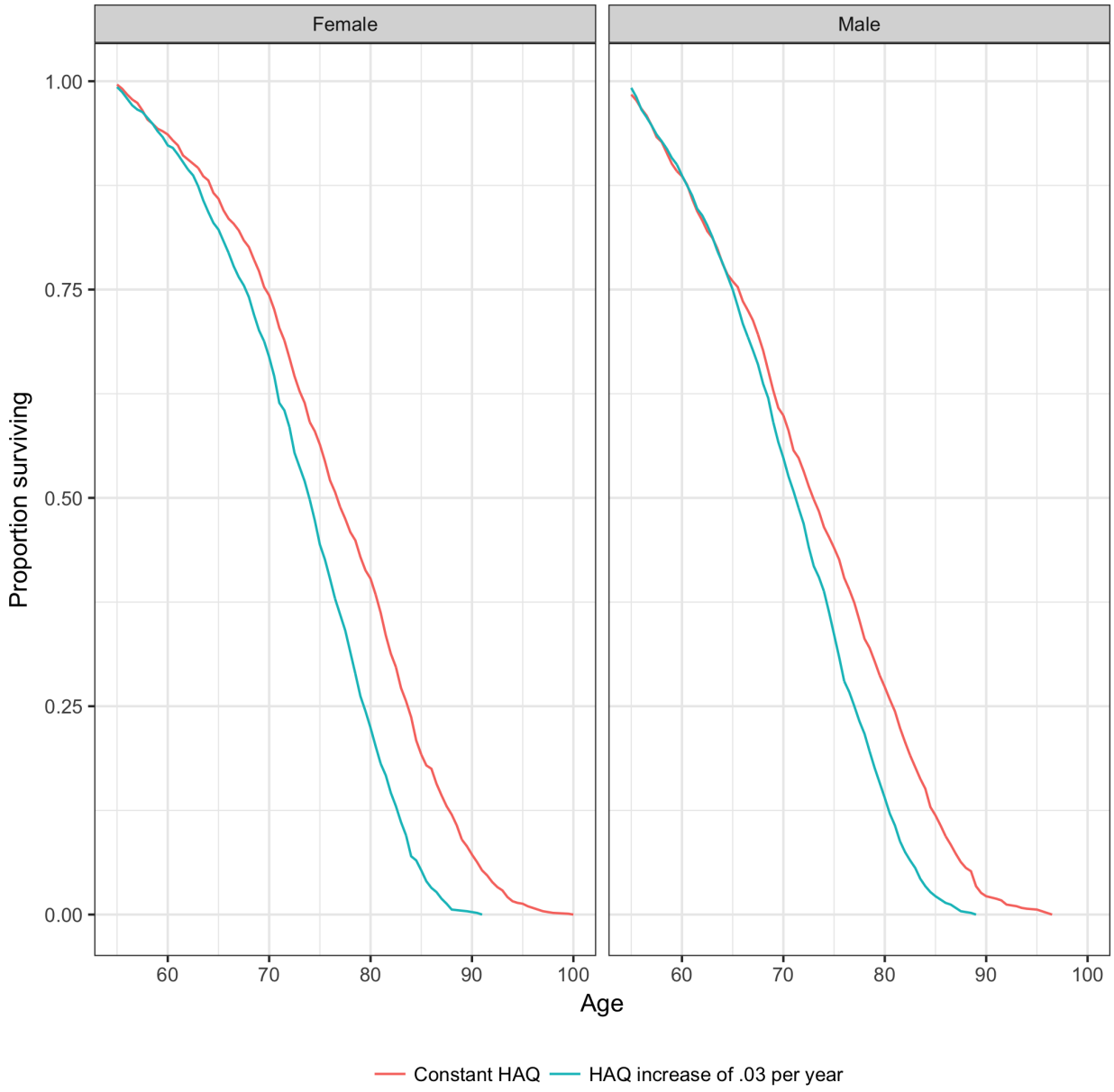

Figure 6: Simulated survival curve for a patient age 55 with a baseline HAQ of 1 by change in HAQ per year

Table 9: Mortality parameters

	Estimate	95% CI Lower	95% CI Upper	Reference
Impact of baseline HAQ on mortality				
Log odds of mortality	0.798	0.586	1.009	Wolfe et al. (2003)
Impact of change in HAQ from baseline on mortality				
Log hazard ratio 0-6 months	0.113	0.077	0.157	Michaud et al. (2012)
Log hazard ratio >6-12 months	0.148	0.104	0.191	Michaud et al. (2012)
Log hazard ratio >12-24 months	0.148	0.095	0.191	Michaud et al. (2012)
Log hazard ratio >24-36 months	0.191	0.131	0.247	Michaud et al. (2012)
Log hazard ratio >36 months	0.174	0.104	0.239	Michaud et al. (2012)

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

3.9 Cost

Drug costs are based on WACs; discounts can be applied by reducing WACs for specific bDMARDs. Costs related to physician visits, chest X-rays, tuberculosis tests, and inpatient hospital days are based on Claxton et al. (2016). The annual number of hospital days relates to the HAQ score according to 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. 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. These estimates are inflated to 2016 dollars.

Table 10: Resource use parameters

	Estimate	95% CI Lower	95% CI Upper	Reference
Days in hospital per year				
HAQ: 0-<0.5	0.260	0.000	1.725	Carlson et al. (2015)
HAQ: 0.5-<1	0.130	0.000	1.409	Carlson et al. (2015)
HAQ: 1-<1.5	0.510	0.015	1.850	Carlson et al. (2015)
HAQ: 1.5-<2	0.720	0.092	1.979	Carlson et al. (2015)
HAQ: 2-<2.5	1.860	1.013	2.960	Carlson et al. (2015)
HAQ: >2.5	4.160	3.238	5.196	Carlson et al. (2015)
Cost per day in hospital	1,251	904	1,652	Carlson et al. (2015)
General management cost				
Chest x-ray	109	97	121	Claxton et al. (2016)
X-ray visit	53	45	61	Claxton et al. (2016)
Outpatient follow-up	187	159	215	Claxton et al. (2016)
Mantoux tuberculin skin test	30	30	30	Claxton et al. (2016)
Productivity loss				
Linear regression coefficient - HAQ	5,853	2,861	8,845	Wolfe et al. (2005)

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 Claxton et al (2016). 95% confidence interval for productivity loss are calculated using a normal distribution and inflated to 2016 dollars.

Table 11: Drug acquisition and administration cost

Drug	Dose and frequency of administration	Strength and dosage form	Number of doses first 6 months	Number of dosees per year beyond the first 6 months	Wac per unit	Infusion cost	Cost for the first 6 months	Cost per year beyond the first 6 months
Etanercept	50 mg QW	50 mg/0.98 mL syringe or pen injector	26	52	1,110.50	0	28,873	57,746
Adalimumab	40 mg EOW	40 mg/0.8 mL syringe or pen injector	13	26	2,220.62	0	28,868	57,736
Infliximab	3 mg/kg at 0, 2, and 6 weeks, 3mg/kg Q8W, 6 mg/kg Q6W after 6 months	100 mg vial	5	8	1,113.27	164	17,519	51,853
Golimumab	50 mg QM	50 mg/0.5 mL syringe or pen injector	6	12	3,811.18	0	22,867	45,734
Certolizumab pegol	400 mg at weeks 0, 2, 4 then 200 mg Q2W	400 mg kit or syringe kit (200 mg 2)	8	26	3,679.87	0	29,438	47,838
Abatacept IV	750 mg IV at weeks 0, 2, 4 then Q4W	250mg vial	8	13	931.16	164	23,659	38,447
Abatacept SC	125 mg SC QW with IV loading dose	125mg/ml syringe	26	52	957.14	0	24,885	49,771
Tocilizumab	162 mg SC EOW	162 mg/0.9 mL syringe	13	26	898.31	0	11,678	23,356
Rituximab	1000 mg at weeks 0, 2; then Q24 W	500 mg/50ml vial	4	4	4,176.10	164	34,064	36,903
Tofaticinib citrate	5 mg BID	5mg tablet	364	728	63.26	0	23,026	46,053
Methotrexate monotherapy	15mg QW	15 mg injection	26	52	32.42	0	842	1,685
Hydroxychlorquine sulfate	400mg daily	200 mg tablet	182	364	3.18	0	1,157	2,315
Sulfazalazine	1-2 g daily	500 mg tablet	182	364	0.47	0	342	684

Notes: Costs do not include rebates or discounts. Cost for infliximab are calculated by assuming that ‘r male.prop‘% of patients are male and that the weight of men and women are ‘r wtmale‘ kg and ‘r wtfemale‘ 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; WAC = whoesale acqusition cost.

3.10 Patient preferences for treatment attributes

4 Simulation and uncertainty analysis

4.1 Individual patient simulation

The IPS is a discrete-time simulation that simulates individual patients one at a time. Model cycle, denoted by t, were chosen to be 6-months long to be consistent with most RCT and real-world data evidence. Algorithm 1 describes the main components of the IPS for a single patient and a given treatment in a treatment sequence. The full simulation cycles through each treatment in a sequence and through each simulated patient.

Algorithm 1 Main components of the individual patient simulation

1. **Initial treatment effect** ($t = 0$)
   (a) Simulate clinical response (SDAI or EULAR), time to serious infection T_{si}, and death.
      i. **If** no clinical response, **then** stop treatment. Treatment switch caused by a serious infection if time to serious infection occured during cycle 0 (i.e. $T_{si} = 0$). Change in HAQ is assumed to be 0.
      Else if clinical response, **then** continue treatment. Simulate change in HAQ and time to treatment discontinuation T.
      ii. **If** patient died, **then** move to next patient.
2. **Maintenance phase** (for $t > 0$ and $t \leq T$)
   (a) Simulate death (see Section A.2) and change in HAQ.
   (b) **If** patient died, **then** move to next patient.
   (c) **If** $t = T$, **then** switch treatment. Treatment switch caused by a serious infection if time to serious infection occured during or before cycle T (i.e. $T_{si} \leq T$).

4.2 Parameter uncertainty

4.3 Structural uncertainty

4.4 Implementation

5 Model validation

Appendices

A Individual Patient Simulation

A.1 Effect of age on linear HAQ progression

A.2 Simulating death

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. Use the annual probability of death (q_x) from lifetables based on patient age and gender.
2. Adjust the probability of mortality, p_m, using odds of mortality, OR, of a change in baseline HAQ.

$$p_m = \frac{1}{1 + \exp\left[-(\log(q_x) + HAQ \cdot \log(OR))\right]} \tag{1}$$

3. Convert the mortality probability, p_m, into a mortality rate, r_m.

$$r_m = -log(1 - p_m) \tag{2}$$

4. Adjust the mortality rate using the estimated hazard ratio of mortality, HR of a change in HAQ from baseline, Δ HAQ.

$$r_m = r_m \cdot \exp[\log(HR) \cdot \Delta HAQ] \tag{3}$$

5. Convert the mortality rate into a probability given a 6-month cycle length.

$$p_m = 1 - \exp[-r_m * (6/12)] \tag{4}$$

6. Randomly draw a 0/1 death indicator, d, given the probability of death, p_m.

$$d \sim \text{Bin}(1, p_m) \tag{5}$$

A.3 Simulating utility

A.3.1 Mixture model

Utiliy was simulated in a two stages using the mixture model estimated by Alava et al. (2013). 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 $\mu_{pain} = 61.65$ and standard deviation $\sigma_{pain} = 19.10$, while HAQ was reported to have mean $\mu_{haq} = 1.39$ and standard deviation $\sigma_{haq} = 0.59$.

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 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 $\rho = 0.52$, was calculated by rearranging the OLS estimate for the slope, β, of the regression model,

$$\rho = \beta \cdot \frac{\sigma_{haq}}{\sigma_{pain}}. \tag{6}$$

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

$$pain|haq = h \sim N\left(\mu_{pain} + \rho\frac{\sigma_{pain}}{\sigma_{haq}}(h - \mu_{haq}), \sigma^2_{pain}(1 - \rho^2)\right). \tag{7}$$

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 c, the HAQ score for patient i in period t was modeled as,

$$y_{it|C_{it}} = \begin{cases} 1 & \text{if } y^*_{it|C_{it}} > 0.883 \\ y^*_{it|C_{it}} & \text{otherwise} \end{cases} \tag{8}$$

$$y^*_{it|C_{it}} = \alpha_{ic} + x^T_{it}\beta_c + \epsilon_{it} \tag{9}$$

$$\alpha_{ic} = \gamma_{ic} + z^T_i\gamma^0_i + \mu_i \tag{10}$$

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

$$P(C_{it} = c) = \frac{\exp(w^T_{it}\delta_c)}{\sum_{s=1}^{4}\exp(w^T_{it}\delta_s)}. \tag{11}$$

We sampled from the mixture model as follows.

1. For each individual i, sample the error term, $\mu_i \sim N(0, \sigma_\mu^2)$.

A.3.2 Wailoo utility algorithm

B Network Meta-Analysis

B.1 Bayesian NMA for initial treatment effects

B.1.1 Systematic literature review

Population

- Adult (>18 years) patients with moderate to severe RA who have had inadequate response to cDMARDs

Interventions and comparators

- Biologics as monotherapy or in combination with cDMARDs (adalimumab, certolizumab pegol, etanercept, golimumab, infliximab, abatacept, rituximab, tocilizumab, sarilumab, tofacitinib, baricitinib)
- Triple therapy (MTX, HCQ, and SSZ)
- cDMARDs alone or in combination (MTX, HCQ, SSZ or LEF)

Outcomes

- ACR20/ACR50/ACR70
- DAS28
- Total sharp score
- HAQ-DI score
- SF-36 PCS and MCS
- EQ-5D (VAS and utility scores)
- AEs leading to drop-outs
- Randomized controlled trials

Other

- Studies published in English
- Primary study available as full text published manuscript only; no study available as a conference abstract only was included with the exception of abstracts pertaining to investigational products, baricitinib and sarilumab

B.1.2 Criteria for studies to be selected from the systematic literature review and included in the NMA

The following criteria were used to select relevant studies to be included in the NMA:

Population

- Adult (>18 years) patients with moderate to severe RA who have had inadequate response to cDMARDs and are bDMARD-naive

Interventions

- Biologics as monotherapy or in combination with cDMARDs (adalimumab, certolizumab pegol, etanercept, golimumab, infliximab, abatacept, rituximab, tocilizumab, sarilumab, tofacitinib, baricitinib)

Comparators

- cDMARDs
- Any active comparator that allows for an indirect comparison between the bDMARDs of interest

Outcomes

- ACR20/ACR50/ACR70 at 6 months follow-up

B.1.3 Identified evidence base

Figure 7 summarizes the study identification and selection process. Of the 181 studies included in the large systematic literature review, 79 studies concerned the bDMARD-naive population (table NMA studies). There were 66 studies evaluating 36 interventions for which ACR response criteria were reported at 6 months (with a tolerability window of ±4 weeks). The corresponding evidence network is presented in Figure 8. For the network meta-analysis the following were deemed to be clinically equivalent and were pooled.

- “INF 3mg/kg q8w” or “INF 5mg/kg q8w” or “INF 6mg/kg q8w”
- “ETN 50mg qw” or “ETN 25mg biw”
- “ABA 10mg/kg q4wa or”ABA SC 125mg qw"
- “CER 200mg q2w+MTX” or “CER 400mg q4w+MTX
- DMARDs including methotrexate, sulfasalazine, hydroxychloroquine, leflunomide at any dosage; studies which only described DMARD therapy as conventional or nonbiologic

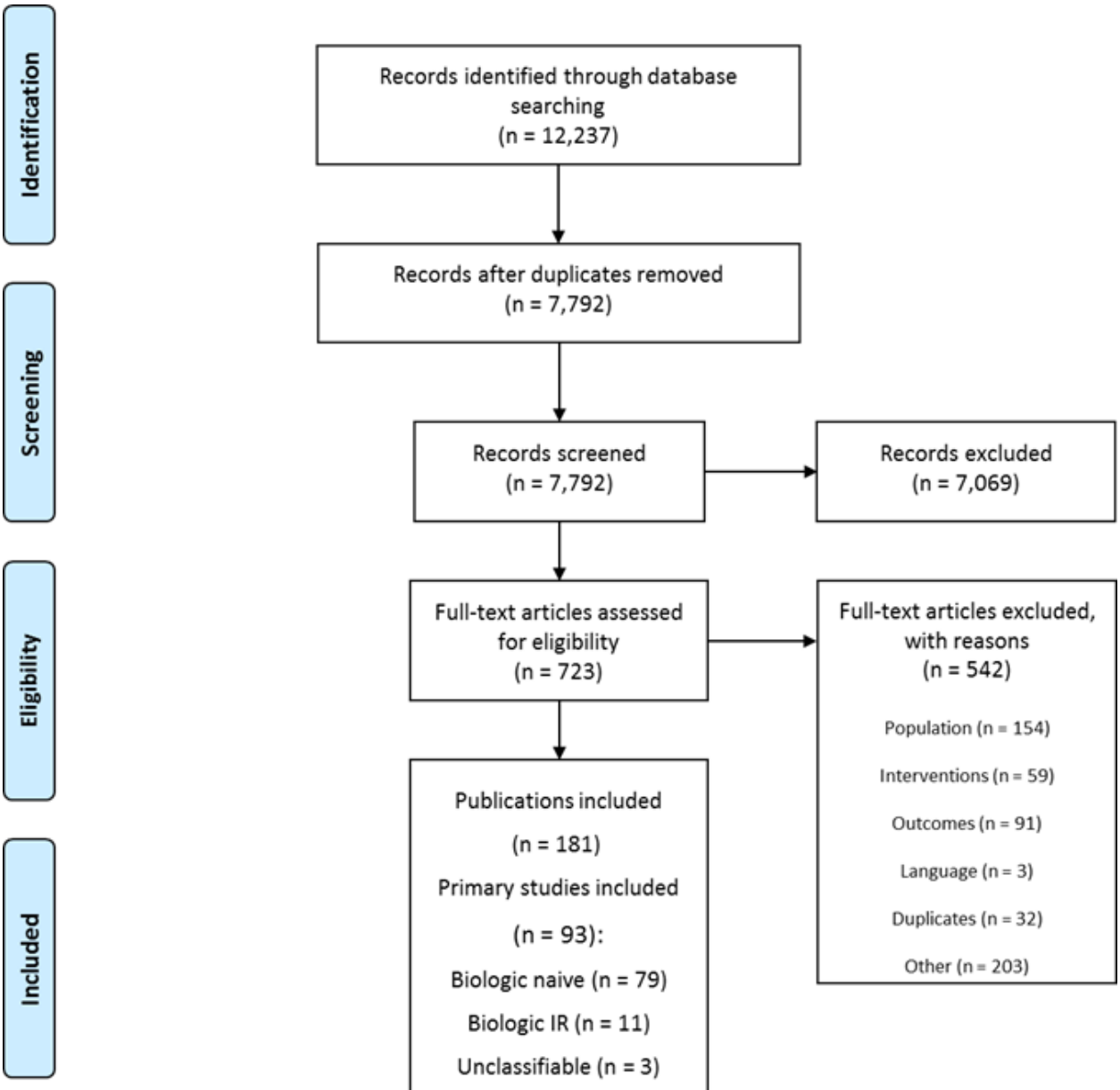

Figure 7: Study identification and selection

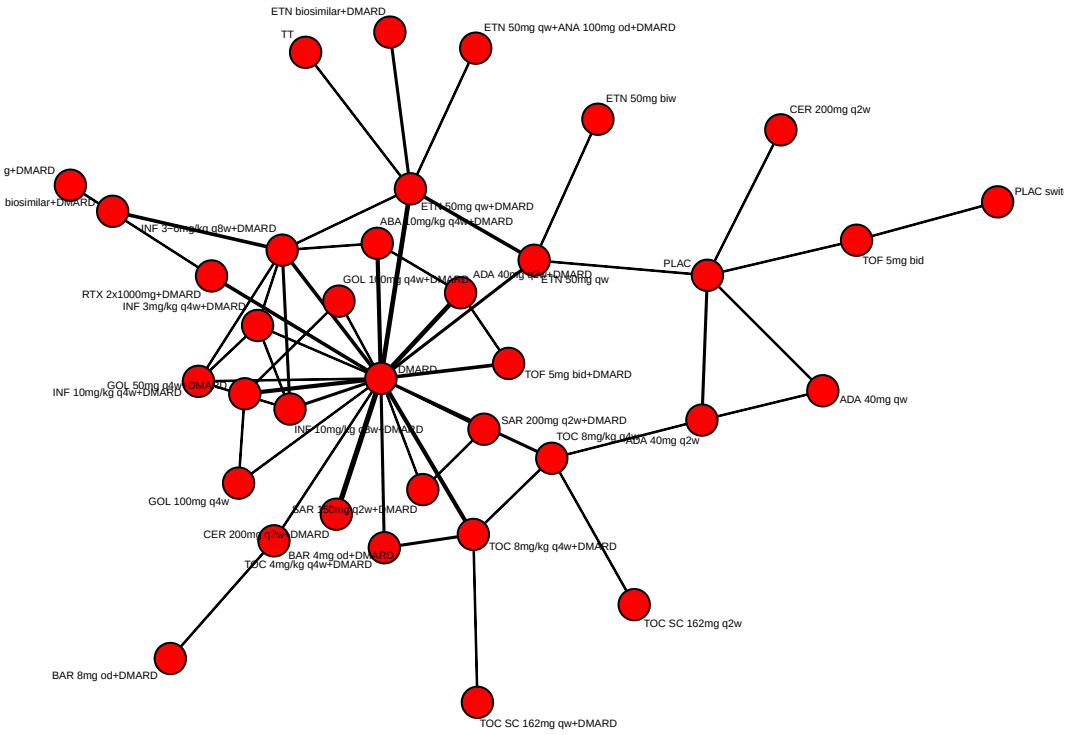

Figure 8: Bayesian random effects NMA network diagram for patients naive to bDMARDs

B.1.4 Network meta-analysis to obtain ACR 20/50/70 response

The probability of ACR20/50/70 responses was estimated using a Bayesian (random effects) network meta-analyses model 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. To avoid influencing the observed results by prior belief, uninformative prior distributions were used for the estimated model parameters. The relative treatment effects for each bDMARD versus cDMARDs estimated on the probit scale were transformed into absolute probabilities of the nonoverlapping ACR response categories by combining them with the average results for cDMARDs. The posterior distributions of parameters of interest were summarized by the median as a reflection of the point estimate and 95% credible intervals, constructed from the 2.5 and 97.5 percentiles. Analyses were performed with the Markov chain Monte Carlo method using the JAGS software package (http://mcmc-jags.sourceforge.net/).

B.2 Network meta-analysis to obtain HAQ

References

Alava, M. H., Wailoo, A., Wolfe, F., and Michaud, K. (2013). The relationship between eq-5d, haq and pain in patients with rheumatoid arthritis. *Rheumatology*, 52(5):944–950.

Arias, E. (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.

Carlson, J. J., Ogale, S., Dejonckheere, F., and Sullivan, S. D. (2015). Economic evaluation of tocilizumab monotherapy compared to adalimumab monotherapy in the treatment of severe active rheumatoid arthritis. *Value in Health*, 18(2):173–179.

Claxton, L., Jenks, M., Taylor, M., Wallenstein, G., Mendelsohn, A. M., Bourret, J. A., Singh, A., Moynagh, D., and Gerber, R. A. (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):1088–1102.

Dias, S., Sutton, A. J., Ades, A., and Welton, N. J. (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):607–617.

Guyot, P., Ades, A., Ouwens, M. J., and Welton, N. J. (2012). Enhanced secondary analysis of survival data: reconstructing the data from published kaplan-meier survival curves. *BMC medical research methodology*, 12(1):9.

Madan, J., Ades, T., Barton, P., Bojke, L., Choy, E., Helliwell, P., Jobanputra, P., Stein, K., Stevens, A., Tosh, J., et al. (2015). Consensus decision models for biologics in rheumatoid and psoriatic arthritis: recommendations of a multidisciplinary working party. *Rheumatology and Therapy*, 2(2):113–125.

Michaud, K., Vera-Llonch, M., and Oster, G. (2012). Mortality risk by functional status and health-related quality of life in patients with rheumatoid arthritis. *The Journal of rheumatology*, 39(1):54–59.

Michaud, K., Wallenstein, G., and Wolfe, F. (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):366–372.

Norton, S., Fu, B., Scott, D. L., Deighton, C., Symmons, D. P., Wailoo, A. J., Tosh, J., Lunt, M., Davies, R., Young, A., 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*, volume 44, pages 131–144. Elsevier.

Norton, S., Sacker, A., Dixey, J., Done, J., Williams, P., and Young, A. (2013). Trajectories of functional limitation in early rheumatoid arthritis and their association with mortality. *Rheumatology*, page ket253.

Oppong, R., Kaambwa, B., Nuttall, J., Hood, K., Smith, R. D., and Coast, J. (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):197–209.

Ramiro, S., Sepriano, A., Chatzidionysiou, K., Nam, J. L., Smolen, J. S., van der Heijde, D., Dougados, M., van Vollenhoven, R., Bijlsma, J. W., Burmester, G. R., 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*, pages annrheumdis–2016.

Sarzi-Puttini, P., Fiorini, T., Panni, B., Turiel, M., Cazzola, M., and Atzeni, F. (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):18.

Singh, J. A., Wells, G. A., Christensen, R., Tanjong Ghogomu, E., Maxwell, L. J., MacDonald, J. K., Filippini, G., Skoetz, N., Francis, D. K., Lopes, L. C., et al. (2011). Adverse effects of biologics: a network meta-analysis and cochrane overview. *The Cochrane Library*.

Stevenson, M., Archer, R., Tosh, J., Simpson, E., Everson-Hock, E., Stevens, J., Hernandez-Alava, M., Paisley, S., Dickinson, K., Scott, D., 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):1–610.

Tosh, J., Brennan, A., Wailoo, A., and Bansback, N. (2011). The sheffield rheumatoid arthritis health economic model. *Rheumatology*, 50(suppl 4):iv26–iv31.

Wolfe, F. and Michaud, K. (2010). The loss of health status in rheumatoid arthritis and the effect of biologic therapy: a longitudinal observational study. *Arthritis research & therapy*, 12(2):R35.

Wolfe, F., Michaud, K., Choi, H. K., and Williams, R. (2005). Household income and earnings losses among 6,396 persons with rheumatoid arthritis. *The Journal of rheumatology*, 32(10):1875–1883.

Wolfe, F., Michaud, K., Gefeller, O., and Choi, H. K. (2003). Predicting mortality in patients with rheumatoid arthritis. *Arthritis & Rheumatism*, 48(6):1530–1542.