PROGRESSION OF DISEASE IN PEOPLE WITH RHEUMATOID ARTHRITIS TREATED WITH NON BIOLOGIC THERAPIES

REPORT BY THE DECISION SUPPORT UNIT

February 2015

Laura Gibson, Mónica Hernández Alava, Allan Wailoo

School of Health and Related Research, University of Sheffield

Decision Support Unit, ScHARR, University of Sheffield, Regent Court, 30 Regent Street
Sheffield, S1 4DA

Tel (+44) (0)114 222 0734
E-mail dsuadmin@sheffield.ac.uk
Website www.nicedsu.org.uk
Twitter @NICE_DSU

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) is a collaboration between the Universities of Sheffield, York and Leicester. We also have members at the University of Bristol, London School of Hygiene and Tropical Medicine and Brunel University. The DSU is commissioned by The National Institute for Health and Care Excellence (NICE) to provide a research and training resource to support the Institute's Technology Appraisal Programme. Please see our website for further information www.nicedsu.org.uk

The production of this document was funded by the National Institute for Health and Care Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

Acknowledgements

The authors would like to acknowledge the contributions of Sofia Ajeganova, Alan Brennan, Chris Deighton, Jenny Dunn, Zoe Garrett, Frank McKenna, Kaleb Michaud, David Scott, Mark Strong, Frances Sutcliffe, Paul Tappenden, Annette HM van der Helm-van Mil, Hanna van Steenbergen, Suzanne Verstappen, Adam Young, Fred Wolfe.

Statement of competing interests:

Allan Wailoo and Monica Hernandez are co-authors of the Assessment Group report for the current NICE appraisal. Allan Wailoo has previously published on the cost effectiveness of drug therapies for rheumatoid arthritis in the US and in the UK, and on HAQ progression rates in Rheumatoid Arthritis.

EXECUTIVE SUMMARY

Estimates of cost effectiveness for biologic therapies are based on long term estimates of differences in costs and effects. One value that has been shown to have a significant impact on those estimates is the rate and trajectory of progression of disease, measured in terms of functional disability via the Health Assessment Questionnaire (HAQ), for patients when they are treated with conventional Disease Modifying Anti Rheumatic Drugs (DMARDs).

There is no study that reports how HAQ progresses in rheumatoid arthritis (RA) patients on non-biologic therapies that coincides entirely with the requirements of the cost effectiveness analysis. The cost effectiveness (CE) model requires estimates of HAQ progression over a patient's lifetime from the point at which they would be eligible for biologic therapies (which under current NICE Guidance is having failed two DMARDs and having a high Disease Activity Score (DAS) but also under consideration are those that have failed two DMARDs and have a moderate DAS score, or those with a high DAS score that have not yet failed any DMARDs). The model also requires such estimates of HAQ progression in patients once they have exhausted a sequence of biologic therapies. It is unsurprising that such evidence does not exist.

The purpose of this report is to identify related evidence that helps in the required estimates of how HAQ progresses in these different circumstances. It aims to assess:

1) The extent to which evidence suggests a constant linear rate of HAQ progression is appropriate.
2) The evidence for the concept of 4 latent classes of HAQ trajectory within a broad population of RA patients.
3) How can the subgroups of patients, relevant to the NICE decision problem in the context of biologic DMARDs, be allocated to these latent classes?
4) What is the support for the concept of HAQ a) continuously rising (as is the case if using a single annual rate worsening), b) rising at an increasing rate (as is the case in previous NICE appraisals) and c) rising at a decreasing rate (as in the AG base case)?
5) Can methods be employed to model the impact of dropout from observational datasets?
6) Are there subgroups of patients with faster/slower rates of HAQ progression than the average, and can these groups be identified?

Existing evidence

Some previous cost effectiveness studies have pooled average annual rates of HAQ progression together from diverse study sources. We suggest that this will lead to inappropriate estimates because studies have different lengths of follow-up, different times of follow-up and different frequencies of recording patient data.

Evidence from nine studies of patients with established disease and follow up of more than eight years was reviewed. Limited evidence was identified that allows an assessment of the long term trajectory of HAQ. In those studies that permit an assessment of the shape of HAQ trajectory, there was evidence that HAQ does not progress at a linear rate.

Individual level data analysis

We identified five studies of RA patients from different countries that had long term follow up of patients including their HAQ scores. We obtained patient level data for each of these studies and analysed them using different methods.

Descriptive data for two of those datasets exhibit a trend of rising HAQ over time (Early Rheumatoid Arthritis Study (ERAS) and Better Anti Rheumatic PharmacOTherapy (BARFOT)). ERAS is the only dataset that shows a rate of worsening in HAQ that is higher than the 0.045 simple rate used in previous NICE appraisals (0.054 from years 2-15). The BARFOT data is substantially lower.

Longitudinal studies of this type inevitably feature substantial dropout. Model predictions based on fewer and fewer observations at greater time points become more uncertain.

We replicated a latent class growth model method. We confirm our preferred model comprises four latent classes and a cubic specification in the ERAS data. In this model, the rate of worsening is faster for all the subgroups of interest, during the early part of diseases (years two to eight) but this rate slows over time.

The finding that HAQ rapidly deteriorates in the relevant patient groups but that this worsening slows over time is further supported by analysis of the ERAS data using an alternative modelling approach (the auto-regressive latent trajectory (ALT) model). Indeed, this is a consistent feature of the findings throughout this report.

We applied four different methods for accounting for attrition, assuming data are Not-Missing-At-Random (NMAR), and found that results continued to support the general findings of the original latent class analysis. Using the Roy-Muthen method for dealing with data NMAR, we identify three dropout subclasses within each of four latent trajectory classes. These provide credible estimates of the course of HAQ in the absence of dropout. We propose these serve as an appropriate upper bound for considerations of the plausible course of HAQ over time.

Our preferred analyses, described above, make use of all available data and adjust for covariates that distinguish the patient subgroups of relevance for the cost effectiveness analyses from the broader RA populations recruited into these studies. An alternative approach we explored is to limit the analysis of data only to those patients that meet, or more closely meet, the criteria for receipt of biologic therapies. We found that there were insufficient data for analyses where samples were restricted to those that had failed two DMARDs and also had a DAS>5.1. We did conduct subgroup analyses on those that had failed two DMARDs.

The National Databank for Rheumatic Diseases (NDB) and Early Rheumatoid Arthritis Network (ERAN) studies both suggest there is a relatively slow worsening of HAQ over time and this reduces over time. The ERAS data also supports this view in general though the latent class analysis does differ from the analyses conducted using the full dataset in suggesting the rate of HAQ continues to rise, albeit at a slower rate, particularly for those in the highest latent class. Overall these rates still suggest a lower overall rate of worsening than 0.045 per annum though the predictions for the severe disease subgroup are very close to this.

CONTENTS

1. INTRODUCTION 9
1.1. BACKGROUND 9
1.2. ASSESSMENT GROUP ANALYSIS 9
1.3. AIMS 11
2. CONTEXT AND LITERATURE 12
3. DATASETS AND METHODS 18
3.1. DATASETS 18
3.1.1. The Leiden Early Arthritis Clinic (LEC) Cohort 18
3.1.2. Better AntiRheumatic PharmacOTherapy (BARFOT) 19
3.1.3. National Databank for Rheumatic Diseases (NDB) 19
3.1.4. Early Rheumatoid Arthritis Network (ERAN) 20
3.1.5. Early Rheumatoid Arthritis Study (ERAS) 20
3.2. METHODS 21
3.2.1. Latent Class Growth Analysis 21
3.2.2. Auto-Regressive Latent Trajectory Models 22
3.2.3. Analysis of missing data 23
4. RESULTS 24
4.1. DESCRIPTIVE STATISTICS 24
4.1.1. Baseline Characteristics 24
4.1.2. HAQ Trajectories 28
4.1.3. Missing Data and Attrition 33
4.2. LATENT CLASS GROWTH MODELS (MAR) 41
4.2.1. ERAS 41
4.2.2. ERAN 43
4.2.3. BARFOT 45
4.2.4. NDB 47
4.2.5. Leiden 48
4.2.6. Additional analyses using ERAS data 48
4.2.7. Comparisons of models 49
4.3. ALT MODELS (MAR) 50
4.3.1. ERAS 50
4.4. SUB ANALYSIS OF PATIENTS AFTER FAILING TWO DMARDS 52
4.5. ACCOUNTING FOR MISSING DATA AND ATTRITION 59
4.5.1. Diggle-Kenward Latent Class Mixture Model 59
4.5.2. Wu and Carroll Method 60
4.5.3. Roy Method 60
4.5.4. Roy-Muthen Method 62
5. DISCUSSION 67
6. REFERENCES 71

TABLES AND FIGURES

Table 1: Characteristics of studies assessing HAQ change over time in established RA, minium 8 years follow up (amended from Norton et al 2014) *14*
Table 2: Patient subgroup characteristics for use in cost-effectiveness analysis. *24*
Table 3: Summary Baseline Characteristics *25*
Table 4: Statistics on failing 2 DMARDs *27*
Table 5: Summary Statistics over all Observation Periods *27*
Table 6: Mean HAQ over time *28*
Table 7: Mean Changes in HAQ *33*
Table 8: Missing Data and Attrition in the ERAN Data (n=1,124) *34*
Table 9: Missing Data and Attrition in the ERAS Data (n=1,430) *35*
Table 10: Missing Data and Attrition in the LEIDEN Data (n=563) *36*
Table 11: Missing Data and Attrition in the BARFOT Data (n=2,595) *37*
Table 12: Missing Data and Attrition in the NDB Data (n=19,462) *37*
Table 13: Cubic and Quadratic Information Criteria for ERAS data *42*
Table 14: Annual Rate of HAQ change – Cubic LCGA ERAS *42*
Table 15: Annual Rate of HAQ change – Quadratic LCGA ERAS *43*
Table 16: Annual Rate of HAQ change – Cubic LCGA ERAN *44*
Table 17: Annual Rate of HAQ change – Quadratic LCGA ERAN *45*
Table 18: Annual Rate of HAQ change – Cubic LCGA BARFOT *46*
Table 19: Annual Rate of HAQ change – Quadratic LCGA BARFOT *47*
Table 20: Annual Rate of HAQ change – Cubic LCGA ERAS with Free Variances *48*
Table 21: Annual Rate of HAQ change – Quadratic LCGA ERAS with Free Variances *49*
Table 22: Log-likelihood and Information Criteria for Cubic and Quadratic Models *50*
Table 23: Summary Statistics after taking Third DMARD (new baseline) *53*
Table 24: Mean Changes in HAQ after receiving third DMARD *55*
Table 25: Cubic and Quadratic Information Criteria for ERAS data *56*
Table 26: Annual Rate of HAQ change – Quadratic LCGA ERAN *57*
Table 27: Annual Rate of HAQ change – Quadratic LCGA ERAS *58*
Table 28: Probability of Latent dropout class by time of dropout. *65*
Table 29: Within sample probabilities of class membership by latent class model type *65*
Table 30: Probability of class membership for decision problem populations *66*

Figure 1: Representation of the role of non-biologic HAQ progression in cost effectiveness models *10*
Figure 2: Four Classes of HAQ trajectory shown in analysis of ERAS and NOAR *11*
Figure 3: HAQ growth curves by disease duration from Krishnan et al. *17*
Figure 4: Average HAQ Trajectories *31*
Figure 5: Average Change in HAQ *32*
Figure 6: Attrition over time *40*
Figure 7: Percentage of patients observed over time *41*
Figure 8: Cubic Replication of Analysis on ERAS by Norton et al. *42*
Figure 9: Quadratic Replication of Analysis on ERAS by Norton et al. *43*
Figure 10: Cubic LCGA of ERAN data *44*
Figure 11: Quadratic LCGA of ERAN data *45*
Figure 12: Cubic LCGA of BARFOT data *46*
Figure 13: Quadratic LCGA of BARFOT data *46*
Figure 14: Quadratic LCGA of NDB data *47*
Figure 15: Cubic LCGA of ERAS data with free variances *48*
Figure 16: Quadratic LCGA of ERAS data with free variances *49*
Figure 17: ALT model of ERAS data *51*
Figure 18: Average HAQ Trajectories *54*
Figure 19: Average Change in HAQ *55*
Figure 20: Quadratic LCGA from Third DMARD in ERAN data *57*
Figure 21: Quadratic LCGA from Third DMARD in ERAS data *57*
Figure 22: LCGA from Third DMARD in NDB data *58*
Figure 23: Quadratic LCGA using ERAS Data and Roy Latent Dropout Adjustment *61*
Figure 24: Cubic LCGA using ERAS Data and Roy Latent Dropout Adjustment *62*
Figure 25: Latent Dropout Classes within latent trajectories *64*

ABBREVIATIONS AND DEFINITIONS

ACR	American College of Rheumatology
AG	Assessment Group
AIC	Akaike Information Criterion
ALT	Auto-regressive Latent Trajectory
BARFOT	Better Anti Rheumatic PharmacOTherapy
BIC	Bayesian Information Criterion
BSRBR	British Society for Rheumatology Biologics Register
CE	Cost Effectiveness
DAS	Disease Activity Score
DMARDs	Disease Modifying Anti Rheumatic Drugs
ERAN	Early Rheumatoid Arthritis Network
ERAS	Early Rheumatoid Arthritis Study
HAQ	Health Assessment Questionnaire
HAQ-DI	Health Assessment Questionnaire Disability Index
ICER	Incremental Cost Effectiveness Ratio
LCGM	Latent Class Growth Models
LEC	The Leiden Early arthritis Clinic cohort
MAR	Missing At Random
MNAR	Missing Not At Random
NDB	National Databank for Rheumatic Diseases
NMR	Not Missing At Random
NOAR	Norfolk Arthritis Register
RA	Rheumatoid Arthritis
RADAI	Rheumatoid Arthritis Disease Activity Index
RF	Rheumatoid Factor

1. INTRODUCTION

1.1. BACKGROUND

NICE is conducting an appraisal of several biologic therapies for the treatment of Rheumatoid Arthritis (RA): TA537 "Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying anti-rheumatic drugs (DMARDs) and after the failure of conventional DMARDs only".

The Independent Assessment Group (AG) identified that one of the key drivers of cost effectiveness is the rate of progression of disease, measured in terms of the Health Assessment Questionnaire (HAQ), for patients that are not treated with biologic therapies. The purpose of this report is to provide additional evidence for the Appraisal Committee relating to this single parameter.

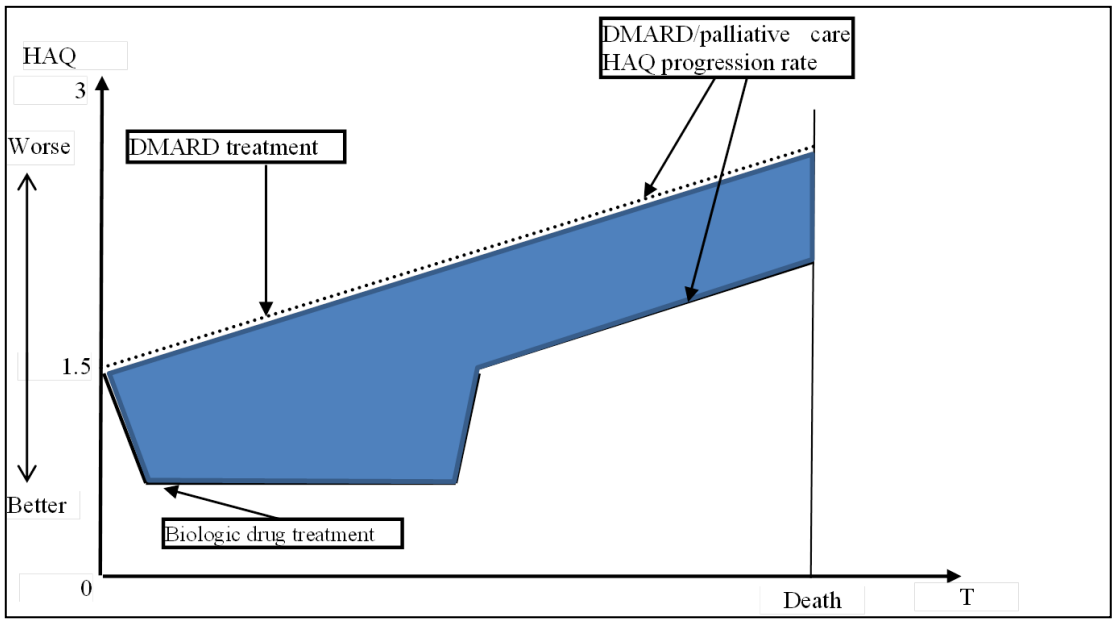
1.2. ASSESSMENT GROUP ANALYSIS

The AG analysis showed substantial differences in the estimated cost effectiveness of biologic therapies depending on whether the rate of HAQ progression when not on biologics was assumed equivalent to the rates assumed in previous NICE appraisals of whether the rate was estimated from data provided by a UK longitudinal study (the Early Rheumatoid Arthritis Study – ERAS).

The parameter features in several places within the cost effectiveness analysis. Figure 1 provides a stylised representation of the issue. Where patients are simulated within the model to be treated without biologic therapies, the rate of HAQ worsening over time governs the entire pathway for that patient until death. Simulated patients that are treated with biologic therapies (or a sequence of therapies) may exhaust those treatments prior to death and therefore experience some worsening in disease from that point onwards.

In both situations the rate of HAQ worsening is likely to consist of a period of time whilst patients receive non biologic Disease Modifying Anti Rheumatic Drugs (DMARDs) but could include periods of treatment with other non-biologic therapies or no active therapy.

Figure 1: Representation of the role of non-biologic HAQ progression in cost effectiveness models



In previous NICE appraisals of biologic therapies for RA (TA130, TA186, TA225, TA195, TA247 and TA280), there has been a consistent use of rates of progression first used in TA130 which distinguish rates of progression for biologic therapy, DMARD therapy and palliative care:

> *"It was assumed that patients remaining on TNF inhibitors experience a worsening (increase) in HAQ equivalent to the general population. Based on the study by Krishnan and colleagues, this was set a progression of 0.03 per year… It was assumed that TNF inhibitors halve the general worsening in HAQ, so that patients on palliation have a progression rate of 0.06 per year….. For conventional DMARDs, an intermediate progression rate of 0.045 per year was assumed …… These assumptions were varied in sensitivity analysis." (Chen et al., 2006, p.100).*[1]

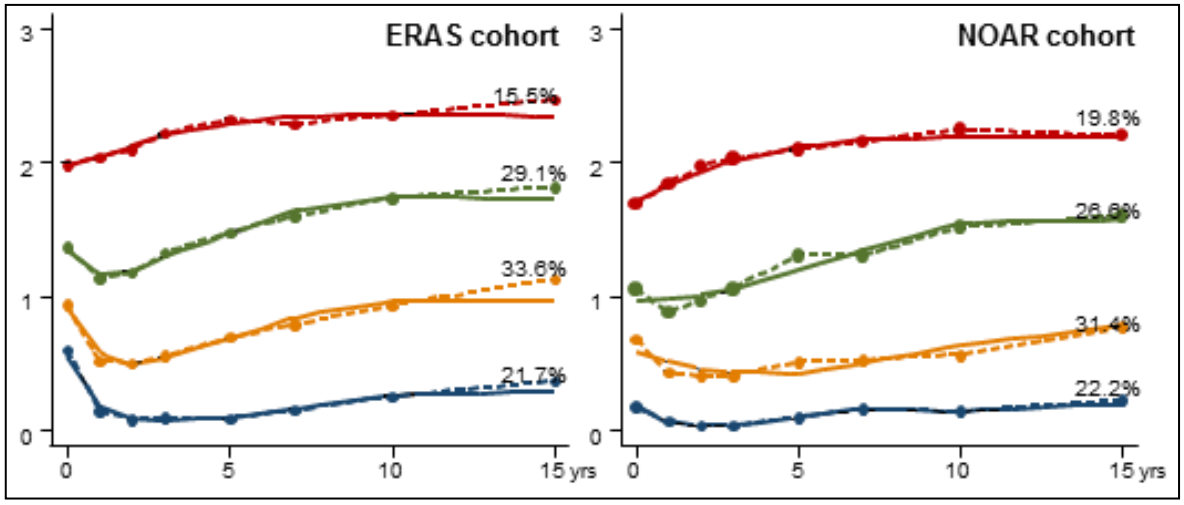
Since it is assumed that palliative care is provided only once the sequence of available DMARDs has been exhausted, these assumptions imply an increase in the rate of worsening for patients over time.

The AG's base case model for the current appraisal took an alternative approach to the estimation of HAQ progression. Their analysis was based on a published analysis of data from the ERAS study.[2] ERAS is a UK based, longitudinal inception cohort study of consecutive patients thought to have RA by a consultant rheumatologist, in the outpatient

clinics of nine UK rheumatology departments. Patients were enrolled between 1986 and 1997. Patients were enrolled early in disease: they had symptom duration of less than 2 years and were DMARD naïve.

The analysis used latent class growth models (LCGM) to try to identify different trajectories of HAQ progression for distinct groups of patients. The approach identified four distinct classes within the ERAS cohort, based on up to 15 years of follow up data. The notion of there being four classes has proven robust in further analyses undertaken in the Early Rheumatoid Arthritis Network (ERAN) and Norfolk Arthritis Register (NOAR) datasets. The analysis for the AG utilised this model but included covariates within the class probabilities that coincide with those used to define patients within the cost effectiveness model. Figure 2 shows the four latent classes identified in the Norton analysis.[3]

Figure 2: Four Classes of HAQ trajectory shown in analysis of ERAS and NOAR



The use of the HAQ progression estimates from the ERAS study rather than those used historically in previous NICE appraisals is an important driver of cost effectiveness. The Assessment Group reported changes in the Incremental Cost Effectiveness Ratios (ICERs) that from approximately £33k when using the historical HAQ progression rates to £57k when using the ERAS based estimates.

1.3. Aims

The purpose of this report is to provide additional information on the rate of HAQ progression in patients with rheumatoid arthritis whilst being treated with non-biologic therapies.

First an overview of existing literature is provided. Second we identify a series of datasets that provide relevant data additional to that included in the original assessment report. Third we conduct analyses that replicate the latent class growth model in these datasets. We then consider alternative methods of modelling these data. Based on the results we discuss and illustrate potential subgroups that have faster/slower HAQ progression rates beyond those identified to date in the four latent class model.

Specific issues addressed in this report are as follows:

- What is the evidence for the concept of four latent classes of HAQ trajectory within a broad population of RA patients?
- How can the subgroups of patients, relevant to the NICE decision problem in the context of biologic DMARDs, be allocated to these latent classes?
- What is the support for the concept of HAQ a) continuously rising (as is the case if using a single annual rate worsening), b) rising at an increasing rate (as is the case in previous NICE appraisals) and c) rising at a decreasing rate (as in the AG base case)?
- Can methods be employed to model the impact of dropout from observational datasets?
- Are there subgroups of patients with faster/slower rates of HAQ progression than the average, and can these groups be identified?

2. CONTEXT AND LITERATURE

It is quite common for cost effectiveness analyses of therapies for RA to assume a linear progression of HAQ for patients whilst on conventional DMARDs.[4,5] This is not a feature that is confined to analyses that have been undertaken for NICE appraisals. In most cases, these constant rates have been estimated by reference to a range of cross sectional studies. One source of estimates for such a constant rate is a study by Scott and Garood (2000)[6] which reported the mean annual HAQ progression rate from 12 cross sectional studies of varying duration of follow up. The raw mean from these studies is 0.031 and the weighted average is 0.042.

There are however, several limitations of using such an approach. If the true rate is not constant over time then it is not valid to pool estimates from different time periods.

Extrapolation will result in serious bias. In addition, there are substantial differences between studies in terms of rates of dropout, characteristic of patients, frequency of observations and study design, all of which will make crude pooling of annual rates an inaccurate approach. Even in more complex studies with longer follow up there is a risk that either the observed HAQ progression is linear within that timeframe, or that the choice of analysis method imposes linearity.

Norton *et al*.[2] undertook a recent systematic review of studies that reported on HAQ change over time in patients with rheumatoid arthritis or undifferentiated polyarthritis for studies published up to and including 2012. The review only included studies with follow up of at least three years. We report here those studies identified that had follow up of eight years or greater, in a cohort of patients with established RA (disease duration of greater than two years) since it is only this longer time frame that is likely to provide additional information on the trajectory of HAQ over the period relevant for the cost effectiveness model.

Table 1 presents key features of these studies from Norton *et al*.

Table 1: Characteristics of studies assessing HAQ change over time in established RA, minium 8 years follow up (amended from Norton *et al* 2014)

Publication	N	Age	Female gender	Disease duration	Follow-up duration	HAQ baseline	Analysis method	Progression HAQ-score	HAQ scores / change over time
Sherrer *et al*, 1986 (Can)	681	62 (13)	72 %	10 (6) yrs	12 (6) yrs	Cross-sectional: not available	Mean HAQ score at end study	Annual rate in first few yrs and after 15 yrs	0.10 and 0.02
Hawley, 1992 (US)	157	50.8 (12.5)	75%	6.7 (8.2) yrs	Mean 9.8 SD 0.75)	0.5 (0.5)	Cross-sectional at different time-points	Mean (effect size from baseline) baseline, 2, 5, 10 yrs	0.5 (NA); 0.5 (-0.01); 1.3 (-1.63); 1.6 (-2.39)
Leigh *et al*, 1992 (US)	L: 209 D: 54	L: 52 (14) D: 66 (10)	L; 86% D: 63%	L: 12 (9) yrs D: 18 (9) yrs	8 yrs	L: 1.09 (0.83) D: 1.75 (0.88)	Cross tabulation (L) at follow-up and all (A) patients including deceased Multiple regression pooling data and including all time-points. Different models (duration, duration2, duration3)	Annual rate	L cohort: 0.018/yr L (0-10 yrs) W vs M 0.017 vs -0.003 L (10-20 yrs) W and M: 0.016 vs -0.010. A (0-10 yrs) W and M: 0.032 vs 0.063. A (10-20 yrs) W and M: 0.029 vs 0.079: Linear model: $\beta = 0.0518$
Ward, 1993 (US)	188 94	Married (ma): 54.6 (11.7) Unmarried (unm): 54.0 (13.4)	78% 94%	13.5 (9) yrs 15.0 (9.2) yrs	~9.5 yrs	1.1 (0.8) 1.3 (0.9)	Pooled time-series linear regression analysis	Mean adjusted annual rate (ma vs unm); all patients; patients with complete follow-up; men and women	$\beta= 0.01$ vs $\beta= 0.03$; $\beta= 0.007$ vs $\beta= 0.02$; $\beta= 0.006$ vs $\beta= 0.03$; $\beta= 0.01$ vs $\beta= 0.03$
Leigh, 1993 (US)	L: 209 D: 54 LFU: 67	L: 52 (14) D: 66 (1) LFU: 55 (12)	L: 86% D: 63% LFU: 85%	L: 12 (9) yrs D: 18 (9) LFU: 14 (9)	L: 8 yrs D: until last visit LFU: until last visit	L: 1.16 (0.81) D: 1.75 (0.88) LFU: 1.20 (0.90)	Using all valid observations during follow-up: 1) Linear regression 2) Tobit regression 3) OLS fixed effects By cohort (I = 0-9, II = 9-19 and II =>19 yrs disease duration at baseline) 4) Cohort OLS 5) Cohort OLS fixed	Annual slope	1) $\beta = 0.014$ 2) $\beta = 0.014$ 3) $\beta = 0.019$ 4) I, $\beta = 0.003$; II, $\beta = 0.0001$; III, 0.017 5) I, $\beta = 0.0210$; II, $\beta = 0.0103$; III, $\beta = 0.0293$

							effects		
Lassere *et al*, 1995 (AU)	358	61. (12.7)	73.2 %	13.6 (10.4) yrs	≤3 yrs >3 to ≤6 yrs >6 to ≤12 yrs >12 to ≤18 yrs >18 yrs	Median [IQR] 0.250 [0.781] 0.625 [1.188] 0.875 [1.25] 1.125 [0.75] 1.375 [1.25]	Median difference Percentile curves using the weighted average method.	Median difference (95%CI) from 3 yrs: 3-6; 6-12; 12-18; and >18 yrs NA	-0.25 (-0.500,0.001); -0.375 (-0.625, -0.125); -0.625 (-0.875, -0.375); -0.875 (-1.125, -0.500) NA
Ward *et al*, 1998 (US)	282	52.5 (11.7) a 52.6 (11.4) aa	84% 77%	13.6 (9.1) yrs 14.0 (8.6) yrs	10.5 (3.8) yrs 10.0 (4.1) yrs	1.03 (0.8) 1.00 (0.8)	Pooled time series regression analysis	Adjusted annual rate	β= 0.015 (95%CI 0.012, 0.018) a β = 0.019 (95%CI 0.014 to 0.024) aa
Krishnan, 2004 (US and CA)	6436	58.5 [48.0-67.4]	74%	8.0 [2.3-14.0] yrs	20 yrs	1.13 [0.5-1.8]	Percentile curves	Smoothed growth curves, men and women separately per age group	See paper
Odegard, 2007 (NO)	149	50.2 (12.5)	76%	2.2 (1.2) yrs	10 yrs	0.86 (0.61)	Cross-sectional at different time-points	Mean (SD) score at 1 yr, 2 yrs, 5 yrs and 10 yrs	0.86 (0.61); 0.85 (0.62); 0.85 (0.65); 0.86 (0.60); 0.91 (0.70)

Data on age, disease duration, follow-up and HAQ are mean and standard deviation. L= alive at follow-up; D = died, LFU = lost to follow-up; NA = not available; W = women; M = men; MA = married; UNM = unmarried.

Sherrer[7] is a study conducted in RA patients in Canada recruited in 1966 and 1974 who were followed up in 1982. They show a rapid worsening in the early stages of disease (approximately 0.1 per year in the first few years) followed by a period where the rate of worsening in HAQ decreases (a rate of less than 0.02 after 15 years). 13% of the study had HAQ scores in excess of 2.5 at the end of the study.

Hawley and Wolfe[8] report the results from a US observational study with patients first recruited in 1976. It shows that there is a rapid increase in HAQ between baseline and five years of clinic treatment (from a mean of 0.5 to 1.3) and this rate of increase slows between years five and ten (1.3 to 1.6).

Leigh *et al.*[9] studied a longitudinal sample of patients with RA in the US. They selected groups of males and females with differing lengths of disease duration (0-10, 10-20, 20+ years) and followed them up for eight years between 1981 and 1989. They report that those with greater than 20 years' disease duration experienced faster deterioration than those with less than 20 years' duration.

The focus of the study by Ward[10] is differences between married and unmarried groups in terms of the rate of change of HAQ. This is a US study of RA patients enrolled between 1978 and 1981. It is difficult to ascertain a clear picture of changes in the rates either over time or in relation to disease duration. However, it does seem that the overall rate of change in most periods and for both groups is very low and substantially less than 0.045 per annum.

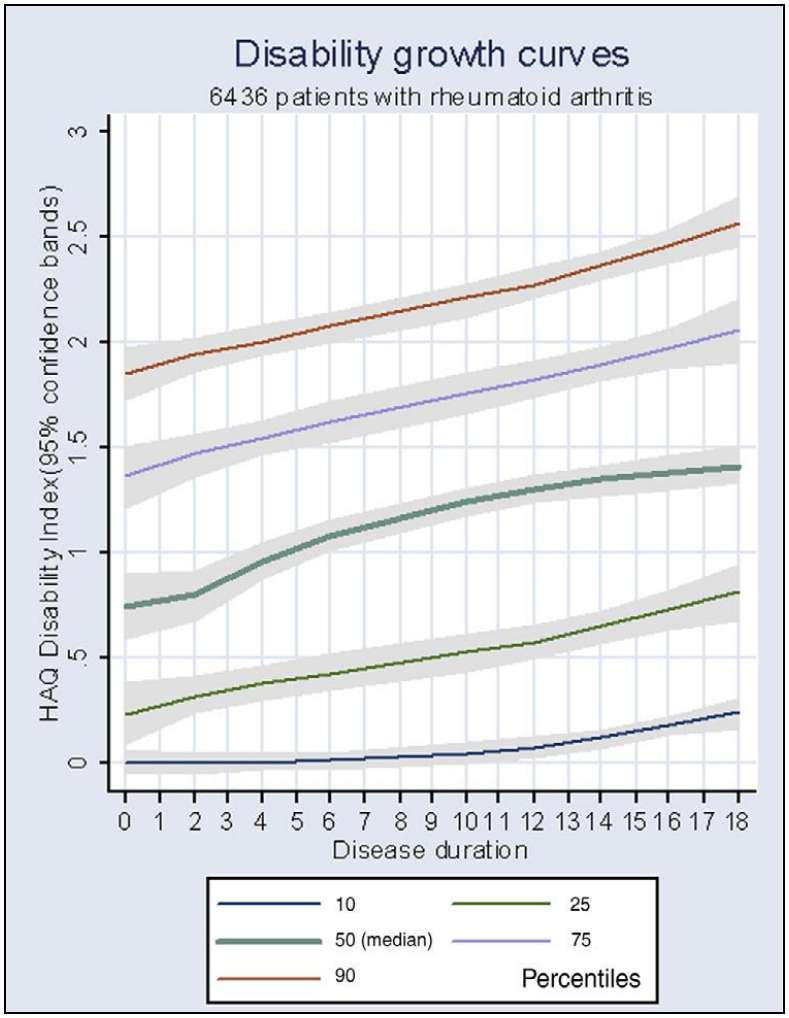
Leigh *et al.*[11] is a study comparing different econometric models to estimate HAQ as a function of disease duration (and other variables). It only reports on models that reflect disease duration in a linear form so it is not possible to draw conclusions on the plausibility of the constantly increasing HAQ concept. The study is based on RA patients recruited to the ARAMIS study with data collected between 1981 and 1989. The reported rates of worsening are substantially lower than 0.045.

Lassere[12] compared 358 RA patients in an Australian sample using median percentile curves with disease duration as an explanatory variable. The results are not reported in a form directly informative to the current study.

Ward *et al.*[13] aim to compare HAQ between RA patients treated in managed care settings with those treated in fee-for-service practices. Patients were followed for up to 13 years. The paper reports a single adjusted annual rate for each group which does not inform the shape of the HAQ trajectory. However, the rates reported are substantially lower than 0.045.

Krishnan[14] reports on approximately 6,000 RA patients followed between 1981 to 2000 from multiple databases in the US and Canada. They report that very few patients were treated with biologics. They plot the median HAQ growth curve (and other percentiles) against disease duration, shown below in Figure 3. Whilst this is the median as opposed to the mean, it does demonstrate a more rapid initial worsening in function from year two, followed by a slowing in the rate of worsening. Interesting, the distribution of HAQ across all observations does show a non-negligible number of observations in excess of 2.5, including the maximum level of disability described by the HAQ (three).

Figure 3: HAQ growth curves by disease duration from Krishnan *et al.*



Ødegard *et al*.[15] report a Norwegian cohort of 238 patients followed at one, two, five and ten years. The study did not focus on modelling HAQ change over time (its focus is pain, anxiety and depression). However, the paper does report mean HAQ at each timepoint. It shows no change from baseline to five years in this population that had a mean disease duration at baseline of 2.2 years. There is an increase in HAQ of 0.05 from years five to ten.

3. DATASETS AND METHODS

The aim of any new analyses undertaken was to provide estimates of the long term trajectory of HAQ for RA patients and to provide evidence to validate the existing estimates used in the AG model. Therefore, we attempted to identify data from studies of patients with RA (or a substantial proportion of patients with RA that could be identified from a broader population), treated either exclusively with non-biologic DMARDs, or with follow up that included biologic DMARDs but that was clearly recorded, and had a minimum of five years of follow up. The study had to include regular assessments of the HAQ instrument since this is the measure of disease progression used in the cost effectiveness model, and additional information on the number of DMARDs failed and DAS score in order to be able to link results to the relevant subpopulations in the NICE decision problem.

We discussed these requirements with three clinical advisors. Potential data sources identified from a non-systematic review of literature were also considered. We contacted owners of datasets in order to establish the details of studies and then decided whether to include them. As a result of this process the datasets listed in Section 3.1 were included.

Our suggested data sources were discussed and presented in two workshops held in 2014, attended by clinical experts, economists with experience of this area, as well as patient and manufacturer stakeholders.

3.1. DATASETS

3.1.1. The Leiden Early Arthritis Clinic (LEC) Cohort

The LEC cohort is an observational study comprising a population based, inception cohort of patients with RA managed at the Leiden University Medical Centre, Netherlands.[16] Patients with early RA were referred to the LEC from a substantial (>400,000 inhabitants), semi-rural

area. The study started in 1993. HAQ, and other variables, were recorded at baseline (first visit with the rheumatologist), and then again at 3 months and yearly intervals from then on. The diagnosis of RA was established in cases where patients fulfilled the 1987 American College of Rheumatology (ACR) criteria for RA. Follow-up continues for as long as the patient is being seen by the rheumatologist. Follow up ends when the patients are discharged either because of sustained remission or death. The dataset supplied to us contains 563 patients and a maximum follow up of nine years.

3.1.2. Better AntiRheumatic PharmacOTherapy (BARFOT)

BARFOT is a long term, multicentre, observational study of patients with early RA at baseline in Southern Sweden. Patients have a diagnosis of RA according to the 1987 revised ACR criteria and all available patients are included, provided they are seen within one year of symptom onset.[17,18] Data (including HAQ and DAS28) were recorded at baseline, six months, one, two, five, eight and 15 years with the study commencing in 1995. In most cases, treatment began at baseline. The dataset supplied to us contains 2,595 patients and a maximum follow up of 15 years.

3.1.3. National Databank for Rheumatic Diseases (NDB)

The National Databank for Rheumatic Diseases (NDB) is a not-for-profit rheumatic disease research databank in which patients complete detailed self-report questionnaires at six month intervals.[19] Patients in the NDB are recruited from two sources: 1) non-selected patients from the practices of US rheumatologists and 2) patients enrolled as part of pharmaceutical company sponsored registries. Eligible patients in this study were those with RA who had completed a biannual survey for events occurring from July 1 1998 onwards. Patients were referred by more than 1,000 rheumatologists dispersed throughout the US. More than 90% of rheumatologists were in private practice and not full time university physicians. The diagnosis of RA was made by the patients' rheumatologists.

At each assessment, demographic variables were recorded including sex, age, ethnic origin, education level, current marital status, medical history and total family income. Patients also complete the Health Assessment Questionnaire Disability Index (HAQ-DI), EuroQol, SF-6D and a VAS QOL scale. Patients describe all medications used and provide information regarding medical treatments, physician visits and hospitalizations. Note that the NDB does

not include DAS and therefore has a significant limitation when trying to predict how results translate to the relevant RA populations for the NICE Technology Appraisal.

The NDB attracts participants that are not necessarily representative of the RA community, either in the US or in the UK NHS. NDB participants tend to be from higher income backgrounds, are less likely to come from an ethnic minority and are better educated than the general US RA population. Nevertheless, the NDB is one of the richest sources of data for the study of RA patients in the US if not the world. The dataset supplied to us contains 19,462 patients and a maximum follow up of 15 years.

3.1.4. Early Rheumatoid Arthritis Network (ERAN)

ERAN is a UK and Ireland based study reporting an inception cohort of newly diagnosed RA patients drawn from a network of rheumatology departments that began recruiting in 2002. It is similar in design to the ERAS study (multicentre, prospective, observational) but is intended to be larger in terms of patient numbers and draw from a wider geographical area. Patients are enrolled based on diagnosis of RA by the treating rheumatologist, there is no requirement for fulfilment of the ACR 1987 criteria. Data are recorded at baseline, three-six months, one year and annually thereafter.[20,21] The dataset supplied to us contains 1,124 patients and a maximum follow-up of 11 years.

3.1.5. Early Rheumatoid Arthritis Study (ERAS)

ERAS recruited patients thought to have RA by their treating rheumatologist between 1986 and 1998 from nine UK hospitals. Where the diagnosis changed subsequently, patients were excluded. Each participating centre recorded clinical, radiological, laboratory and genetic features of all consecutive patients with RA of less than two years duration and prior to any second line (disease modifying) treatment. Patients are reviewed at one, three, six, twelve months and yearly thereafter. Our analysis excluded all observations for patients that took biologic therapies, though this was a small proportion of patients given the time the data for the study were collected.

This study was the source of data for the analysis underpinning the Norton *et al.* latent class model that is used in the base case for the AG cost effectiveness base-case model. Note that the AG did not have access to the ERAS data but used the results reported by Norton *et al.*

We obtained data from ERAS for the purposes of the DSU report. The dataset supplied to us contained 1,430 patients and a maximum follow up of 15 years.

It should be reiterated here that the original analysis performed by Norton *et al*. using the ERAS dataset,[2] has been validated in terms of the appropriate number of latent classes and the general trajectories for those classes in the ERAN and Norfolk Arthritis Register (NOAR) study datasets.

3.2. METHODS

This report has several aims and therefore we present a number of different results for the different datasets we identified. We aim to validate the concept of there being distinct latent classes, in terms of HAQ trajectory over time, within a broad population of patients with rheumatoid arthritis. We also aim to establish the extent to which sub-populations of that broad population, those defined by the NICE decision problem, can be allocated to those different classes.

3.2.1. Latent Class Growth Analysis

First, we replicate and extend the latent class growth analysis that underpins the results reported by Norton *et al*.[2]

In order to compare our results with those found by Norton *et al.* (2012) as was used in the Assessment Group base-case analysis, we initially used latent class growth analysis (LCGA) with the ERAS data. In this model, explanatory variables influence class membership, but not HAQ trajectory. HAQ is initially assumed to have a cubic relationship with time. This could create problems when predicting future HAQ beyond the sample period (15 years in the case of ERAS) because the shape of the extrapolated curve may exhibit a tendency to increase, or decrease, rapidly. For this reason, we censor HAQ so it is always positive. However, the problem remains that predicted HAQ could increase above three.

Formally, the model is

$$y^*_{itc} = \eta_{0ic} + \eta_{1ic}x_t + \eta_{2ic}x_t^2 + \eta_{3ic}x_t^3 + \varepsilon_{it}$$

where y^*_{itc} is a latent dependent variable representing HAQ for patient i, at time t and in class c. The random coefficients, η_{tic} ($\eta_{tic} = \eta_{tc} + \xi_{it}$) for the intercept, slope, quadratic and

cubic terms (η_{0ic}, η_{1ic}, η_{2ic} and η_{3ic} respectively) have a full covariance matrix, i.e. the model allows them to be correlated. In line with previous studies we initially restrict the variances of the intercept, slope, quadratic and cubic terms to zero ($\xi_{it} = 0$). The x_t are the time scores, which are used to impose the trend (in this case $x_0 = 0$, $x_1 = 1$, $x_2 = 2$, *etc.*). The ε_{it} are a normally distributed error terms with zero mean and variance $\sigma^2_{\varepsilon t}$ which varies over time.

$$y_{itc} = \begin{cases} y^*, & if\ y^*_{itc} > 0 \\ 0, & if\ y^*_{itc} \leq 0 \end{cases}$$

where y_{itc} is the observed HAQ score for patient i at time t in class c and the probabilities of class membership are estimated using a multinomial logit model:

$$\Pr(C_{it} = c|z_{it}) = \frac{e^{z_{it}\beta c}}{\sum_{s=1}^{4} e^{z_{it}\beta s}}$$

where z contains the covariates which predict class membership. It is this which allows us to use the data from the entire ERAS dataset and then use the estimated model to predict the expected HAQ trajectory for patients with differing characteristics.

3.2.2. Auto-Regressive Latent Trajectory Models

Second, we use an auto-regressive latent trajectory (ALT) model which combines two traditional panel data methods; an auto-regressive model and a latent growth curve model. The ALT model benefits from the auto-regressive component which allows present HAQ to be predicted by past HAQ, while also benefitting from separate trajectories over time. These trajectories can be non-linear and therefore more flexible than the standard auto-regressive model.

The LCGA discussed previously estimated the latent trajectories of four latent classes. However, we are more interested in the trajectories for patients with specific characteristics rather than the differences between those latent classes. The LCGA assumes that trajectories are set from baseline and the HAQ trajectory does not differ by individual characteristics. For these reasons, autoregressive latent trajectory (ALT) models (see Bollen & Curran 2004[22]) were also used to analyse the data. These models combine elements of autoregressive models with those of latent trajectories and allow a more flexible estimation than either a latent trajectory model or an autoregressive model on their own. The model uses the random intercept and slope of a latent trajectory model and also includes an AR process. This AR process allows HAQ to be influences by HAQ in the previous observation

period, so patients who start with a higher HAQ might be more likely to have a higher HAQ throughout their trajectory. In this model, patient characteristics have some influence the intercept and the slope of the HAQ trajectory, allowing patients trajectories to vary.

The initial HAQ score, y_{i1}, depends on baseline covariates z, with coefficients vector β_y, so that

$$y_{i1} = v_1 + \beta_y z_i + \varepsilon_{i1}$$

and the ALT equation for the trajectory of HAQ is

$$y_{it} = \eta_{0i} + \eta_{1i} x_t + \rho_t y_{it-1} + \varepsilon_{it} \qquad t = 1, \ldots, T$$

Where η_{0i} and η_{1i} are the random intercept and random slope, respectively and allow linear or non-linear trajectories. The x_t are the time scores for a non-linear trend. The x_t must be fixed for at least two time point for identification; the time scores that we fix are different for the different datasets and are decided depending on which model had the best fit. The random intercept, η_{0i} and the random slope, η_{1i} have means μ_0 and μ_1, respectively and depend on individual baseline characteristics z_i with coefficients γ_0 and γ_1, respectively, so that

$$\eta_{0i} = \mu_0 + \gamma_0 z_i + \zeta_{\alpha i}$$
$$\eta_{1i} = \mu_1 + \gamma_1 z_i + \zeta_{\beta i},$$

where error terms ζ_{0i} and ζ_{1i} are normally distributed error terms with full covariance matrix. We assume that $E(\varepsilon_{it}) = 0$, $COV(\varepsilon_{it}, y_{it-1}) = 0$, $COV(\varepsilon_{it}, \eta_{0i}) = 0$, $COV(\varepsilon_{it}, \eta_{1i}) = 0$ and $E(\varepsilon_{it}, \varepsilon_{jt}) = 0$ for all t and $i \neq j$ and $E(\varepsilon_{it}, \varepsilon_{it}) = \sigma^2_{\varepsilon_t}$ for all t. We also assume that all residuals have a mean of zero and are uncorrelated with all exogenous variables.

3.2.3. Analysis of missing data

Two different assumptions relating to missing data are considered. First, we assume that all missing data and attrition is "Missing At Random" (MAR), that it is unrelated to the dependent variables, in this case the HAQ score. Under this assumption, missingness can be related to the independent variables which are accounted for in the models. Second, we assume that observations which are lost to attrition are "Missing Not At Random" (MNAR), and we use a number of different techniques to adjust the HAQ trajectories in an attempt to determine whether attrition in the data is a source of bias.

We illustrate the results for each model by comparing the observed and predicted data. We also provide model predictions for three separate subgroups related to the cost effectiveness analysis. These are i) the mean characteristics of the full, UK treated biologics population from the British Society for Rheumatology Biologics Register (BSRBR) that formed the sampling frame for the AG cost effectiveness model, ii) "severe active" group: the mean characteristics for patients treated with biologics, (Jan 2010 – June 2014), and with a DAS>5.1 from the BSRBR register and iii) "moderate active" group: the mean characteristics for patients treated with biologics, (Jan 2010 – June 2014), and with a DAS≤5.1 and >3.2.

These characteristics are as follows:

Table 2: Patient subgroup characteristics for use in cost-effectiveness analysis.

	i) BSRBR		**ii) Severe**		**iii) Moderate**	
	Mean	**SD (%)**	**Mean**	**SD (%)**	**Mean**	**SD (%)**
Age	56.2	12.2	57.3	12.5	58.0	13.6
Proportion female	0.8	0.4	0.8	0.2	0.7	0.2
Disease duration (yrs)	13.3	9.6	9.4	9.3	10.2	10.5
DAS28	6.6	1.0	6.2	0.8	4.4	0.6
Previous DMARDSs	3.9	1.6	2.8	1.0	2.9	1.0
HAQ	2.0	0.6	1.6	0.7	1.5	0.8
Weight (kg)	73.1	17.6	78.8	19.6	76.1	19.1

4. RESULTS

4.1. DESCRIPTIVE STATISTICS

4.1.1. Baseline Characteristics

Table 3 shows summary statistics at time of recruitment for each of the included datasets. These summary statistics represent the datasets after patients with missing HAQ values and those less than eighteen years of age at the time they were recruited to the study have been removed. In all datasets, the average age of the patients was mid to late fifties and around two thirds are female. The average HAQ at recruitment is just over one in all datasets but is slightly higher in the ERAS data, possibly because patients were recruited only if they had never taken any DMARDs which might have helped lower their HAQ at recruitment. The average DAS score in the ERAN, ERAS and LEIDEN datasets suggest moderate disease activity (means of 4.5 to 4.8), whereas in the BARFOT data, the average DAS is high. In the

NDB data, the Rheumatoid Arthritis Disease Activity Index (RADAI) score is used instead of the DAS because DAS was unavailable; the average RADAI score suggests moderate disease activity. The average disease duration is given, in months, from symptom onset to recruitment into each dataset. Patients in the NDB dataset have substantially longer disease durations when entering the study because this is not an inception cohort. Disease duration was not available for patients in the LEIDEN data.

Table 3: Summary Baseline Characteristics

	MEAN (STANDARD DEVIATION)				
	ERAN	**ERAS**	**LEIDEN**	**BARFOT**	**NDB**
HAQ OBSERVATIONS	1091	1421	498	2455	19,292
AGE	56.72 (13.85)	55.13 (14.37)	55.82 (16.07)	58.06 (15.73)	58.88 (13.38)
GENDER	0.3190 (0.4663)	0.3329 (0.4714)	0.3173 (0.4659)	0.3308 (0.4706)	0.2205 (0.4146)
HAQ	1.085 (0.7692)	1.143 (0.7653)	1.020 (0.7113)	1.032 (0.6542)	1.041 (0.7257)
DAS	4.561 (1.575)	4.501 (1.057)	4.806 (1.064)	5.261 (1.235)	-
RADAI	-	-	-	-	2.875 (1.652)
DISEASE DURATION (MONTHS)	8.403 (6.920)	8.167 (6.077)	-	5.997 (3.089)	12.26 (11.26)
ACR FULFILMENT FOR RA	0.5263 (0.4995)	0.7023 (0.4574)	*	*	-
POSITIVE RHEUMATOID FACTOR	0.6257 (0.4842)	0.6306 (0.4828)	-	-	-
FAILED TWO DMARDS	0.0110 (0.1043)	0 (-)	-	-	0.3118 (0.4632)

*Fulfilment of 1987 ACR criteria for RA was a requirement for entry to these studies

The ERAN and ERAS data also provided information on whether or not a patient fulfilled the American College of Rheumatology's (ACR) criteria for RA. In the ERAN and ERAS datasets 53% of patients and 70% of patients fulfilled these criteria, respectively. Similarly, the ERAN and ERAS datasets contained data on whether patients were classed as having a positive rheumatoid factor (RF) or not at recruitment. Around 63% of patients in both the ERAN and ERAS had a positive RF at the time they were recruited. Patients were considered to have failed two DMARDs if they were on their third or more DMARD. At the time of recruitment only around 1% of patients in the ERAN data had failed two DMARDs. In the NDB data this was much higher, around 31%, because the patients were not necessarily recruited close to their diagnosis and so had a much more varied disease history. No patients in the ERAN data had failed two DMARDs at recruitment because only patients who had never had DMARDs were recruited. The data available to us from the LEIDEN and BARFOT datasets did not include sufficient information on DMARDs to determine the proportions of patients who had failed two DMARDs but since these were also inception cohorts, this figure must be very low.

Table 4 shows the proportion of patients who were observed to have failed two DMARDs at any time during the studies, where data was available. It also gives the mean disease duration from symptom onset to failing the second DMARD and the proportion of patients who are observed to meet NICE criteria for biologics at any time during each study. These data are for patients while they were not receiving biologics; for those patients that receive biologics, they were removed from analysis at that point.

Table 4: Statistics on failing 2 DMARDs

	Mean (standard deviation)		
	ERAN	**ERAS**	**NDB**
Proportion failing 2 DMARDs during observation	0.28 (0.45)	0.26 (0.44)	0.37 (0.48)
Mean time of failing 2 DMARDs from recruitment (months)	27.77 (22.42)	64.53 (40.34)	30.71 (31.95)
NICE criteria for biologics: Proportion observed to fail 2 DMARDs with high DAS	0.11 (0.31)	0.10 (0.30)	0.08 (0.28)
Proportion observed to fail 2 DMARDs with medium DAS	0.10 (0.30)	0.12 (0.32)	0.21 (0.40)

Table 5 presents summary statistics for patients across all observation periods.

Table 5: Summary Statistics over all Observation Periods

	Mean (standard deviation)				
	ERAN	**ERAS**	**LEIDEN**	**BARFOT**	**NDB**
HAQ Observations	5,418	13,234	1,737	9,880	105,678
Age	56.63 (13.47)	58.84 (14.08)	54.83 (15.27)	57.16 (15.37)	66.05 (14.00)
Gender	0.3162 (0.4650)	0.3308 (0.4705)	0.3299 (0.4703)	0.3280 (0.4695	0.2237 (0.4167)
HAQ	0.9776 (0.7897)	0.9860 (0.8168)	0.8129 (0.6861)	0.7311 (0.6592)	0.9588 (0.7232)
DAS	3.681 (1.639)	3.786 (1.211)	3.933 (1.204)	3.689 (1.638)	-
RADAI	-	-	-	-	2.446 (1.601)
Disease duration (months)	8.670 (7.002)	8.393 (6.170)	-	6.034 (3.088)	53.15 (45.37)
Failed two DMARDs	0.1731 (0.3784)	0.1647 (0.3709)	-	-	0.3495 (0.4768)

4.1.2. HAQ Trajectories

Table 6 shows the mean HAQ for each observation period in each data set. It also gives the range of HAQ score for each period.

Table 6: Mean HAQ over time

	MEAN HAQ (HAQ RANGE)				
TIME T (YEARS)	**ERAN**	**ERAS**	**NDB**	**BARFOT**	**LEIDEN**
0	1.085 (0-3)	1.143 (0-3)	1.040 (0-3)	1.032 (0-2.88)	1.020 (0-3)
0.5	1.000 (0-3)	0.884 (0-3)	0.987 (0-3)		
1	0.957 (0-3)	0.848 (0-3)	0.979 (0-3)	0.619 (0-3)	0.710 (0-3)
1.5	-	-	0.970 (0-3)		
2	0.945 (0-3)	0.825 (0-3)	0.958 (0-3)	0.604 (0-3)	0.711 (0-2.75)
2.5	-	-	0.939 (0-3)		
3	0.922 (0-3)	0.898 (0-3)	0.939 (0-3)		0.747 (0-3)
3.5	-	-	0.937 (0-3)		
4	0.885 (0-3)	0.919 (0-3)	0.919 (0-3)		0.677 (0-2.875)
4.5	-	-	0.919 (0-3)		
5	0.937 (0-3)	0.989 (0-3)	0.924 (0-3)	0.645 (0-3)	0.723 (0-2.875)
5.5	-	-	0.916 (0-3)		
6	0.986 (0-3)	1.043 (0-3)	0.907 (0-3)		0.8431 (0-3)

	MEAN HAQ (HAQ RANGE)				
TIME T (YEARS)	**ERAN**	**ERAS**	**NDB**	**BARFOT**	**LEIDEN**
6.5	-	-	0.908 (0-3)		
7	0.966 (0-3)	1.048 (0-3)	0.893 (0-3)		0.7187 (0-2.625)
7.5	-	-	0.889 (0-3)		
8	0.980 (0-3)	1.077 (0-3)	0.870 (0-3)	0.6693 (0-3)	0.9951 (0-2.375)
8.5	-	-	0.895 (0-3)		
9	0.877 (0-2.625)	1.139 (0-3)	0.941 (0-3)		0.725 (0-1.875)
9.5	-	-	0.886 (0-3)		
10	0.926 (0-2.5)	1.137 (0-3)	0.908 (0-3)		
10.5	-	-	0.866 (0-3)		
11	0.500 (0-2.375)	1.000 (0-3)	0.891 (0-3)		
11.5		-	0.880 (0-3)		
12		1.043 (0-3)	0.901 (0-3)		
12.5		-	0.862 (0-3)		
13		1.100 (0-3)	0.831 (0-2.875)		
13.5		-	0.845 (0-2.75)		
14		1.080 (0-3)	0.825 (0-2.875)		

	MEAN HAQ (HAQ RANGE)				
TIME T (YEARS)	**ERAN**	**ERAS**	**NDB**	**BARFOT**	**LEIDEN**
14.5		-	0.774 (0-2.875)		
15		1.165 (0-3)	0.795 (0-2.875)	0.761 (0-3)	

Figure 4 shows the mean HAQ trajectories of patients in each dataset. The error bars show the standard errors of the means of HAQ score at each observation period. These error bars increase over time in all of the datasets but to differing extents. The errors get particularly large towards the later observations in the ERAN and LEIDEN datasets where there are very small numbers of observations. In the majority of the datasets, there is an initial drop in HAQ, creating a J-shaped curve. This is because in all datasets except the NDB, patients were recruited at or shortly after diagnosis and generally put onto therapy at that point or shortly after. This is particularly obvious in the ERAS data, where patients were required not to have taken any DMARDs before joining the sample, so the full effect of the initial DMARDs, or other therapy, is apparent here. After this, there appears to be a slow but steady increase in the mean HAQ score over time in the ERAS and BARFOT datasets. The ERAN and LEIDEN datasets do not show this same steady increase, they have fewer observations and larger error bars. The ERAS and BARFOT data also have a lower dropout rate than the other datasets. The NDB data shows a slow steady decrease in HAQ.

Figure 4: Average HAQ Trajectories

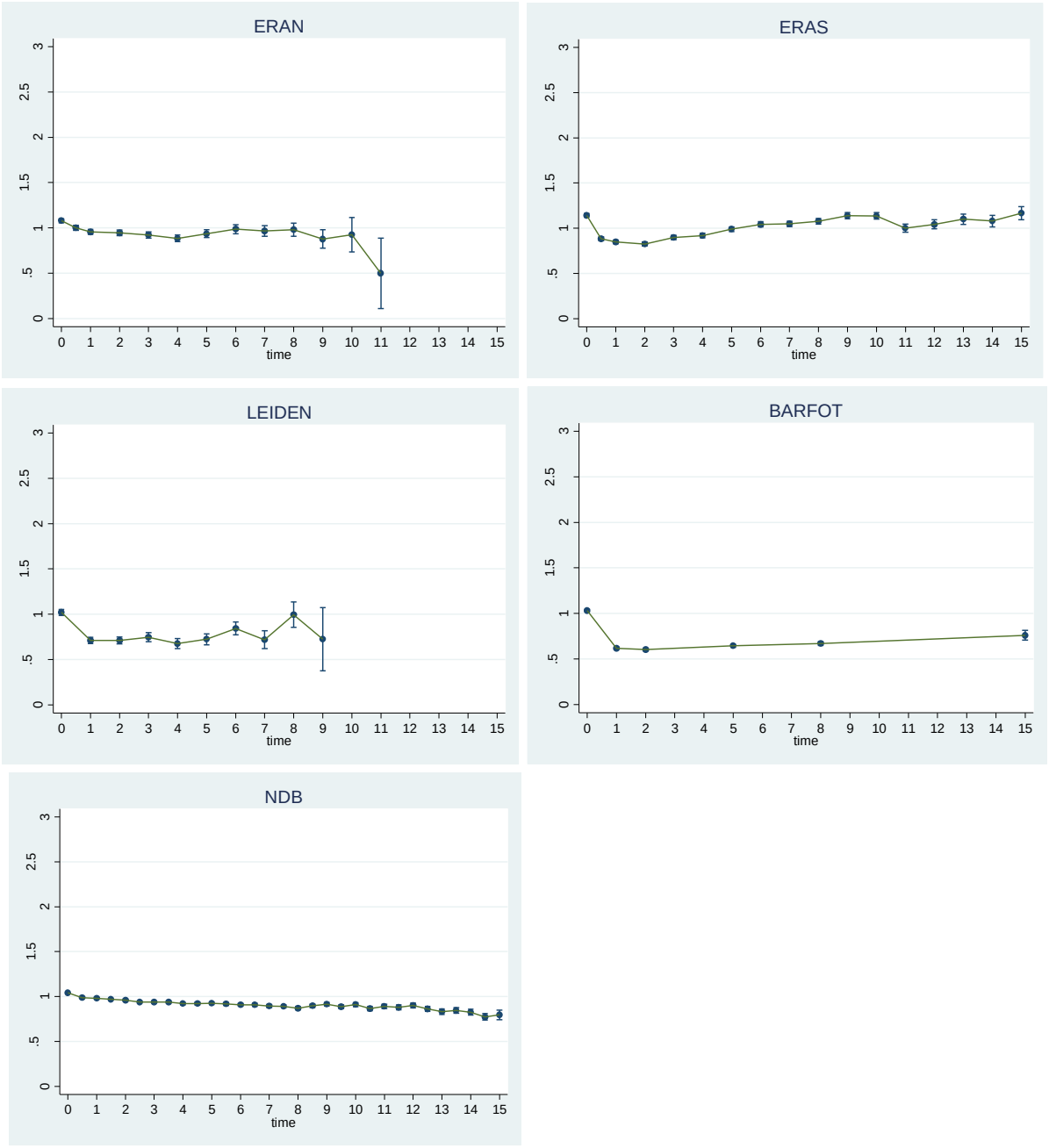
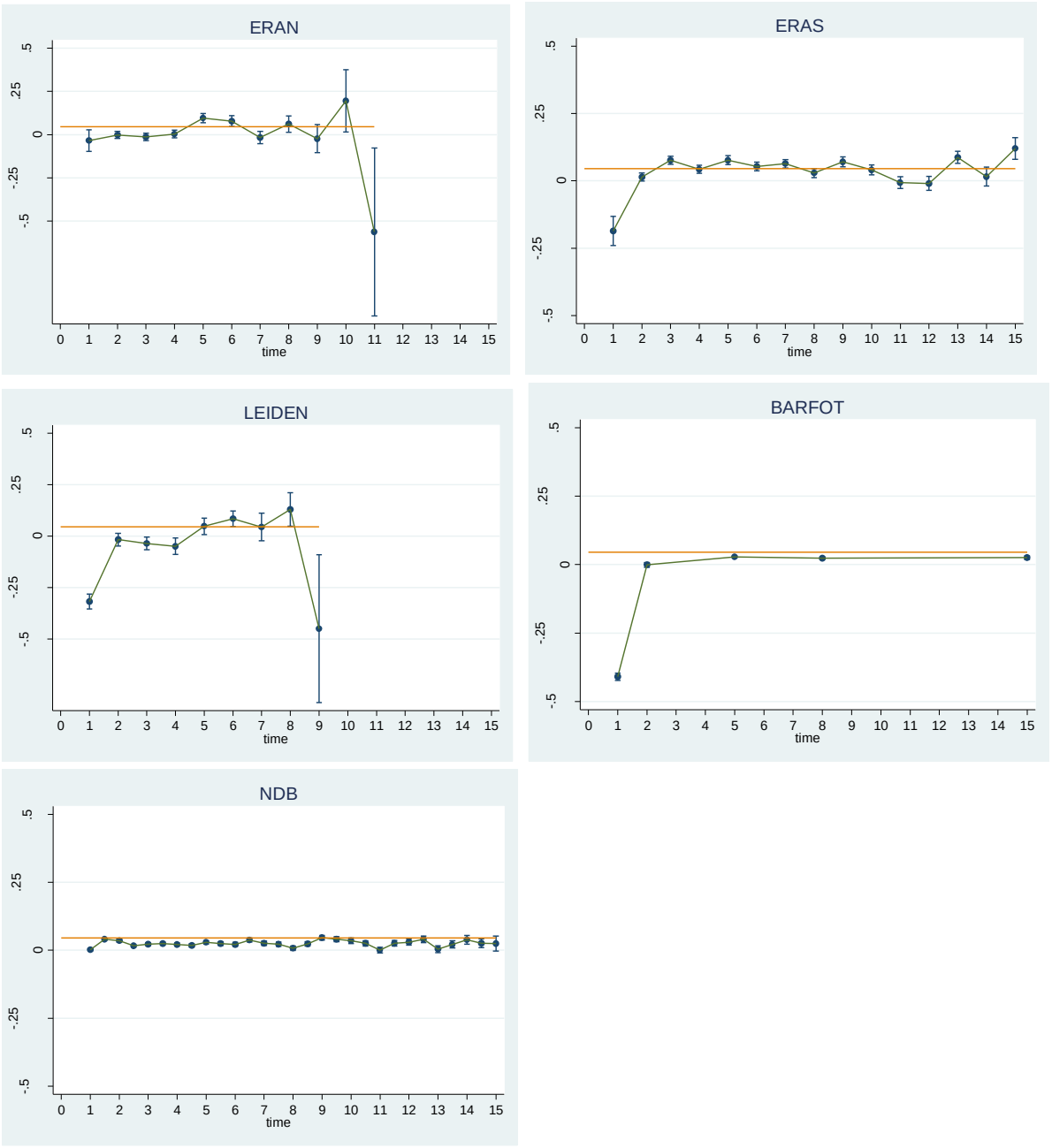


Figure 5 shows the mean annual change in HAQ over time for each of the datasets. Again, the error bars show the standard error of the mean and get larger with time because of the reduced number of patients in the study. In all of the datasets except the NDB, it is possible to see the initial decrease in HAQ immediately after recruitment.

Figure 5: Average Change in HAQ



The horizontal line on each of these graphs shows a change in HAQ of 0.045, the current annual HAQ change assumed for previous appraisals after the initial dip in HAQ observed during the first two years.

Table 7 gives the mean annual change in HAQ from year two to the end of each study and also from year two to year eight (since this is where generally there is greater certainty due to larger numbers of observations).

Table 7: Mean Changes in HAQ

MEAN CHANGE	ERAN	ERAS	LEIDEN	BARFOT	NDB
$t = 2$ TO END	0.0115	0.0542	-0.0016	0.0151	0.0269
$t = 2$ TO $t = 8$	0.0124	0.0568	0.0010	0.0147	0.0255

Based on the raw data alone, the observed annual rate of change of HAQ is lower than 0.045 in all datasets except ERAS.

4.1.3. Missing Data and Attrition

Longitudinal studies of this type inevitably suffer both from missing data and attrition. Table 8 to Table 12 show the number of observations at each observation period in each dataset. They also show the number of missing HAQ scores at each time and the number of observations which have left each dataset due to attrition.

There are a number of possible different reasons for attrition. Attrition could be caused because patients do not respond or do not fill in the questionnaires, some patients die, some patients leave the samples because they have adverse reactions to the drugs or move onto other biologic therapies or patients might go into remission, among other reasons. The dataset available to us had missing or incomplete data on the reasons for attrition. Remission is an important cause of dropout because for several studies this is an inherent part of the study design: BARFOT and the LEIDEN studies withdrew patients from the study once the patient was no longer under the care of the rheumatologist.

Table 8 to Table 12 show the number of observations, missing observations and attrition rates of the ERAN, ERAS, LEIDEN, BARFOT and NDB datasets, respectively. In each of the datasets patients were not included because there were insufficient data in all observation periods; there were 112, 37, 48, 13 and 4,580 patients removed for this reason in the respective datasets.

Table 8: Missing Data and Attrition in the ERAN Data (n=1,124)

TIME T (YEARS)	HAQ OBSERVATIONS	MISSING AT TIME T	ATTRITION
	NUMBER OF OBSERVATIONS (%)		
0	1,091 (97.06%)	33 (2.94%)	0 (0.00%)
0.5	857 (76.25%)	220 (19.57%)	47 (4.18%)
1	696 (61.92%)	271 (24.11%)	157 (13.97%)
2	585 (52.05%)	235 (20.91%)	304 (27.05%)
3	478 (42.53%)	217 (19.31%)	429 (38.17%)
4	367 (32.65%)	179 (15.93%)	578 (51.42%)
5	259 (23.04%)	159 (14.15%)	706 (62.81%)
6	167 (14.86%)	144 (12.81%)	813 (72.33%)
7	98 (8.72%)	144 (12.81%)	882 (78.47%)
8	110 (9.79%)	126 (11.21%)	888 (79.00%)
9	64 (5.69%)	171 (15.21%)	889 (79.09%)
10	22 (1.96%)	2 (0.18%)	1,100 (97.86%)
11	6 (0.53%)	0 (0.00%)	1,118 (99.47%)

Table 9: Missing Data and Attrition in the ERAS Data (n=1,430)

TIME T (YEARS)	# HAQ OBSERVATIONS	MISSING AT TIME T	ATTRITION
	NUMBER OF OBSERVATIONS (%)		
0	1,421 (99.37%)	9 (0.63%)	0 (0.00%)
0.5	1,244 (92.31%)	181 (3.22%)	5 (0.35%)
1	1,320 (86.99%)	46 (12.66%)	64 (4.48%)
2	1,165 (81.47%)	120 (8.39%)	145 (10.14%)
3	1,068 (74.69%)	123 (8.60%)	239 (16.71%)
4	906 (65.52%)	220 (8.53%)	304 (21.26%)
5	937 (63.36%)	122 (15.38%)	371 (25.94%)
6	767 (53.64%)	209 (14.62%)	454 (31.75%)
7	753 (52.66%)	156 (10.91%)	521 (36.43%)
8	691 (48.32%)	157 (10.98%)	582 (40.70%)
9	615 (43.01%)	86 (6.01%)	729 (50.98%)
10	382 (26.71%)	60 (4.20%)	988 (69.09%)
11	272 (19.02%)	77 (5.38%)	1,081 (75.59%)
12	210 (14.69%)	66 (4.62%)	1,154 (80.70%)
13	161 (11.26%)	49 (3.43%)	1,220 (85.31%)

TIME T (YEARS)	# HAQ OBSERVATIONS	MISSING AT TIME T	ATTRITION
14	124 (8.67%)	6 (0.42%)	1,300 (90.91%)
15	35 (2.45%)	0 (0.00%)	1,395 (97.55%)

Table 10: Missing Data and Attrition in the LEIDEN Data (n=563)

TIME T (YEARS)	HAQ OBSERVATIONS	MISSING AT TIME T	ATTRITION
	NUMBER OF OBSERVATIONS (%)		
0	498 (88.45%)	65 (11.55%)	0 (0.00%)
1	359 (63.77%)	79 (14.03%)	125 (22.20%)
2	259 (46.00%)	64 (11.37%)	240 (42.63%)
3	185 (32.86%)	57 (10.12%)	321 (57.02%)
4	149 (26.47%)	46 (8.17%)	368 (65.36%)
5	119 (21.14%)	25 (4.44%)	419 (74.42%)
6	97 (17.23%)	12 (2.13%)	454 (80.64%)
7	40 (7.10%)	9 (1.60%)	514 (91.30%)
8	26 (4.62%)	1 (0.18%)	536 (95.20%)
9	5 (0.89%)	0 (0.00%)	558 (99.11%)

Table 11: Missing Data and Attrition in the BARFOT Data (n=2,595)

TIME T (YEARS)	# HAQ OBSERVATIONS	MISSING AT TIME T	ATTRITION
	NUMBER OF OBSERVATIONS (%)		
0	2,455 (94.61%)	140 (5.39%)	0 (0.00%)
1	2,332 (80.73%)	120 (5.01%)	143 (5.51%)
2	2,095 (80.73%)	130 (5.01%)	370 (14.26%)
5	1,756 (67.67%)	91 (3.51%)	748 (28.82%)
8	1,059 (40.81%)	12 (0.46%)	1,524 (58.73%)
15	183 (7.05%)	0 (0.00%)	2,412 (92.95%)

Table 12: Missing Data and Attrition in the NDB Data (n=19,462)

TIME T (YEARS)	HAQ OBSERVATIONS	MISSING AT TIME T	ATTRITION	CENSORED
	NUMBER OF OBSERVATIONS (%)			
0	19,294 (98.58%)	277 (1.42%)	0 (0.00%)	0 (0.00%)
0.5	12,740 (65.10%)	1,519 (7.76%)	5,176 (26.45%)	136 (0.69%)
1	10,461 (53.45%)	1,229 (6.28%)	7,666 (39.17%)	215 (1.10%)
1.5	8,464 (43.25%)	1,056 (5.40%)	9,745 (49.79%)	306 (1.56%)
2	6,981 (35.67%)	931 (4.76%)	11,296 (57.72%)	363 (1.85%)
2.5	5,871 (30.00%)	799 (4.08%)	12,460 (63.67%)	441 (2.25%)

Time t (years)	HAQ Observations	Missing at time t	Attrition	Censored
	Number of observations (%)			
3	5,011 (25.60%)	704 (3.60%)	13,373 (68.33%)	483 (2.47%)
3.5	4,162 (21.27%)	808 (4.13%)	14,079 (71.94%)	522 (2.67%)
4	3,769 (19.26%)	557 (2.85%)	14,670 (74.96%)	575 (2.94%)
4.5	3,274 (16.73%)	521 (2.66%)	15,154 (77.43%)	622 (3.18%)
5	3,001 (15.33%)	364 (1.86%)	15,554 (79.47%)	652 (3.33%)
5.5	2,519 (12.87%)	461 (2.36%)	15,911 (81.30%)	680 (3.47%)
6	2,435 (12.44%)	235 (1.20%)	16,192 (82.73%)	709 (3.62%)
6.5	2,044 (10.44%)	328 (1.68%)	16,455 (84.08%)	744 (3.80%)
7	1,935 (9.89%)	182 (0.93%)	16,684 (85.25%)	770 (3.93%)
7.5	1,722 (8.80%)	191 (0.98%)	16,885 (86.28%)	773 (3.95%)
8	1,622 (8.29%)	138 (0.71%)	17,032 (87.03%)	779 (3.98%)
8.5	1,439 (7.35%)	129 (0.66%)	17,218 (87.98%)	785 (4.01%)
9	1,293 (6.61%)	105 (0.54%)	17,367 (88.74%)	806 (4.12%)
9.5	1,147 (5.86%)	120 (0.61%)	17,483 (89.33%)	821 (4.19%)
10	1,042 (5.32%)	107 (0.55%)	17,582 (89.84%)	840 (4.29%)
10.5	964 (4.93%)	80 (0.41%)	17,665 (90.26%)	862 (4.40%)

TIME T (YEARS)	HAQ OBSERVATIONS	MISSING AT TIME T	ATTRITION	CENSORED
	NUMBER OF OBSERVATIONS (%)			
11	885 (4.52%)	64 (0.33%)	17,751 (90.70%)	871 (4.45%)
11.5	774 (3.95%)	54 (0.28%)	17,855 (91.23%)	888 (4.54%)
12	670 (3.42%)	60 (0.31%)	17,931 (91.62%)	910 (4.65%)
12.5	616 (3.15%)	46 (0.24%)	17,993 (91.94%)	916 (4.68%)
13	560 (2.86%)	41 (0.21%)	18,048 (92.22%)	922 (4.71%)
13.5	474 (2.42%)	33 (0.17%)	18,118 (92.58%)	946 (4.83%)
14	408 (2.08%)	13 (0.07%)	18,176 (92.87%)	974 (4.98%)
14.5	315 (1.61%)	8 (0.04%)	18,260 (93.30%)	988 (5.05%)
15	161 (0.82%)	0 (0.0%)	18,342 (93.72%)	1,068 (5.46%)

In Table 12, showing the attrition rate and missing observations for the NDB, some of the patients are censored rather than lost to attrition. Due to the observational nature of the NDB data patients were continually recruited to the study throughout the observation period of fifteen years. In this report, we analyse the data using the time of patient recruitment to the NDB as baseline; therefore patients who are recruited later in the study have a limited number of follow up observations and their follow up might stop before 15 years because they had been enrolled for less than 15 years. Patients who do not have a full 15 years of data but who were remaining in the NDB study when it ended are censored rather than lost to attrition. For the purposes of our analysis the censored observations will be treated in the same way as attrition.

Figure 6 shows the percentage of patients who are lost from the study due to attrition over the years of the study. Patients who were missing but return to the study are included in all periods that they were observed. The ERAS and the BARFOT data have the lowest attrition rates but attrition is a problem in all of the datasets, as is the case with most longitudinal studies. The attrition rate of the NDB is given with and without censored observations.

Figure 6: Attrition over time

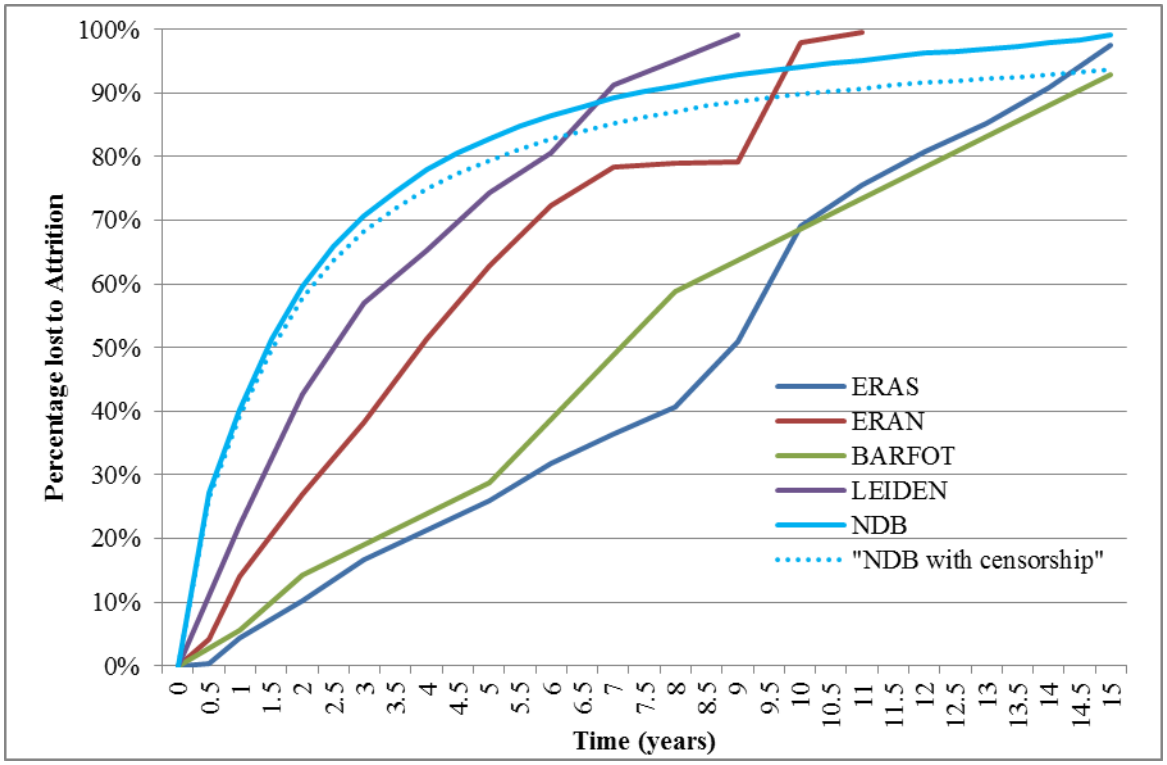
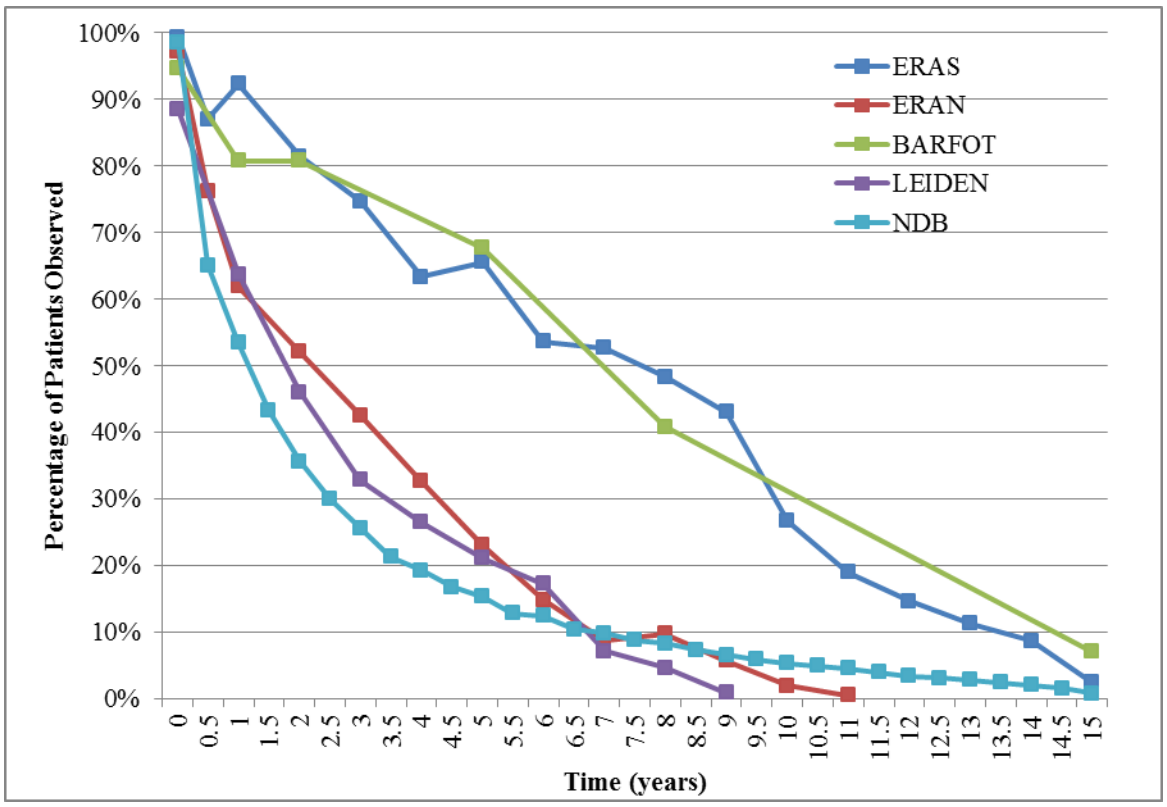


Figure 7 shows the percentage of patients who are observed at each time point. This percentage does not include patients who are missing during an observation period but later return to the study. Again, ERAS and BARFOT datasets show the best patient attendance.

Figure 7: Percentage of patients observed over time



4.2. LATENT CLASS GROWTH MODELS (MAR)

4.2.1. ERAS

The covariates used to predict latent class membership are the same as in Norton *et al*:[3] age, gender, DAS, symptom duration, rheumatoid factor, and fulfilment of ACR criteria, all at baseline.

We find that a model with four distinct latent classes has the best fit, in accordance with Norton *et al*.[3] The goodness of fit statistics, namely the Akaike and Bayesian information criterion (AIC and BIC, respectively) suggest that four classes is better than three classes, as shown in Table 13. We had an insufficient number of observations in the data available to model five classes using all observation periods.

Table 13: Cubic and Quadratic Information Criteria for ERAS data

Specification	Number of Classes	
Cubic	**3 Classes**	**4 Classes**
AIC	24862.426	24007.085
BIC	25087.989	24291.682
Quadratic	3 Classes	4 Classes
AIC	24980.963	24133.792
BIC	25190.789	24397.307

Figure 8 and Figure 9 show the cubic and quadratic LCGA models respectively using the ERAS data.

Figure 8: Cubic Replication of Analysis on ERAS by Norton *et al*.

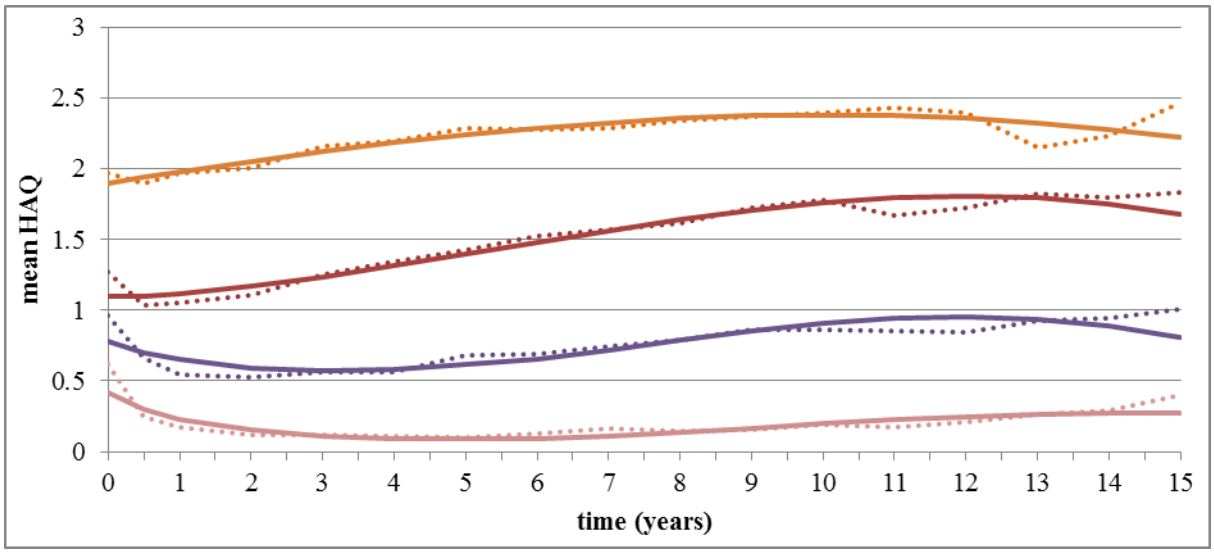


Figure 8 shows the results estimated using cubic LCGA for the ERAS data. Again, the class with the highest HAQ scores at baseline remain highest throughout the observation period and have a shallower J-shape after recruitment.

Table 14: Annual Rate of HAQ change – Cubic LCGA ERAS

Rate of Change	BSRBR	High DAS	Moderate DAS
After $t = 2$	0.0326	0.0293	0.0264
$t = 2$ to $t = 8$	0.0617	0.0551	0.0480

Figure 9 shows the results estimated using a quadratic LCGA for the ERAS data. This uses the same formula at the cubic example outlined above but restricts the coefficient of the cubic term to be zero ($\eta_{3ic} = 0$), leaving only the intercept, slope and quadratic terms,

$$y^*_{itc} = \eta_{0ic} + \eta_{1ic}x_t + \eta_{2ic}x_t^2 + \varepsilon_{it};$$

everything else remains the same as the previous, cubic model. Again, we initially restrict the variances of the intercept, slope, quadratic and cubic terms to zero.

Figure 9: Quadratic Replication of Analysis on ERAS by Norton *et al*.

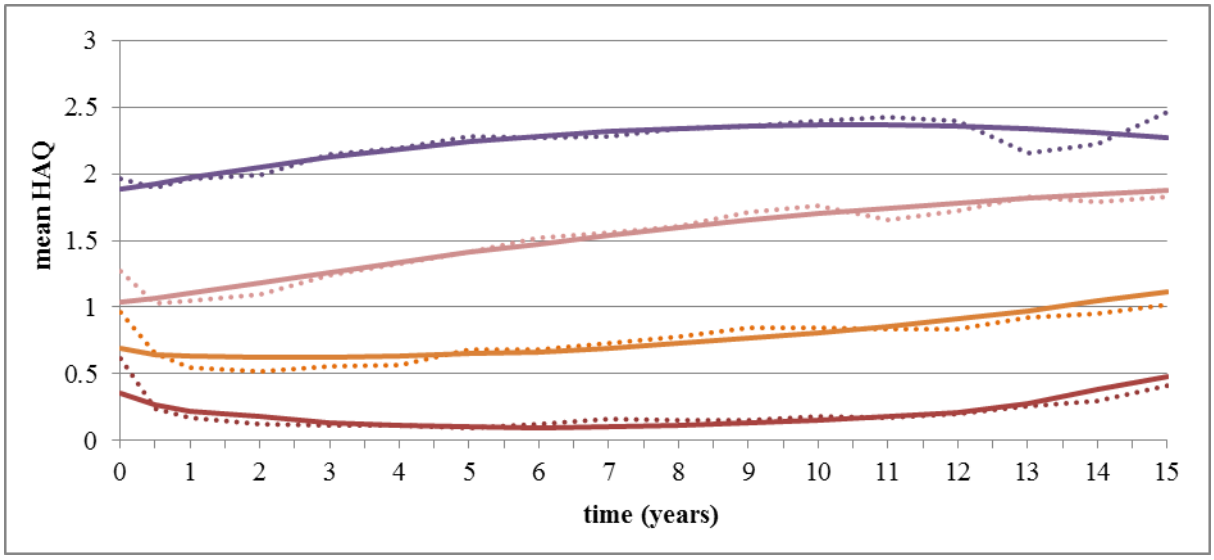


Table 15: Annual Rate of HAQ change – Quadratic LCGA ERAS

Rate of Change	BSRBR	High DAS	Moderate DAS
After $t = 2$	0.0377	0.0393	0.0412
$t = 2$ to $t = 8$	0.0273	0.0316	0.0338

Similar to the findings of Norton *et al*. (2014), in both the cubic and quadratic models shown here, the class with the highest HAQ scores at baseline, remain highest throughout the observation period; this class also has a flatter J-shape after recruitment suggesting that patients in this latent class do not achieve a substantial response to initial, nor indeed subsequent, DMARDs. Both the cubic and quadratic models predict a decrease in HAQ towards the end of the observation period, particularly in the latent classes with the highest HAQ values. However, this is not representative of the observed values which do not show this dip in HAQ.

4.2.2. ERAN

Similar results were found using the other datasets and four classes was consistently the optimal number of classes. Figure 10 to Figure 14 show LCGA for the remaining data sets accounting for the following covariates: age, gender, DAS and symptom duration all at

baseline. In these analyses, less explanatory variables are used because they are available in all datasets; this makes little difference to the predicted values. Again, the variances of the intercept, slope, quadratic and cubic term, where applicable, are restricted to zero in accordance with the Norton *et al*. papers.

Figure 10: Cubic LCGA of ERAN data

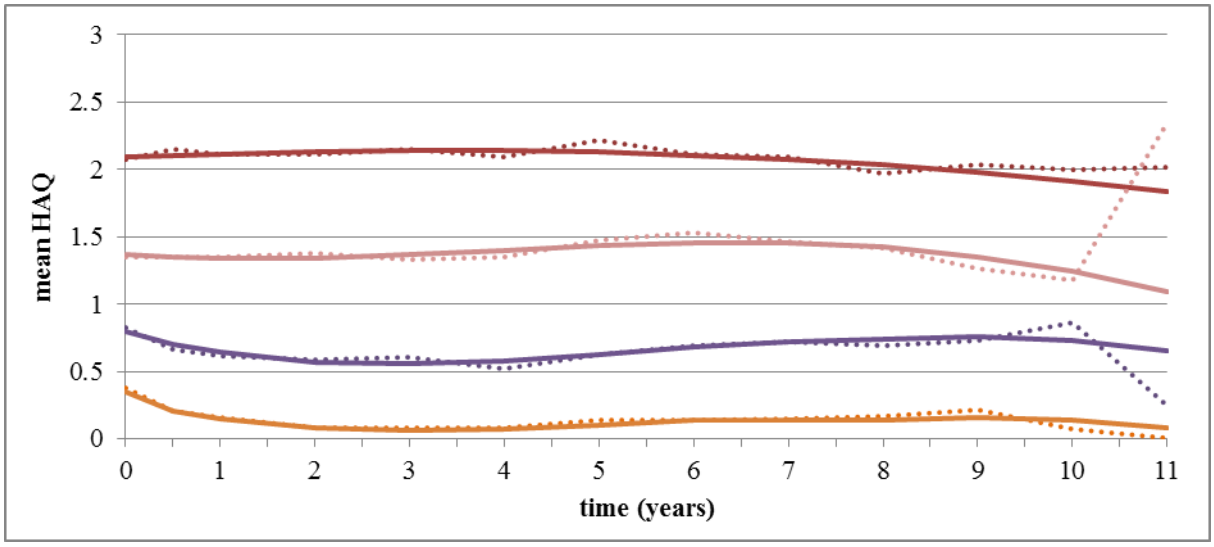


Table 16: Annual Rate of HAQ change – Cubic LCGA ERAN

Rate of Change	BSRBR	High DAS	Moderate DAS
After $t = 2$	-0.0866	-0.0812	-0.0554
$t = 2$ to $t = 8$	0.0095	0.0085	0.0074

Both Figure 8 and Figure 9 show that the ERAN and ERAS data give similar results. The ERAN data does not show as steep a J-shaped curved immediately after recruitment but the trajectories are very similar. The analysis on the ERAN data does not use all of the covariates which are included in the ERAS analysis, but this seems to make little difference to the trajectories.

Figure 11: Quadratic LCGA of ERAN data

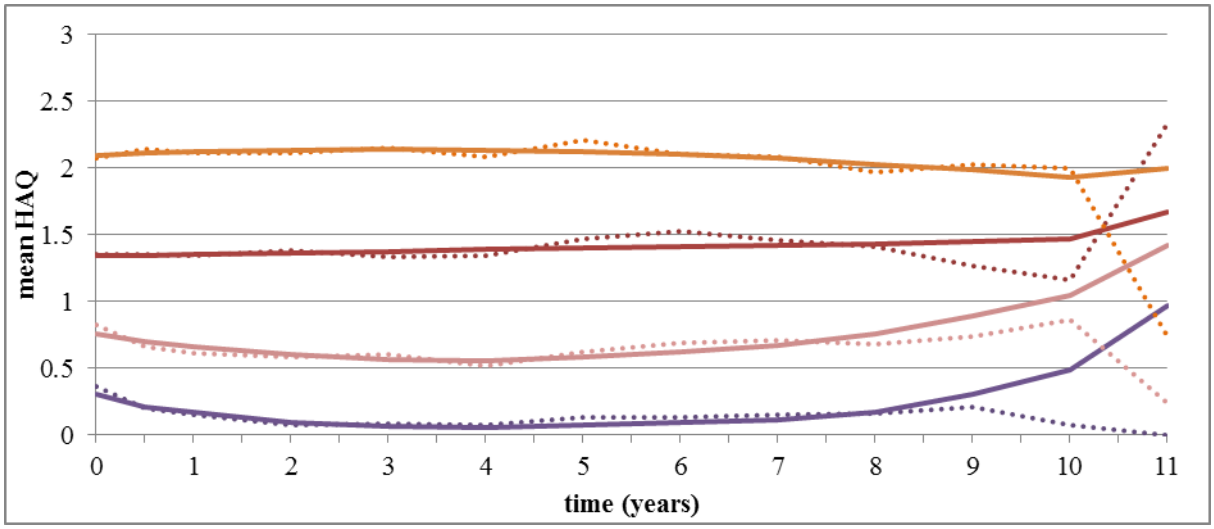


Table 17: Annual Rate of HAQ change – Quadratic LCGA ERAN

Rate of Change	BSRBR	High DAS	Moderate DAS
After $t = 2$	0.0525	0.0548	0.0677
$t = 2$ to $t = 8$	0.0087	0.0079	0.0069

4.2.3. BARFOT

Figure 12 and Figure 13 show the cubic and quadratic LCGA for the BARFOT data. Here, the same covariates have been used as those used in the analysis on the ERAN data; namely age, gender, DAS and symptom duration at baseline. The number of previous DMARDs is also used as a predictor of class membership.

Figure 12: Cubic LCGA of BARFOT data

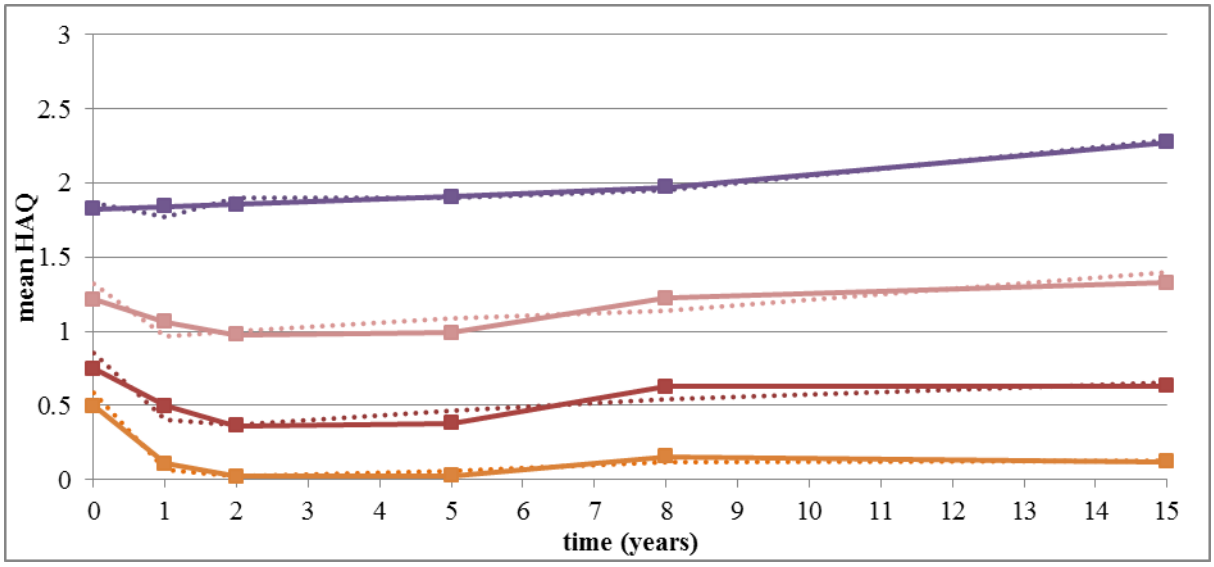


Table 18: Annual Rate of HAQ change – Cubic LCGA BARFOT

Rate of Change	BSRBR	High DAS	Moderate DAS
After $t = 2$	0.0187	0.0187	0.0187
$t = 2$ to $t = 8$	0.0226	0.0225	0.0225

Figure 13: Quadratic LCGA of BARFOT data

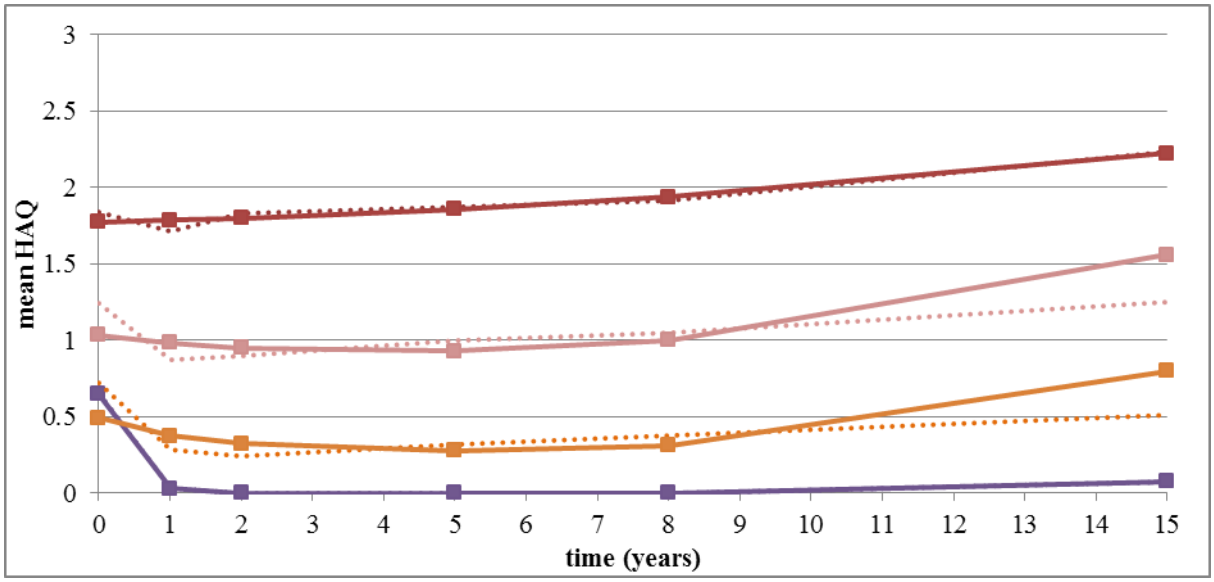


Table 19: Annual Rate of HAQ change – Quadratic LCGA BARFOT

Rate of Change	BSRBR	High DAS	Moderate DAS
After $t = 2$	0.0412	0.0408	0.0405
$t = 2$ to $t = 8$	0.0026	0.0022	0.0019

The analysis on the BARFOT data is limited due to the lack of observation periods: observations are not annual unlike the other datasets.

4.2.4. NDB

Figure 14 shows the quadratic model using data from the NDB data. Here, the covariates used to predict class membership are age, gender, RADAI score and symptom duration all at baseline. RADAI score was used as a substitute for DAS because DAS is not available in the NDB data. These patients were not recruited as close to their symptom onset as those in other datasets and therefore patients are more likely to have tried a range of therapies at their time of recruitment; the data includes lots of patients who had already moved onto their third DMARD. This is the reason for the lack of the J-shape in the curve immediately after recruitment.

Figure 14: Quadratic LCGA of NDB data

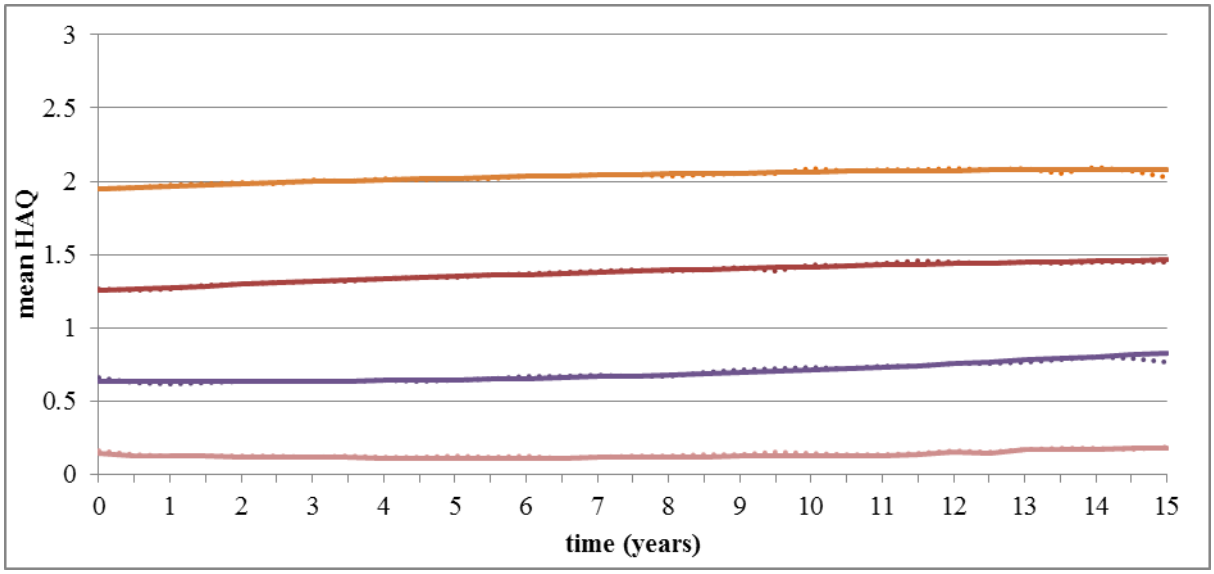


Figure 14 shows a slow steady increase in HAQ overtime for the NDB data. Due to the absence of information on DAS score it is not possible to illustrate these results for the subgroups of patients relevant to the NICE decision problem.

Due to the very large data in the NDB it was not feasible to run the cubic LCGA on the NDB data due to limitations on processing time.

4.2.5. Leiden

Attempts to estimate HAQ trajectory using LCGA on the LEIDEN data were unsuccessful. The models would not converge using any specification. This is believed to be due to a smaller number of observations.

4.2.6. Additional analyses using ERAS data

Figure 15 shows a cubic LCGA with four classes using the ERAS data but allowing the variances of the random intercept, slope and quadratic term to be free. The variance of the cubic term is still fixed at zero in an attempt to prevent the predicted trajectories from continuing to increase above a HAQ of three. However the latent class represented in blue shows a more exaggerated curve than the other classes.

Figure 15: Cubic LCGA of ERAS data with free variances

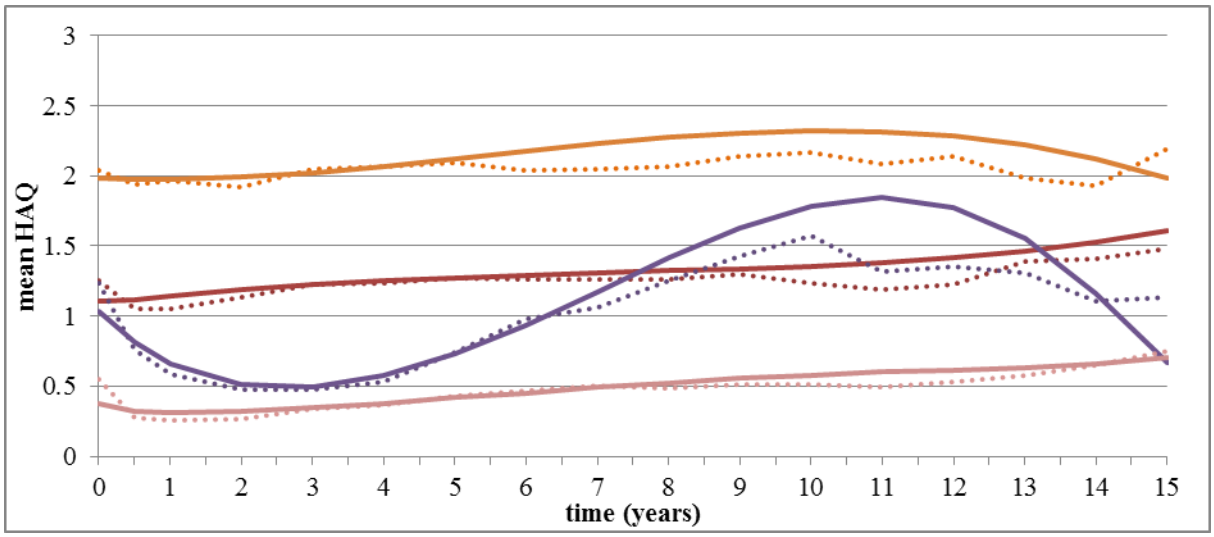


Table 20: Annual Rate of HAQ change – Cubic LCGA ERAS with Free Variances

Rate of Change	BSRBR	High DAS	Moderate DAS
After $t = 2$	0.0329	0.0326	0.0327
$t = 2$ to $t = 8$	0.0257	0.0263	0.0262

Figure 16: Quadratic LCGA of ERAS data with free variances

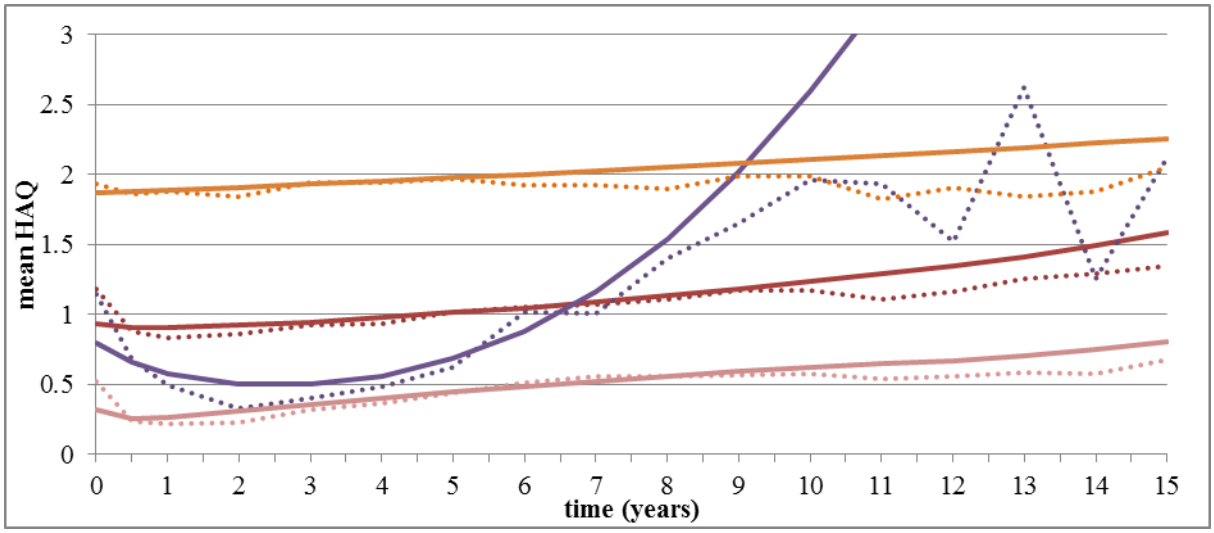


Table 21: Annual Rate of HAQ change – Quadratic LCGA ERAS with Free Variances

Rate of Change	BSRBR	High DAS	Moderate DAS
After $t = 2$	0.0478	0.0475	0.0480
$t = 2$ to $t = 8$	0.0326	0.0326	0.0330

Figure 16 shows a quadratic LCGA with four classes using the ERAS data but allowing the variances of the random intercept, slope and quadratic term to be free. This graph highlights the problems that can occur when using quadratic and cubic terms; some of the HAQ trajectories shoot upwards and increase above the maximum feasible HAQ of three. This is particularly problematic when trying to predict the HAQ score of patients in the future. However, it should be noted that for the application to the three patient groups of interest in this report, there was a probability of zero of being in the class that increases to a HAQ of three at year ten.

4.2.7. Comparisons of models

Table 22 shows the AIC, BIC and log-likelihood values of the cubic and quadratic models for each dataset. Likelihood ratio tests using the log-likelihood values indicate that the cubic models consistently have a significantly better fit than the quadratic specifications.

Table 22: Log-likelihood and Information Criteria for Cubic and Quadratic Models

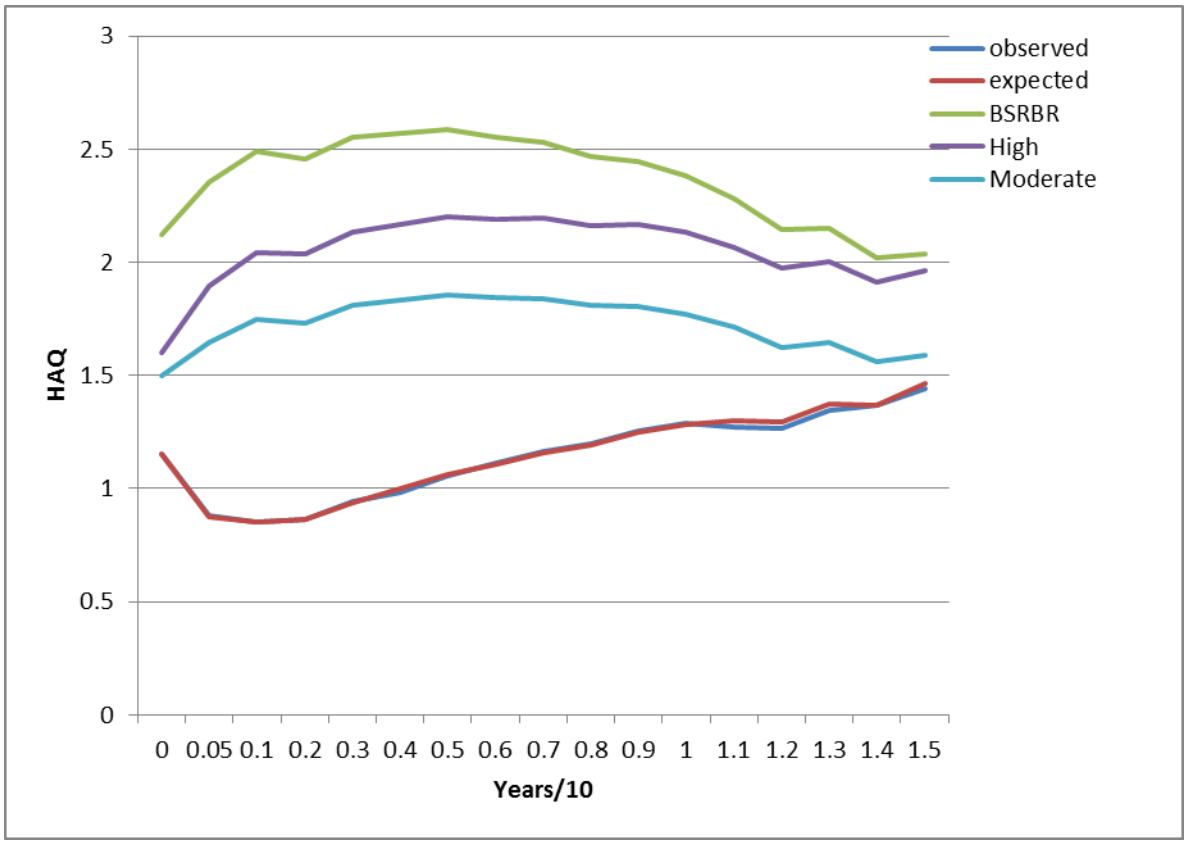
	Specification	
ERAS	Cubic	Quadratic
AIC	*24007.085*	24133.792
BIC	*24291.682*	24397.307
Log-likelihood	-11949.543	-12016.896
ERAN	Cubic	Quadratic
AIC	*9551.462*	9570.955
BIC	9788.899	*9788.184*
Log-likelihood	-4728.731	-4742.477
BARFOT	Cubic	Quadratic
AIC	*17225.061*	17644.210
BIC	*17458.247*	17836.588
Log-likelihood	-8572.531	-8789.105
NDB	Cubic	Quadratic
AIC	-	229412.184
BIC	-	229913.228
Log-likelihood	-	-114645.092
ERAS with free variance	Cubic	Quadratic
AIC	*21592.140*	21850.809
BIC	*21906.879*	22144.566
Log-likelihood	-10736.070	-10869.404

4.3. ALT MODELS (MAR)

4.3.1. ERAS

We fitted the ALT model to the ERAS data in order to identify the extent to which this alternative modelling approach did, or did not, coincide with the estimates made using the latent class analysis. Figure 17 displays the results from this exercise.

Figure 17: ALT model of ERAS data



We found very close alignment between the observed and predicted data. The graph shows that there is virtually no difference between the observed and prediced mean HAQ at each year, all the way to 15 years. This shows a trajectory that is J-shaped with a lower degree of flattening than was predicted in the separate classes of the latent class analysis.

The model was then used to predict the expected HAQ over time for patients with the characteristics of the three patient subgroups (BSRBR, severe and moderate DAS groups). This demonstrates a different pattern. There is a predicted rapid worsening in HAQ which decreases over time and then falls. The flexibility of the ALT model does not impose this shape but rather this is casued by the characteristics of the subgroups. These groups have much higher starting HAQs than the ERAS early RA population.

Within the ERAS data there is a negative correlation between disease duration and baseline HAQ. This stands to reason as one might expect that those with more aggressive disease, manifesting itself in functional disability, would have a more rapid worsening of their symptons causing them to seek help at an early time than the average early RA patient, and

therefore reciving a first visit with a rheumatologist earlier. However, this is a relationship observed within the small range of disease durations seen amongst the ERAS baseline population. Once this is applied to disease durations exceeding that by a factor of ten this relationship is questionable and is a substantial contributory factor in the predicted HAQ reducing at longer time periods. This demonstrates a limitation of extrapolation beyond the data. The model results are considered to lend support to the qualitative finding in the latent class analyses of rapid worsening in HAQ followed by a period where this slows and potentially flattens.

4.4. Sub analysis of Patients After Failing Two DMARDs

All analyses to this point have been conducted on the entire datasets provided to us, recognising that these included observations from patients at different stages of disease and with different characteristics to those that are eligble, or potentially eligible, for biologic drug treatment under NICE guidelines. We have used the full datsets and attempted to adjust for these differences within the modelling. An alternative view is to restrict analysis to those data drawn from relevant patients.

Here we investigate the possibility of estimating HAQ trajectories in patients drawn solely from the relevant (or at least, more relevant) groups. In this section, we investigate HAQ trajectories of patients who have already failed two DMARDs. We created new datasets where $t = 0$ was the first period in which a patient was observed on their third (or greater) non biologic DMARD. This was possible for the ERAN, ERAS and NDB datasets which each had information on the number of DMARDs at each observation period and regular observation periods; the longer times between observation periods mean that we cannot know the exact year that the third DMARD is received. Although the BARFOT data had sufficient data on DMARDs, the irregularity of the observation periods meant that it was difficult to tell when patients were moved onto their third DMARD. The LEIDEN data did not have sufficient data on DMARDs. For the ERAN and ERAS data, the observations at six months from recruitment were removed.

Table 23 shows the summary statistics for the data after removing patients who had not yet failed two DMARDs. The number of observations in these new datasets is, of course, much smaller than the original datasets. Again, the average age of patients at the new baseline is in

their mid to late fifties. The number of female patients at the new baseline is slightly higher than at the original baseline, suggesting that female patients are more likely to have failed two DMARDs, possibly because they tend to have higher DAS and HAQ scores. The average HAQ score at the new baseline is higher than at the previous baseline, particularly in the ERAS data.

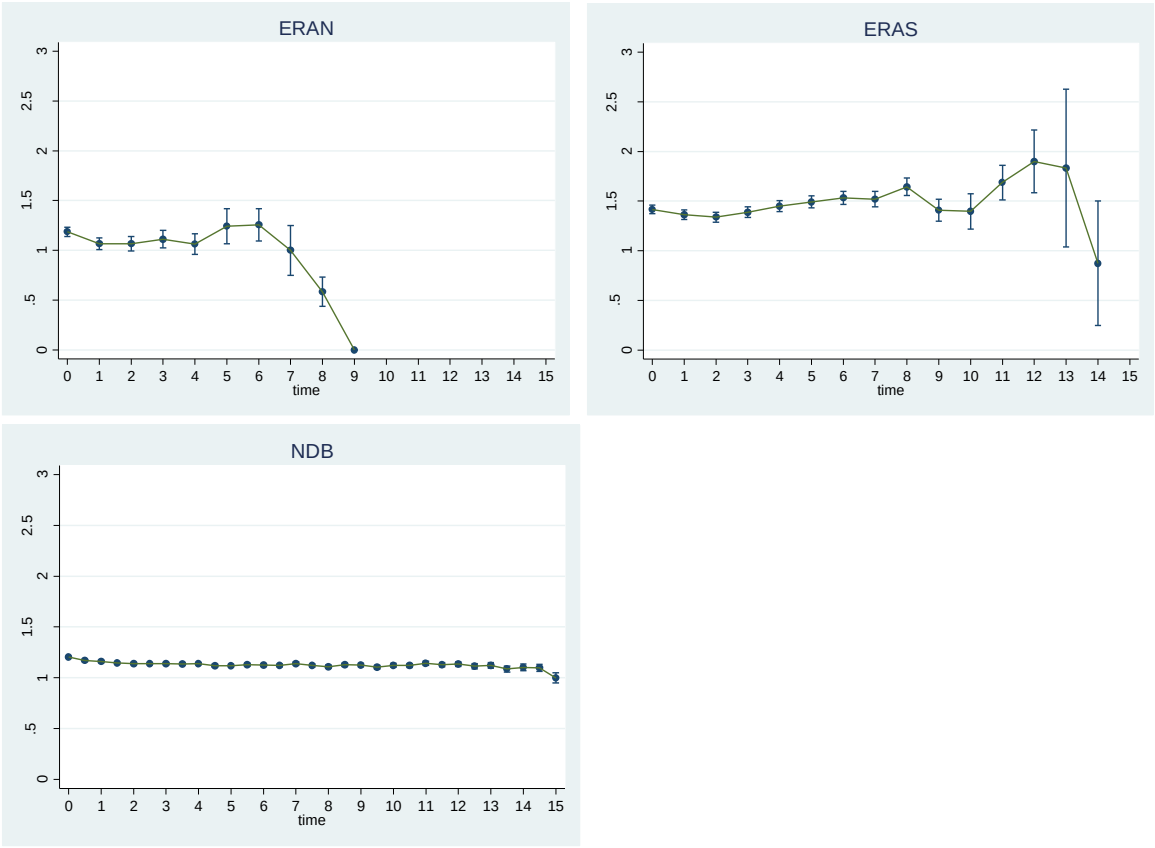
Table 23: Summary Statistics after taking Third DMARD (new baseline)

	ERAN	ERAS	NDB
# HAQ Observations	285	380	7,353
Age at 3rd DMARD	54.55 (12.52)	55.54 (14.12)	58.67 (12.88)
Gender	0.2912 (0.4551)	0.2579 (0.4381)	0.1922 (0.3940)
HAQ	1.186 (0.8058)	1.414 (0.8046)	1.191 (0.7183)
DAS	3.958 (1.726)	4.317 (1.154)	-
DAS>5.1	0.2561 (0.4373)	0.2421 (0.4289)	-
RADAI	-	-	3.003 (1.609)
Disease duration (months)	38.78 (22.42)	68.78 (42.64)	19.69 (20.27)
Positive Rheumatoid Factor	0.6245 (0.4852)	0.6772 (0.4681)	-

As all patients in this new sample have failed at least two DMARDs, those patients who meet the NICE criteria for biologic drugs are simply those who have a DAS score higher than 5.1 at the baseline. The proportion of patients fulfilling these criteria is similar in the ERAN and ERAS datasets. Approximately one quarter of patients fulfilled the current NICE criteria for biologics when they were observed to have started their third DMARD, i.e. a quarter of patients who fail two DMARDs have a high DAS score. It is worth noting that we do not know the exact time that the patients started their third DMARD and some of them could have responded to the third DMARD with lower DAS and HAQ scores. The proportion of patients with a high DAS and therefore, the proportion who meet NICE criteria for biologics, may have been higher if we had the DAS values immediately after they failed their second DMARD.

Figure 18 shows the average HAQ trajectories after failing two DMARDs using the ERAN, ERAS and NDB datasets. Here, each of the different datasets shows a different pattern of HAQ over time.

Figure 18: Average HAQ Trajectories



The ERAN data show a decrease in average HAQ over time. However, the standard error bars are wide and there are very few observations towards the later years. The ERAS data also has wide error bars towards the later observations but the data shows a steady increase in average HAQ score until 13 years after starting the third DMARD; this drop may be due to attrition bias, as mentioned previously. The NDB data shows a relatively stable average HAQ over time. HAQ remains slightly above one throughout the observation periods. The J-shaped curve seen before in Figure 4 and Figure 5 is not seen here because the benefits of the initial DMARDs given shortly after diagnosis are not seen.

Figure 19 shows the average change in HAQ over time from starting the third DMARD. The patterns are consistent with those shown in Figure 18. Again, the ERAN data demonstrates periods of rising HAQ followed by falling HAQ but this is subject to very wide confidence

limits due to attrition. The ERAS data shows a general pattern of worsening in HAQ. The NDB shows worsening in HAQ but at a very low rate.

Figure 19: Average Change in HAQ

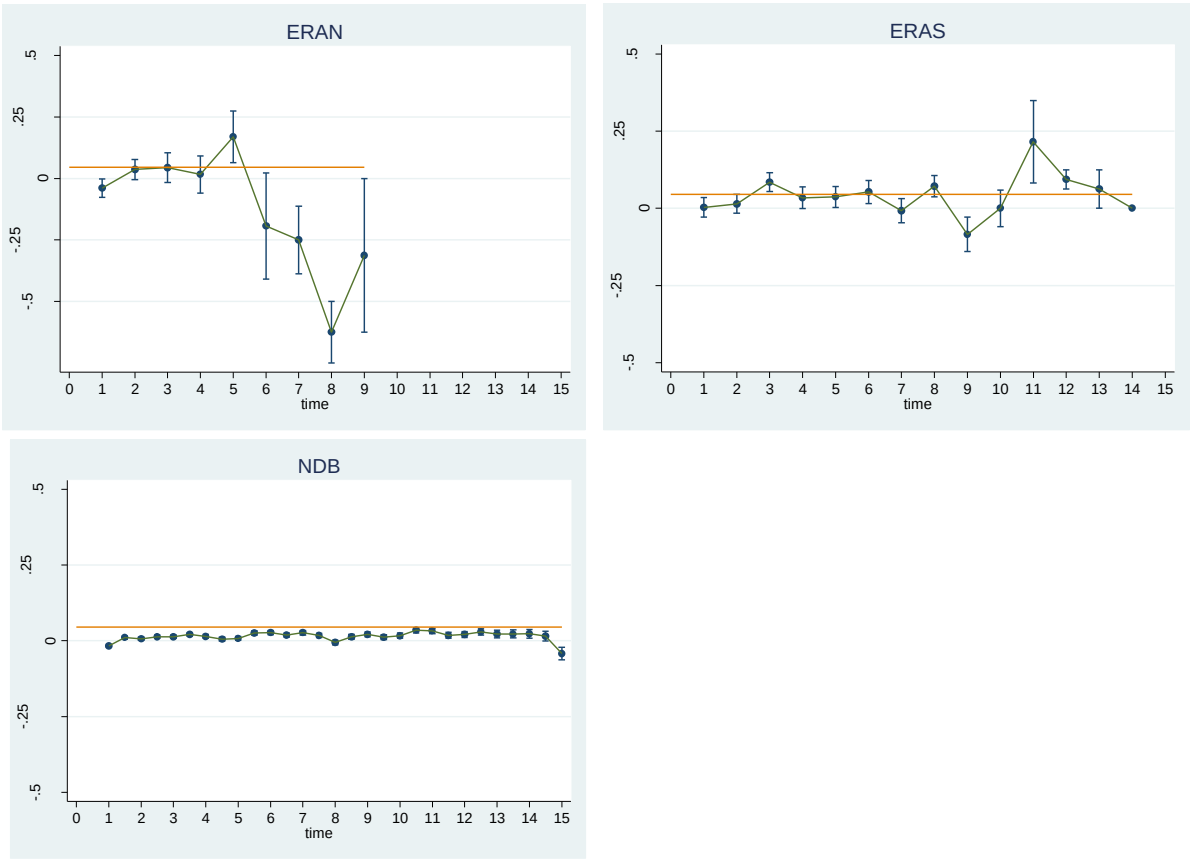


Table 24 shows the mean change in HAQ observed in each of these data sets. The first year of data is not included to remove the potential initial treatment response from the third DMARD. The table gives both the mean HAQ change to the end of the observation period and to eight years after taking the third DMARD. This allows comparisons with the previous rates of change.

Table 24: Mean Changes in HAQ after receiving third DMARD

Mean change	ERAN	ERAS	NDB
$t = 2$ to end	0.0313	0.0412	0.0175
$t = 2$ to $t = 8$	0.0334	0.0453	0.0157

Again, ERAS gives the highest rate of annual HAQ change and the NDB gives the lowest rate of change. The ERAN data gives a much higher rate of change after a patient receives

their third DMARD that when including patients on any number of DMARDs. Conversely, the ERAS and NDB data show that failing two DMARDs lowered the rate of HAQ change.

As before, we used the LCGA to analyse the new samples of data. However, due to the smaller number of patients in these new samples, there was insufficient data to use the cubic LCGA; therefore we only used the quadratic models and kept variances of the random intercepts, slopes and quadratic terms fixed at zero. As discussed previously, the model allows patient characteristics to influence class membership but not to influence the subsequent HAQ trajectories. Similarly, to the previous samples, four latent classes provide a better fit that three latent classes and there is insufficient data to run the analysis with five latent classes, as illustrated in Table 25.

Table 25: Cubic and Quadratic Information Criteria for ERAS data

Specification	**Number of Classes**	
ERAS	3 Classes	4 Classes
AIC	3435.948	3263.098
BIC	3564.189	3422.355
ERAN	3 Classes	4 Classes
AIC	1397.282	1354.366
BIC	1490.551	1472.745
NDB	3 Classes	4 Classes
AIC		101610.863
BIC		102035.703

Figure 20 shows the quadratic LCGA for the ERAN data. The patient characteristics which were allowed to influence class membership here are age, DAS and disease duration. Gender was found to almost perfectly predict some class memberships (almost all patients in the class with the highest HAQ were female) and so gender was not included in the analysis.

Figure 20: Quadratic LCGA from Third DMARD in ERAN data

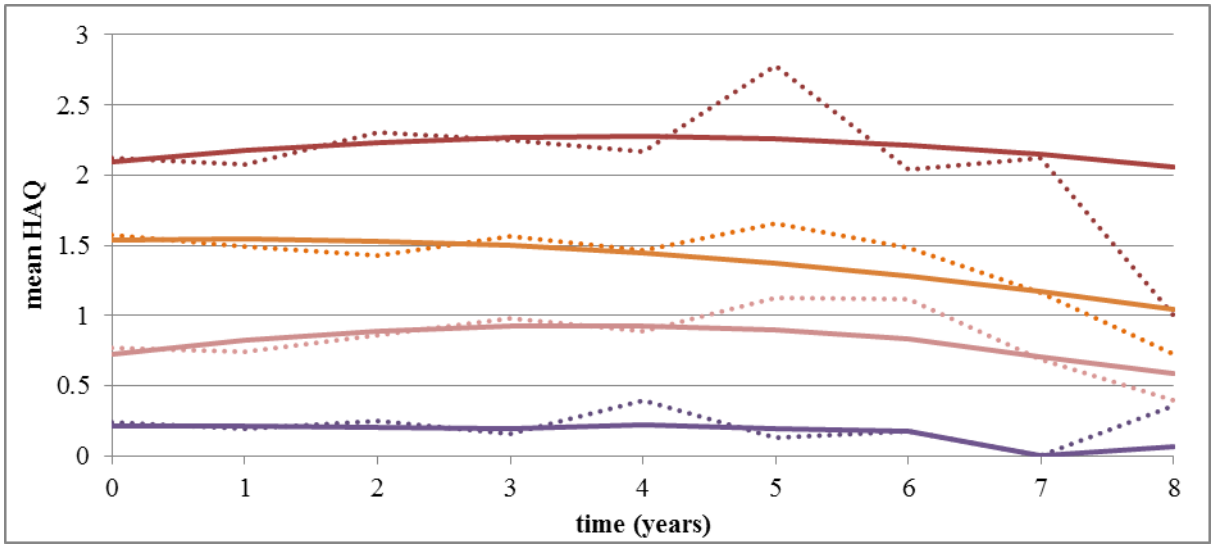


Table 26: Annual Rate of HAQ change – Quadratic LCGA ERAN

Rate of Change	BSRBR	High DAS	Moderate DAS
After $t = 2$	-0.0250	-0.0313	-0.0352
$t = 2$ to $t = 8$	-0.0250	-0.0313	-0.0352

Figure 21 shows the same analysis performed on the ERAS data, this time including gender as a covariate but there was insufficient data to predict the HAQ trajectories after thirteen years, $t = 13$. The model predicts reasonably well until nine years after baseline when the observed and predicted values start to differ more.

Figure 21: Quadratic LCGA from Third DMARD in ERAS data

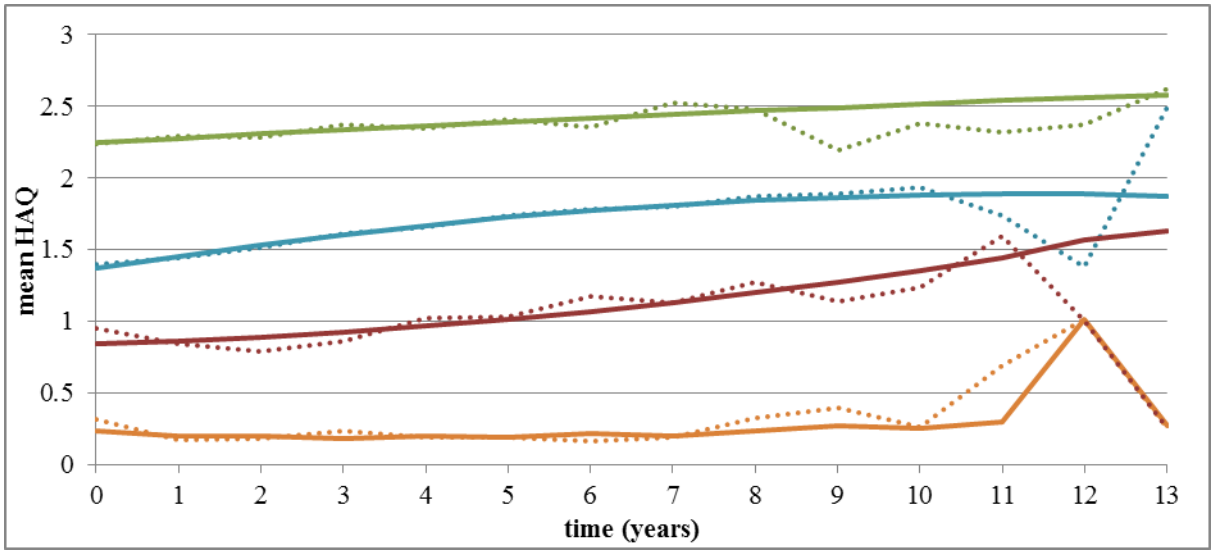


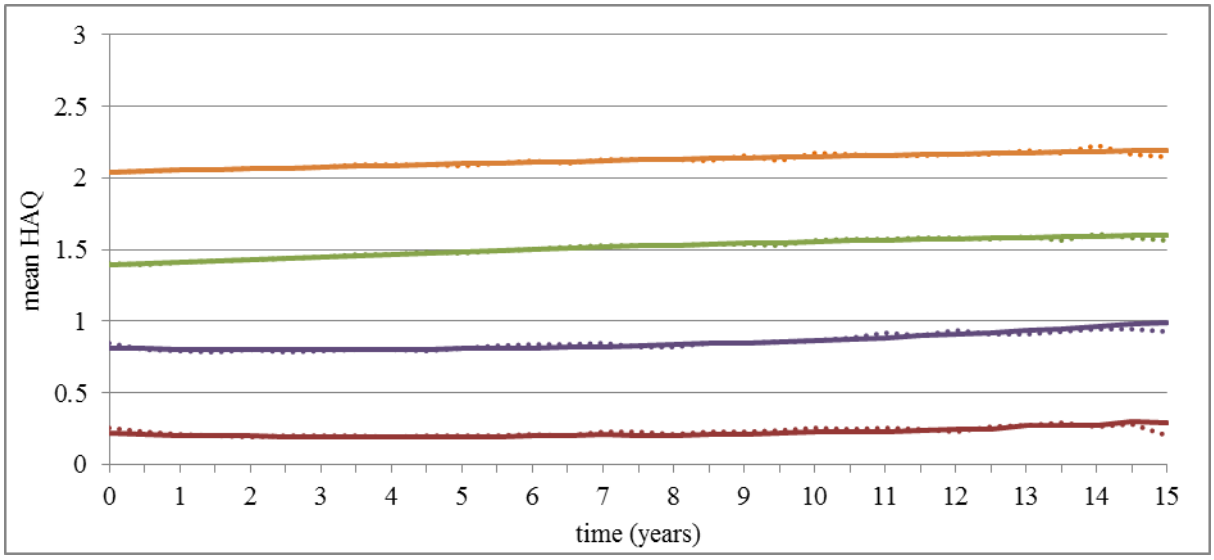
Table 27: Annual Rate of HAQ change – Quadratic LCGA ERAS

Rate of Change	BSRBR	High DAS	Moderate DAS
After $t = 2$	0.0401	0.0406	0.0362
$t = 2$ to $t = 8$	0.0494	0.0489	0.0389

In this quadratic analysis, there is again a relatively high rate or worsening in the early years of the modelling exercise, particularly in the more severe population subgroups. This rate of increase decreases after that point.

Figure 22 shows the quadratic LCGA for the NDB dataset. The analysis predicts the observed values very well, although the predicted and observed values do start to differ slightly in later observations. The patient characteristics used here to predict class membership are age, gender, disease duration and RADAI score. Again, variances are fixed at zero. The NDB shows a slow but steady increase in HAQ score across all latent classes.

Figure 22: LCGA from Third DMARD in NDB data



There were not enough observations in the ERAN and ERAS data to predict the HAQ trajectories after failing two DMARDs using the ALT model. For example, in the ERAN data we could only predict up to five years from the baseline with the available data; any longer and the model would not converge. Convergence problems relating to the computational burden of the very large dataset in the NDB prohibited estimation of the ALT model.

4.5. ACCOUNTING FOR MISSING DATA AND ATTRITION

So far, the models we have described assume that all missing data are MAR. In this section we discuss different approaches of accounting for attrition. We assume that observations which are missing in a finite number of periods but later return to the sample remain MAR. However, we assume that patients who leave the sample entirely, due to death, moving to biologics, remission or any other reason, are not missing at random (NMAR).

We focus our attention on the ERAS datat. This is because it has a larger number of observations than the LEIDEN and ERAN data and has regular observation periods unlike the BARFOT data, making analysis more simple. The ERAS data was also collected before the use of biologic drugs was common, meaning that less of the patients are removed from the sample because they were on biologics. The NDB, has regular observations, however, the characteristics of the patients were very different to the other datasets. For all analysis which assumes that attrition is NMAR we use the original ERAS data from recruitment to the study, on or soon after diagnosis.

Each of the methods used here are extensions of selection models or pattern-mixture models which are used with the LCGA model discussed previously to account for attrition bias.

We applied the Diggle-Kenwood and Wu and Carroll methods using the ALT model described in section 4.3 above. We were unable to apply these methods to the latent class analyses because of the relatively limited size of the ERAS dataset. We found that the estimates obtained once accounting for attrition in the data were not noticeably different to those described in section 4.3.1.

4.5.1. Diggle-Kenward Latent Class Mixture Model

The Diggle-Kenward model is a selection model which we use to extend the LCGA (see Muthen *et al*. 2011[23]) to account for attrition. Observations which are missing but where the patient returns to the study at a later time are considered to be missing at random (MAR) but patients who are missing due to attrition are considered missing not at random (MNAR). Here, dropout due to attrition is influenced by latent class and by the HAQ outcomes from that period and the previous period. That is,

$$log\left[\frac{P\left(d_{it}=1|_{y_{it},y_{it-1}}\right)}{P\left(d_{it}=0|_{y_{it},y_{it-1}}\right)}\right]|_{c_i=k}=\beta_{0tk}+\beta_{1tk}y_{it}+\beta_{2tk}y_{it-1}$$

where d_{it} are binary survival indicators with the value 0 before dropout, 1 in the period of dropout and missing thereafter. The logistic regression slopes β, are allowed to vary between latent classes in the case of the LCGA.

Once the Diggle-Kenward extension was added to the LCGA, there was insufficient data in the ERAS to produce a sensible result. This was true for both the quadratic and the cubic specifications.

The Diggle-Kenward extension was also used to account for attrition in the ALT model, again using the ERAS data. Similar to the previous extension,

$$log\left[\frac{P\left(d_{it}=1|_{y_{it},y_{it-1}}\right)}{P\left(d_{it}=0|_{y_{it},y_{it-1}}\right)}\right]=\beta_{0tk}+\beta_{1tk}y_{it}+\beta_{2tk}y_{it-1}$$

where the dropout again depends on the current and previous HAQ score.

4.5.2. Wu and Carroll Method

Unlike the Diggle-Kenward Latent Class Mixture Model in which dropout is directly related to the HAQ score, the Wu and Carroll method uses the random intercept and random slope to predict dropout in each period; the individual growth trajectories influence the probability of attrition.

4.5.3. Roy Method

Another method we used to account for attrition was the Roy latent dropout mixture modelling method. Here, dropout influences latent class membership, rather than the other way around. This means that dropout can influence HAQ outcomes through their effect on class membership.

This method is similar to a conventional pattern-mixture model but rather than allowing dropout patterns to directly influence the random intercept, slope etc., it instead allows them to indirectly influence them through their effect on class membership. Formally, we have

$$P(c_i = k|d_{1i}, \ldots, d_{Ti}) = \frac{e^{\gamma_{0k}+\sum_{t=1}^{T}\gamma_{tk}d_{it}}}{\sum_{s=1}^{K} e^{\gamma_{0s}+\sum_{t=1}^{T}\gamma_{ts}d_{it}}}.$$

Initially, the variances of the random intercept, slope, quadratic and cubic terms are fixed at zero.

Figure 23 shows the quadratic LCGA of the ERAS data using the Roy adjustment. The results are very similar to the LCGA without accounting for attrition. In the latent class with the highest HAQ, the model predicts that HAQ starts to decrease towards the end of the observation period.

Figure 23: Quadratic LCGA using ERAS Data and Roy Latent Dropout Adjustment

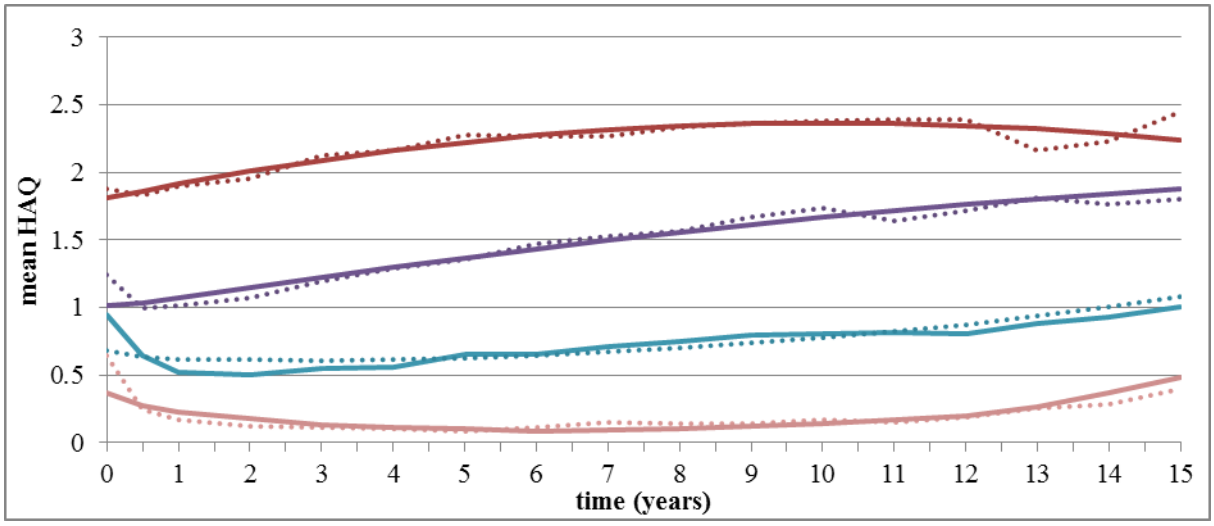
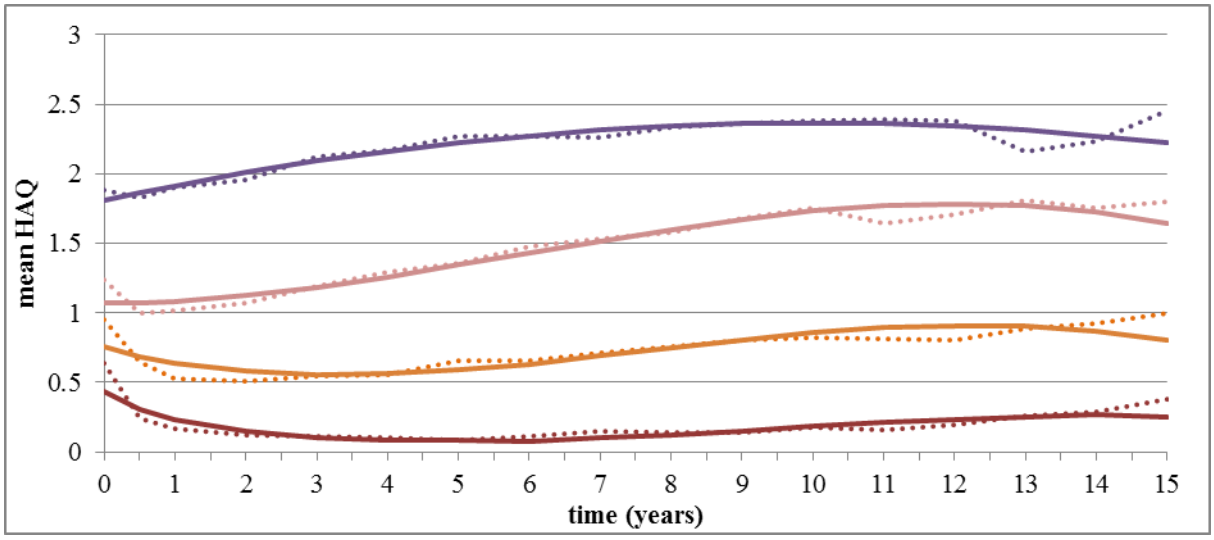


Figure 24 shows a similar pattern using a cubic specification.

Figure 24: Cubic LCGA using ERAS Data and Roy Latent Dropout Adjustment



4.5.4. Roy-Muthen Method

One potential criticism of the previous method is that the dropout classes might be contaminated if there are inherent classes in the sample for other reasons. Since we know this may be the case, we considered an additional extension of pattern mixture modelling to account for NMAR attrition: Roy-Muthen modelling with latent classes. This method uses latent dropout classes as well as latent trajectory classes. These are

$$P(cd_i = k|d_{1i}, \dots, d_{Ti}) = \frac{e^{\gamma_{ok} + \sum_{t=1}^{T} \gamma_{tk} d_{it}}}{\sum_{s=1}^{K} e^{\gamma_{os} + \sum_{t=1}^{T} \gamma_{ts} d_{it}}}$$

from the Roy method and

$$P(cy_i = l|d_{1i}, \dots, d_{Ti}) = \frac{e^{\gamma_{yok}}}{\sum_{s=1}^{K} e^{\gamma_{yos}}}$$

defining the trajectory type. Combining these, we have:

$$P(cd_i = k, cy_i = l|d_{1i}, \dots, d_{Ti}) = \frac{\exp(\gamma_{odk} + \gamma_{oyl} + \gamma_{oydkl} + \sum_{t=1}^{T} \gamma_{tk} d_{it})}{\sum_{k,t} \exp(\gamma_{odk} + \gamma_{oyl} + \gamma_{oydkl} + \sum_{t=1}^{T} \gamma_{tk} d_{it})};$$

see Muthen *et al*.[23] for full details of this method.

Figure 25 shows the results from this model. There are four separate figures presented which align with those identified from the ERAS data as reported in section 4.2.1. The extension to the model identifies, within each latent trajectory class, three sub-classes based on the latent

dropouts. These classes show the predicted course of HAQ over time had the patients not dropped out. Table 28 shows that there is a strong relationship between membership of these three classes and the time of dropout. Those that drop out early are likely to be in class 1. For those with the lowest starting HAQ these early dropouts are predicted to move to a very low HAQ. This group seems compatible with those that dropout due to disease remission. Class 1 in the two high HAQ latent classes have the highest rate of worsening in HAQ, which then flattens over time. These fast worsening groups may be those that withdraw due to death or simply because of disease severity. Dropouts between years 4 and 8 are more split between dropout classes 1 and 3. Later dropouts are more mixed but with increasing probability of class 2.

Figure 25: Latent Dropout Classes within latent trajectories

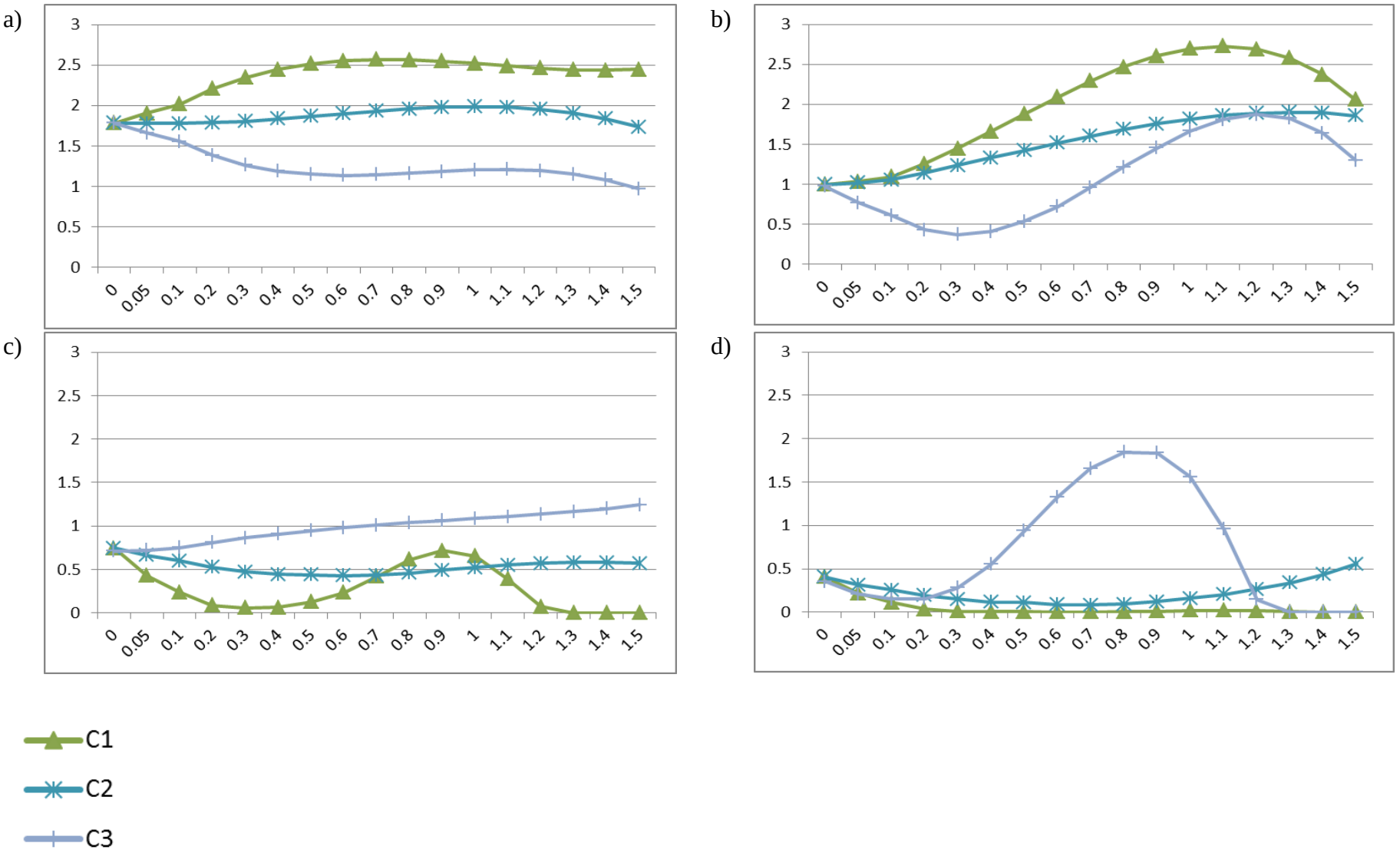


Table 28: Probability of Latent dropout class by time of dropout.

		Probability		
		Class 1	***Class 2***	***Class 3***
drop out at yr	0.5	1.0000	0.0000	0.0000
	1	0.7365	0.0000	0.2635
	2	0.6318	0.0000	0.3682
	3	0.5534	0.0000	0.4466
	4	0.4533	0.1427	0.4039
	5	0.3091	0.1919	0.4990
	6	0.4920	0.1552	0.3527
	7	0.4038	0.1725	0.4238
	8	0.3506	0.3589	0.2905
	9	0.2387	0.3839	0.3774
	10	0.1972	0.4759	0.3269
	11	0.1993	0.5326	0.2681
	12	0.2048	0.4461	0.3491
	13	0.1708	0.5577	0.2716
	14	0.1480	0.5046	0.3474
	15	0.1539	0.4743	0.3718

Table 29: Within sample probabilities of class membership by latent class model type

	Latent Class Model type		
	NMAR	**MAR**	**Norton**
Class 1*	0.248	0.174	0.155
2	0.263	0.296	0.291
3	0.292	0.307	0.336
4 (bottom)	0.197	0.223	0.217

* Classes 1, 2, 3 and 4 correspond to a), b), c) and d) of Figure 25.

Table 29 compares the within sample probabilities of the Roy-Muthen method of adjusting for dropouts with those that assume MAR in this report and in the Norton analysis. It shows that the MAR analysis in this report is very similar to that reported by Norton *et al* with differences explained by slight changes to the ERAS samples used. Accounting for missingness not at random allocates more observations to the highest class.

Table 30 calculates these same probabilities for the NMAR model for each of the three patient subgroups referred to throughout the report. For the more severe patient subgroups the probability of being in the highest class is greater, with only very low probabilities of being allocated to the lowest two classes.

Table 30: Probability of class membership for decision problem populations

	Population		
	BSRBR	HIGH	MOD
Class 1*	0.847	0.702	0.418
2	0.141	0.259	0.399
3	0.009	0.020	0.067
4 (bottom)	0.003	0.019	0.116

* Classes 1, 2, 3 and 4 correspond to a), b), c) and d) of Figure 25.

The model again provides support for the concept of four latent classes and for their general shape, that is, a period of worsening of HAQ which slows over time. The model uses a sophisticated method for dealing with potential bias arising from dropout. It identifies subgroups within the modelling according to the pattern of dropout. In this situation, we find evidence that the projected pattern of HAQ progression can be explained by some degree by the timing of dropout.

We suggest that this modelling approach provides important information for use in sensitivity analysis for the cost effectiveness analysis. The highest latent dropout subgroups within each trajectory can be taken as a credible upper bound of the HAQ trajectory for any subgroup of patients, that is, class 1 within the latent classes a) and b) displayed in Figure 25, and class 3 in latent classes c) and d). In latent class b) the improvement in the estimate is very closely related to the actual observed data and since that is limited at the observations beyond 11 years (where the curve begins to fall) we make the additional assumption that HAQ is flat from that point onwards in implementation in the CE model.

We also assume in implementation of these results in the CE model, that HAQ does not continue to rise beyond 15 years for the lowest two latent classes. This makes no difference to the severe populations because the probability of being in these low HAQ classes is very

small. For the moderate DAS group this assumption may be more relevant because there is a 0.18 probability of being in either class.

Note that this sensitivity analysis does not identify a priori who these patients are.

5. DISCUSSION

There is no study that reports how HAQ progresses in RA patients on non-biologic therapies that coincides entirely with the requirements of the cost effectiveness analysis. The CE model requires estimates of HAQ progression over a patient's lifetime from the point at which they would be eligible for biologic therapies (which is currently having failed two DMARDs and having a high DAS but also under consideration are those that have failed two DMARDs and have a moderate HAQ score, or those with a high DAS score and have not yet failed any DMARDs). The model also requires such estimates of HAQ progression in patient once they have exhausted a sequence of biologic therapies. It is unsurprising that such evidence does not exist.

The purpose of this report, together with the estimates that have been used in previous NICE appraisals and by the AG in TA537, is to identify related evidence that helps in the required estimates of how HAQ progresses in these different circumstances.

In many previous cost effectiveness analyses in this area, analysts have chosen to assume a constant annual rate of HAQ progression. Whilst the rationale provided for this in earlier NICE appraisals is not based on any coherent assessment of the evidence, others have made such estimates on the basis of empirical evidence and these are sometimes very similar to those used in previous NICE appraisals. Further consideration of these methods highlights significant limitations. Pooling average annual rates of HAQ progression drawn from studies that have different lengths of follow-up, different times and frequencies of follow-up and patient characteristics will be entirely inappropriate unless the true rate of HAQ progression is constant.

Evidence from nine studies of patients with established disease and follow up of more than eight years was reviewed. In those studies that permit an assessment of the shape of HAQ

trajectory, there was evidence that HAQ does not progress at a linear rate. Most studies suggest rapid worsening initially followed by a period of slower worsening, although one study with particularly long follow up suggests regained rapid worsening after 20 years.
We identified five studies of RA patients from different countries that had long term follow up of patients including their HAQ scores. We obtained patient level data for each of these studies and analysed them using different methods.

When considering the raw data alone, two of those datasets exhibit a trend of rising HAQ over time (ERAS and BARFOT). ERAS is the only dataset that shows a rate of worsening in HAQ that is higher than the 0.045 simple rate used in previous NICE appraisals (0.054 from years 2-15). The BARFOT data is substantially lower. However, all studies are of course affected by dropout and include a much broader sample of patients than those that are candidates for biologic therapies in the NICE TA. Reliance on the raw data is therefore not advisable.

Dropout is inevitably substantial in observational studies with long term follow up. Models predictions based on fewer and fewer observations at greater time points become more uncertain.

We replicated the Norton *et al* analysis. We confirm the preferred model comprises four latent classes and a cubic specification in the ERAS data. In this model, the rate of worsening is faster for all the subgroups of interest, during the early part of diseases (years two to eight) but this rate slows over time. There is a suggestion that HAQ continues to rise if a quadratic specification is selected, both in analyses of ERAS and other datasets. However, the cubic specification is consistently preferred based on likelihood ratio tests and AIC/BIC.

The finding that HAQ rapidly deteriorates in the relevant patient groups but that this worsening slows over time is further supported by analysis of the ERAS data using an alternative modelling approach (the ALT model). Indeed, this is a consistent feature of the findings throughout this report.

We applied four different methods for accounting for attrition, assuming NMAR, and again found that results continued to support the general findings of the original latent class analysis. Using the Roy-Muthen method for dealing with data NMAR, we identify three

dropout subclasses within each of four latent trajectory classes. These provide credible estimates of the course of HAQ in the absence of dropout. We propose these serve as an appropriate upper bound for considerations of the plausible course of HAQ over time.
Our preferred analyses, described above, make use of all available data and adjusts for covariates that distinguish the patient subgroups of relevance for the cost effectiveness analyses from the broader RA populations recruited into these studies. An alternative approach we explored is to limit the analysis of data only to those patients that meet, or more closely meet, the criteria for receipt of biologic therapies. We found that there were insufficient data for analyses where samples were restricted to those that had failed two DMARDs and also had a DAS>5.1. We did conduct subgroup analyses on those that had failed two DMARDs.

These analyses of course result in much reduced sample sizes, particularly towards the end of the follow up period but it does allow more meaningful comparisons to be made including the NDB study because these analyses now treat all patients as having a common baseline (the time they are observed to have started a third DMARD). The NDB and ERAN data both suggest there is a relatively slow worsening of HAQ over time and this reduces over time. The ERAS data also supports this view in general though the latent class analysis does differ from the analyses conducted using the full dataset in suggesting the rate of HAQ continues to rise, albeit at a slower rate, particularly for those in the highest latent class. Overall these rates still suggest a lower overall rate of worsening than 0.045 per annum though the predictions for the severe disease subgroup are very close to this.

Some important caveats must be considered in relation to these subpopulation analyses. The sample size is of greater concern and the relevance of the very large error bars for the ERAS and ERAN studies should be acknowledged. The additional uncertainty in ERAS is added to by the fact that the mean time to start the third DMARD is longer than in ERAN (65 vs 28 months), leading to a reduction in follow up time for relevant patients. These reduced sample sizes mean that the models here are based on quadratic modelling: the cubic specification could not be run. Given we have established the cubic as superior in the full dataset this may be an important consideration for the extrapolation element of any analysis. Similarly, the NMAR analyses could not be run. It should also be noted that there remains some discrepancy between the timing of the revised baseline in these sub analyses and the time relating to potential biologic use. These datasets provide us with the ability to identify the

first time a patient is observed to have started their third DMARD. But the time of failure of the second DMARD could be a substantial time before that given these studies have six monthly or yearly data collection. Disease may be substantially worse at the time a therapy is deemed to have "failed" compared to a period some time after a new therapy has been initiated.

6. REFERENCES

1. Chen Y.F., Jobanputra P., Barton P., Jowett S.M., Bryan S., Clark W. et al. A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. *Health Technology Assessment* 2006; 10(42):1-266.

2. Norton S., Fu B., Scott D.L., Deighton C., Symmons D.P., Wailoo A.J. et al. Health Assessment Questionnaire disability progression in early rheumatoid arthritis: Systematic review and analysis of two inception cohorts. *Seminars in Arthritis and Rheumatism* 2014; 44(2):131-144.

3. Norton S., Fu B., Verstappen S.M.M., Symmons D.P.M., Lunt M., Davies R. et al. Common trajectories of HAQ disability progression over 15-years in the Early Rheumatoid Arthritis Study and the Norfolk Arthritis Register. *ANNALS OF THE RHEUMATIC DISEASES* 2012; 71:508.

4. Tosh J, Wailoo AJ, Scott D, Deighton C. Cost-effectiveness of combination non-biologic disease modifying anti rheumatic drug strategies in patients with early RA. *Journal of Rheumatology* 2011; 38:1593-1600.

5. Wailoo A.J., Bansback N., Brennan A., Michaud K., Nixon R.M., Wolfe F. Biologic drugs for rheumatoid arthritis in the medicare program: A cost–effectiveness analysis. *Arthritis & Rheumatism* 2008; 58(4):939-946.

6. Scott D.L., Garrood T. Quality of life measures: use and abuse. *Best Practice & Research Clinical Rheumatology* 2000; 14(4):663-687.

7. Sherrer Y.S., Bloch D.A., Mitchell D.M., Young D.Y., Fries J.F. The development of disability in rheumatoid arthritis. *Arthritis & Rheumatism* 1986; 29(4):494-500.

8. Hawley D.J., Wolfe F. Sensitivity to change of the Health Assessment Questionnaire (HAQ) and other clinical and health status measures in rheumatoid arthritis results of short–term clinical trials and observational studies versus long–term observational studies. *Arthritis & Rheumatism* 1992; 5(3):130-136.

9. Leigh J.P., Fries J.F., Parikh N. Severity of disability and duration of disease in rheumatoid arthritis. *The Journal of Rheumatology* 1992; 19(12):1906-1911.

10. Ward M.M., Paul Leigh J. Marital status and the progression of functional disability in patients with rheumatoid arthritis. *Arthritis & Rheumatism* 1993; 36(5):581-588.

11. Leigh J.P., Fries J.F. Tobit, fixed effects, and cohort analyses of the relationship between severity and duration of rheumatoid arthritis. *Social Science & Medicine* 1993; 36(11):1495-1502.

12. Lassere M., Wells G., Tugwell P., Edmonds J. Percentile curve reference charts of physical function: rheumatoid arthritis population. *The Journal of Rheumatology* 1995; 22(7):1241-1246.

13. Ward M.M., Lubeck D., Leigh J.P. Longterm health outcomes of patients with rheumatoid arthritis treated in managed care and fee-for-service practice settings. *The Journal of Rheumatology* 1998; 25(4):641-649.

14. Krishnan E., Tugwell P., Fries J.F. Percentile benchmarks in patients with rheumatoid arthritis: Health Assessment Questionnaire as a quality indicator (QI). *Arthritis Res Ther* 2004; 6(6):R505-R513.

15. Odegard S., Finset A., Mowinckel P., Kvien T.K., Uhlig T. Pain and psychological health status over a 10-year period in patients with recent onset rheumatoid arthritis. *ANNALS OF THE RHEUMATIC DISEASES* 2007; 66(9):1195-1201.

16. de Rooy D.P., van der Linden M.P., Knevel R., Huizinga T.W., van der Helm-van A.H. Predicting arthritis outcomes – what can be learned from the Leiden Early Arthritis Clinic? *Rheumatology* 2011; 50(1):93-100.

17. Svensson B., Schaufelberger C., Teleman A., Theander J. Remission and response to early treatment of RA assessed by the Disease Activity Score. *Rheumatology* 2000; 39(9):1031-1036.

18. Svensson B., Andersson M.L., Bala S.V., Forslind K., Hafstrom I. Long-term sustained remission in a cohort study of patients with rheumatoid arthritis: choice of remission criteria. *BMJ Open* 2013; 3(9):e003554.

19. Wolfe F., Michaud K. A brief introduction to the National Data Bank for Rheumatic Diseases. *Clinical and Experimental Rheumatology* 2005; 23(5):S168.

20. Kiely P., Walsh D., Williams R., Young A. Outcome in rheumatoid arthritis patients with continued conventional therapy for moderate disease activity – the early RA network (ERAN). *Rheumatology* 2011; 50(5):926-931.

21. Garwood W. The Early Rheumatoid Arthritis Network (ERAN). *Musculoskeletal Care* 2004; 2(4):240-244.

22. Bollen K.A., Curran P.J. Autoregressive latent trajectory (ALT) models a synthesis of two traditions. *Sociological Methods & Research* 2004; 32(3):336-383.

23. Muthen B., Asparouhov T., Hunter A.M., Leuchter A.F. Growth modeling with nonignorable dropout: alternative analyses of the STAR* D antidepressant trial. *Psychological Methods* 2011; 16(1):17.