Biologic therapies and the need for hip and knee replacement amongst patients with rheumatoid arthritis: A population-based epidemiology study using electronic medical records and registry data

Samuel Hawley

Lady Margaret Hall

Michaelmas Term 2018

Doctor of Philosophy – Musculoskeletal Science

ABSTRACT

Biologic therapies and the need for hip and knee replacement amongst patients with rheumatoid arthritis: A population-based epidemiology study using electronic medical records and registry data

Samuel Hawley, Lady Margaret Hall, Michaelmas 2018

Registry reports indicate over 200,000 hip or knee replacements are performed annually in England, Wales and Northern Ireland, with approximately 1-2% carried out for inflammatory conditions such as rheumatoid arthritis (RA). The aim of this DPhil was to estimate the impact of biologic therapies on the need for major joint replacement amongst RA patients using observational health data.

An interrupted time-series analysis was used to estimate the impact of NICE approval of biologics on population-level temporal trends of total hip replacement (THR) and total knee replacement (TKR) amongst RA patients in England and Wales. Similar analyses were repeated for Denmark and Ontario. Overall, these studies indicated a decrease in TKR but not THR for RA patients following the introduction of biologics. When the rates in non-RA patients were taken into account (in Denmark and Ontario only), there was an inferred reduction in both THR and TKR for RA patients within the biologic era. There was a lack of guidance on sample size planning for such analyses, so a simulation study was also conducted to estimate power in various time-series scenarios.

A patient-level analysis was then conducted using UK registry data, applying various novel methodologies to account for the inherent problem of confounding by indication. The results suggested no significant impact of biologics on rates of joint replacement, although in age-stratified analyses biologics was associated with a 40% reduction in THR rates amongst patients $\geq$60 years old.

To conclude, a reduction is observed in population-level rates of THR and TKR amongst RA patients (compared to non-RA patients) following the introduction of biologics. Patient-level analyses confirmed a favourable impact on THR rates amongst older patients, but otherwise no significant associations. More patient-level analyses are required to confirm and/or further elucidate the impact of biologic therapies on the need for joint replacement in RA.

ACKNOWLEDGMENTS

The central clinical questions this thesis set out to answer were originally proposed by Prof Daniel Prieto-Alhambra as part of an NIHR funded Clinician Scientist award (CS-2013-13-012). I am therefore greatly indebted to Prof Prieto-Alhambra for sharing some of these ideas with me and providing excellent supervision throughout the DPhil.

I was very fortunate to have the additional co-supervision of Prof Andrew Judge and Dr M. Sanni Ali. I've learned a great deal in terms of conducting and communicating epidemiological research in general from Prof Judge, and in pharmaco-epidemiological methods in particular from Dr Ali. I would sincerely thank Prof Judge and Prof Prieto-Alhambra for employing me as a research assistant, and to everyone in the wider G.24 crowd whose sense of fun and passion for science got me into epidemiology in the first place.

Throughout the DPhil I've been grateful for the salient advice and suggestions received following my presentations at numerous advisory committee meetings, notably from Profs Cyrus Cooper, Christopher Edwards and Nigel Arden. It was primarily for these inputs, in addition to their contribution to the aforementioned Clinician Scientist proposal and/or acquisition of data that these individuals appear as co-authors on several publications arising from this work (as listed in the preface). I would particularly thank Prof Arden for nominating me as Marjorie Ridley Scholar during these last three years at Lady Margaret Hall. Likewise, I've been appreciative for the helpful comments at my transfer and confirmation of status assessments carried out by Profs Dominic Furniss and Gary Collins.

I've been privileged to work closely with excellent colleagues in Denmark during the validation component of this thesis. It's been a delight to collaborate with Dr René Cordtz and Prof Lene Dreyer, who whilst leading on their own projects, kindly welcomed significant input from myself in terms of planning, conducting and reporting of interrupted time series analysis. I've greatly valued the opportunity to learn simulation methodology under the supervision of Dr Ali and have enjoyed working with fellow DPhil student Klara Berencsi on the coding component. Klara assisted in the development of my functional albeit clumsey Stata simulation programme into an efficient and elegant product worthy of sharing. Dr Antonella Delmestri was responsible for carrying out the initial data management and cleaning of the CPRD data used.

I'm hugely grateful for my family, especially my parents, who provided amazing support during the time of writing up and who were key in me starting and finishing this undertaking, as they have been in many other undertakings. I'm very thankful too for encouraging and understanding friends. I hope I will be able to emulate these many kindnesses shown to me.

Finally, to quote Florence Nightingale (1820-1910), in order to acknowledge something of the underlying motivation for me throughout this DPhil and my academic endeavours in general, "*To understand God's thoughts we must study statistics".*

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

Preface

Several of the chapters included in this thesis are associated with published papers, submitted papers accepted for publication or some elements that have involved receiving assistance from others. Where this is the case, this is stated in the acknowledgment section of the thesis in addition to the relevant thesis chapters. All co-authors (including my supervisors) give their permission for me to include this work in my thesis. Publications arising from this work are:

(1) Hawley S, Cordtz R, Lene D, Edwards CJ, Arden NK, Delmestri A, Silman A, Cooper C, Judge A, Prieto-Alhambra D. Association between NICE guidance on biologic therapies with rates of hip and knee replacement among rheumatoid arthritis patients in England and Wales: An interrupted time-series analysis. Semin Arthritis Rheu. 2018;47(5):605-10.

(2) Hawley S, Ali MS, Cordtz R, Lene D, Edwards CJ, Arden NK, Cooper C, Judge A, Hyrich K, Prieto-Alhambra D. Impact of TNF inhibitor therapy on joint replacement rates in rheumatoid arthritis: a matched cohort analysis of UK registry data. Rheumatology (Oxford). *Accepted for publication*

(3) Hawley S, Ali MS, Berencsi K, Judge A, Prieto-Alhambra D. Sample size and power considerations for ordinary least squares interrupted time-series analysis: a simulation study. Clinical Epidemiology. *Accepted for publication*

The publisher of paper (1) (above), Elsevier, states that within the publishing agreement authors retain the right to include material for inclusion within a thesis or dissertation, and therefore further permission was not required.

THESIS OVERVIEW

The thesis begins (chapter 1) with a review of the literature on the management of rheumatoid arthritis (RA) and the existing data on the impact of pharmacotherapy, particularly biologic therapy, on a number of clinical outcomes. It describes the growing body of observational data that suggest the number and/or rate of joint replacement amongst RA patients is in decline across the developed world. Existing data on the estimated association between biologics and joint replacement in RA is described at the population-level and patient-level. Possible study designs for investigating this research question are explored and the challenges of using routinely collected health data for comparative effectiveness research is discussed. The methodological advancements within the field of pharmacoepidemiology are introduced, particularly in the form of quasi-experiments. The chapter ends with stating the key aims and objectives of the DPhil.

Chapter 2 describes the data sources used throughout the main experimental chapters of the thesis (chapters 3-7). These consist of routinely collected health data from the UK, Denmark and Ontario (Canada). The RA specific drug register data used from the UK is also described in detail. An overview is given of the role of collaborators in extracting and managing these data.

The first experimental chapter (chapter 3) introduces the concept that although the natural history of RA is well known, less is known of the descriptive epidemiology of the comparatively long-term outcome of joint replacement. The chapter proceeds to report and discuss analyses carried out in order to better elucidate the descriptive epidemiology of total hip replacement (THR) and total knee replacement (TKR) within UK RA patients. The validity of these outcomes within the primary care setting (against hospital procedure codes) is presented, and key epidemiological parameters are subsequently estimated. These measures include 10-year and 20-year cumulative incidence, median time-to-outcome and incidence rates. Stratified incidence rates are also reported across a number of key demographic characteristics: age, sex, body mass index (BMI), smoking status, geographic region and deprivation score.

Chapter 4 moves from descriptive epidemiology to the explicit estimation of the population-level impact of the introduction of biologics (in the form of NICE approval in 2002) on subsequent rates of THR and TKR amongst UK RA patients. This chapter introduces and utilises interrupted time-series (ITS) methodology in order to adjust for the level and trend of outcome prior to the introduction of biologics. ITS is the first of several quasi-experimental methods used throughout the DPhil and the strengths of the method are described.

Chapter 5 builds upon the analyses within chapter 4 by using "big data" external to the UK in order to provide some form of validation of results. It reports on the application of ITS to data from Denmark and Ontario and how this methodology was extended through

the analysis of control populations not exposed to biologics, thus permitting the use of a difference-in-differences approach. The results from an age-stratified ITS analysis using the Canadian data are reported, as are those from an ITS analysis of prescriptions data of both conventional and biologic therapies.

Chapter 6 is something of a methodological tangent and focusses on the issue of statistical power in ITS. The lack of existing guidance on sample size planning for ITS is discussed and a proposed solution is presented in the form of a simulation study. Specifically, results are presented from a Monte-Carlo simulation in which power was estimated as a function of a range of parameters such as number of timepoints, mean sample-size per timepoint and location of intervention within the time-series. The simulation used the analysis of TKR in chapter 4 as a case-study and a Stata programme is presented, allowing researchers to enter values for numerous parameters in order to generate their own estimates of required sample size for proposed future ITS studies.

The central clinical question of the patient-level impact of biologics (specifically, tumour necrosis factor inhibitor (TNFi) therapy) versus conventional therapy (without TNFi) on subsequent rates of THR and TKR amongst UK RA patients is then addressed in chapter 7. The British Society for Rheumatology Biologics Register (BSRBR) drug registry is leveraged, with the use of additional quasi-experimental methodologies to minimise the bias inherent in estimation of comparative effectiveness using observational data. Results from the main analysis using propensity score matching is reported, as are sensitivity analyses using firstly an instrumental variable approach and secondly a

regression discontinuity approach. These various methods are evaluated, and the potential mechanisms/explanations of findings are discussed.

Chapter 8 describes the main findings of the overall DPhil and draws together the numerous pieces of evidence. A discussion of the comparative strengths and limitations of the different approaches used throughout different chapters is included and explanations of findings are offered. Potential clinical implications of these findings are proposed as are future research priorities. Final conclusions are made concerning the question of the impact of biologics on need for joint replacement in RA patients.

1. LITERATURE REVIEW

Rheumatoid arthritis (RA)

RA is a chronic autoimmune disease that is clinically characterised by persistent joint inflammation and progressive damage to cartilage and bone (1, 2). Classification criteria to differentiate RA from other forms of joint disease first emerged from the late 1950s although these have evolved considerably (1, 3). Patients experience stiffness, swelling and pain in multiple joints (3) that are often symmetrically affected bilaterally (4). In addition to symptoms at the joint, the natural course of RA involves increased risk of developing various secondary conditions including depression, asthma, cardiovascular events, malignancies and COPD (5). Related to these is the increased risk of mortality (1). RA significantly reduces quality of life (6-9) and leads to many patients stopping work (10, 11). Prevalence of RA is highest at ~0.5% in Western Europe, North America and Australasia (12), although UK prevalence has previously been reported at ~1% (13, 14). It is globally ranked as the 42nd highest contributor to global disability, just below malaria (12).

RA aetiology / risk factors

The exact causes of RA remain unknown (15, 16) but it's generally described as being the consequence of genetic susceptibility combined with a sufficiently forceful although as yet unknown external trigger (17, 18). It's been estimated that between 50%-60% of susceptibility for developing RA can be attributed to genetic factors (6, 19), of which 50%

are associated with the HLA region of the short arm of chromosome 6 (20). Risk associated with various HLA DRB1 alleles have been reported, ranging from odds ratio (OR) = 1.0 [95% CI: 0.8 to 1.5] to OR = 3.5 [95% CI: 1.8 to 6.8] for different variants (21). RA is more prevalent in women, with several hormonal risk factors described including prolactin which is likely to explain the significant 11-fold increase in risk of developing RA after a woman's first pregnancy (6, 21). Linked to this is that men with low testosterone levels are at increased risk of RA (21) as are obese individuals (although this is less clear), a possible explanation being heightened oestrogens in the case of high BMI (21). The differential risk by sex is reduced in older age which may again be explained by hormonal changes (6). Although there is evidence that vaccination against infectious agents may in particular cases precede and likely trigger RA incidence (18), this must constitute a very small number of cases (6). Many infectious agents have also been studied as possible triggers but none of these investigations have identified strong evidence for any particular organism (22). Some dietary factors have been linked with lower rates of RA although this is not clear, in contrast to fairly consistent data that indicates smoking is a significant risk factor (21).

RA pathophysiology and joint damage

While the synovium of healthy joints normally allows minimal cellular migration, in the state of RA there is an influx of immune cells due to upregulation of pro-inflammatory cytokines (including tumour necrosis factor (TNF)), which induces a state of synovitis (15, 23). Associated with this is increased cartilage degradation and bone resorption leading to juxta-articular bone erosions (23, 24). A very high percentage of early RA patients have

such bone erosions in hand and wrist joints (25), with recent data indicating an approximate 70% prevalence of at least one erosion detectable from ultra-sound within 12 months of RA diagnosis (26). The longitudinal radiological assessment of patients' joints has over time given rise to a progressive evolution of measurements to quantify joint destruction in the form of number and/or severity of erosions and/or joint space narrowing (due to cartilage loss) (27, 28). The most usual pattern is progression and proliferation joint damage in a linear fashion over time (25, 29) at approximately 2% of maximal damage (as per the scoring system used) per year (30), with an estimated 40% of possible maximum damage occurring in smaller joints within 20 years (25). Other patterns are also common, including a significant minority without erosive progression at all (31) and others with accelerated progression in the early years (32). Clinical manifestations of joint damage include joint swelling and tenderness, while patient-based symptoms include considerable pain, disability and the increased risk of various comorbidities (1). Despite a high comorbidity prevalence rate in RA (5), it is estimated that joint destruction accounts for approximately 25% of disability in established disease (25).

RA joint damage and joint replacement

The progressive and often permanent nature of radiographic joint damage, in conjunction with associated pain, loss of function and failure to adequately respond to therapeutic options are strong indications for eventual joint replacement surgery, which can be considered as the ultimate marker of joint destruction (33, 34). While progression of structural damage in RA has been well described (25, 30), long-term clinical outcomes

such as the incidence of joint replacement remains less well studied (35). RA-related joint damage has been shown to evolve most in patient's smaller joints (25), with feet being eroded earlier than hands (31, 36). However, RA-related joint replacement procedures most commonly carried out are for the hip and knee (37, 38). This preponderance of major lower limb replacements most likely reflects surgeons' expectation of predictably good outcomes (37), although replacement of other joints such as shoulder, elbow and ankle are also prevalent (38). For example, in the US approximately 25% of elbow joint replacements performed in 2002 were RA-related (38). Conversely, RA-related hip and knee replacements only constituting approximately 3% of the total number of hip and knee replacements carried out in the US study, yet these were still by far the most common joint replacement procedures provided to RA patients considering that the overall total incidence overall was high at 31 and 65 per 100,000 people for hip and knee replacement, respectively (38). For context, the incidence of elbow replacement was only 0.3 per 100,000 in the overall population (38). Likewise, recent National Joint Registry (NJR) estimates from England, Wales and Northern Ireland indicate inflammatory arthritides account for only 1-2% of all THR and TKR procedures performed but that this is still a substantial number given >200,000 hip and knee replacements were carried out across all indications in 2016 alone (39, 40).

Management of RA

The American College of Rheumatology (ACR) began publishing guidance for the management of RA in the mid 1990s (41) and early 2000s (42), which primarily focussed on describing acceptable practice for use of pharmacologic treatment (43). Guidance of

a more prescriptive nature on the holistic and multi-faceted approach to the management of RA was provided by the British Society for Rheumatology (BSR), firstly pertaining to the care of early RA patients (44) and later to the longer-term management in more established RA (45). This BSR guidance referred to a "*window of opportunity*" (44) following onset of RA, and stressed the need for early diagnosis, streamlined transference to specialist multi-disciplinary care, intensive education and immediate initiation and subsequent monitoring of disease modifying therapy. Also recommended was use of analgesics and low dose non-steroidal anti-inflammatory steroids (NSAIDS), assessment/management of sleep disturbance, availability of hydrotherapy and encouragement of exercise. Various specialists should be available including rheumatologists, nurses, physiotherapists, occupational therapists, orthopaedic surgeons and health professionals to discuss sexuality and relationship issues (44). Whilst 24 recommendations were made in total, one of the few termed "grade A" was that patients should be initiated a disease modifying medication as soon as possible with aggressive escalation/progression of therapies where necessary. This was expanded in the subsequent guidance pertaining to patients with more established disease, in that these therapies should be considered "*medium- to long-term treatments… …, whose withdrawal usually results in flare and disease progression*" (45). Many of these recommendations were similarly enshrined into routine clinical practice in England and Wales through National Institute for Health and Care Excellence (NICE) clinical guideline 79 (46).

Therapies for RA

Treatments for RA have evolved markedly over the approximate 150 years since the disease was recognised within the medical literature (47), and particularly over the last 30 years (48). There have been at least two watershed moments in this evolution: the introduction of methotrexate in the late 1980s and the introduction of biologics in the late 1990s (48). Each of these have been described as treatment revolutions (49, 50).

Traditional therapies can be broadly categorised as those primarily targeting symptoms and those that modify the actual disease process through slowing clinical and/or radiographic progression (50). Common examples of the former are NSAIDS (e.g. ibuprofen) and analgesics which address swelling and pain, respectively. While glucocorticoid steroids can have structural benefits, they're not considered disease modifying (23) but are fast-acting suppressors of inflammation which are useful in early RA or in managing a flare of disease activity (45). Use of steroids is only recommended over the short-term due to serious side-effects including increased risk of myocardial infarction and osteoporosis, in addition to many less serious unfavourable outcomes (e.g. acne) (51, 52). In terms of conventional synthetic disease modifying anti-rheumatic drugs (csDMARDs), the three most commonly used are methotrexate, sulfasalazine and leflunomide (17, 53) which have various modes of action (54) although precise mechanisms of effect is not known (23). The disease modifying efficacy of methotrexate has been clearly demonstrated in the randomised controlled trial (RCT) setting since 1985 (55, 56) and has thereafter entered widespread usage, being described as "*the cornerstone in the treatment in RA*" (54). It is recommended by NICE as first-line therapy

for RA in the UK, in conjunction with other csDMARDs (46) and is recommended by ACR and European League Against Rheumatism (EULAR) as the first DMARD to be prescribed to DMARD naïve patients with symptomatic early/active RA (57, 58). Sulfasalazine has been used widely for RA since the 1980s (55), with RCT efficacy data following on reducing various measures of disease activity (59, 60). Likewise, Leflunomide has been shown to have efficacy on numerous clinical and structural outcomes, and to an extent that is equivalent to methotrexate (58, 61). Other options generally considered csDMARDs are antimalarials hydroxychloroquine and chloroquine, which although lacking the potency of other csDMARDs in reducing disease activity (58) nonetheless have a relatively strong safety profile and are therefore still used for milder disease (62). Combinations of csDMARDS are often used, e.g. methotrexate, sulfasalazine and hydroxychloroquine as "triple therapy" (57). The exact treatment choice will factor in the patient specific benefit-risk ratio, contraindications, drug interactions and poor response (57, 58).

The introduction of biological (b)DMARDs in the late 1990s demonstrated a shift away from empirical approaches to RA treatment (i.e. the use of therapies for reasons almost solely based on observed beneficial effects) and towards the use of drugs of known mechanism and therapeutic target (63, 64). In the case of bDMARDs, these are proteins derived from human genes to target some specific step within the auto-inflammatory disease process and so to ideally prevent ensuing structural damage (50). Feldmann and colleagues provided a necessary pre-cursor to this revolution by demonstrating that TNF played a key role in upregulating downstream pro-inflammatory cytokine cascades that

were known to be continually expressed in RA synovial tissue and were associated with joint damage (63). Furthermore, they subsequently conducted the first clinical trial of TNF inhibitor (TNFi) therapy for RA and reported favourable clinical and laboratory responses, both over the short-term (65) and long-term (66). Unequivocal data from large RCTs soon followed that clearly revealed TNFi therapy as an effective treatment for RA (67, 68). Guidelines for the use of TNFi for RA within the UK were developed by the BSR in 2001 (69) and subsequently adopted in their entirety by NICE in March 2002 (70). These recommended TNFi for patients with continuing severe RA (disease activity score (DAS) >5.1) who had already failed to respond to two 6-month long regimens of csDMARDs. The proviso was also made that any prescribing consultant must enrol the patient within the BSR Biologics Registry for Rheumatoid Arthritis (BSRBR-RA) (71), based in Manchester, UK. Interestingly, this guidance is now at odds to the general international consensus of leading rheumatology societies, in that these advocate initiation of bDMARDs for patients with a DAS >3.6 following csDMARD therapy (72).

Although blocking/inhibition of TNF was the start of the biologic-era for RA treatment, since then many other bDMARDs have been developed with different targeted mechanisms of action (50). These include inhibition of various interleukin receptors (e.g. Tocilizumab, Anakinra), B-cell depletion/inhibition (e.g. Rituximab), T-cell co-stimulation (e.g. Abatacept) and RANKL inhibition (e.g. Denosumab) (54).

Biologics and joint damage

It has become axiomatic in the rheumatology community that early and effective management of disease activity is key to minimising long-term irreversible joint damage, (64) and the emphasis of many clinical guidelines across the world has been on early and aggressive treatment (44, 45, 57, 58, 73, 74). Some studies suggest that existent joint destruction can even be reversed with potent TNFi therapy regimens (75-77), although this is considered rare (16, 24). In terms of the UK, BSR guidance (44) states that, "*The main aim of management of RA is to control the symptoms and signs of disease, maintain function and foster self-efficacy. All three are most likely to be achieved if inflammation is supressed… The aim should be remission*". (44) The subsequent 'treat-to-target' guidelines, developed by international experts, were even more specific, "*The primary goal of treating the patient with rheumatoid arthritis is to maximise long-term health-related quality of life through control of symptoms, prevention of structural damage, normalisation of function and social participation. …Abrogation of inflammation is the most important way to achieve these goals*" (78).

Following the early days of biologics, there has been a wealth of RCT studies demonstrating the efficacy of bDMARDs for RA (50, 79, 80). The primary outcomes of most of these trials are the ACR response criteria (80-82), defined (in the case of ACR20) as "*20% improvement in tender and swollen joint counts and 20% improvement in 3 of the 5 remaining ACR core set measures: patient and physician global assessments, pain, disability, and an acute-phase reactant*" (80). ACR20, ACR50 and ACR70 are all commonly used (80). The impact of biological therapies on patients' DAS is less well known but the

few trials investigating this outcome have reported favourable responses that mirror ACR criteria (50). Improvement in DAS following exposure to biologic therapies has been demonstrated in the observational setting (83). Numerous RCTs have demonstrated a favourable impact of biologic therapies on radiographic progression, although there have been mixed results (79, 84-88), most likely reflective of heterogeneity in the comparators used across these various trials (89, 90). For example, one meta-analysis demonstrated the structure-sparing effect after 1-year of bDMARD exposure compared to csDMARD monotherapy, reporting a 30% increased relative risk of non-progression of joint damage (radiographically) (88). However, another large meta-analysis reported that whilst radiographic progression was significantly lower at 1-year among those on bDMARD versus placebo, an almost identical reduction was achieved amongst those on a combination csDMARD regimen (85).

Temporal trends in joint replacement for RA

Over recent years there has been a growing body of evidence suggesting that the need for joint replacement has been in decline for RA patients, leading to speculation regarding some role of bDMARDs and their joint structure-sparing effects in these downward trends (91-94). The existing data on trends in joint replacement for RA is entirely observational in nature with a great deal of heterogeneity in study designs employed. The simplest of these are numerous reports on trends in count data of joint replacement for which RA is the main reason, and/or trends in the prevalence (proportion) of RA amongst all joint replacements performed for a given population. Such studies include the NJR for England, Wales and Northern Ireland which reported in

2004 that an indication of RA accounted for 3.1% of all THRs performed and 1.5% of all TKRs (95). In 2007 this was only 1% of THRs (percentage among TKRs not reported) (96) whilst in 2012 this was only 1% of TKRs (percentage among THRs not reported) (39). Similar data have been reported from the United States (US) where David *et al*. interrogated the very large Nationwide Inpatient Sample and found the raw number of all THR and TKR procedures carried out - for which RA was a primary diagnosis – significantly declined from 1993 to 2008. David *et al*. proceeded to estimate that the introduction of TNFi was associated with a 32% and 24% reduction in the prevalence of a primary RA diagnosis amongst all TKRs and THRs performed, respectively. Interestingly, when the same US study sample was recently interrogated for similar purposes, there was found to be a negligible increase from 2002 to 2012 in the prevalence of RA among all THRs and TKRs performed, probably explained by this latter study considering any diagnosis of RA – both primary and secondary (38).

Other studies have analysed trends of the burden of joint replacement actually within RA patients. This includes Harty *et al*. who performed a study of 57 public hospitals from the republic of Ireland (RoI) (94) and reported a "*profound*" reduction from 1995 to 2010 in the number of musculoskeletal surgical procedures in RA patients. The authors reported that despite numbers of THR and TKR increasing dramatically amongst the RoI general population, numbers of TKR decreased amongst patients with an RA diagnosis although numbers of THR remained approximately stable. Although not reporting on specific THR/TKR trends, another previous analysis of two UK inception cohorts of RA patients (97) found the 5-year and 10-year cumulative incidence of major joint

interventions (including joint replacement at any site) remained stable from 1986 to 2006. Likewise, whilst not explicitly investigating THR/TKR, Leon *et al*. (93) performed a retrospective medical record review of over 40 Spanish rheumatology units in order to compare the prevalence of orthopaedic surgery (since RA diagnosis) amongst a sample of 1,272 RA patients in 2010 to that in a similar sample taken ten years previously. The authors concluded this proportion had declined from 25.9% to 7.4% from 2000 to 2010. A similar methodology was used by Shourt *et al*. (98) who reported that the 10-year cumulative incidence of THR and TKR within 813 RA patients in Minnesota was lower throughout 1995-2007 relative to 1980-1994, although neither of these reached statistical significance. A study from a single institute in Japan indicated the number of prosthetic joint replacements per 1,000 RA outpatients peaked at approximately 16/1,000 RA outpatients in 2003 but had declined to approximately 9/1,000 in 2012 (99).

A slight variation on study design has been to investigate the incidence of RA-related joint replacement amongst the general population. Mertelsmann-Voss *et al*. (92) analysed 2,839,325 joint replacement procedures in the US, categorising them by whether they were performed for inflammatory or non-inflammatory reasons. They used census data to estimate incidence rates, showing that incidence of inflammatory-related joint replacements remained stable at 5.1/100,000 of the population from 1991 to 2005, despite the incidence of non-inflammatory-related joint replacement increasing from 124.5 to 247.5/100,000 of the population over the same timeframe. Similarly, Nystad *et al*. found the incidence of RA-related THR and TKR significantly reduced from approximately 4/100,000 inhabitants in 1994 to approximately 3/100,000 inhabitants in

2012. In an almost identical approach, Jamsen *et al*. analysed 245,854 joint replacements and found the incidence of all RA-related procedures to significantly decline from 18.5/100,000 in 1995 to 10.7/100,000 in 2010. Related to this approach is a study of joint replacement amongst an estimated RA population in California (100), using a combination of hospitalization data and RA prevalence estimates from survey data. This study suggested rates of THR and TKR (considering all ages together) had remained stable at approximately 300/100,000 RA patients and 450/100,000 RA patients from 1983 to 2007.

Finally, and possibly of highest quality in terms of methodology are those studies that have estimated temporal trends in the actual incidence (in person-years) of joint replacement in RA patients. These studies include that of Rodriguez-Rodriguez *et al*. (101) and Hekmat *et al*. (102). The former of these showed the 5-year incidence of THR and/or TKR among 1812 RA patients attending a single hospital in Madrid, Spain to decrease from 10.8 [95% CI: 7.3 to 16.2] / 1,000PYs to 6.9 [95% CI: 4.5 to 10.6] /1,000 PYs from 1994-1999 to 2005-2009. Hekmat and colleagues followed up 2,164 RA patients and reported incidence of THR fell from 12.6/1,000 PYs to 6.6/1,000 PYs, comparing the time periods 1998-2001 versus 2002-2006; (rate ratio (RR) = 0.52 [95% CI: 0.35 to 0.76]. The same estimates for TKR were 4.8/1,000 PYs and 6.8/1,000 PYs; RR=1.43 [95% CI: 0.89 to 2.31].

Overall, these findings tend to suggest some reduction in joint replacements for RA patients over latter years, although by no means is this unanimous as not all studies

report a decline and various limitations to the different methodologies used undermine the findings to a lesser or greater extent. While these issues are discussed and addressed in detail in chapter 4, it suffices to comment that none of the existing studies offer an explicit estimate of the trend-adjusted impact of introducing biologics on joint replacement rates within RA patients.

Other observational data on biologics and joint replacement

Apart from temporal trends of joint replacement, there are a small number of observational studies describing the burden of joint replacement in RA patients according to bDMARD exposure, although these are generally low in methodological quality. Several studies merely report on either the prevalence or incidence of joint replacement amongst RA biologic users versus non-users. One such study (103) found that the annual prevalence of orthopaedic surgery across years 2004 to 2007 amongst a Japanese RA cohort was on the whole greater (range 5.9% to 12.0%) for those receiving a biologic than those who did not (range 4.7% to 6.2%). Koenig *et al*. (104) likewise reviewed the US MarketScan database and identified a cohort of 90,545 incident, biologic-naïve RA patients. The authors reported that ~11% of those who subsequently initiated a biologic also subsequently underwent THR/TKR surgery, while this proportion was only ~7% amongst those who did not initiate a biologic. Ollendorf *et al*. (105) conducted another US database study comparing joint surgery interventions amongst RA patients after receiving Leflunomide, stratified by those also initiated etanercept, infliximab or not receiving a biologic. They found approximately equivalent 1-year

cumulative incidence of joint replacement: 12.2% (etanercept), 9.1% (infliximab) and 10.8% (no biologic).

Other studies have yielded more controlled estimates of the impact of exposure to biologics on subsequent need for joint replacement in RA. Two of these utilise large population-based health databases in Canada (106, 107). That by Widdifield *et al*. studied two large elderly (≥65 years old) cohorts of RA patients, one from Ontario (n=20,918) and the other from Quebec (n=6,754), in order to estimate the effect of early intensive therapy in general on joint replacement rates. Exposure to anti-TNF therapy over follow-up was entered as a time-dependent covariate into survival models adjusted for many other factors, which yielded a HR of 2.50 [95% CI: 0.72-8.65] and 1.13 [95% CI: 0.91-1.39] per year of exposure for Ontario and Quebec, respectively. The study by Moura *et al*. (106) investigated a very similar research question for all-aged incident RA patients in Quebec who had at least 1 year of follow-up. They found that anti-TNF use prior to follow-up was not associated with subsequent joint replacement surgery (HR of 1.00 [95% CI: 0.35 to 2.85]) nor was drug use during follow-up: HR of 1.13 [95% CI: 0.99 to 1.29] and HR of 1.37 [95% CI: 0.61 to 3.08] for exposure to anti-TNF and other biologics, respectively. One previous study has attempted to estimate the patient-level impact of biologics versus csDMARD amongst RA patients using drug registry data from Finland (108). This study found higher rates of primary joint replacement in bDMARD users (3.90 [95% CI: 3.42 to 4.43] / 100 PYs) compared to those only on csDMARD (2.64 [95% CI: 2.35 to 2.95) / 100 PYs], although that study was relatively small and failed to adjust for many important characteristics such as disease severity, prior medications and comorbidities,

which likely confounded their reported associations. Indeed, confounding is one of the key challenges to address in such comparisons of observational data and is one that the field of pharmacoepidemiology is continually seeking to better overcome.

Quasi-experiments in epidemiology

One of the fundamental challenges being specifically faced within the field of pharmacoepidemiology is that of the issue of confounding by indication. This biases estimation of treatment-outcome relationships, owing to the reasons (i.e. indications) for the treatment under study (109). In the case of biologic therapy and subsequent need for joint replacement, it is likely that users of biologics are - on average - at significantly higher baseline risk of joint replacement than non-users. This may be due, for example, to a higher disease severity (34) or greater degree of prior joint destruction. Failing to account for these systematic differences between biologic treated and untreated patients in an observational comparative effectiveness study would render results unreliable (110). Conversely, the RCT is often described as the gold standard of medical research (111, 112) given that treatment arms only differ according to exposure status and are otherwise balanced. That is, in a sufficiently large and well-conducted RCT the random allocation of treatment leads to equivalent groups allowing comparison of a primary outcome from the study data in such a way that is unconfounded by either measured or unmeasured characteristics (113). This is an enormous benefit and provides a means for making causal statements, for example about treatment efficacy.

However, there hasn't been an RCT investigating the impact of biologics on the need for joint replacement in RA despite there being appetite for such data, for example by health economists seeking to better model the cost-effectiveness of these therapies. Reasons for the dearth of good quality evidence are probably because of the large sample size and long follow-up that would be required to study such an outcome using an RCT (and associated costs) and likely ethical implications of randomly withholding biologic therapy to eligible patients. There are other hypothetical reasons to be wary of over reliance on RCT data. Inclusion criteria for most RCTs are often stringent and lead to study samples that are different to the "real world", often selecting healthier and younger participants with less co-morbidities and co-medications (114, 115). This often generates findings that have high internal validity (owing to random treatment assignment) but with potentially very low generalizability (114, 116).

In this context, there has been increasing development and use of novel methodologies within the field to better minimise the bias resulting from the use of population-based observational datasets and the inherent confounding by indication problem (117, 118). These include various methods that are termed "quasi-experimental" (119) and which offer the opportunity for causal inference where an RCT is likely to be unfeasible (112, 116). The underlying principle of these methods consists of having a natural or derived variable in the analytical dataset that is 'exogenous', i.e. not dependent on other variables in the system and the value of which is not controlled by the investigator (120, 121). This gives rise to the possibility of a research opportunity where "quasi-randomisation" can be achieved through careful consideration of who and/or when to

measure rather than who and/or when to allocate (119). Several of these methods can therefore yield estimates that account for not only measured but unmeasured confounding (122) and so have unique strengths due to high internal validity (higher than conventional regression-stratification approaches to observational data analysis), but simultaneously possessing much greater external validity compared to the typical RCT setting (116). However, the individual quasi-experimental methodologies each comes with a set of assumptions that are typically much stronger than those required in an RCT, some of which are generally untestable yet if invalidated then may lead to biased estimation (122). Important quasi-experimental examples include interrupted time-series (ITS), instrumental variable (IV) and regression discontinuity design (RDD) analyses (118, 122). These will each be described in subsequent chapters.

Aims and objectives of DPhil

In the light of this introduction, the aims of this thesis are set out below. The overarching purpose is to provide clinically useful findings on the relationship between TNFi therapy and need for joint replacement in RA, whilst advancing research methodologies as used where beneficial and as feasible. The project will use large health datasets and simulated data in conjunction with quasi-experimental approaches to address limitations surrounding analysis of observational data.

- **Aim 1**: **Describe the epidemiology of THR and TKR in RA, from a UK perspective**. Objectives include: estimate validity of THR/TKR codes in primary care data; estimate

overall incidence rates (per 1,000 PYs) and cumulative incidence; estimate stratified incidence rates and cumulative incidence according to key demographic variables; estimate overall average time-to-event from RA diagnosis.

- **Aim 2: Estimate the population-level impact of the historical introduction/approval of TNFi on rates of THR and TKR among RA patients in England and Wales**. Objectives include: estimate incidence rates over study period; estimate the impact of TNFi introduction on level/trend of incidence using interrupted time-series approach

- **Aim 3**: **Validate findings from aim 2 in external datasets for RA populations in countries (and healthcare systems) outside the UK.** Objectives include: estimate incidence rates in RA patients over study period; estimate the impact of TNFi introduction on level/trend of incidence in RA patients using interrupted time-series approach; estimate incidence rates and changes in level/trend of incidence in non-RA individuals using interrupted time-series approach; estimate changes in the incidence delta between RA patients and non-RA patients following introduction of TNFi using interrupted time-series approach

- **Aim 4**: **Estimate the patient-level impact of TNFi use vs. csDMARD use on subsequent need for THR and TKR surgery for RA patients in the UK.** Objectives include: overcome issue of missing data in UK drug registry data; match TNFi users to csDMARD users using propensity scores; estimate rates of joint replacement; perform survival analysis to estimate hazard ratios; stratify analyses according to

potential effect modifiers; perform instrumental variable and regression discontinuity design sensitivity analyses to explore impact of unmeasured confounding

- **Aim 5: Develop the relatively novel interrupted time-series methodology used for aim 2, as feasible within the scope of the project.** Objectives: develop simulation approach to estimate power in interrupted time-series; explore impact of various scenario-specific parameters on power; provide tool for estimating required sample size to achieve adequate power for interrupted time-series analysis.

2. DATA USED

Introduction

As the title of this thesis indicates, and as has been introduced in chapter 1, the question to be addressed in forthcoming chapters is the impact of TNFi on need for THR/TKR. This will be estimated using electronic medical records and/or administrative databases (from the UK, Denmark and Canada) and observational drug registry data (from the UK). This relatively brief chapter provides an overview of the various data sources used throughout the project. Each dataset is described in turn and finally the role of collaborators is provided.

Clinical Practice Research Datalink (CPRD) and the Hospital Episode Statistics (HES)

CPRD GOLD was used for analyses in chapters 3 & 4. CPRD is a database of UK primary care health records. the UK National Health Service (NHS) primary care is offered/provided free of charge through general practitioners (GPs), with almost the entirety of the population registered to a local GP practice (123). GPs act as the 'gatepost' to the UK NHS, providing free consultations to registered patients and providing subsequent non-emergency healthcare, prescriptions, diagnostic tests or referral to secondary care where necessary. The GP is required to make notes following consultations, which may include patient symptoms, test results, diagnoses, clinical measurements (e.g. BMI and smoking status) and medicines prescribed. These notes are almost invariably in digital format which provides a self-contained longitudinal medical

record for each patient. Furthermore, where patients are seen in secondary care then letters and/or discharge sheets are sent from such secondary care providers to the local GP for appropriate action and inclusion in the patient's medical record. CPRD is a sample of such GP routinely collected medical records and as of 2013 cumulatively contained data from over 600 GP practices and covered approximately 11 million people (123). It's been previously shown to be representative of the UK in terms of age, sex and ethnicity (123).

Symptoms, diagnoses, operations, procedures and other health-related information is recorded into CPRD using Read/Oxford Medical Information Systems (OXMIS) coding (124). A bespoke dataset of RA patients was requested, along with a number of baseline characteristics and outcome identifier using Read/OXMIS code lists that are included in appendix file 1 & 2, and as used previously (125, 126). Raw data were subsequently provided by CPRD GOLD for these patients, which also included linked Office for National Statistics (ONS) mortality data, index of multiple deprivation (IMD) and secondary care hospital data (Hospital Episode Statistics [HES]). These raw data were provided in the form of multiple text files that were then cleaned and merged by a senior data manager to create a pre-analytical dataset. Any proposed study of CPRD leading to publication must be submitted to the Independent Scientific Advisory Committee (ISAC) for review and this had previously been carried out with permission received for the current DPhil project.

Danish National Patient Register (DNPR)

The DNPR was used in chapter 5 and is an internationally regarded resource of secondary care health data, set up in 1977 and covering the entirety of in- and out-patient hospital visits across Denmark (127). It contains data on each hospital department visited, date of admission, diagnoses given, operations performed, and treatments provided. Surgical procedures were coded, as of 1996, using the NOMESCO Nordic Classification of Surgical Procedures (NCDP). Since the year 2000 the DNPR has formed the basis of payment to public hospitals, and therefore is assumed to be a complete account of all health-care provided.

Civil Registration System (CRS)

The CRS was used in chapter 5 and is a population-wide administrative database containing information such as date of birth, sex, citizenship, civil status and vital status on all individuals in Denmark (128). It's great usefulness in epidemiology research has been previously described given that it contains the civil person register number, which can then be used to link to a host of other Danish registers and databases. It was here used to identify an age, sex and municipality matched control cohort to compare to that of the RA cohort identified through the DNPR. Furthermore, it was used to censor patients given the vital status not only includes death but migration out of the country.

Ontario Rheumatoid Arthritis administrative Database (ORAD)

ORAD was used in chapter 5 and is an administrative database of RA patients covering a denominator population of approximately 14 million people (as of 2014) and is designed

to contain information on all RA patients in Ontario, Canada. As of 2010 ORAD included nearly 100,000 RA patients (129) but was estimated prior to data access to have grown to 150,000 patients. Along with its linked datasets, it contains information on all visits to both primary and secondary care, in addition to demographic and (publicly funded) pharmacy data (129). It has previously been validated and shown to have high sensitivity and specificity when compared to either primary-care (130) or secondary-care (131) gold standards. Patients are included into the ORAD if they fulfil either of the following conditions (129, 131). Firstly, if they've ever had an RA hospitalization code as according to either the Canadian Institute for Health Information Discharge Abstract Database (covering all hospital admissions) or the National Ambulatory Care Reporting System (covering all ambulance activity). Secondly, if they have three of more RA-related insurance claims to the Ontario Health Insurance Plan (OHIP) database (with more than one being from a specialist), which all Ontario physicians must use for reimbursement purposes. RA-related claims are determined using ICD-8 or ICD-9 codes associated with each claim in order to identify the main reason for the consultation.

The OHIP Registered Persons Database was also used to identify non-RA controls for matching purposes. This is a collection of demographic data (including vital status, age, sex and place of residence) on all individuals living in Ontario who are eligible for OHIP coverage. ORAD is also linked to the Ontario Drug Benefit Program (ODBP) database and this linkage was used to identify pharmacy claims on all individuals included in the study who were $\geq$65 years old (132). The reason why ODBP only contains pharmacy data on seniors is because these are by default covered by the Ontario public drug insurance

program, whereas those under 65 must meet various conditions to be covered for medication and is not usual.

British Society for Rheumatology Biologics Register - Rheumatoid Arthritis (BSRBR-RA)

The BSRBR-RA was used in chapter 7 and contains prospectively collected observational data on over 20,000 RA patients recruited from 2001 onwards, primarily in order to evaluate the real-world safety of TNFi. The initial UK NICE guidance on use of TNFi stipulated clinicians initiating such therapy must register the patient into the BSRBR-RA, and recruitment continued originally until the target of at least 4,000 patients per TNFi cohort (etanercept, infliximab or adalimumab) was reached (71) (last "original" patient recruited 2008) but reopened recruitment to these three originator drugs from 2010 onwards. NICE guidance restricts NHS prescribing of TNFi to patients with a sustained 28-joint disease activity score (DAS28) >5.1 who have failed to adequately respond to two csDMARDs, with each treatment lasting ≥6 months. The BSRBR-RA contains an additional comparator cohort of non-biologic treated RA patients on csDMARDs, entry into which was dependent on having active disease (guide DAS28 >4.2). Recruitment to this cohort closed in 2008.

Participants in all study cohorts are followed up using physician questionnaires sent from BSRBR-RA to the patients' rheumatology clinic. These were sent every six months for a patients' first three years of follow-up, and annually thereafter. In addition to collecting data on changes to therapy and disease details, physicians were asked what serious and adverse events had occurred since last follow-up date. In addition, patients were asked

to complete a health diary every six months for the first three years in the study in which they detailed any hospital admissions. These free text responses were coded by BSRBR-RA staff using the MedRA hierarchy (Medical Dictionary for Regulatory Activities (133)). Mortality data for participants were obtained from the Health and Social Care Information Centre (HSCIC) (now merged into NHS Digital) via alerts generated from ONS records. The BSRBR steering committee provided approval for the project, granting data access in 2016. All patients provided written informed consent, no further approvals were necessary.

Collaborator roles in data extraction/management

The various data linkages and extractions from the above databases were performed by collaborating partners (data managers) with expertise in such large-scale data management of the respective databases. The senior database manager for the Oxford pharmaco- and device-epidemiology research group performed the data extraction from CPRD-provided raw data and merged individual variables into a combined pre-analytical dataset, using a data dictionary as developed in collaboration with my supervisors and myself, and applying pre-established health big data standard operating procedures (see appendix 3). In the case of Danish data, this pre-analytical processing was carried out in a fashion to replicate the form of the CPRD dataset although this was performed remotely almost entirely independently from myself (I attended a meeting in Copenhagen to facilitate this). In the case of the Canadian dataset, the pre-analytical data processing was carried out by collaborators in Ontario, working to a protocol as developed by my supervisor and myself. I visited the BSRBR-RA management team in

Manchester to explore data availability and submitted the application for data. These were subsequently provided in the form of baseline, follow-up and medication tables. I merged these and derived the required variables for the required analytical dataset.

3. DESCRIPTIVE EPIDEMIOLOGY OF THR AND TKR IN RA

INTRODUCTION

As introduced in chapter 1, although the natural history of structural damage in RA has been previously described (25, 30), long-term clinical outcomes such as the incidence of joint replacement remains less well studied (35). Notwithstanding the prior studies reporting on the prevalence of RA among larger samples of joint replacement procedures (38, 91, 92, 134, 135), good quality data pertaining to the incidence of joint replacement following RA diagnosis is arguably more useful in terms of better understanding the longer term natural history of RA. However, these data are still only emerging (35, 93, 94, 97, 98, 102, 106, 108, 136, 137). Generalizable, population-based estimates of joint replacement incidence among RA patients and the potential influence of demographic and clinical characteristics are scarce. Such data would provide clinicians and other stake holders a more thorough understanding of the long-term prognosis for specific RA patients and allow a greater ability to plan healthcare resource utilisation. This is a worthwhile consideration given the UK and US estimates mentioned above of 1-3% of hip and knee replacements being carried out for RA, and the recently estimated cost of between £6,000 and £7,000 per THR/TKR operation (138).

The aim of the work reported in this chapter was therefore to describe the epidemiology, from a UK perspective, of THR and TKR following a diagnosis of RA using population-

based electronic medical record data. Objectives included the validation of THR/TKR coding in primary care data, estimation of overall incidence rates (per 1,000 PYs) and 10- and 20-year cumulative incidence of THR/TKR, estimation of stratified incidence rates and 10-year cumulative incidence according to key demographic variables, and estimation of overall average time-to-THR/TKR from RA diagnosis.

METHODS

Data and participants

Primary care health data were obtained from the UK Clinical Practice Research Datalink (CPRD) for the period April 1995 to September 2014. These are described in detail in chapter 2. In brief, incident RA patients with at least 1 year of CPRD GOLD data prior to index (diagnosis) date were identified using a pre-defined READ code list (appendix file 1) as developed elsewhere (125), with the date of first recorded RA considered as diagnosis date. Data on gender, age, BMI, smoking status and index of multiple deprivation (IMD) were extracted from CPRD. The values for these variables were taken as recorded either on the date of diagnosis or otherwise the last prior value before date of RA diagnosis. Patients diagnosed with RA before the study period were excluded as were those with either a prior or subsequent diagnosis of a different inflammatory arthritis (lupus, ankylosing spondylitis, psoriatic arthritis or crystal arthropathy) due to possible diagnosis or coding errors. Patients aged <18 years old were also excluded.

Outcome

Outcome of interest was first occurrence of THR or TKR following RA diagnosis. These were identified using CPRD Read codes (appendix file 2) as used previously in the published literature (126, 139). THR and TKR were considered separately so patients could potentially have both outcomes of interest. Patients were followed up from date of RA diagnosis until the first date of either outcome event, death or loss-to-follow up.

Outcome Validation

As a preliminary step, the validity of THR and TKR Read codes as recorded in CPRD were assessed by comparing them to linked secondary care HES data. The following epidemiological measures were estimated: sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) (140). THR and TKR as recorded in HES was considered the reference standard as it's been shown to be comparable to the NJR, which itself has a near 100% compliance against levy returns for the number of implants sold (141). Given that some patients may have bilateral hip or knee replacement (the likelihood of which increases over follow up) and that laterality data was not available in CPRD, a time window was introduced around the first coded event in either database around which to 'look' for an event in the reciprocal data source. This time window was a-priori set to 60 days, with 30 days and 180 days explored as sensitivity definitions. Only the first occurrence of THR or TKR (analysed separately) during follow up in either data source was available. In this way, if an event was identified in CPRD with no corresponding HES event within the time window then it was considered a 'false positive' in CPRD. Conversely, if an event was identified in HES with no event in CPRD within the

time window then a 'false negative' in CPRD was considered to have occurred.

Statistical analysis

In addition to baseline patient characteristics, standard epidemiological descriptive statistics were calculated for the whole study sample. These were: number of patients included in analyses, total number of person-years contributed by all included patients, median follow-up time (in years) per patient with associated IQR, number of outcome events, median time to outcome event (in years) with associated IQR, incidence rate (per 1,000 person years) of outcome event with associated 95% CI and cumulative probability (accounting for death as a competing risk) at 10 and 20 year timepoints with associated 95% CI. To explore data graphically, cumulative incidence function plots were generated.

Similarly, number of patients, number of outcome events, incidence rates (per 1,000 PYs) with associated 95% CI and cumulative probability at 10 years with associated 95% CI were generated according to a number of pre-determined stratification variables. Characteristics explored in this fashion were: gender, age group, BMI category, smoking status, IMD quintile and geographic region. Univariate Cox models were used to test for proportional differences in hazard rates/curves (i.e. the instantaneous probability of event of interest occurring at a given time of follow-up, conditional on surviving event-free up until that time) across levels of stratification variables. *p*-values from these models were used to assess the conformity of data to what would be expected under the null hypothesis of no difference in rates between strata. Specifically, *p*-values from these Cox models were used to assess for differences between binary stratification

variables while *p*-trend was used to assess evidence of linear trend across ordinal categorical variables. Geographic region and smoking were considered nominal categorical variables (i.e. neither binary nor ordinal) and *p*-values for equality were generated using a chi-squared test.

RESULTS

Outcome validation

The total CPRD sample consisted of 27,607 incident RA patients, of whom 15,215 (55.1%) had HES linkage. In validation analyses constricted to this subset, there were a recorded 548 THR procedures and 816 recorded TKR procedures within the linked HES dataset. Of these procedures, 441 (80.5%) and 590 (83.6%) had a corresponding event recorded in CPRD (within a period of $\pm$60 days) (Table 3.1). This yielded a sensitivity of 80.5 and 83.6 for THR and TKR, respectively. Likewise, specificity was 98.6 and 98.5 for THR and TKR, respectively. Corresponding values for PPV were 68.5 and 72.5, whilst those for NPV were 99.3 and 99.2 (Table 3.2). All estimates were similar in analyses varying the length of time widow for agreement to 30 days and 180 days (appendix table 3.1). Given this apparent good validity of CPRD coding, the full study sample of 27,607 incident RA patients was used in subsequent analyses.

TABLE 3.1: CROSS TABULATION OF IDENTIFIED OUTCOME EVENTS DURING FOLLOW-UP ACCORDING TO CPRD VERSUS HES (ALLOWING 60-DAY TIME WINDOW FOR AGREEMENT): FOR THR AND TKR

Total Hip Replacement

		HES		
		Positive	Negative	Total
CPRD	Positive	441	203	644
	Negative	107	14,464	14,571
	Total	548	14,667	15,215

Total Knee Replacement

		HES		
		Positive	Negative	Total
CPRD	Positive	590	224	814
	Negative	116	14,285	14,509
	Total	706	14,401	15,215

TABLE 3.2: VALIDATION MEASURES OF OUTCOME EVENTS DURING FOLLOW-UP ACCORDING TO CPRD VERSUS HES (ALLOWING 60-DAY TIME WINDOW FOR AGREEMENT): FOR THR AND TKR

	Sensitivity	Specificity	PPV	NPV
HIP	80.5	98.6	68.5	99.3
KNEE	83.6	98.5	72.5	99.2

Baseline Characteristics

Of the total 33,044 RA patients identified, 5,437 were excluded (figure 3.1). The subsequent study sample consisted of 27,607 incident RA patients (Table 3.3) with a mean age of 61 [IQR: 50, 72] and of whom 70.6% were female. Mean BMI was 27.3 kg/m^2 and 37% had never smoked while 22% were ex-smokers and 19% were current smokers

(Table 3.3). 23% of patients from England resided in an area ranked within the least deprived quintile of the country, while 14% resided in an area ranked within the most deprived quintile (Table 3.3). The number of RA patients diagnosed per region ranged from 565 (2% of the total sample) in the North East to 3,330 (12.1%) in the North West (Table 3.3).

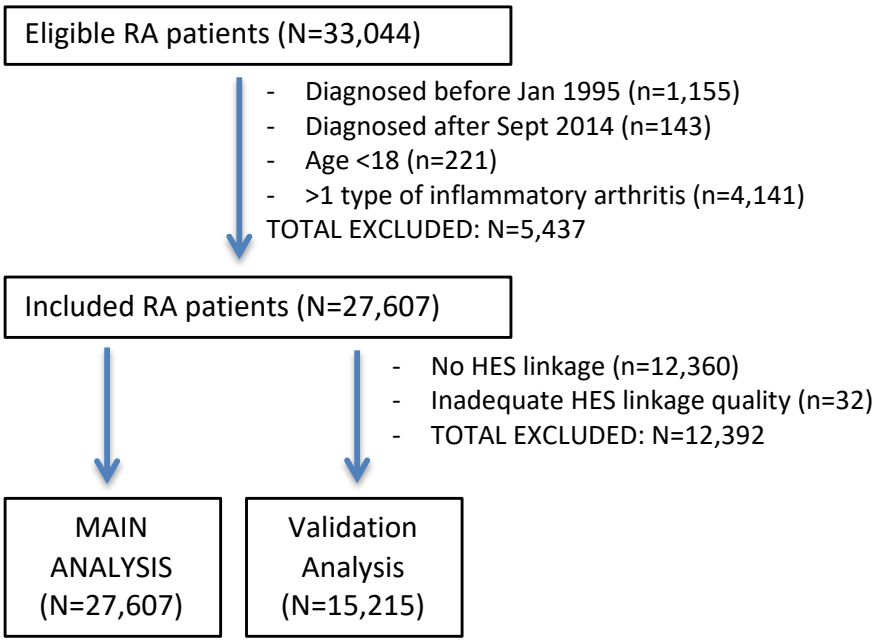


FIGURE 3.1: POPULATION FLOW DIAGRAM

Overall incidence

Overall, 1,028 patients received a primary THR subsequent to their RA diagnosis (Table 3.3) during a total of 161,211 person-years (PYs) and median follow-up of 4.9 [IQR: 2.2, 8.7] years. This yielded an incidence rate of 6.38 [95% C.I. 6.00 to 6.78] per 1,000 PYs

TABLE 3.3: BASELINE CHARACTERISTICS, OUTCOME EVENTS AND INCIDENCE RATES DURING FOLLOW-UP: FOR THR AND TKR

	No. patients		THR		
	N	%	No. Outcome	IR (95% CI) / 1,000 Pys	10-year cumulative % probability (95% CI)
Overall	27,607	100	1,028	6.38 (6.00 - 6.78)	5.2 (4.9 - 5.6)
Gender					
Men	8,107	29.4	239	5.25 (4.63 - 5.96)	4.2 (3.6 - 4.8)
Women	19,500	70.6	789	6.82 (6.36 - 7.31)	5.7 (5.2 - 6.1)
P- value				<0.001	
Age					
<45	4,305	15.6	58	2.12 (1.64 - 2.74)	2.2 (1.6 - 3.0)
45-54	4,979	18.0	128	3.85 (3.24 - 4.58)	3.8 (3.1 - 4.6)
55-64	6,755	24.5	296	6.81 (6.07 - 7.63)	6.2 (5.4 - 7.0)
65-74	6,194	22.4	360	10.46 (9.43 - 11.60)	7.9 (7.0 - 8.7)
≥75	5,374	19.5	186	8.20 (7.10 - 9.46)	4.4 (3.8 - 5.1)
P- trend				<0.001	
BMI					
Under weight (<18.5)	454	1.6	10	4.66 (2.52 - 8.66)	2.4 (1.2 - 4.4)
Normal weight (18.5 - 24.9)	6,852	24.8	222	5.92 (5.19 - 6.76)	4.8 (4.1 - 5.6)
Over weight (25 -29.9)	6,689	24.2	249	7.13 (6.29 - 8.07)	5.9 (5.1 - 6.7)
Obese / Severly Obese (30 - 34.9)	3,207	11.6	115	7.03 (5.86 - 8.84)	6.3 (5.1 - 7.7)
More obese (≥35)	1,835	6.7	33	3.64 (2.59 - 5.12)	3.3 (2.2 - 4.8)
Unknown	8,570	31.0	399	6.52 (5.91 - 7.19)	5.3 (4.8 - 5.9)
P- trend				0.54	
IMD*					
Least deprived	3,899	22.8	171	7.49 (6.45 - 8.70)	6.5 (5.5 - 7.6)
Less deprived	3,834	22.4	176	7.88 (6.80 - 9.13)	6.3 (5.3 - 7.3)
Mid deprived	3,609	21.1	125	5.91 (4.96 - 7.04)	5.0 (4.1 - 6.1)
More deproved	3,348	19.5	121	6.20 (5.19 - 7.41)	4.7 (3.8 - 5.6)
Most deprived	2,442	14.3	71	5.07 (4.02 - 6.40)	4.2 (3.2 - 5.3)
Unknown	10,475	21.6	149	5.56 (4.73 - 6.53)	5.0 (4.3 - 5.8)
P- trend				0.001	
Region					
London	2,511	9.1	78	5.63 (4.51 - 7.02)	4.6 (3.5 - 5.8)
North West	3,330	12.1	118	5.63 (4.70 - 6.75)	4.4 (3.6 - 5.3)
Yorkshire & Humber	1,228	4.5	39	5.48 (4.00 - 7.50)	4.6 (3.2 - 6.4)
East Midlands	1,246	4.5	52	6.85 (5.22 - 8.99)	6.0 (4.5 - 7.8)
West Midlands	2,492	9	89	6.16 (5.00 - 7.58)	5.3 (4.2 - 6.6)
East of England	2,681	9.7	111	7.20 (5.98 - 8.67)	5.8 (4.7 - 7.1)
South West	2,469	8.9	93	6.85 (5.59 - 8.40)	5.7 (4.5 - 7.0)
South Central	2,547	9.2	101	6.95 (5.72 - 8.45)	5.4 (4.3 - 6.6)
North East	565	2	13	3.82 (2.22 - 6.57)	3.1 (1.7 - 5.2)
South East Coast	2,786	10.1	119	7.58 (6.34 - 9.07)	6.5 (5.3 - 7.8)
Northern Ireland	944	3.4	41	6.82 (5.02 - 9.27)	5.4 (3.9 - 7.4)
Scotland	2,410	8.7	102	7.40 (6.10 - 8.99)	5.9 (4.7 - 7.2)
Wales	2,398	8.7	72	4.86 (3.86 - 6.13)	4.1 (3.1 - 5.2)
P- value				0.11	
Time period					
1995-2004	11,699	42.4	645	6.59 (6.10 - 7.12)	5.4 (5.0 - 5.9)
2005-2014	15,908	57.6	383	6.05 (5.47 - 6.68)	4.9 (4.2 - 5.7)
P- value				0.16	

Continued on next page...

...Continued from prior page

	No. patients		TKR		
	N	%	No. Outcome	IR (95% CI) / 1,000 Pys	10-year cumulative % probability (95% CI)
Overall	27,607	100	1,366	8.57 (8.12 - 9.04)	7.0 (6.6 - 7.4)
Gender					
Men	8,107	29.4	382	8.54 (7.73 - 9.44)	6.5 (5.8 - 7.2)
Women	19,500	70.6	984	8.58 (8.06 - 9.14)	7.2 (6.7 - 7.7)
P- value				0.92	
Age					
<45	4,305	15.6	87	3.18 (2.58 - 3.93)	3.0 (2.2 - 3.8)
45-54	4,979	18.0	202	6.15 (5.36 - 7.06)	5.6 (4.7 - 6.5)
55-64	6,755	24.5	437	10.22 (9.31 - 11.22)	9.2 (8.3 - 10.1)
65-74	6,194	22.4	398	11.69 (10.60 - 12.90)	8.9 (8.0 - 9.9)
≥75	5,374	19.5	242	10.82 (9.54 - 12.27)	6.1 (5.3 - 6.9)
P- trend				<0.001	
BMI					
Under weight (<18.5)	454	1.6	16	7.64 (4.68 - 12.47)	5.1 (2.7 - 8.5)
Normal weight (18.5 - 24.9)	6,852	24.8	217	5.80 (5.08 - 6.62)	4.6 (4.0 - 5.4)
Over weight (25 -29.9)	6,689	24.2	286	8.21 (7.31 - 9.22)	6.9 (6.0 - 7.9)
Obese / Severly Obese (30 - 34.9)	3,207	11.6	167	10.39 (8.93 - 12.09	8.3 (7.0 - 9.9)
More obese (≥35)	1,835	6.7	116	13.33 (11.11 - 15.99)	11.1 (9.0 - 13.4)
Unknown	8,570	31.0	564	9.36 (8.62 - 10.17)	7.5 (6.9 - 8.2)
P- trend				<0.001	
IMD*					
Least deprived	3,899	22.8	205	9.06 (7.90 - 10.39)	7.2 (6.2 - 8.3)
Less deprived	3,834	22.4	197	8.87 (7.72 - 10.20)	7.4 (6.4 - 8.6)
Mid deprived	3,609	21.1	191	9.21 (7.99 - 10.61)	7.8 (6.6 - 9.0)
More deproved	3,348	19.5	150	7.76 (6.62 - 9.12)	5.9 (4.9 - 7.0)
Most deprived	2,442	14.3	101	7.30 (6.00 - 8.87)	6.0 (4.8 - 7.3)
Unknown	10,475	21.6	226	8.52 (7.48 - 9.70)	7.3 (6.5 - 8.3)
P- trend				0.041	
Region					
London	2,511	9.1	96	6.96 (5.70 - 8.50)	6.0 (4.7 - 7.4)
North West	3,330	12.1	137	6.59 (5.57 - 7.79)	5.6 (4.6 - 6.6)
Yorkshire & Humber	1,228	4.5	57	8.13 (6.27 - 10.54)	6.3 (4.7 - 8.3)
East Midlands	1,246	4.5	71	9.43 (7.47 - 11.90)	7.9 (6.1 - 10.0)
West Midlands	2,492	9	146	10.34 (8.79 - 12.16)	8.1 (6.8 - 9.6)
East of England	2,681	9.7	127	8.29 (6.97 - 9.87)	6.5 (5.4 - 7.8)
South West	2,469	8.9	123	9.14 (7.66 - 10.91)	7.5 (6.2 - 9.1)
South Central	2,547	9.2	138	9.58 (8.11 - 11.32)	7.3 (6.0 - 8.6)
North East	565	2	31	9.32 (6.55 - 13.25)	8.3 (5.5 - 11.8)
South East Coast	2,786	10.1	144	9.28 (7.88 - 10.92)	7.1 (5.9 - 8.4)
Northern Ireland	944	3.4	43	7.16 (5.31 - 9.65)	5.7 (4.0 - 7.8)
Scotland	2,410	8.7	119	8.70 (7.27 - 10.42)	7.6 (6.2 - 9.2)
Wales	2,398	8.7	134	9.28 (7.84 - 10.99)	7.8 (6.4 - 9.3)
P- value				0.034	
Time period					
1995-2004	11,699	42.4	858	8.89 (8.31 - 9.50)	7.2 (6.7 - 7.7)
2005-2014	15,908	57.6	508	8.08 (7.41 - 8.82)	6.6 (5.4 - 7.8)
P- value				0.091	

* Values unknown for Wales, Scotland and Northern Ireland

(Table 3.3), with median time from diagnosis to THR being 3.2 [IQR: 1.3, 6.5] years. Similarly, there were 1,366 patients who received a primary TKR during a total of 159,384

PYs at an incidence rate of 8.57 [95% C.I. 8.12 to 9.04] per 1,000 PYs (Table 3.3). Median follow-up in the TKR analysis was 4.9 [IQR: 2.1, 8.6] overall and median time from RA until TKR was 3.5 [IQR: 1.5, 6.7] years. The overall 10-year cumulative percentage probability for THR and TKR were 5.2% [95% C.I. 4.9, 5.6] and 7.0% [95% C.I. 6.6, 7.4], respectively (Table 3.3). At 20 years these values were 8.4% [95% C.I. 7.3, 9.7] and 11.1% [95% C.I. 10.0, 12.4], respectively (Figure 3.2).

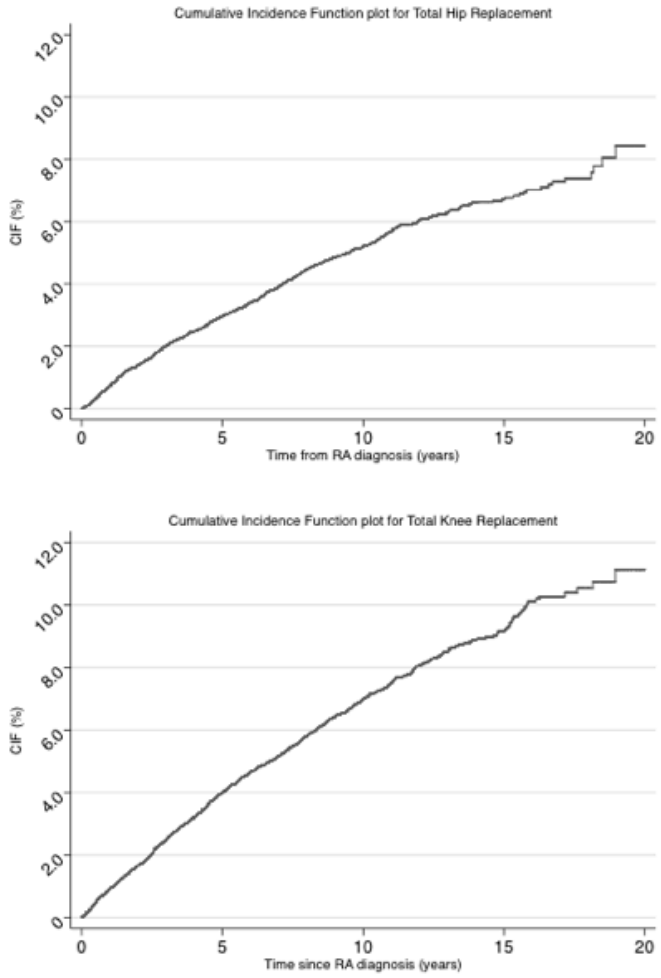


FIGURE 3.2 **CUMULATIVE INCIDENCE FUNCTION PLOTS FOR THR AND TKR OVER STUDY FOLLOW-UP**

Incidence by sex and age

Among those receiving a THR, the proportion of females was 76.8% and mean age at surgery was 68.4 years. Among the TKR cohort, the proportion of females was 72.0% and mean age was 67.6 years. While rates of THR were significantly higher in women (6.82 [95% C.I. 6.36 – 7.31] per 1,000 PYs) than in men (5.25 [95% C.I. 4.63 – 5.96] per 1,000 PYs), the rates of TKR were approximately equal between genders (Table 3.3, Figure 3.3). Rates were lowest among those aged <45 years of age (2.12 [95% C.I. 1.64 – 2.74] per 1,000 PYs for THR and 3.18 [95% C.I. 2.58, 3.93] per 1,000 PYs for TKR), with rates rising significantly with increasing age up to those aged 65-74 years old (10.46 [95% C.I. 9.44 – 11.64] per 1,000 PYs for THR and 11.69 [95% C.I. 10.60 – 12.90] per 1,000 PYs for TKR) (Table 3.3, Figure 3.4). However, rates then began to decline in those ≥75 years (Table 3.3, Figure 3.4).

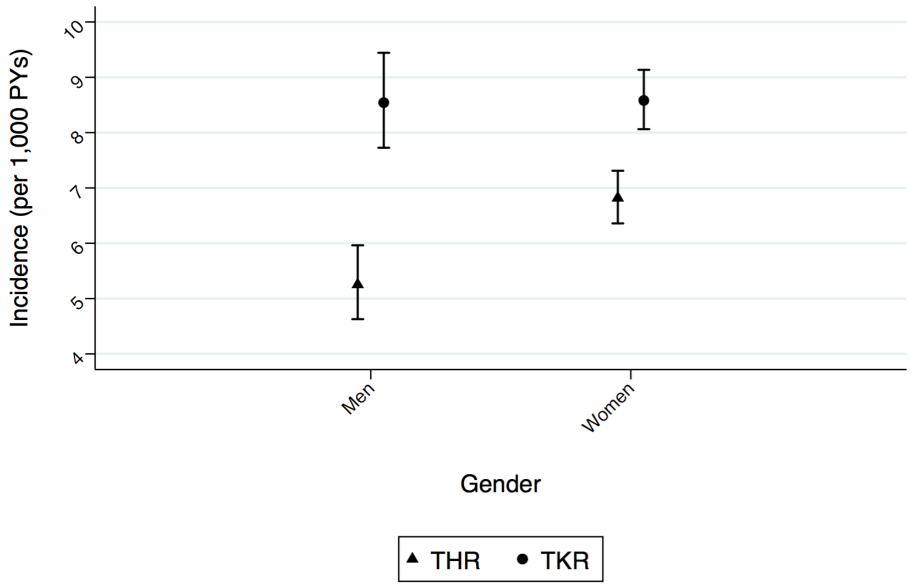


FIGURE 3.3: INCIDENCE RATE OF THR AND TKR: STRATIFIED BY GENDER

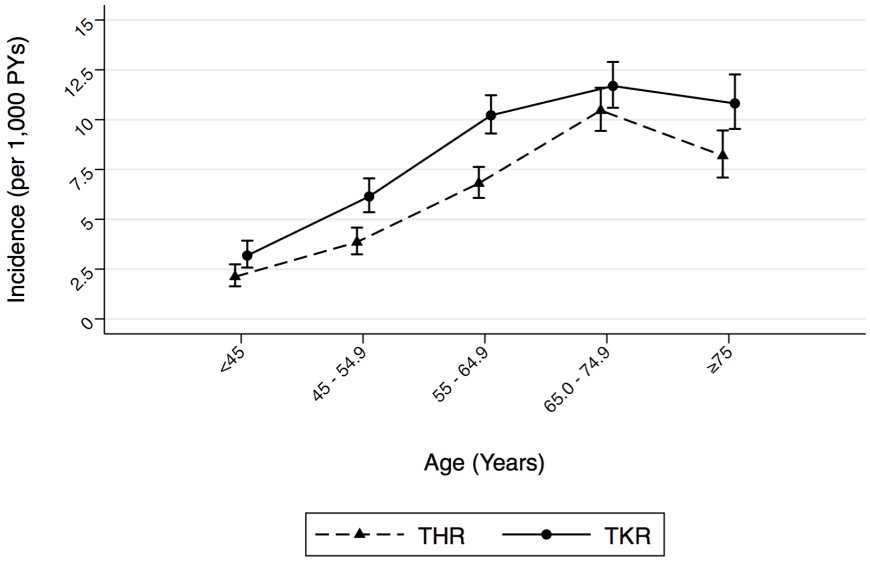


FIGURE 3.4: INCIDENCE RATES OF THR AND TKR: STATIFIED BY AGE

Incidence by BMI and smoking

BMI had little impact on rates of THR, with incidence rates approximately constant. However, a significant upward trend (p<0.001) was seen in TKR procedures with increasing BMI, with incidence rates lowest among those at a normal weight (5.80 [95% C.I. 5.08 – 6.62] per 1,000 PYs) (Table 3.3, Figure 3.5). Those with a BMI ≥30 had an incidence of TKR nearly twice as large as those with a normal weight (11.42 [95% C.I. 10.17 – 12.83] per 1,000 PYs) (Table 3.3, Figure 3.5). BMI was not known for 31% of the sample, within whom the incidence of THR and TKR was 6.52 [95% C.I. 5.91 – 7.19] and 9.36 [95% C.I. 8.62 – 10.17] per 1,000 PYs, respectively (Table 3.3). There were significant differences in incidence according to smoking status (Never, Ex-, current), with lowest rates observed among the "current smoker" category (Table 3.3, Figure 3.6). However, 22% of the sample had unknown smoking status, in whom rates of THR and TKR were

6.65 [95% C.I. 5.96 – 7.42] and 9.96 [95% C.I. 4.42 – 10.90] per 1,000 PYs, respectively (Table 3.3).

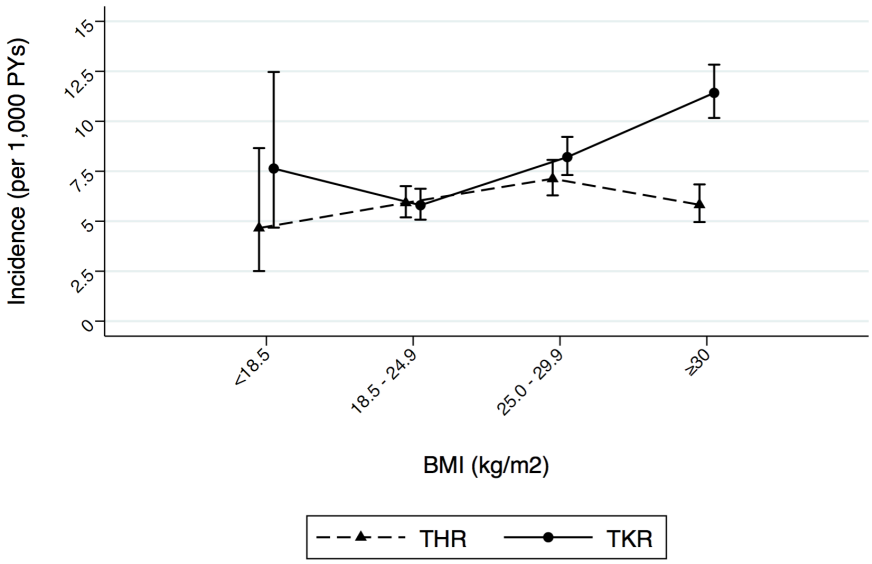


FIGURE 3.5: INCIDENCE RATES OF THR AND TKR: STRATIFIED BY BMI

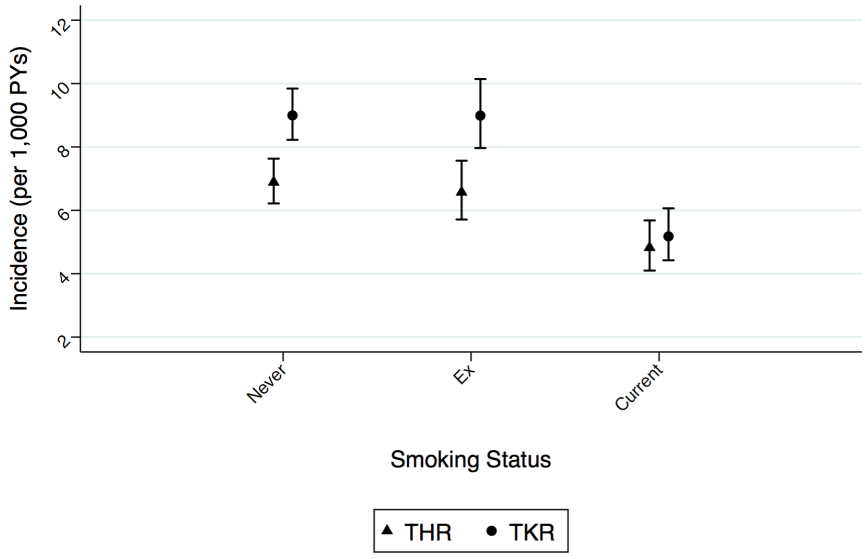


FIGURE 3.6: INCIDENCE RATES OF THR AND TKR: STRATIFIED BY SMOKING STATUS

Incidence by deprivation and geographic region

A significant downward trend in joint replacement rates with greater deprivation was observed for both THR and TKR, *p*-trend=0.001 and *p*-trend=0.041, respectively (Table 3.3, Figure 3.7). Estimated rates of THR were >30% lower in those most deprived versus least deprived (5.07 [95% C.I. 4.02 – 6.40] versus 7.49 [95% C.I. 6.45 – 8.71] per 1,000 PYs) (Table 3.3, Figure 3.7). For TKR there was an approximate 20% decrease in incidence among those most deprived versus least deprived (7.30 [95% C.I. 6.00 – 8.87] versus 9.06 [95% C.I. 7.90 – 10.39] per 1,000 PYs) (Table 3.3, Figure 3.7). There was little evidence for regional variation of THR (*p*-value=0.11), with incidence being lowest in the North East (3.82 [95% C.I. 2.22 – 6.57] per 1,000 PYs) and highest on the South East Coast (7.58 [95% C.I. 6.34 – 9.07] per 1,000 PYs) (Table 3.3, Figure 3.8). Weak evidence was found for variation of TKR (*p*=0.034), with incidence being lowest in the North West (6.59 [95% C.I. 5.57 – 7.79] per 1,000 PYs) and highest in the West Midlands (10.34 [95% C.I. 8.79 – 12.16]) (Table 3.3, Figure 3.8).

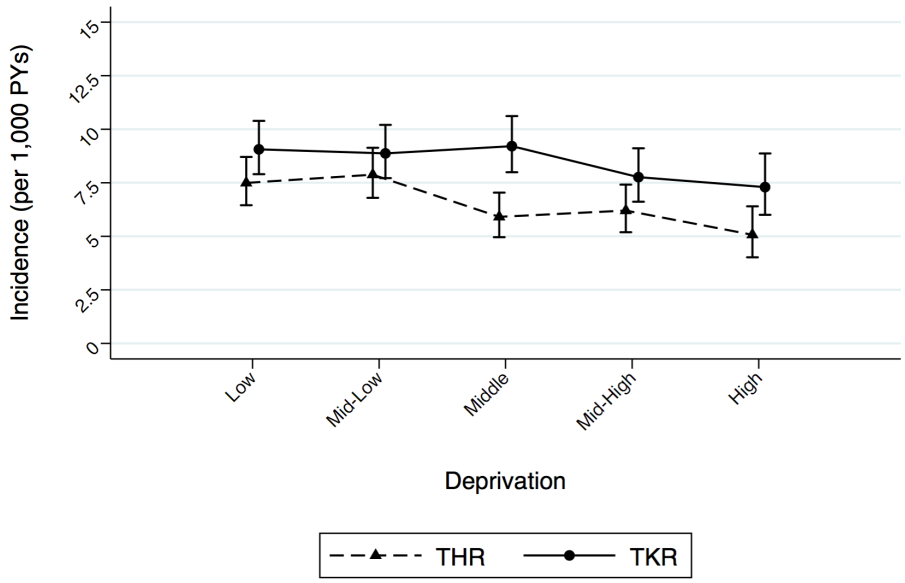


FIGURE 3.7: INCIDENCE OF THR AND TKR: STRATIFIED BY INDEX OF MULTIPLE DEPRIVATION

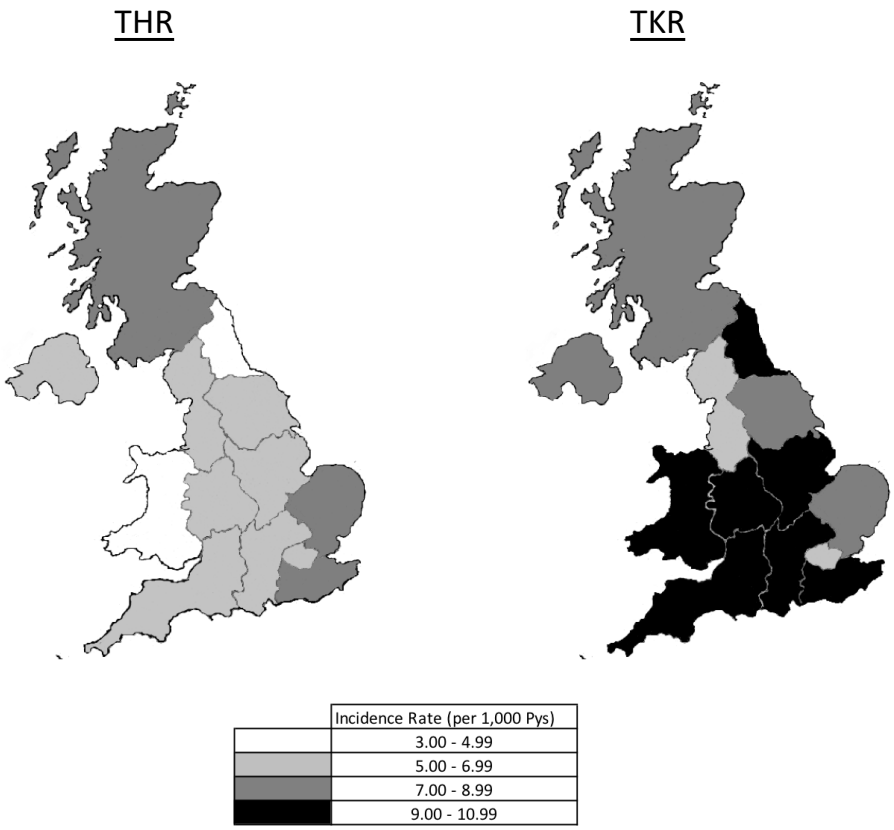


FIGURE 3.8: INCIDENCE OF THR AND TKR: STRATIFIED BY GEOGRAPHIC REGION

DISCUSSION

Outcome Validation

The Read code list used to identify THR and TKR has been developed and used in prior CPRD investigations studying THR/TKR (126, 139). The present validation exercise appears to be the first time that the accuracy of these code lists has been evaluated and addresses a limitation of previous studies using these lists given the primary care nature of CPRD and that joint replacement is carried out in secondary care, thus giving rise to

the potential of miscoding. However, the high sensitivity found for THR (80.5%) and TKR (83.6%) indicate that the majority of "true" events as recorded in HES are also recorded in CPRD, whilst the very high specificity (>98.0%) indicates the vast majority of patients THR/TKR negative as recorded in HES are also THR/TKR negative in CPRD. The PPV estimates indicate that for a given THR/TKR event identified in CPRD, there is 68.5% and 72.5% probability that these events are correctly identified (as indicated in HES) for THR and TKR, respectively. On the other hand, the NPV of 99.6 and 99.2 indicates that where an individual is identified in CPRD as being THR/TKR negative then there is >99.0% probability that this is correct (as indicated in HES). While these results demonstrate acceptable agreement, the lower values for PPV indicate there is a significant minority of procedures identified in CPRD which are not identified in HES. One probable explanation for this is that only NHS-funded procedures are required to be entered into HES. So, while patients receiving a privately funded operation are still likely to visit their GP post-operatively and therefore appear in CPRD, they may not have a corresponding HES record. Indeed, this has previously been found to be the exact reason for 9% of TKRs in NJR not appearing in HES (142).

Summary of Main Findings

This study demonstrates from a large population-based sample that there is a moderate incidence of major lower limb joint replacement among UK RA patients, and that risk develops in an approximately linear fashion over time (Figure 3.2). From RA diagnosis, the 10-year cumulative % probability of THR and TKR was approximately 5% and 7%, respectively, while at 20 years these values were 8% and 11%, respectively. Rates of

THR/TKR increased with age (except for the very elderly) but decreased with greater degrees of deprivation status and greater exposure to smoking. Women experienced higher rates of THR but not TKR than men. Those with a BMI ≥30 at time of RA diagnosis had a subsequent TKR incidence rate nearly twice as large as those with a normal weight. Regional variation was relatively minor, although there was evidence for differences in TKR rates by geography, specifically that rates were generally higher in Southerly and Midland regions while being lower in London and the North West. There were significant downward trends in THR and TKR rates according to small-areas with greater relative deprivation.

Strengths and limitations

One of the key strengths of the analysis is the large study sample with a breadth of demographic variables and long follow-up available. This allowed PY incidence rates to be estimated not only for the overall RA population but for various sub-groups of interest. The generalisability of CPRD also greatly adds to the value of the investigation, with the resource having been demonstrated to be broadly representative to the wider UK in terms of age, sex, ethnicity and BMI (123). The linkage to ONS mortality data meant that the follow-up of each patient accurately took into account when a patient died. Furthermore, the linkage to the HES database here allowed the accuracy of the defined THR/TKR Read code lists to be evaluated against in-hospital secondary care data, which showed a high degree of agreement. HES was here treated as the reference standard as the NJR has been reported to be 93% compliant with HES and the Patient Episode Database Wales (143). While entries into the NJR are very comparable with levies raised

on protheses sold, in a NJR case-study of NHS-funded TKRs performed in England for the period 2003 – 2011, 4% of the total identified in the NJR were not identified in HES and had no known reason for not being so (142). This therefore introduces a slight limitation in that while good agreement was here demonstrated between CPRD and HES, it is likely that HES itself has some (albeit minor) degree of incompleteness.

A standard approach to survival analysis is the Kaplan-Meier method, as used for example by the NJR for England, Wales and Northern Ireland. This estimates the probability of survival at each time interval of follow-up in which an event occurs, conditional on survival just before that time interval. The strength of this approach over a standard cross tabulation of outcome being that the length of follow-up time for each patient is used. However, in the Kaplan-Meier context 'survival' merely means still being active in the study and not having had the event of interest. This has been shown to overestimate the incidence of an outcome over time where there is the competing risk of an alternative event (e.g. death) that precludes the outcome of interest (144, 145). This was here addressed by generating cumulative incidence functions incorporating death as a competing risk, i.e. calculating the cumulative probability of THR/TKR conditional on being alive (not just uncensored) (145).

There are a number of other limitations. Although it has been previously demonstrated that RA can be identified with high sensitivity and specificity from CPRD (125), various definitions have been used in the past, of which the one implemented here is relatively simple. For example, the requirement to have been prescribed a DMARD was not

included in the definition. However, patients with a prior or subsequent diagnosis of a different inflammatory arthritis were excluded due to possible misdiagnosis. Added to this issue is that other researchers have reported on a likely delay between onset of RA symptoms and coding of the disease in a patient's record (146). Although this should not have a large impact on the subsequent rates of THR/TKR as reported here, it does mean that the estimates of time from diagnosis to event should not be interpreted overly precisely. Likewise, the cumulative incidence estimates over 10- and 20-year time frames should not be interpreted overly precisely either. Despite these issues, it is reassuring (as discussed below) that incidence rates are comparable with studies from elsewhere in Europe and that the 10-year % probability estimates accord with previous UK inception cohorts, both indicators of good external validation.

Approximately 30% of the sample were missing BMI data, approximately 20% missing smoking status and approximately 20% missing a value for IMD. Imputation of missing data was not carried out here as the number of variables with which to predict missing values would likely lead to unreliability, and that the impact of missing data appeared relatively minimal given the purpose of assessing univariate stratified rates. Inevitably however there is a resultant risk of bias, for example if patients from more deprived areas were systematically more likely to have missing IMD values, a phenomenon which may have some support given the distribution of IMD values among the present study sample. Even were that to be the case though, the rates of joint replacement among patients with missing data were not significantly different to other strata for any of the variables with missing data (with the exception of the 'unknown' versus 'current'

categories of smoking status). Another important caveat to the estimates are that they are all unadjusted given the aim of providing "real-life" burden of disease estimates rather than controlled for confounding factors in order to investigate causality or make comparisons to external populations. This means restraint is required so as to not make interpretations that are not strictly supported by the data given the lack of adjustment in the analysis. While such an undertaking would be a logical next step, unfortunately the data availability here is not sufficient to build a clinically meaningful multivariable model with which to predict joint replacement.

Overall incidence

This seems to be the first UK population-based study to estimate PY incidence rates following RA diagnosis. As such it provides a better understanding of long-term prognosis for RA patients, both overall and for specific sub-groups. Previous knowledge of the prevalence of RA among all THRs/TKRs performed (for example, from the NJR) is of little use to a clinician or newly diagnosed RA patient wanting a better understanding of the likely subsequent need for a major lower limb joint replacement.

UK data examining the occurrence of joint replacement in RA are emerging, which mainly focus on cumulative incidence estimates (97, 136, 137). These are discussed below although it is difficult to make direct comparisons to the present study due to differences in study design, length of follow-up and exact measurements used. The overall cumulative % probabilities for THR and TKR are consistent with an analysis by Nikiphorou *et al.* combining two UK inception cohorts of RA patients, in which the authors there

reported a 10-year cumulative incidence of major joint replacement of 11.7% [95% CI 10.4 - 13.4%] (97). Although the 10-year estimates from the present analysis for THR (5.2% [95% CI 4.9 – 5.6]) and TKR (7.0% [95% CI 6.6 – 7.4]) would total 11.8%, this should not be directly compared to the Nikiphorou paper given that some patients in the present analysis have both a THR and TKR and as such would contribute twice to such an amalgamation. A more valid comparison to the Nikiphorou estimate would be using results from an analysis of a combined THR/TKR outcome (considered together), which for CPRD at 10-years was 11.1% [95% CI: 10.7 – 11.6] (results not shown). Although this is almost identical to the 11.7% from the inception cohorts, it should be noted that they also included elbow and shoulder events in their estimate.

Continuing the comparison to previous UK data on the subject, less informative estimates have previously been generated simply expressing the proportion of RA patients at or near diagnosis who receive joint replacement during a given period of follow-up (136, 137). While these are useful, there is no accounting for the loss to follow-up or mortality that inevitably happens during what can be a long follow-up period for a given study (137). James *et al.* analysed 1,064 early RA patients, following them up for 5 years and reporting that 17% of them had at least one orthopaedic procedure and 7% had a major (including hip, knee, elbow and shoulder) joint replacement surgery (136). This would be comparable with the present CPRD data in which the cumulative % probability of a combined THR/TKR outcome (as discussed in the preceding paragraph) at 5 years was 6.6% [95% CI: 6.3 – 6.9]. Although the similarity in rates is reassuring, it should be noted that the estimate using CPRD accounts for censoring while the estimate

by James *et al*. does not, hence their estimate is likely to underestimate the actual percentage of people at baseline who would proceed to receive a joint replacement during follow-up. The fact our estimates are practically identical despite this issue may be due to their 'major joint replacement' outcome including shoulder and elbow which compensated for any underestimation. Similar to this, Gwinnutt *et al*. have recently published data from the Norfolk Arthritis Register, in which 17% of early RA patients underwent a major joint surgery over the subsequent period of 20 years (137). Specifically, of all the major joint surgeries carried out among this 17% of patients, 68 (37.7%) were hip replacements and 58 (32.2%) knee replacements. These estimates are somewhat lower than the present CPRD study, however a major factor for this is again likely to be that over half their original sample either died or were lost to follow-up over the 20 year period but their estimates do not account for this attrition.

In terms of studies from outside the UK, there is growing data on the incidence of joint replacement among RA patients (93, 101, 102, 106-108). The rates found in the present analysis accord well with those of a 2013 study utilising data from a Swedish RA register, which reports multiple incidence rates for THR and TKR for different calendar time periods, the most recent of which (2002-2006/7) were 6.6 per 1,000 PYS for THR and 6.8 per 1,000 PYs for TKR (102). Similar rates were reported by Aaltonen and colleagues for joint replacement rates among RA patients in Finland on different therapy regimens, with incidence (reported per 100 PYs) among DMARD users being 0.89 [95% CI: 0.73 – 1.08] and 1.11 [95% CI: 0.93 – 1.32] for hip and knee replacements, respectively (108). Elsewhere from Europe, a recent Spanish investigation performed by Rodriguez-

Rodriguez *et al* analysed 1,812 newly diagnosed RA patients and reported an incident rate of 9.2 [95% CI: 7.2 – 11.6] per 1,000 PYs for a combined THR/TKR outcome (101). While this is lower than the estimates from the current CPRD investigation, Rodriguez-Rodriguez only followed patients up during the first 5 years from RA diagnosis (101) which may partly explain the lower rates. Conversely, another study using data on 1,272 patients from 47 specialist units around Spain reported a rate of 2.25 per 100 PYs (i.e. 22.5 per 1,000 PYs), which is relatively high although this estimate includes all joint replacement procedures identified (not just hip and knee) (93). Canadian studies have reported similar incidence rates (considering joint replacement at any site) among large cohorts of RA patients to be 1.09 (106), 1.4 (107) and 2.0 (107) per 100 PYs.

It is interesting to note that the 10-year risk of THR and TKR has been previously reported for the UK general population using CPRD, by Culliford and colleagues (147). Although direct comparisons should be carried out cautiously given differences in study design, it is still noteworthy that Culliford's 10-year risk estimates for people aged 60 years old were 3.5% [95% CI: 2.8 – 4.1] and 2.2% [95% CI: 1.7 – 2.7] for THR in females and males, respectively. This is approximately half that of the 5.7% [95% CI: 5.2 – 6.1] and 4.2% [95% CI: 3.6 – 4.8] THR estimates for the RA patients here analysed for females and males, respectively. Likewise, Culliford's estimates for TKR in the general population were 3.1% [95% CI: 2.5 – 3.7] and 2.6% [95% CI: 2.0 – 3.1] for females and males, respectively. This is substantially lower than among the present RA sample: 7.2% [95% CI: 6.7 – 7.7] and 6.5% [95% CI: 5.8 – 7.2] for females and males, respectively. These differences are also

somewhat conservative given the fact the present analysis corrected for the over-estimation of the Kaplan-Meier method whereas Culliford's did not.

Incidence by age

The increase in THR/TKR incidence according to age is something of an expected finding given that joint destruction is progressive among patients with insufficiently managed RA, leading to the accumulation of pain and loss of function over time and an increase in need for a joint replacement. Although the incidence of osteoarthritis which increases with age (148) is also a likely factor. The recent UK analysis by Gwinnutt *et al*. found that age at RA diagnosis was predictive of undergoing a joint replacement during their study follow-up, with each year increase in baseline age associated with a 2% increase in risk of subsequent surgery (137). Likewise, in the UK inception cohort studied by James *et al*., 11.1% of patients aged >60 years had a major operation during 5-years of follow-up whereas this estimate was only 3.5% amongst patients <45 years old (136). Richter *et al*. recently reported age was significantly predictive of large joint surgery in RA patients in their study of 1,077 RA patients in Olmsted County, Minnesota (OR=1.29 [95% CI: 1.17 – 1.29]) per 10-years increase in age (149).

The decline in joint replacement rates seen here among patients over 75 is not surprising given the heightened risks of complications (150, 151) and mortality (152, 153) that are known to exist following major surgery for elderly frail patients. It may be that the surgeon and/or patient perceive that these risks outweigh the benefits of surgery, which are also known to be diminished in those who are elderly and less mobile (154, 155), for

example Gordon *et al*. found improvement in post-THR quality of life to be blunted in those approximately 70 years or older (156). It is a finding consistent with the previous CPRD study by Culliford *et al*. who reported 10-year risk of THR or TKR among the general population to increase according to age up until 80 years old, at which point 10-year risk declined (147).

Incidence by sex

Previous studies that have explored potential predictors of joint replacement among RA patients are not unanimously clear on whether patient sex plays a role. James *et al*. found a significant odds ratio (OR) of 3.2 [95% CI: 1.3 – 7.6] for female gender in relation to subsequent intermediate joint replacement (136), although this used logistic regression and did not account for censoring. When only larger joint replacements were considered, this was attenuated and was no longer significant (136). Likewise, Leon *et al*. studied 1,272 RA patients and found female gender was associated with subsequent orthopaedic surgery (OR = 1.83 [95% CI: 1.05 – 3.19]; p=0.03) but not total joint replacement ($p \geq 0.05$). Richer *et al*. also found that women had a higher rate of small joint surgery (OR=1.73 [95% CI: 1.17 – 2.99]) but not large joint surgery (OR =1.06 [95% CI: 0.79 – 1.43[) (149). Gwinnutt *et al*. reported that within their study the proportion of patients who were women was approximately the same among those who underwent subsequent joint surgery (66%) and those who did not (71%) (137).

An observation of higher THR rates among females is supported by prior findings from a large international study of over 6,000 RA patients that found females had on average higher disease activity than men and achieved remission less frequently (149). It is also in accord with previous estimates of THR in the UK general population, where rates are also generally higher among females (126). This finding is also supported by the NJR report, in that 60% of all hip replacements are carried out for females (40). However, this is most likely due to the higher prevalence rates of osteoarthritis in women than men (148), which could also explain why females in the general population also receive a higher proportion (56%) of all knee replacements, according to the NJR report (40). The incidence of TKR here was almost identical between the genders (8.54 versus 8.58 per 1,000 PYs). This raises the question as to why there is divergence in THR but not TKR rates between male and female RA patients. Although this is a subject for further research, one very tentative explanation could be that within the inflammatory processes of RA, the knee is more effected than the hip and therefore more equivalence in joint destruction and subsequent need for joint replacement is seen between the genders for the knee, whereas the hip may possibly be more influenced than the knee by secondary osteoarthritis (the prevalence of which is highest in females (148)). While this is a possibility, it should also be noted that the 10-year percentage probability of TKR was slightly higher among women (8.3% [95% CI: 7.7 – 8.9]) than men (7.8% [95% CI: 7.0 – 8.8]), although this was not statistically significant.

Incidence by BMI

The increase in TKR rates with increasing BMI (*p*-trend <0.001) is a pattern previously reported by Leyland *et al*. for osteoarthritis patients (157), where being overweight or

obese was associated with >40% and 100% increased risk of knee replacement, relative to patients with normal weight. Furthermore, Leyland *et al*. describe increasing risk of TKR according to BMI to be an approximately linear relationship, except some plateauing of risk with very high values of BMI among those >68 years old (157). Increasing BMI is strongly associated with increased risk of developing incident knee or hip osteoarthritis (158), so greater degrees of co-existing osteoarthritis with higher BMI values could at least partially explain the increase in TKR rates. However, it is also very biologically plausible that greater weight bearing would likely exacerbate the process of inflammatory-induced cartilage and bone destruction within effected knee joints of RA patients. Although the slightly higher rate of TKR among underweight patients does not have a ready explanation, the confidence intervals were wide for this estimate and it may be linked to these patients being perceived as lowest risk of adverse outcomes, for example it's previously been observed that these patients have the lowest mortality rates following TKR (159)).

The increase in TKR risk with increasing BMI was not here repeated for THR risk (*p*-trend = 0.67). Findings from Culliford *et al*. related to the UK general population similarly indicate that the prevalence of obesity among all TKRs performed is higher than that among all THRs performed, and that this discrepancy is expected to become more pronounced in the future (126). The reason for the drop-off in THR rates among obese RA patients may reflect the fact that, although it's known that high BMI is not a limiting factor for clinically meaningful improvement in terms of pain or function after either TKR (154) or THR (160), complications are known to be higher after THR surgery for patients

with higher BMI (155). For example, a >4-fold increase in dislocation rates is reported for those undergoing THR with a BMI ≥35 kg/m^2 relative to <25 kg/m^2 (161). While increased post-operative mortality among obese THR inpatients may also possibly be a concern for surgeons (given this heightened risk of complications), the literature actually suggests something of an 'obesity paradox' where patients with higher BMI have a slightly lower mortality rate than those with normal BMI (162). The recent study by Richter *et al*. reported that obesity (BMI ≥30 kg/m^2) was associated with a large increase in need for large joint surgery amongst RA patients (OR=2.03 [95% CI: 1.53 – 2.70]), although they did not differentiate between hip and knee surgery (nor replacement versus non-replacement) (149).

Incidence by smoking

The stratified incidence estimates according to smoking status may initially appear counter-intuitive. Smoking is a risk factor for developing RA (6) and therefore could rightly be expected to contribute to increased disease activity and more joint destruction. However, it is also a strong predictor of both complications and mortality following joint replacement (163) (e.g. OR for death = 3.08 [95% CI: 2.21 – 4.29] (159)). These known risks most likely influence surgeon preference for not operating on RA patients who are current smokers, therefore leading to less provision for these patients. Smoking has previously been found to be independent of the need for large joint surgery in RA (149).

Incidence by deprivation

The finding of significant linear trends of decreasing THR and TKR rates according to lower IMD ranking is an interesting finding. Similar trends have been reported for THR/TKR rates in the general UK population (164). Furthermore, Judge *et al*. reported a decline in equity (provision of THR/TKR relative to need) among the general population for those residing in areas of the UK with greater deprivation, with 70% less provision relative to need among people living in the most deprived areas compared with least deprived areas (165).

There is existing evidence that UK deprivation generally predicts worse outcome of TKR (154) and that better outcomes have been observed following osteoarthritis-induced THR among those with higher socio-economic status (155). While surgeon preference to operate could potentially be influenced by such a perceived association between deprivation and worse outcome, it's previously been suggested that a more likely explanation is that those from more deprived areas are more likely to tolerate greater degrees of ill health and are therefore less likely to demonstrate healthcare-seeking behaviour (154). However, it could also be the case that despite visiting their GPs, individuals residing in more deprived areas may still be less likely to receive surgery due to substandard provision (166), despite their willingness to consider arthroplasty (167). While further research is needed to disentangle these issues, given that these are univariate associations in the current study, it cannot be ruled out that IMD is here acting more as a proxy for other confounding characteristics, either at the patient-level or small-area-level that may be the primary drivers of the trend.

Incidence by geographic region

The differences in actual numbers of RA patients diagnosed in each geographic region approximately reflects the distribution of GP practices contributing data to CPRD (123). For example, within the study sample there is a relatively small proportion of RA patients residing in the North East (2%), however this has the lowest density of GP practices within CPRD, at only 12-19 in total (123). Whereas the North West had 12% of the study sample, however this reflects the fact it has between 80-89 contributing practices (123). The lack of significant variation in THR across regions is reassuring, however the regional variation in TKR rates is concerning and is possibly a reflection of previous findings highlighting a disparity between need and provision within the UK (165), which is greater for knee replacement than hip (165, 168, 169). Specifically, although somewhat dated data now, in the late 1990s the total deficit of hip replacement surgery in England (compared to estimated need) was 2,500 per year (169), whereas similar estimates for knee replacement showed an annual deficit that was approximately 10-fold larger (26,500) (168). It must be noted though that the current study does not demonstrate that the regional variation in TKR rates among the RA patients included is necessarily due to unmet need, for example it could be the product of regional variation in management of RA and/or the consequence of other confounding factors, e.g. BMI. Although considering the previous data on unmet need of TKR within the UK, it would seem these regional variations for TKR (at the most extreme: 6.6 versus 10.3 per 1,000 PYs) are of interest and warrant further investigation in future studies.

Conclusions

This chapter reports on the good validity of THR and TKR Read coding in CPRD as compared to HES. It describes the overall incidence rates of THR (6.38 per 1,000 PYs) and TKR (8.57 per 1,000 PYs) among a large UK cohort of newly diagnosed RA patients over a maximum of 20 years, which accords well with estimates from elsewhere in Europe. An informal comparison of 10-year risk from RA patients in the present study to previous estimates for the general population suggest a greater need for THR and TKR in RA. Average time-to-event, from RA diagnosis, was 3.2 [IQR: 1.3, 6.5] and 3.5 [IQR: 1.5, 6.7] years for THR and TKR, respectively. Stratum specific rates indicate significant differences and trends in rates of THR and TKR exist within a number of key demographic variables. Further research is needed to further explore some of these associations, with variation according to deprivation and geographic region (for TKR) a particularly interesting topic. These findings on rates of major lower limb joint replacement offer clinicians, patients and commissioners a better understanding of the long-term prognosis of RA in the UK.

4. POPULATION-LEVEL IMPACT OF THE INTRODUCTION OF BIOLOGICS ON SUBSEQUENT RATES OF THR AND TKR WITHIN ENGLAND AND WALES

INTRODUCTION

As discussed in chapter 1, there has been a growing body of evidence of changing temporal trends of RA-related joint replacement over recent years. Overall, these findings tend to suggest a reduction of joint replacements for RA patients over latter years, although not all such studies report a decline and various degrees of granularity exist in terms of which joints were included in analyses.

Some of these prior studies have inferred some role of biologic therapies in this reduction in joint replacement numbers/rates amongst RA patients (91, 93, 102, 135), although few have formally tested such a hypothesis. Most report descriptive statistics concerning the overall change in numbers/rates over time (or even before vs. after the introduction of biologic therapies), however this has several limitations in relation to suggesting a role for biologic therapy. Firstly, gradual changes in temporal trends of joint replacement are likely to be the product of various potential factors of which biologics is just one. This means a general description of a decreasing secular trend, or a comparison of rates at the beginning versus the end of a long study period fails to provide a valid quantified estimate of the effect attributable to the introduction of biologics per se.

Secondly (although related to the first reason), where some studies have provided an estimate of the difference in rates before versus after the introduction of biologics (102), this has been in the form of the difference in some average estimate over several years leading up to the introduction of biologics compared to some average estimate over the biologic era. While these estimates are interesting and informative, they fail to account for changes in secular trends during the study period that may well be quite pronounced and are likely to confound any comparison of averages before vs. after the introduction of biologics. A simple illustration of this fact could be where a constant linear decrease in joint replacement rates is hypothetically in place over many years prior to biologics being developed, which continues unchanged for many years into the biologic era. Here, a simple comparison of average pre- versus post-rates would indicate rates were indeed lower in the biologic era, potentially with a considerable magnitude. This however would be a biased estimate of the 'causal effect' of biologics, such an analysis is greatly confounded by a pre-existent secular trend.

Furthermore, differences in study design is an important consideration when interpreting these data. For example, the reported decrease in the proportion of all joint replacements that have RA as a primary indication may be driven by a decreasing incidence of RA (or equally an increasing incidence of osteoarthritis (OA)) rather than particularly demonstrating improved therapeutic management in RA has caused a reduction in joint destruction. On the other hand, a decrease in the incidence of joint replacement amongst RA patients is more supportive of an actual reduction in need of these procedures for RA patients over recent decades.

In this context of increasing availability of routinely collected health-care data on joint replacement rates amongst RA patients, coupled with the real limitations of traditional observational studies and the likely infeasibility of an RCT to determine the impact of biologics, it is worth considering the strengths of quasi-experimental alternatives. As introduced in chapter 1, quasi-experimental study designs implement strong methods that facilitate causal inference regarding an exposure (170, 171). They take advantage of the possibility to choose who and/or when to measure as opposed to in an RCT where one chooses who and/or when to expose (119, 172). Interrupted time-series is such a study design. This method is being increasingly used in epidemiology to evaluate the population-level impact of health policy changes (170, 173, 174) or other naturally occurring phenomena of interest (175, 176). It essentially entails modelling an outcome measure repeatedly over time – both before and after some well-defined intervention – and thereby providing a quantified estimate of the intervention impact on the outcome whilst controlling for general secular trends. It is also a flexible method given it is possible to make various extensions to a basic study design in order to better control for possible confounding factors and strengthen the case for causal inference (119, 170, 171).

In this framework, the introduction of biologic therapies into routine clinical practice in the UK is something of a "naturally" occurring event. This formally occurred for England and Wales with the first national institute for health and care excellence (NICE) approval for TNFi in March 2002, entitled "*Guidance on the use of etanercept and infliximab for the treatment of rheumatoid arthritis*". This was in the form of the publication

"technological appraisal 36" (TA36). Although these biologic therapies recommended in this official guidance had been licensed for RA in 2000, no NHS rheumatologist was obliged to prescribe them, and indeed it was initially a selective "postcode lottery" given they were still considered experimental and prohibitively expensive (69). On the other hand, NICE technological appraisals function to "*ensure that all NHS patients have equitable access to the most clinically – and cost-effective treatments that are available*" (177). Treatments recommended within a technological appraisal should henceforth be available throughout the NHS to any patient as recommended within 3 months of such guidance being published (177).

The NICE TA36 publication provided guidance to NHS clinicians on use of etanercept and infliximab for rheumatoid arthritis patients. It recommended that these therapies should be used as per the guidance issued by the British Society for Rheumatology in April 2001 (69). This stated that these therapies were recommended options (infliximab only to be used concurrently with methotrexate) for adults with persistent active RA as indicated by two DAS28 measurements of >5.1 at least 1 month apart, who had already failed to adequately respond to two csDMARD therapy regimens, each lasting at least 6 months (69).

The aim therefore of the work as reported in this chapter was to estimate the impact of the introduction of biologics, as defined by publication of NICE TA36, on population-level temporal trends of THR and TKR amongst early RA patients in England and Wales.

METHODS

Data and participants

Primary care health data were obtained from the UK Clinical Practice Research Datalink (CPRD) for the period April 1995 to September 2014. These are described in chapter 2, in addition to details of collaborator contributions in pre-analytical data acquisition and processing. In brief, incident RA patients were identified using a pre-defined READ code list, with the date of first recorded RA considered as diagnosis date. Data on gender, age, BMI, smoking status and Charlson comorbidity score were extracted from CPRD at date of RA diagnosis. Patients diagnosed with RA before the study period were excluded as were those with either a prior or subsequent diagnosis of a different inflammatory arthritis (lupus, ankylosing spondylitis, psoriatic arthritis or crystal arthropathy) due to possible diagnosis or coding errors. Patients aged <18 years old were also excluded, as were patients registered to a GP practice outside England and Wales (i.e. patients registered in Scotland or Northern Ireland). The restriction of the sample to only England and Wales was because the current study question concerned evaluating the impact of NICE guidance, compliance to which is only mandatory for these countries.

Intervention

The intervention of interest was the first NICE approval for biologics, defined as the publication of TA36 in March 2002. NICE is a non-departmental public body which aims to reduce variation in the availability and quality of treatments and care within the NHS. It is accountable to the government department of health and social care. NICE technological appraisals are developed following a detailed and independent review of

both the clinical and economic effectiveness of a given therapy or group of therapies, the process of which is outlined on NICEs website (177). A brief overview of this process is as follows. NICE identifies a broad range of consultees prior to appraisal, including the company that produced the technology, patient representatives, the department of health and clinical commissioning groups. Commentators are also appointed, such as relevant academic research groups and companies of comparator technologies. These consultees and commentators have the opportunity to input evidence and comments at various points during the appraisal process. NICE drafts the scope of the appraisal and commissions an independent academic centre to review the evidence and prepare a report. An independent advisory committee (that meets in public) then reviews and discusses all submitted evidence and prepares a document containing provisional recommendations. After further rounds of consultation and possible appeals is implemented before final recommendations are documented and the guidance is formally issued.

Outcome

Outcome of interest was first occurrence of THR or TKR following RA diagnosis. These were identified using CPRD Read codes as used previously in the published literature (126, 139) and as used and validated for analyses reported in chapter 3. THR and TKR were considered separately so patients could potentially have both outcomes of interest.

Statistical analysis

Descriptive statistics were generated per calendar year throughout the study period. The outcome time-series was derived by calculating 5-year incidence rates of THR and TKR among newly diagnosed RA patients within each six months between 1995-2009. For example, in the first six months (Apr-Sep 1995) all incident RA diagnoses were identified and the 5-year incidence rate of THA for those RA patients was calculated by dividing the number of THAs occurring within 5 years from RA diagnosis by the total number of person-years contributed before any censoring event (death, loss to follow-up, last GP data upload, event of interest or end of 5 years follow-up) (appendix figure 4.1). This was then repeated for each 6-monthly period in the study (table 4.2 & 4.3). Incidence rates were only based on the first five years after diagnosis so that the composition of patient follow-up was approximately constant throughout the study period to allow valid rate comparisons irrespective of when a patient was diagnosed. Otherwise, patients diagnosed at the start of the study period would have much longer follow-up than those diagnosed at the end, which would be a likely source of bias if degree of exposure to biologics or the risk of joint replacement was different over different lengths of follow-up from diagnosis. Incidence rates from the last five years of the study period (i.e. the last 10 biannual proportions) were not included in the final analysis to account for the fact these patients did not get a full five years follow-up (which may otherwise artificially attenuate rates in the latter study years). The 5-year time window was chosen as this was approximately the median follow-up, and because this captured the "early" years of disease in which pro-active management of the disease is very much recommended. It also meant the resultant time-series had an approximately equal number of biannual timepoints before and after the intervention. In order to minimise confounding resulting

from changes to the age and sex structure of the RA population over the study period, the 5-year incidence rates measured at each six months were age and sex standardised using the entire RA sample (1995 – 2009) as the reference population. This entailed calculating the 5-year incidence rates within each of the specific sex-age strata (approximate tertiles were used for age categories [<55, 55-69, ≥70]) at each timepoint then calculating a weighted average according to the age-sex structure of the overall study sample (140).

A segmented linear regression was performed on the aggregated standardised time-series to estimate two parameters of interest associated with the publication of NICE TA 36: change in subsequent level of outcome and change in subsequent trend (171). Given that the NICE guidance recommended that only patients having a failed response to two (6 month) trials of csDMARD therapy should initiate a TNFi, a 1-year lag period following March 2002 was decided upon a-priori in which data for the outcome was removed from the time-series. The regression model was specified as following: $Y_t = \beta_0 + \beta_1 * \text{time}_t + \beta_2 * \text{intervention}_t + \beta_3 * \text{post\_intervention\_time}_t + e_t$. Here, Y_t is the 5-year incidence among RA patients diagnosed at time point (i.e. 6-monthly period) t. β_0 estimates the baseline level of the outcome just before the beginning of the time series. β_1 estimates the pre-intervention trend, β_2 the change in level between the time point immediately before vs. after the lag period and β_3 the change in trend occurring immediately after the lag period. Analyses were based on 14 pre-intervention data points (Apr 1995 – Mar 2002) and 13 post-intervention data points (Apr 2003 – Sept 2009). Final model specification was derived using a backward-stepwise approach (p-entry 0.049; p-exit

0.20) to remove non-significant regression terms in order to maximise statistical power, although results from full models were also reported. The Durbin-Watson statistics was generated following final model fitting and compared to significance tables (178) to inspect the assumption of no significant autocorrelation. Here the test statistics is assessed in relation to a lower and upper bound of the critical value for the specific combination of timepoints and regressors – with significant auto-correlation being rejected if the test statistic falls above the upper bound. The null hypothesis of no auto-correlation is not rejected if the statistics falls below the lower bound, and is deemed inconclusive if falling between the lower and upper bound. The normality of residuals was assessed using a Q-Q plot of quantiles of the distribution of model residuals against those from a normal distribution and homoskedasticity by conducting the Breusch-Pagan / Cook-Weisberg test. Linearity of effect was visually assessed by inspecting plots of the time-series data and from plots of residual versus fitted values.

Given that the intervention may impact either or both of: (i) level of outcome, (ii) trend of outcome, it is arguably easier to interpret an overall measure of impact that combines both of these changes. Such an overall estimate of the intervention effect was calculated using regression coefficients to predict what would have been observed post-intervention (i.e. counterfactual rates) had pre-intervention levels and/or trends continued uninterrupted, and comparing this with what was modelled using observed post-intervention data (171, 179). The midpoint of the post-intervention period was used to calculate the average difference over the post-intervention time points. Confidence intervals for this average absolute difference in outcome over the post-intervention time

points were generated using the standard error of the level and/or trend change coefficients, according to methods described by Zhang *et al.* (179, 180). The point estimate was also expressed as a relative change (compared to the midpoint of the counterfactual rates).

Sensitivity analyses

Due to likely delay in implementation of NICE recommendations and possible heterogeneity in speed of uptake across the country, a 2-year lag period following publication of NICE TA36 was used in a sensitivity analysis. A data-driven (with no pre-specified time point/intervention) approach was also conducted in order to identify where – if at all - any changes in linear trend occurred using the Joinpoint Regression Program (version 4.3.1.0, National Cancer Institute) (181). An uncorrelated errors model was specified given that no significant autocorrelation was previously detected. The grid search method was used with a minimum number of joinpoints to be identified set at zero and maximum number of joinpoints to be identified set at one in order to determine whether the trend change associated with the introduction of TNFi (as per the main ITS approach) would be identified. The required setting for the minimum number of observations before and after each joinpoint was set at eight. Although the specification of eight observations constrained the model to only focus on the approximate middle half of the timeseries, eight observations pre- and post-intervention has previously been recommended as a minimum for conducting an ITS (182). Model selection was carried out using permutation tests (181) with a significance level set to 0.05 with Bonferroni correction for any multiple testing. The permutation method involves testing for

significant inflexions (i.e. joinpoints) at every possible location in the time-series dataset. Here the null hypothesis was that of the minimum number of joinpoints (i.e. zero) and the alternative hypothesis was that of the maximum number of joinpoints (i.e. one). Monte Carlo simulation was used (n=5,000) to resample the original data for the 'null hypothesis' model in order to test whether adding a joinpoint statistically improved the model fit.

The reason for using Joinpoint regression as a sensitivity analysis was to test whether any impact detected in the segmented regression (primary analysis) was still detectable in a purely data driven approach which was hypothesis-free as to the location of any inflexion. Whilst a powerful method, Joinpoint still relies on several assumptions being made a-priori - e.g. concerning the number of joinpoints and number of timepoints before and after any joinpoint - and can only detect changes in slope but not abrupt level changes. For these reasons, and given I primarily wanted to evaluate the impact of NICE TA36 (rather than merely detect all possible changes in trend of outcome) I used segmented linear regression as the primary analysis as this is frequently used in the literature to evaluate the impact of well-defined interventions (170, 171, 174), the strengths of which include the accounting for secular trends and the ability to estimate both step and slope changes in outcome.

RESULTS

There was a total of 17,505 incident RA patients diagnosed in CPRD between 1995-2009 as per inclusion/exclusion criteria (figure 4.1). Patient characteristics are presented in table 4.1. Mean age at RA diagnosis increased slightly from 58.7 in 1995 to 60.3 in 1999 (p=0.065), whilst the gender ratio remained fairly stable over the same time frame (70.4% to 66.3% female; p=0.12). Prevalence of obesity (BMI ≥30), as reported at time of RA diagnosis, increased from 11.5% in 1995 to 22.1% in 2009 (p=<0.001), whilst concurrent smoking status decreased from 27.9% to 20.9% (p=0.011). Overall there were 465 THRs and 650 TKRs occurring within 5-years of RA diagnosis (figure 4.2), yielding a crude incidence rate of 6.16/1,000 person years (PYs) [95% CI: 5.63 to 6.75] and 8.65/1,000 PYs [95% CI: 8.01 to 9.34], respectively.

Patient characteristics

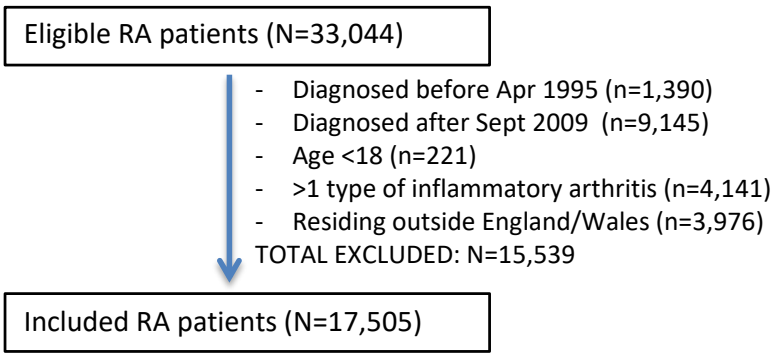


FIGURE 4.1: POPULATION FLOW CHART OF EXCLUDED/INCLUDED PATIENTS

TABLE 4.1: PATIENT CHARACTERISTICS AT TIME OF RA DIAGNOSIS: STRATIFIED BY YEAR

Year ^	Index RA diagnoses	age (mean, S.D.)	sex, female (%)	current smoking (%)	Obesity (%)	Charlson Comorbidity Score (%)			
						None (0)	Mild (1-2)	Moerate (3-4)	Severe (≥5)
1995	601	58.7 (14.7)	423 (70.4)	105 (27.9)	69 (11.5)	449 (74.7)	104 (17.3)	37 (6.16)	11 (1.8)
1996	661	60.3 (14.9)	486 (73.5)	121 (30.7)	63 (9.5)	492 (74.4)	103 (15.6)	59 (8.9)	7 (1.1)
1997	810	60.5 (14.9)	556 (68.6)	114 (25.3)	77 (9.5)	640 (79.0)	100 (12.4)	58 (7.2)	12 (1.5)
1998	772	60.7 (15.5)	554 (71.8)	124 (29.2)	73 (9.5)	592 (76.7)	117 (15.2)	52 (6.7)	11 (1.4)
1999	969	60.6 (14.9)	685 (70.7)	158 (30.1)	105 (10.8)	762 (78.6)	131 (13.5)	62 (6.4)	14 (1.4)
2000	1,160	60.8 (14.9)	820 (70.7)	162 (29.4)	111 (9.6)	961 (82.8)	125 (10.8)	60 (5.2)	14 1.2)
2001	1,491	61.2 (14.9)	1,072 (71.9)	203 (29.0)	161 (10.8)	1,225 (82.2)	183 (12.3)	63 (4.2)	20 (1.3)
2002	1,524	60.8 (14.7)	1,100 (72.2)	254 (31.1)	179 (11.8)	1,242 (81.5)	170 (11.2)	99 (6.5)	13 (0.9)
2003	1,576	61.3 (15.1)	1,088 (69.0)	290 (28.5)	200 (12.7)	1,234 (78.3)	190 (12.1)	123 (7.8)	29 (1.8)
2004	1,634	61.5 (14.9)	1,189 (72.8)	331 (25.9)	257 (15.7)	1,283 (78.5)	189 (11.6)	122 (7.5)	40 (2.5)
2005	1,478	61.0 (15.2)	1,060 (71.7)	331 (25.9)	259 (17.5)	1,108 (75.0)	180 (12.2)	142 (9.6)	48 (3.3)
2006	1,432	60.3 (15.4)	1,047 (73.1)	305 (26.3)	284 (19.8)	1,054 (73.6)	148 (10.3)	171 (12.0)	59 (4.1)
2007	1,369	59.9 (14.8)	968 (70.7)	294 (22.9)	314 (22.9)	959 (70.1)	134 (9.8)	209 (15.3)	67 (4.9)
2008	1,367	59.4 (15.0)	927 (67.8)	316 (24.2)	326 (23.9)	978 (71.5)	153 (11.2)	163 (11.9)	73 (5.3)
2009*	661	60.3 (15.7)	438 (66.3)	136 (20.9)	146 (22.1)	477 (72.2)	63 (9.5)	90 (13.6)	31 (4.7)
Overall	17,505	60.6 (15.0)	12,412 (70.9)	3,244 (26.3%)	4,264 (17.9)	13,456 (76.9)	2,090 (11.9)	1,510 (8.6)	449 (2.6)

^ Apr to Mar

* Only data Apr to Sept available/included

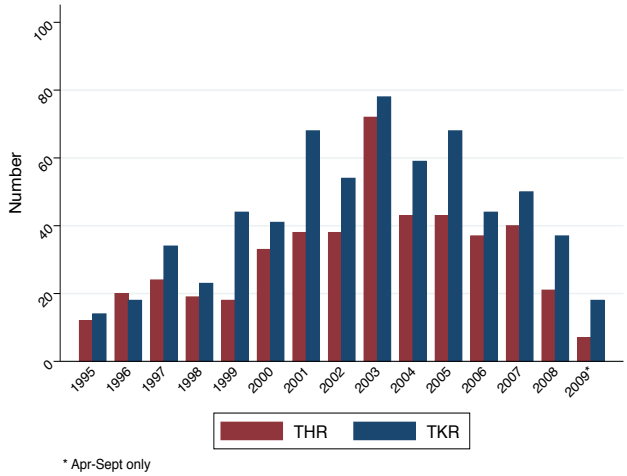


FIGURE 4.2: NUMBER OF RA PATIENTS UNDERGOING THR OR TKR SURGERY WITHIN 5 YEARS OF RA DIAGNOSIS: STRATIFIED BY YEAR OF DIAGNOSIS (APR-MAR)

TABLE 4.2: THR TIME-SERIES: 5-YEAR INCIDENCE RATE

timepoint	year	No. RA diagnoses	No. with subsequent THR (5-year)	Person-years of follow-up	Crude 5-year cumulative incidence of THR	Standardised 5-year cumulative incidence of THR	5-year cumulative incidence of THR lower 95% C.I.	5-year cumulative incidence of THR upper 95% C.I.
1	1995	314	6	1,419	4.23	4.49	0.85	8.13
2	1995	287	6	1,319	4.55	4.70	0.90	8.51
3	1996	331	9	1,487	6.05	5.71	1.96	9.46
4	1996	330	11	1,472	7.47	7.21	2.94	11.48
5	1997	355	11	1,582	6.95	6.76	2.76	10.76
6	1997	455	13	2,023	6.43	6.53	2.99	10.07
7	1998	374	10	1,676	5.97	5.22	1.93	8.51
8	1998	398	9	1,754	5.13	5.17	1.78	8.56
9	1999	418	8	1,831	4.37	4.45	1.37	7.53
10	1999	551	10	2,430	4.12	3.99	1.50	6.48
11	2000	498	14	2,151	6.51	6.42	3.04	9.80
12	2000	662	19	2,898	6.56	6.72	3.70	9.74
13	2001	711	15	3,133	4.79	4.64	2.29	6.99
14	2001	780	23	3,411	6.74	6.75	4.00	9.50
15	2002	750	25	3,373	7.41	7.47	4.55	10.39
16	2002	774	13	3,426	3.80	3.78	1.73	5.83
17	2003	753	31	3,324	9.33	9.28	6.01	12.54
18	2003	823	41	3,629	11.30	11.60	8.08	15.12
19	2004	798	25	3,460	7.23	7.03	4.28	9.78
20	2004	836	18	3,626	4.96	4.86	2.61	7.10
21	2005	784	20	3,477	5.75	5.66	3.18	8.13
22	2005	694	23	3,033	7.58	7.67	4.55	10.80
23	2006	772	14	3,371	4.15	4.00	1.90	6.09
24	2006	660	23	2,817	8.17	8.28	4.91	11.64
25	2007	720	19	3,177	5.98	6.41	3.53	9.28
26	2007	649	21	2,756	7.62	7.67	4.40	10.93
27	2008	670	15	2,828	5.30	5.69	2.77	8.61
28	2008	697	6	3,052	1.97	1.94	0.37	3.51
29	2009	661	7	2,823	2.48	2.38	0.61	4.15

TABLE 4.3: TKR TIME-SERIES: 5-YEAR INCIDENCE RATE OUTCOME

timepoint	year	No. RA diagnoses	No. with subsequent TKR (5-year)	Person-years of follow-up	Crude 5-year cumulative incidence of TKR	Standardised 5-year cumulative incidence of TKR	Standardised 5-year cumulative incidence of TKR lower 95% C.I.	Standardised 5-year cumulative incidence of TKR upper 95% C.I.
1	1995	314	9	1,407	6.40	6.82	2.33	11.30
2	1995	287	5	1,323	3.78	4.97	0.62	9.33
3	1996	331	12	1,488	8.07	8.05	3.48	12.63
4	1996	330	6	1,482	4.05	4.26	0.84	7.69
5	1997	355	15	1,595	9.40	9.48	4.67	14.29
6	1997	455	19	2,010	9.45	9.92	5.48	14.35
7	1998	374	10	1,668	6.00	5.25	1.96	8.54
8	1998	398	13	1,759	7.39	7.61	3.48	11.74
9	1999	418	20	1,819	11.00	10.97	6.19	15.75
10	1999	551	24	2,395	10.02	10.28	6.14	14.43
11	2000	498	20	2,151	9.30	9.06	5.10	13.03
12	2000	662	21	2,892	7.26	7.76	4.46	11.06
13	2001	711	38	3,091	12.29	12.12	8.28	15.96
14	2001	780	30	3,414	8.79	8.95	5.76	12.13
15	2002	750	25	3,382	7.39	7.30	4.44	10.16
16	2002	774	29	3,409	8.51	8.66	5.49	11.83
17	2003	753	31	3,326	9.32	8.88	5.75	12.00
18	2003	823	47	3,609	13.02	13.18	9.44	16.92
19	2004	798	23	3,458	6.65	6.37	3.77	8.96
20	2004	836	36	3,596	10.01	9.72	6.55	12.89
21	2005	784	44	3,435	12.81	12.46	8.80	16.11
22	2005	694	24	3,022	7.94	7.99	4.81	11.17
23	2006	772	26	3,359	7.74	7.57	4.66	10.47
24	2006	660	18	2,817	6.39	6.33	3.41	9.25
25	2007	720	28	3,164	8.85	9.03	5.69	12.37
26	2007	649	22	2,754	7.99	7.89	4.60	11.19
27	2008	670	15	2,831	5.30	5.07	2.49	7.64
28	2008	697	22	3,016	7.29	6.98	4.03	9.92
29	2009	661	18	2,808	6.41	6.46	3.47	9.46

Time-series diagnostics

THR and TKR time-series rates are presented in tables 4.2 & 4.3. Four assumptions of the linear regression model were checked: linearity of relationship, normality of residuals, homoskedasticity and no auto-correlation. Time-series plots and plots of residual versus fitted values (appendix figure 4.2) indicated both THR and TKR time-series followed a linear relationship. QQ and kernel density plots of regression residuals indicated residuals were approximately normally distributed, although this distribution was 'flatter' than expected (appendix figure 4.2). The Breusch-Pagan / Cook-Weisberg test supported the null hypothesis of constant variance (p=0.38 and p=0.26 For THR and TKR

time-series, respectively). Durban-Watson statistics were 1.69 and 2.75 for THR and TKR time-series, respectively. These were both over the lower (1.24) and upper (1.56) bounds of the critical value for significance as per Durbin-Watson significance tables (178), indicating (at the 5% level) that no significant first-order auto-correlation was present.

THR time-series

Age and sex standardised 5-year incidence of THR at the start of the study period was 5.63/1,000 PYs [95% CI: 4.74 to 6.54], which remained unchanged during the pre-TNFi period (Apr 1995 – Mar 2002) (table 4.4; figure 4.3A). Immediately following the intervention lag period there was a level increase in the incidence rate by 3.97/1,000 PYs [95% CI: 1.79 to 6.14] but a subsequent downward trend of -0.47/1,000 PYs [95% CI: -0.71 to 0.22] per 6 months for the remainder of the post-NICE TA36 time period. Based on these coefficients, the estimated incidence of THR at the mid-point of the post-NICE TA36 time period was 6.57/1,000 PYs (table 4.5). The incidence at the same time point estimated solely by extrapolating the pre-NICE TA36 level and trend was 5.63/1,000 PYs, therefore translating to no significant average change in rates (0.95/1,000 PYs [95% CI: -2.66 to 4.56]). Results were unchanged in full models containing all time-series terms (i.e. including estimation of a pre-intervention trend coefficient [appendix table 4.3]).

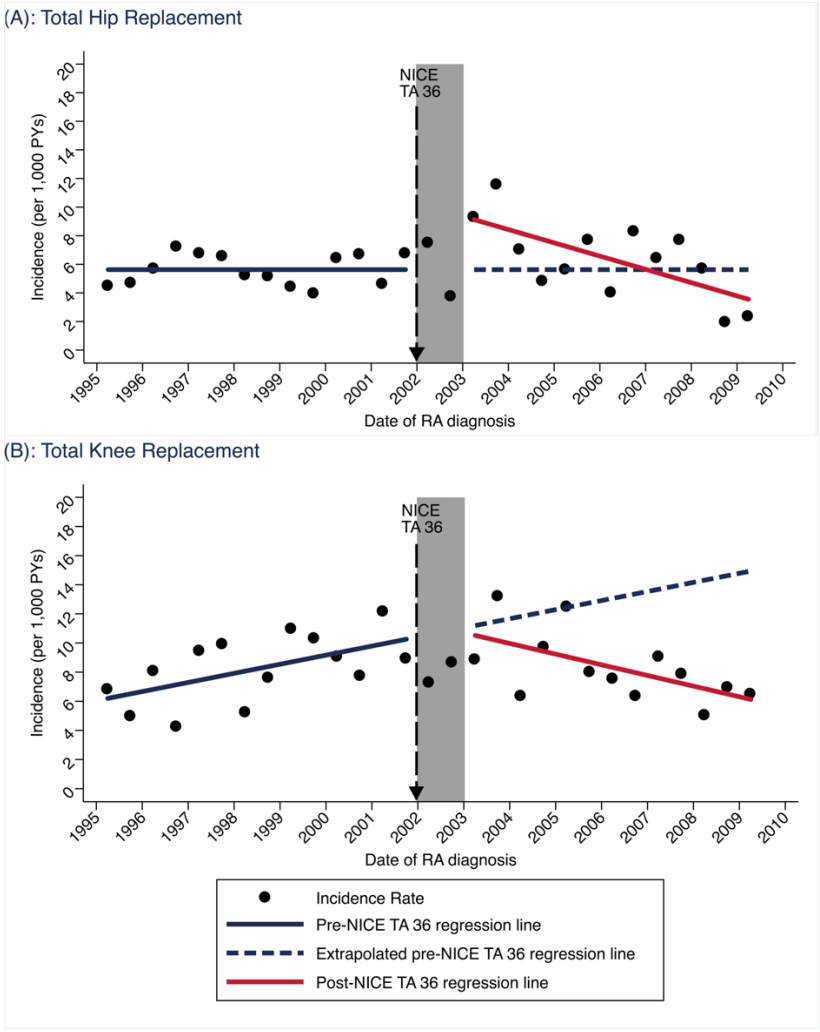


FIGURE 4.3: AGE AND SEX STANDARDISED INCIDENCE RATES OF JOINT REPLACEMENT WITHIN 5 YEARS OF RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS) WITH 1-YEAR LAG POST-NICE TA 36: (A) THR AND (B) TKR

TABLE 4.4: TEMPORAL TRENDS IN 5-YEAR THR AND TKR INCIDENCE RATES AMONG INCIDENT RHEUMATOID ARTHRITIS PATIENTS DIAGNOSED FROM 1995 TO 2009 (PER 1,000 PERSON-YEARS)

Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
Total Hip Replacement				
Intercept	5.63	4.73	6.52	<0.001
Trend [1]	-	-	-	-
Level change after NICE TA 36	3.97	1.79	6.14	0.001
Trend change after NICE TA 36	-0.47	-0.71	-0.22	0.001
Total Knee Replacement				
Apr-Oct 1995 Incidence Rate	5.89	3.83	7.94	<0.001
Trend [1]	0.31	0.11	0.51	0.004
Level change after NICE TA 36	-	-	-	-
Trend change after NICE TA 36	-0.68	-1.08	-0.28	0.002

- = P≥0.2

[1] per 6 months

TABLE 4.5: ESTIMATED IMPACT OF NICE TA36 IN TERMS OF INCIDENCE RATES (PER 1,000 PERSON YEARS) OF THR AND TKR

	Post-NICE TA 36 midpoint estimate (June 2006)		Absolute Difference*		
	Without intervention	With intervention	estimate	lower 95% CI	upper 95% CI
THR	5.63	6.57	0.95	-2.66	4.56
TKR	12.92	8.51	-4.41	-6.88	-1.94

* calculated by comparing estimated values for June 2006 (midpoint of post-intervention period) to counterfactual values for the same timepoint (i.e. those estimated/expected for June 2006 but extrapolated from the pre-intervention level and trend). Estimated from parsimonious models specified using backward-stepwise selection (p-entry 0.049 and p-exit 0.20).

TKR time-series

The incidence of TKR was 5.89/1,000 PYs [95% CI: 3.83 to 7.94] at the start of the study period, which increased by 0.31/1,000 PYs [95% CI: 3.83 to 7.94] per 6 months during the pre-intervention period (table 4.4; figure 4.3B). Immediately following the 1-year lag period there was a significant downward change in the prior upward trend by -0.68/1,000 PYs [95% CI: -1.08 to -0.28] per 6 months. Based on these coefficients the modeled incidence for the midpoint of the post-intervention period was 8.51/1,000 PYs,

which was significantly 4.41/1,000 PYs [95% C.I. 6.88 to 1.98] lower compared to that estimated, had pre-intervention trends continued uninterrupted (table 4.5). This equated to an approximate relative 34% reduction. Results were unchanged in full models containing all time-series terms (i.e. including estimation of a post-intervention step-change coefficient [appendix table 4.3]).

Sensitivity analysis

In sensitivity analyses using a 2-year lag, THR rates remained flat during the study period while results for TKR remained unchanged from the main analysis (table 4.6, figure 4.4). Conversely, Joinpoint analysis identified significant inflexions in upward trends in the incidence of both THR and TKR, at the time points spanning Oct 2005 – Mar 2006 (p=0.034) and Apr 2001 – Sep 2001 (p=0.036), respectively (figure 4.5).

TABLE 4.6: TEMPORAL TRENDS IN 5-YEAR INCIDENCE RATES OF THR AND TKR AMONG 17,505 INCIDENT RHEUMATOID ARTHRITIS PATIENTS DIAGNOSED FROM 1995 TO 2009 (PER 1,000 PERSON-YEARS): SENSITIVITY ANALYSIS USING 2-YEAR LAG

Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
Total Hip Replacement				
Intercept	5.62	4.96	6.27	<0.001
Trend [1]	-	-	-	-
Level change after NICE TA 36	-	-	-	-
Trend change after NICE TA 36	-	-	-	-
Total Knee Replacement				
Apr-Oct 1995 Incidence Rate	6.43	4.44	8.41	<0.001
Trend [1]	0.22	0.03	0.41	0.024
Level change after NICE TA 36	-	-	-	-
Trend change after NICE TA 36	-0.63	-1.09	-0.16	0.011

- = P≥0.2

[1] per 6 months

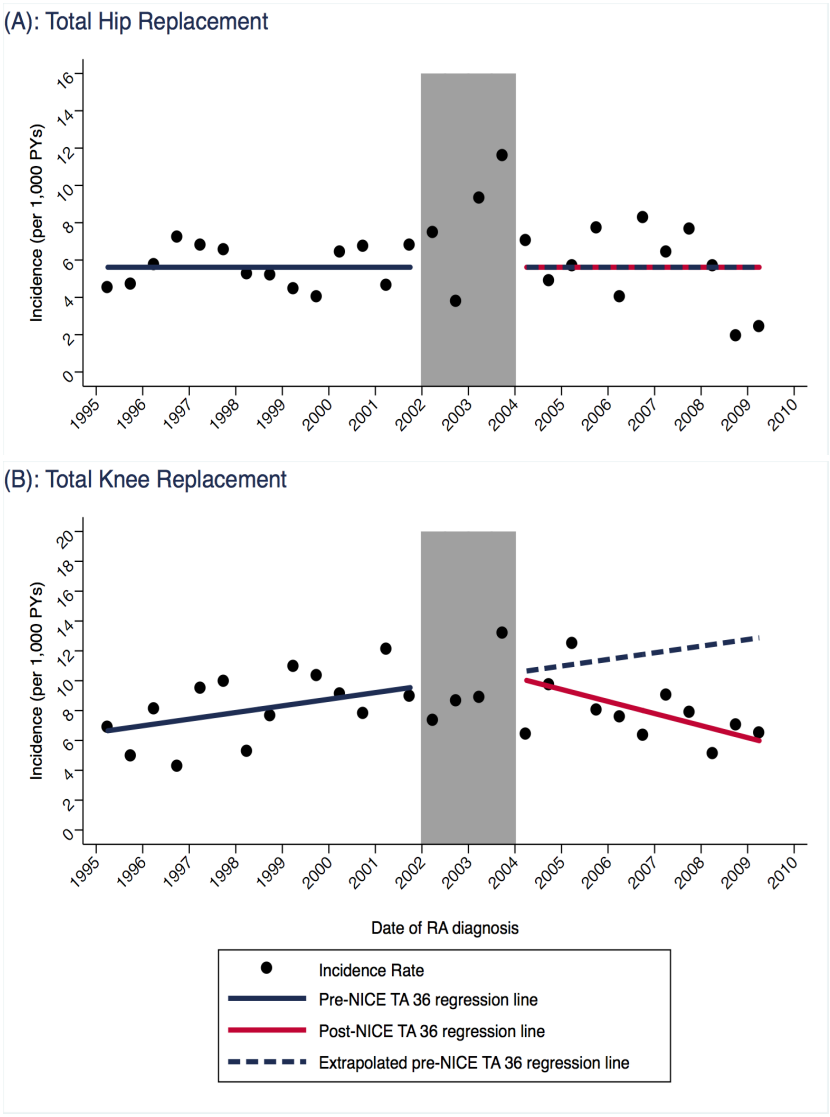


FIGURE 4.4: AGE AND SEX STANDARDISED INCIDENCE RATES OF JOINT REPLACEMENT WITHIN 5 YEARS OF RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS) WITH 2-YEAR LAG POST-NICE TA 36: (A) THR AND (B) TKR

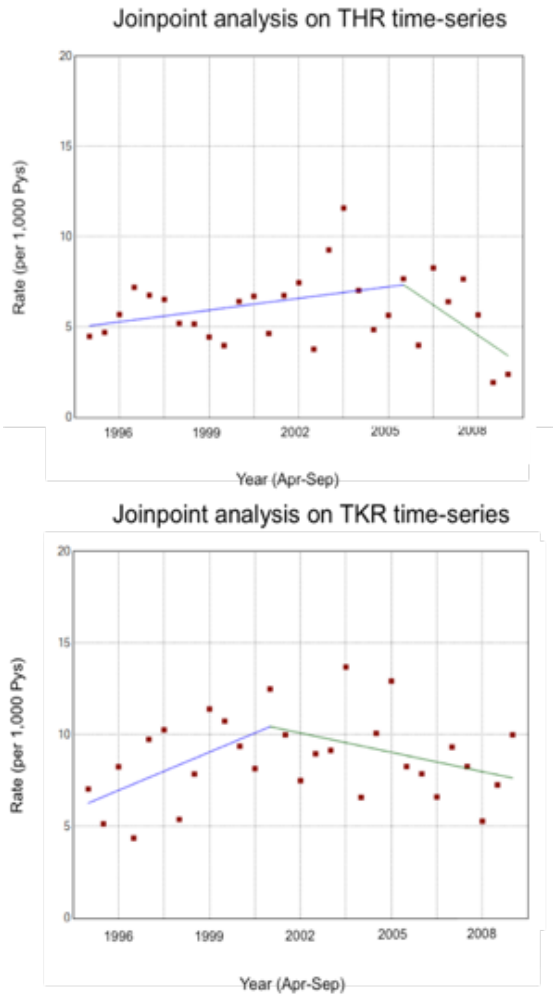


FIGURE 4.5: AGE AND SEX STANDARDISED INCIDENCE RATES OF JOINT REPLACEMENT WITHIN 5 YEARS OF RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS) WITH JOINPOINT REGRESSION LINES: (A) THR AND (B) TKR

DISCUSSION

Summary of main findings

This chapter reports on the population-level impact of national guidance for biologics on THR and TKR rates in a large sample of newly diagnosed RA patients in England and

Wales. The results indicate that the introduction of TNFi therapies for the management of early RA was associated with a significant reduction in 5-year incidence rates of TKR but not THR. Specifically, whilst TKR incidence was increasing prior to TNFi approval, this upward trend was reversed following the start of the biologic era, yielding a relative 34% average reduction in the biologic era compared to counterfactual rates. The relationship between THR incidence and TNFi approval was less clear, with no significant average change in post-NICE TA 36 modelled rates compared to counterfactual values.

Findings in context and possible explanations of effect

The potential impact of TNFi therapy on need for joint replacement is mentioned in the NICE TA 36 document, and inferred in the BSR guidelines upon which the NICE guidance was based (69), although it is stated no such data was available. NICE technological appraisals involve a very robust and thorough review of existing literature, and the evidence-base summarised in TA36 indicates the plausibility of a relationship between TNFi and reduced need for joint replacement. For example, results from a high quality RCT were summarised, in which 31% of patients on methotrexate alone experienced progression of structural damage at 54 weeks, whilst this was only 8% for patients on methotrexate plus infliximab (70). Indeed, the clinical effectiveness of both etanercept and infliximab in terms of reduced erosive damage as compared to methotrexate monotherapy has been clearly demonstrated in the RCT setting (68, 183, 184). Given this context, one major explanation for the lower rates of TKR here observed within the biologic era is the population-level phenomenon of TNFi induced arrest of progression (of joint destruction) amongst severe RA patients. Although data on joint destruction was

not available in the routinely collected health data as used here, supportive of this explanation is the recent report of significantly lower levels of disease activity in RA patients diagnosed in the biologic era as opposed to pre-biologic (185).

The overall incidence of THR was lower than TKR, which is suggestive that there may be less RA involvement at the hip (186) and which may partly explain why TNFi approval was found to be associated with a reduction in TKR incidence but not THR. Previous estimates of hip joint synovitis in early RA range between approximately 20-40% (187, 188) which is considerably lower than the 60% estimated for the knee (189). Also worth mentioning is that while the knee joint featured in the 1987 ACR diagnostic criteria and Disease Activity Score 28 (which itself has been shown to be a reliable indicator of disease activity (190)), the hip joint did not (4, 191). Were this to be the case (i.e. greater RA involvement at the knee over hip), then the lack of reduction in THR may be more a feature of secondary osteoarthritis and/or that a reduction in hip destruction may require a longer follow-up period that 5-years.

The decrease found in TKR rates is consistent with data from the UK NJR which indicate the percentage of knee replacements for which RA was an indication fell from 3% to 2% (2004 to 2010) (95, 192) and down to 1% by 2012 (39). The data also complements and builds on a prior study from the Republic of Ireland where the overall number of THR and TKR surgeries was found to increase dramatically from 1995 to 2010, whilst the number of those amongst those with a diagnosis of RA significantly decreased for TKR but remained stable for THR (94). On the other hand, Nikiphorou *et al.* have previously

reported on the 5-year and 10-year cumulative incidence of joint surgeries within two UK RA inception cohorts, where rates of major joint surgery stayed approximately stable across 1986 – 2011 (97). One reason that the authors offer for this observation, despite a significant decline in small joint surgeries is that routine clinical assessment/imaging of patient's larger joints is less frequently performed. This could potentially give rise to undetected progression of joint destruction in larger joints of some patients who remain untreated given controlled disease as assessed within their smaller joints. Although an interesting hypothesis, this explanation relies upon a degree of independence between disease control at large and small joints which is a subject that remains to be formally investigated. Comparison to the present analysis is limited as this prior study was not large enough to stratify by joint location, so their major outcome was a composite of joint replacements at the hip, knee, shoulder and elbow, in addition to cervical spine fusion and some other major surgical procedures (97).

Arthroplasty trends similar to those found here have been reported for other European countries. Nystad *et al.* have previously found that the incidence of RA-related joint replacement amongst the general population in Norway declined from 1994 to 2012 ($p<0.001$) despite the incidence of joint replacements for OA dramatically increasing over the same time period ($p<0.001$) (135). Their data indicated hip and knee replacements for RA declined in a very similar fashion to each other, but given the study design used there, those trends are likely to partially be reflective of changes in the incidence of the disease itself (RA versus OA) over and above any specific changes in the incidence within RA. Similar data have emerged from Finland which show the proportion of all hip and

knee replacements that were carried out for RA in 2010 was less than half that carried out in 1995, along with a strong decline in incidence of RA-related hip/knee replacement per 100,000 of the general population (p<0.001) (91). Interestingly, Hekmat *et al.* report that a decline in the incidence of hip replacement amongst RA patients followed TNFi introduction in Sweden, but that knee replacement remained unchanged (102). The findings from the current analysis are therefore not altogether consistent with these data from Sweden, although it must be noted that a limitation of their analysis was that it simply compared pre- versus post-2002 average rates of joint replacement (rather than controlling for secular trend). Incidence of total joint replacements amongst RA patients in a recent Spanish study also failed to identify a significant reduction in 5-year rates among RA patients diagnosed between 1994 – 2009 (101).

In terms of comparing the present data trends to those previously reported for the US, one prior US study found significant reductions following the introduction of biologics in terms of the number of both TKR and THR patients who had RA as the primary diagnosis at surgery, whilst the numbers of TKR and THR for other reasons profoundly increased (134). Conversely, a different US study using the US National Inpatient Sample found the prevalence (%) of RA amongst all THR and TKR procedures to remain stable from 2002-2012 (38). However, it would seem no investigators have reported on the incidence of joint replacement actually within/among incident RA cases.

Limitations

The lack of access to a comparator time-series of THR and TKR rates for the general population is a key limitation to the present study. As such, real caution is required in interpreting the findings. For example, the approximately stable incidence of THR could be interpreted as a favourable outcome were THR incidence in the general population to have undergone a significant concurrent increase. On the other hand, it may be that TKR rates in the general population also decreased over latter years which would weaken the case for a causal inference between declining TKR rates in RA patients and uptake of biologics. Related to this point is that while the outcome (THR/TKR) was validated to HES data (as described in chapter 3), the underlying reason for surgery was not available and therefore the inclusion of at least some surgeries for secondary osteoarthritis is probable. Although these are genuine limitations, it is reassuring that previous NJR data for the general population do indicate an overall year-on-year increase in the frequencies of both THR and TKR procedures carried out from 2003-2012 (193). This highlights the importance of a decline in TKR as observed here among RA patients, and indicates that any bias in the inclusion of THR/TKR procedures for OA may have biased trends towards showing an increase in THR/TKR rates over the time course (but not a decrease).

Furthermore, other factors such as prescription rates of csDMARDs having markedly increased within RA patients may have contributed to a reduced need for joint replacement over the study period (106). This is indeed one of the most likely sources of residual confounding in the present analysis. Early diagnosis and treatment of RA was the dedicated subject of BSR guidance in 2006 (44), and 'treat-to-target'

recommendations were widely endorsed in 2010 (78). The impact of csDMARDs is especially difficult to disentangle owing to the fact that clinicians following NICE TA36 were constrained to prescribe two 6-month-long regimens of csDMARDs prior to initiating a TNFi. It's likely therefore that NICE TA36 guidance gave rise to a tendency to initiate csDMARDs earlier and possibly at a larger/more sustained dose in order to speed up eligibility for TNFi. Conversely, a prior UK study investigating prescribing trends of methotrexate in early RA over recent decades provides little support for increased prescribing of csDMARDs in 2002, although possibly suggests rates of methotrexate were temporarily elevated (194). However, this prior study does clearly show a marked increase in such prescribing from 2006 onwards (associated with BSR guidance on early treatment (44)), suggesting the latter years of the time-series of the present analysis are likely to be confounded to a degree by this increased intensity of csDMARD prescribing. The introduction of other biologics other than TNFi throughout the biologic era may also have been influential.

There are other factors that may have introduced something of an ecologic fallacy here, i.e. speculatively attributing population level associations to the person-level (195). This is underlined by considering the possibility that the patients in whom the reduction in TKR occurred may not have been patients who initiated biologics. For example, other than increasing csDMARD prescribing, improvement in non-therapeutic aspects of RA management may likewise have played a role (196), as may have a gradually declining disease severity or smoking prevalence although these too are likely insufficient to explain the relatively sudden inflexion observed in the TKR trend following NICE

recommendations. The standardisation of the time-series by age and sex controls for changes in these two factors, although some residual confounding may remain owing to the extent of granularity in categorisation used in the standardisation (i.e. tertiles of age) which was limited by the study sample size. Similarly, trends may have been affected by a push by the government to reduce waiting times, for example by the expansion of NHS capacity through independent sector treatment centres/supplementary contracts with private providers to do orthopaedics. All these possible sources of confounding arise from one of the assumptions of the ITS method, being that confounding events do not occur around the same time as the 'intervention' under investigation. To properly resolve these issues, further studies are needed. One insightful avenue would be to replicate this same analysis in different countries with different health-care settings to confirm and/or further elucidate the population-level impact of biologics on THR and TKR rates. Additionally, a patient-level analysis comparing TNFi users versus non-users, whilst controlling for all important confounding factors such as mentioned above would be an equally logical next step, particularly in order to disentangle the effects of co-medication. These additional research avenues will be the subject of chapters 5 and 7.

An alternative explanation for the decline in TKR rates as observed here could be that the mere availability of biologics may have given rise to a greater tendency among rheumatologists to delay referral for joint replacement given an increased optimism over pharmacotherapy. Similarly, the availability of more treatments of greater perceived efficacy may have encouraged clinicians to diagnose RA earlier due to greater hope of successful management. Although the available data do not permit a thorough

exploration of these hypotheses, earlier diagnoses of RA is not supported by the observation that average comorbidity index worsened over the study period and that there was no dramatic increase in incident RA diagnoses associated with the start of the biologic era at the turn of the millennium (table 4.1).

Related to this point is that the biologics here investigated are not administered or captured in the primary care setting and so could not be included in analyses. Although a 1-year lag (and 2-year lag in sensitivity analyses) was used to account for a delay, it may be that uptake was much slower or of varying intensity, for example according to region. It's reassuring to note however that by 2005 approximately 8,500 RA patients had been recruited to the British Society for Rheumatology Biologics Register (BSRBR-RA) due to being administered TNFi, according to one report (71). Given the authoritative and widespread reach of NICE, a population-level increase in the use of etanercept and infliximab (in line with the recommendations) would indeed be expected (197, 198) despite any local regional variation in speed of uptake.

The use of Joinpoint regression indicated a downward inflexion in the TKR time-series was best-placed 12-6 months prior to the publication of NICE TA 36 (figure 4.5B). This may suggest some role for the 2001 BSR guidelines (upon which NICE TA 36 was based), although it is difficult to determine this with much confidence given the close temporal proximity and the limited data to allow much greater granularity. The significant trend change in THR rates which the Joinpoint analysis identified was not consistent with the main analysis and this is more likely the product of Joinpoint only identifying trend

changes (rather than trend and/or step changes) and was driven by the particularly low values for the last two time points of the time-series (figure 4.5A). This reveals a need for further investigation of these methods, particularly the sensitivity to the number of time points in pre- vs. post- intervention periods and the number of outcome events occurring per time point. It maybe that more timepoints (e.g. quarterly or monthly data) or rather less timepoint (e.g. annual data), instead of the bi-annual timepoints may have been more appropriate in terms of power and bias. This precise issue will be the dedicated subject of chapter 6.

Another important caveat is that all these findings pertain to the context of joint replacements within the first 5 years after RA diagnosis. It is therefore probable that any true impact of TNFi therapy maybe underestimated here due to not considering longer-term outcomes. The fixed 5-year time window across the study period was primarily used to prevent THR and TKR rates over time being influenced by underlying variation in the length of follow-up available (i.e. the bias of patients at the beginning of the study period systematically having longer enrolment in the database). Despite this being so, the use of 5-years is still suitable given that this was approximately the median length of follow-up and that prior data show rates of joint replacement in the first 5-years after RA diagnosis is far from negligible (approximately 5% at 5-years amongst those with highest disease activity (34)).

Strengths

Notwithstanding these various limitations, the study has several strengths. Not least of which is the large sample of RA patients identified from a data source generalizable to the UK population in terms of age, sex and ethnicity (123). Rather than describing temporal trends of RA as an indicator for THR/TKR surgery, the approach of using RA patients as the denominator as used here is much more preferable because this accounts for underlying changes in the incidence of RA over time. As described in the introduction, the ITS method is another strength as this quasi-experimental method controls for secular trends in the outcome prior to the intervention (170, 171). The importance of this approach is made clear by considering that a conventional before-after impact analysis would have here yielded no significant difference as there are two almost equal but opposite trends in existence (Figure 2b). That is, pooling rates before biologics and simply comparing them to pooled rates after their introduction would have masked the inflexion.

The use of incidence rates as the repeated measure over time accounts for any changes in rates of mortality or loss-to-follow-up during the study period, which would not have been accounted for had these measures been 5-year cumulative incidence rates (as initially used, although with little difference in results (199) [appendix table 4.14]). As reported in chapter 3, the linkage to HES allowed for an internal validation of CPRD coding of THR and TKR, with the basic five and ten-year cumulative incidence of a combined THR/TKR outcome being very similar to reports from previous UK early RA inception cohorts, at approximately 7% and 11%, respectively.

Conclusions

This chapter reports that approval of TNFi therapies in England and Wales was temporally associated with a significant and clinically meaningful decline in TKR incidence among early RA patients. Specifically, that an estimated 34% relative reduction in rates were observed in 5-year incidence rates compared to what would have been expected had biologics not been introduced. However, there was no significant overall change observed in 5-year THR incidence rates. The use of the ITS method as used here amongst an incident cohort of RA patients has allowed for an explicit estimation of the impact of the introduction of TNFi whilst addressing confounding by secular trend, which has previously not been done in the literature. The lack of a control time-series (not exposed to biologics) in a key limitation, as is the possibility of confounding by earlier/more intense usage of csDMARDs. Further research as reported in chapters 5 and 7 will seek to address these issues.

5. POPULATION-LEVEL IMPACT OF THE INTRODUCTION OF BIOLOGICS ON SUBSEQUENT RATES OF THR AND TKR WITHIN DENMARK AND ONTARIO, CANADA

INTRODUCTION

The previous chapter reported on the population-level analysis of temporal trends of THR and TKR incidence for RA patients within England and Wales. The aim of the analysis was to estimate the impact of the introduction and approval of TNFi on subsequent rates of joint replacement. The main findings were that rates of TKR declined following the introduction of TNFi in 2002, but no difference was observed in rates of THR. One of the strengths of the analysis was the method used. ITS is a strong quasi-experimental approach for the evaluation of population-level interventions, which controls for prior level and trend and allows the sampling unit to act as its own control in a before-after design.

However, not all ITS analyses are equally robust. There is a gradient in terms of the strength of evidence produced and the degree of support provided for causal inference (119, 170). For example, it is possible to investigate a single time-series of a repeated measure of an exposure/outcome as was performed in the previous chapter. Whilst this

is still a stronger approach than a mere comparison of outcome means before vs. after an intervention (which does not account for changing trends), one of the limitations of such an analysis is that other confounding events occurring at the same time as the 'intervention' cannot be ruled out as influencing the changes in outcome observed. An ecological fallacy is a possibility. In the case of the reduced rates of TKR in the biologic era as reported in the previous chapter, it may be that the decrease was a function of more general secular changes to rates of joint replacement for the entire UK population (not just RA patients), for example possibly due to changes in NHS budgetary constraints. One way to strengthen the level of evidence would be to repeat the analysis in a different health care setting outside the UK to determine whether introduction of TNFi therapy in other countries had a similar effect. Although not totally negating the threat to validity by confounding events, it would be reassuring if similar estimates of the population-level impact of TNFi were found in validation cohorts outside the UK.

Furthermore, the ITS method can be substantially strengthened with the use of a negative control population, i.e. repeatedly estimating the outcome measure over time - before and after the intervention – amongst a similar study cohort that was by definition not eligible for exposure to the 'intervention' (171). This provides a means of assessing whether any effect of the intervention in the primary study cohort was absent in the cohort that should be unaffected by the intervention. This builds on the difference-in-differences study design, most often employed in the social sciences, which evaluates whether the difference in means between a treated and control group is itself different post-treatment as compared to pre-treatment (119, 200-202), thus estimating an

average treatment effect on the treated (ATT). This simple pre-post study design was used in the foundational epidemiological investigations of John Snow who sought to estimate the impact of pump handle removal on the transmittance of cholera in 1855 (202). However, the simple pre-post comparison of the difference in outcome between exposure groups may still be confounded by secular trends quite apart from the treatment effect when relatively long time periods are used to assess pre- and post-intervention levels. The use of repeated measures pre- and post-intervention therefore combines the strengths of both the ITS and simple difference-in-differences designs, resulting in what is referred to variously as a controlled ITS (200, 201, 203), an ITS with a non-equivalent no treatment control group (119) or a multiple group time-series (204).

An interesting previous example of how this method has been applied is a study of the impact of direct-to-consumer television advertising of medications in Canada (205). To estimate the impact of television advertising campaigns for etanercept, mometason and tegaserod, the study authors evaluated the impact on trends of prescription rates following the advertising campaigns for these medications within English speaking provinces versus a predominantly French speaking province. All the direct-to-consumer advertising of interest was performed in English on English television and therefore would have little coverage in French speaking provinces. The authors observed a significant increase in sales for some medications following direct-to-consumer advertising in the English-speaking provinces, but not in the French speaking province. While such a negative control time-series would have aided in interpreting the data in chapter 4, for example looking at outcome trends in matched non-RA patients, no

suitable alternative population within the UK was accessible in the available data for this DPhil project.

The analyses reported in the present chapter therefore focuses on strengthening the evidence as provided in chapter 4 through: (i) repeating the same analyses using two different datasets from health-care settings from elsewhere in the world and (ii) making use of negative control time-series within these validation cohorts. The aim of the analyses was to estimate the population-level impact of TNFi on the incidence of THR and TKR in RA for the two populations to which these external datasets pertain, whilst controlling for secular trends and levels of outcome for both RA and non-RA patients.

METHODS

Successful application was made to gain access to two external datasets of incident RA patients covering the entirety of Denmark and for the region of Ontario, Canada. These datasets are described in chapter 2, in addition to details of collaborator contributions in pre-analytical data acquisition and processing. For sake of clarity, the methods are here described separately for the two validation datasets.

DENMARK

Data sources

Danish data were obtained from two sources: the Danish National Patient Register (DNPR) (127) and the Civil Registration System (CRS) (128). Register linkage between the two had previously been made and an extract was obtained for the period 01/01/1996 to 31/05/2016. Over this period the DNPR collected all hospital visits to Danish hospitals and private clinics, with discharge diagnoses recorded using ICD-10 codes and surgeries recorded using Nordic Medico-Statistical Committee (NOMESCO) codes. The CRS contained data on mortality and migrations out of the country for all residents.

Study participants

Two person-level cohorts were extracted. The first consisted of all incident RA cases as identified according to the first ICD10 code for RA (M05-06) within the DNPR. ICD10 coding has previously been shown to have a high validity within the DNPR (206). DNPR contains inpatient data from 1977 which enabled prevalent RA cases (i.e. those with an RA code prior to the study period) to be identified and discounted. A second cohort was assembled consisting of general population controls, matched at date of RA diagnosis on age (year of birth), sex and geographic area. Ten general population controls were matched to each incident RA patient, unless ten could not be identified in which case up to ten were selected. Individuals within either cohort who were <18 years of age were excluded, as were those with a prior THR (for the THR analysis) or TKR (for the TKR

analysis). Exclusions were performed after matching and replacement of exclusions was not carried out.

Intervention

A similar date for the introduction of TNFi was considered as per the UK analysis in chapter 4. The start of 2002 was defined as the beginning of this introduction, with a subsequent 1-year lag period to allow for the changes in TNFi prescribing to become widespread. Reasons to support this choice were that the first Danish Institute for Regional Pharmacotherapy was published in November 2002 (207), but that prior assessment of reports by DANBIO (the Danish registry for biologic therapies) showed negligible use of TNFi prior to 2002 but that a significant increase in prescribing began throughout this year (208).

Outcome

Primary THR and TKR as recorded in DNPR (NOMESCO codes: THR, KNFB, TKR and KNGB) were identified within 5-years of index date (i.e. diagnosis date for RA patients and imputed matched date for controls).

Statistical analysis

As per analysis of UK data, an aggregated time-series of 5-year age and sex standardised incidence rates of THR and TKR was derived consisting of repeated measures among newly diagnosed/matched individuals within each six months of the study period. The

person-year denominator for each of these 5-year incidence rates was calculated by following up individuals from their index date until the first date of either THR/TKR (analysed separately), death, emigration or end of 5 years of follow-up. The last 5-years of the time-series, i.e. from 01/05/2011 onwards, were discounted to allow composition of follow-up and comparability of rates throughout the study period. This yielded a total of 12 pre-intervention timepoints and 17 post-intervention timepoints.

Segmented linear regression was then carried out to estimate change in level and/or trend following the defined introduction of TNFi in 2002. This was performed separately for the RA and non-RA cohorts. Backward stepwise selection was used as in chapter 4 to specify the most parsimonious model given that unnecessary terms in the model can reduce statistical power. The final regression coefficients were also used to estimate the difference between the observed average post-intervention fitted values and those expected had pre-intervention trends continued post-intervention, uninterrupted.

Given the availability of both RA and non-RA cohorts, the first of which was exposed to some degree to biologics and the second unexposed, a difference-in-differences style controlled ITS analysis was also performed (119, 205) to estimate any impact on the RA cohort over and above the non-RA group. This analysis consisted of calculating the difference in outcome between the two exposure cohorts at each timepoint, then conducting a segmented linear regression on these differences as per the individual time-series analysis. Here the null hypothesis was that the introduction of biologics had no impact on these differences, i.e. no difference in the differences.

Given that all work described should be assumed to have been carried out by myself unless otherwise stated (as described in the acknowledgment section), it should be noted that interrupted time-series analysis of the Danish data was initially carried out in Oxford under my supervision by a visiting student from Denmark (Dr René Cordtz). However, I subsequently repeated the analyses to generate all graphs and tables here included, and additionally carried out the difference-in-differences time-series approach.

ONTARIO, CANADA

Data sources

These are described in more detail in chapter 2. The Ontario Rheumatoid Arthritis administrative Database (ORAD) is an administrative database covering an approximate general population of 14 million people (as of 2014) and is designed to contain information on all RA patients in Ontario, Canada (129). The OHIP Registered Persons Database was also used to identify non-RA controls for matching purposes. This is a collection of demographic data (including vital status, age, sex and place of residence) on all individuals living in Ontario who are eligible for OHIP coverage. ORAD is also linked to the Ontario Drug Benefit Program (ODBP) database and this linkage was used to identify pharmacy claims on all individuals included in the study who were ≥65 years old (132).

Study participants

All individuals registered into ORAD as an incident RA patient were initially eligible for inclusion. The final RA cohort used for creating the subsequent time-series was then defined as follows. Only RA patients aged 18 years or older were selected for inclusion, with the first diagnosis date between 1st April 1996 and 31 March 2015 considered their index date. Patients with less than 1-year enrolment to OHIP prior to index-date were excluded to ensure at least a period of 12 months to determine prevalent cases. Furthermore, patients with any non-inflammatory arthritis diagnoses in the 3 years preceding the index date were also excluded to reduce the bias of including misclassified RA which may actually be a different inflammatory condition. Patients with a history of hip or knee replacement in the previous 5 years were also excluded as these patients are either not able to experience the outcome of interest or have a significantly reduced risk of experiencing it. The 5-year lookback was used to increase consistency in this condition, given that otherwise there would be varying degrees of 'look back', particularly for patients of different ages and time period of RA diagnosis.

To each of the ORAD RA patients included, 5 non-RA individuals living in Ontario (and eligible for OHIP) were matched using the OHIP registered persons database. Exclusion criteria as applied to the RA cohort was applied to the OHIP database prior to matching, so that each RA patient was matched to exactly 5 controls. The matching criteria included sex, date of birth (+/- 90 days) and geographic region of residence within Ontario. The

same index date for each RA patient was imputed into their controls. Change in RA status (i.e. controls becoming RA cases) was handled by allowing these patients to switch to the newly diagnosed RA cohort and then be matched accordingly (unless fulfilling one of the exclusion criteria in which case the individual was merely censored at date of RA diagnosis).

Intervention

The time period April 2001 – March 2002 was decided upon as a 1-year intervention period to account for the introduction of biologics into routine clinical practice in treating RA in Ontario. This was discussed with Prof. Gillian Hawker who is a local academic rheumatologist and the reasons for this timepoint were that it incorporates the ACR guidelines for using biologics, published in February 2002 (42), but accounts for the fact that etanercept was nationally approved in December 2000 and infliximab in June 2001 (209), and that previously published data shows the prescribing of biologics had already begun to increase in 2001 which has previously been considered as when biologics were introduced into Ontario (132). Using the 1-year 'phase in' period also accounts for the fact that similar to NICE guidelines, treatment in Canada is restricted to individuals with severe RA who had already failed to adequately respond to csDMARD therapy, although in Canada this only has to be for at least 3 months (210) (rather than 6 months in the UK).

Outcome

The study outcome (for both RA patients and matched controls) was first occurrence of either (i) THR or (ii) TKR (analysed separately) during the first 5-years of follow-up from index date. These were identified using the following Canadian Classification of Intervention (CCI) codes as included in appendix file 4.

Given readily available linkage to pharmacy data in the senior population, i.e. those 66 years or older (previously shown to make up approximately 40% of RA patients (129)), two secondary outcomes were examined: incidence of csDMARD (during the first year of follow-up from index date) and incidence of biologics (during the first 5-years of follow-up from index date). As a sensitivity analysis, the 1-year cumulative incidence of csDMARDs and 5-year cumulative incidence of biologics was modelled as most patients start methotrexate relatively early and therefore incidence rates over short follow-up periods maybe difficult to interpret.

Statistical analysis

As described above (and in chapter 4), incident rates for all outcomes were estimated per each 6 months of the study period – separately for patients in the RA cohort and those in the non-RA matched control cohort. A parsimonious segmented linear regression model fitted to the data, separately for the RA and non-RA cohorts. The final regression coefficients were used to estimate the difference between the observed average post-intervention fitted values and those expected had pre-intervention trends continued post-intervention, uninterrupted. In order to model any differential changes

in rates, following the introduction of biologics, between the RA and non-RA cohorts, a difference-in-differences style approach was also performed. Here, the absolute difference in outcome (incidence in RA cohort - incidence in non-RA cohort) was first calculated for each timepoint, and then a segmented linear regression used to model the impact of the intervention period on these differences, the null hypothesis being that there would be no change in level or trend following the introduction of biologics (i.e. no differential impact among the RA patients).

Combined Population-Level Forest Plot

The estimates of absolute risk difference at the midpoint of the post-intervention period were graphed for the RA cohorts in Denmark, Ontario and England & Wales (chapter 4). This was based on estimates from the parsimonious final models, after removal of non-significant time-series terms (as described above), as this is recommended to maximise power to detect 'true' intervention effects (171). However, full models were used in cases where all time-series terms had been removed due to non-significance. Although estimates derived from full models had larger confidence intervals due to forced entry of all time-series terms, incorporating them in the forest plot seemed preferable to excluding these estimates altogether. Results from the difference-in-difference ITS analyses were graphed in similar fashion.

RESULTS

DENMARK

Baseline characteristics

Baseline characteristics are in table 5.1. There were a total of 29,427 RA patients included in the THR time-series analysis and 29,703 in the TKR time-series. The reason for the two RA cohorts was to minimise unnecessary exclusions, only excluding patients with a prior THR from the THR time-series analysis and excluding patients with a prior TKR from the TKR time-series. The number of matched non-RA controls was 290,778 and 294,806 for THR and TKR, respectively. Mean age at RA diagnosis – rounded to the nearest whole year - was 58 years amongst RA patients (for both THR and TKR cohorts) whilst this was 54 years in the matched controls. The reason for discrepancy in age between matched cohorts in DK was because exclusions were made of prior THR/TKR or other (non-RA) inflammatory conditions after matching, with no replacement of patients excluded. Exclusions were henceforth made prior to matching when using the Ontario dataset. The gender split was 70% female in all cohorts.

TABLE 5.1: COHORT CHARACTERISTICS AT BASELINE

	RA	non-RA controls
Denmark (THR cohort)		
Number of individuals	29,427	290,816
Age, mean (S.D.)	58.3 (15.7)	53.9 (15.4)
Female sex, n (%)	20,612 (70.0)	203,232 (69.9)
Denmark (TKR cohort)		
Number of individuals	29,703	294,813
Age, mean (S.D.)	58.5 (15.7)	54.1 (15.4)
Female sex, n (%)	20,760 (69.9)	205,979 (69.9)
Ontario		
Number of individuals	99,974	499,870
Age, mean (S.D.)	57.50 (16.0)	57.50 (16.0)
Female sex, n (%)	33,321 (33.3%)	166,623 (33.3%)

THR time-series

Results for the THR ITS analysis are depicted in figure 5.1A and described in table 5.2. Age and sex standardised THR incidence (per 1,000 PYs) among RA patients at the start of the study period was 8.72, which decreased by -0.19 [95% CI: -0.32 to -0.06; p=0.006] during the pre-intervention period. The introduction of biologics was however associated with something of an abrupt increase in incidence of 2.23 [95% CI: -0.23 to 4.70; p=0.073]. Conversely, among non-RA patients there was an upward trend in THR incidence prior to biologics (0.05 [95% CI: 0.03 to 0.08; p<0.001]) although a significant downward inflexion occurred post-intervention: -0.04 [95% CI: -0.09 to 0.00; p=0.04] per timepoint. The difference in THR incidence (per 1,000 PYs) between the RA cohort and non-RA cohort at the start of the study period was 5.59 [95% CI: 4.39 to 6.80; p<0.001] (figure 5.1B, table 5.3). This gap between the cohorts was closing at a rate of -0.21 [95% CI: -0.34 to -0.08; p=0.003] per timepoint prior to biologics being introduced, which

continued to be the case post-intervention. However, the introduction of biologics was associated with the delta growing by 1.95 [95% CI: -0.45 to 4.36; *p*=0.11] (table 5.3).

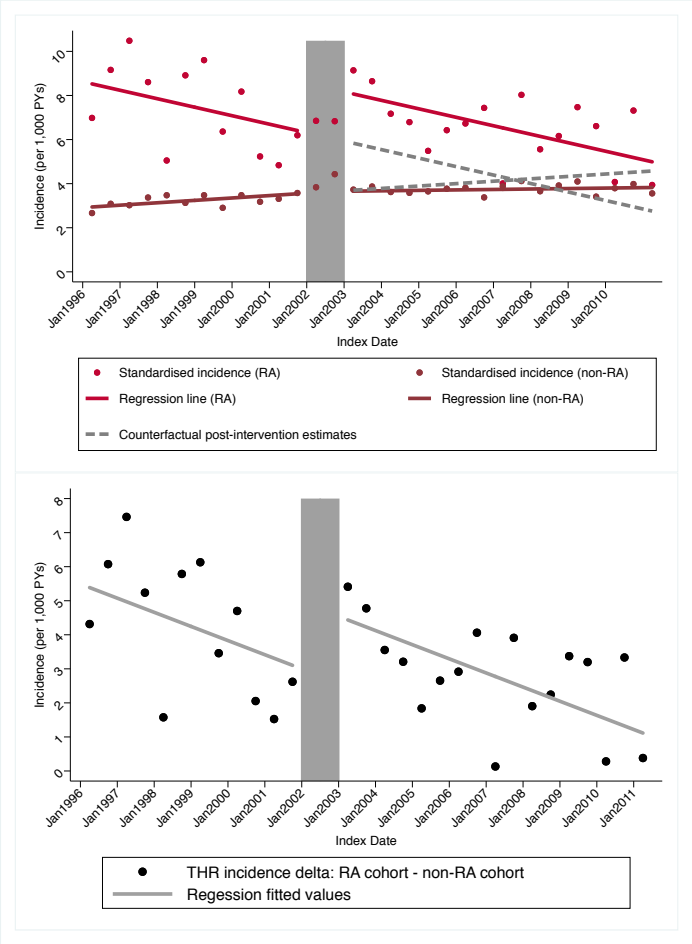


FIGURE 5.1: (A) AGE AND SEX STANDARDISED INCIDENCE RATES WITHIN DENMARK OF THR WITHIN 5 YEARS OF RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS) BEFORE AND AFTER THE INTRODUCTION OF BIOLOGICS: STRATIFIED BY COHORT; (B) DIFFERENCE-IN-DIFFERENCES IN 5-YEAR INCIDENCE RATES OF THR BETWEEN COHORTS

TABLE 5.2: ITS SEGMENTED LINEAR REGRESSION ANALYSIS FOR RHEUMATOID ARTHRITIS AND CONTROL COHORTS WITHIN DENMARK AND ONTARIO: 5-YEAR INCIDENCE RATES OF THR (PER 1,000 PERSON-YEARS)

	Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
Denmark					
RA*	Baseline Incidence Rate	8.72	7.48	9.95	<0.001
	Trend (per 6 months)	-0.19	-0.32	-0.06	0.006
	Level change after Biologics	2.23	-0.23	4.70	0.073
	Trend change after Biologics				
non-RA*	Baseline Incidence Rate	2.89	2.64	3.14	<0.001
	Trend (per 6 months)	0.05	0.03	0.08	<0.001
	Level change after Biologics				
	Trend change after Biologics	-0.04	-0.09	0.00	0.04
Ontario					
RA*	Baseline Incidence Rate	5.29	4.88	5.70	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-0.07	-0.14	0.01	0.093
non-RA*	Baseline Incidence Rate	0.68	0.57	0.79	<0.001
	Trend (per 6 months)	0.04	0.02	0.05	<0.001
	Level change after Biologics				
	Trend change after Biologics	-0.02	-0.04	0.00	0.035

*RA: rheumatoid arthritis

TABLE 5.3: ITS SEGMENTED LINEAR REGRESSION ANALYSIS FOR DIFFERENCES BETWEEN RHEUMATOID ARTHRITIS AND CONTROL COHORTS WITHIN DENMARK AND ONTARIO: 5-YEAR INCIDENCE RATES OF THR (PER 1,000 PERSON-YEARS)

	Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
Denmark					
	Baseline Incidence Rate	5.59	4.39	6.80	<0.001
	Trend (per 6 months)	-0.21	-0.34	-0.08	0.003
	Level change after Biologics	1.95	-0.45	4.36	0.11
	Trend change after Biologics	-	-	-	-
Ontario, Canada					
	Baseline Incidence Rate	4.30	3.72	4.88	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-0.10	-0.17	-0.02	0.017

These changes in THR associated with the intervention period corresponded to an average absolute reduction (compared to counterfactual estimates) over the post-intervention period of 2.24 [95% CI: -0.11 to 4.58] in RA patients and -0.38 [95% CI: -0.71 to -0.04] in non-RA patients (table 5.4). Comparing the average absolute difference-in-differences post-intervention (relative to counterfactual differences) between the RA cohort and non-RA cohort was 1.95 [-0.34 to 4.25] (table 5.4).

TABLE 5.4: ESTIMATED CHANGES IN 5-YEAR THR OUTCOMES FOLLOWING FIRST RA DIAGNOSIS, ASSOCIATED WITH THE INTRODUCTION OF BIOLOGICS IN DENMARK AND ONTARIO, CANADA

	'Post-intervention period' midpoint estimate		Absolute Change*		
	Without intervention	With intervention	estimate	lower 95% CI	upper 95% CI
Denmark					
Rheumatoid Arthritis (per 1,000 PYs)	4.39	6.63	2.24	-0.11	4.58
Non-Rheumatoid Arthritis (per 1,000 PYs)	4.11	3.74	-0.38	-0.71	-0.04
Difference (RA - non-RA) (per 1,000 PYs)	0.93	2.88	1.95	-0.34	4.25
Ontario, Canada					
Rheumatoid Arthritis (per 1,000 PYs)	5.24	4.72	-0.52	-1.11	0.07
Non-Rheumatoid Arthritis (per 1,000 PYs)	1.41	1.23	-0.18	-0.34	-0.02
Difference (RA - non-RA) (per 1,000 PYs)	4.30	3.53	-0.76	-1.35	-0.18

* calculated by comparing estimated values for midpoint of post-intervention period to counterfactual values for the same timepoint (i.e. those estimated/expected taking into account only the pre-intervention level and trend). Estimated from final/parsimonious models specified using backward-stepwise selection (p-entry 0.049 and p-exit 0.20).

TKR time-series

Results for the TKR ITS analysis are depicted in figure 5.2A and table 5.5. Age and sex standardised incidence (per 1,000 PYs) was 5.87 at the start of the study period and this increased by 0.10 [95% CI: -0.05 to 0.24; *p*=0.17] during the pre-intervention period with a subsequent downward trend change of -0.20 [95% CI: -0.43 to 0.03; *p*=0.083] post-

intervention, although neither of these trends were statistically significant at the p=0.05 level (table 5.5). For non-RA patients there was a similar pre-intervention trend as was seen in the RA cohort (0.11 [95% CI: 0.08 to 0.13]; $p<0.001$), likewise the intervention was also associated with a downward trend change (-0.07 [95% CI: -0.11 to -0.03]; $p=0.003$) although this was not as large (table 5.5) and comparison of significance should be treated cautiously given the far larger sample size for the controls. The difference in TKR rates between the cohorts at the start of the study was 5.38 [95% CI: 4.76 to 6.01; $p<0.001$] (figure 5.2B, table 5.6). This difference remained constant through the pre-intervention period but began to significantly close at a rate of -0.15 [95% CI: -0.23 to -0.07; $p=0.001$]) per timepoint following the introduction of biologics (table 5.6). In terms of the estimated overall impact of introducing TNFi, the average absolute difference in post-intervention modelled values compared to counterfactuals was -1.69 [95% CI: -3.54 to 0.15] for RA patients and -0.57 [95% CI: -0.91 to -0.23] for non-RA patients (table 5.7). This yielded a significant average difference-in-differences of -1.24 [95% CI: -1.89 to -0.59] over the post-intervention period (relative to counterfactual differences) (table 5.7).

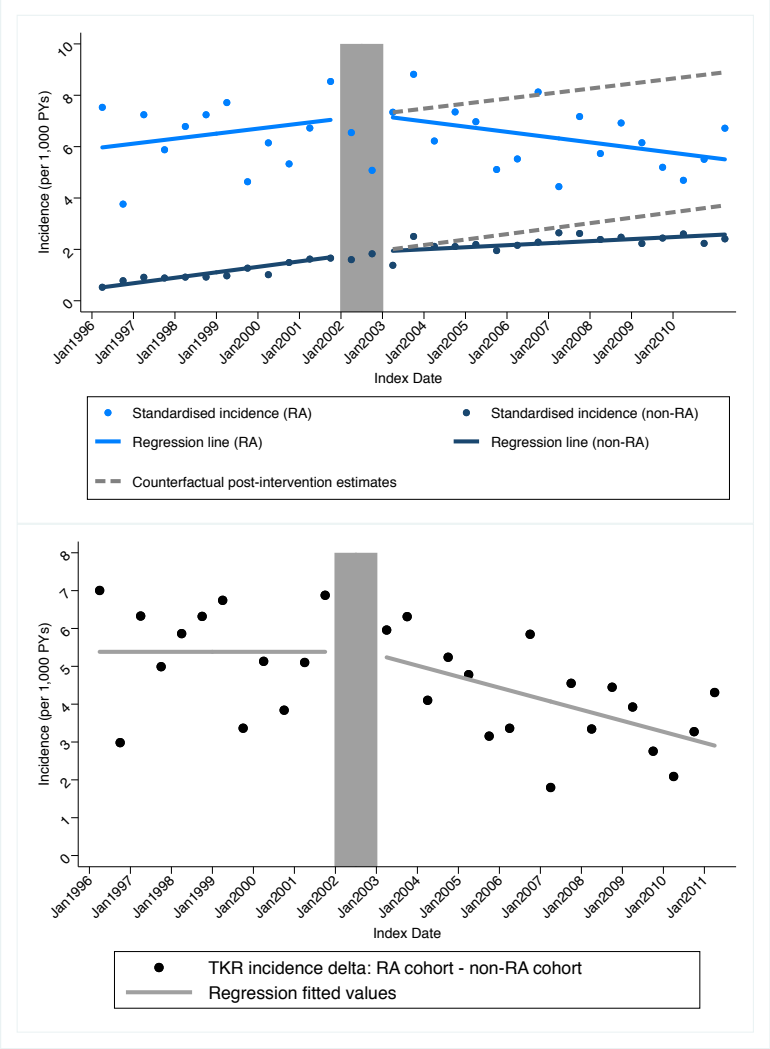


FIGURE 5.2: (A) AGE AND SEX STANDARDISED INCIDENCE RATES WITHIN DENMARK OF TKR WITHIN 5 YEARS OF RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS) BEFORE AND AFTER THE INTRODUCTION OF BIOLOGICS: STRATIFIED BY COHORT; (B) DIFFERENCE-IN-DIFFERENCES IN 5-YEAR INCIDENCE RATES OF TKR BETWEEN COHORTS

TABLE 5.5: ITS SEGMENTED LINEAR REGRESSION ANALYSIS FOR RHEUMATOID ARTHRITIS AND CONTROL COHORTS WITHIN DENMARK AND ONTARIO: 5-YEAR INCIDENCE RATES OF TKR (PER 1,000 PERSON-YEARS)

	Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
Denmark					
RA*	Baseline Incidence Rate	5.87	4.52	7.22	<0.001
	Trend (per 6 months)	0.10	-0.05	0.24	0.17
	Level change after Biologics				
	Trend change after Biologics	-0.20	-0.43	0.03	0.083
non-RA*	Baseline Incidence Rate	0.42	0.17	0.66	0.002
	Trend (per 6 months)	0.11	0.08	0.13	<0.001
	Level change after Biologics				
	Trend change after Biologics	-0.07	-0.11	-0.03	0.003
Ontario					
RA*	Baseline Incidence Rate	6.42	6.12	6.72	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-	-	-	-
non-RA*	Baseline Incidence Rate	0.77	0.68	0.86	<0.001
	Trend (per 6 months)	0.03	0.02	0.04	<0.001
	Level change after Biologics	0.19	0.00	0.37	0.048
	Trend change after Biologics				

*RA: rheumatoid arthritis

TABLE 5.6: ITS SEGMENTED LINEAR REGRESSION ANALYSIS FOR DIFFERENCES BETWEEN RHEUMATOID ARTHRITIS AND CONTROL COHORTS WITHIN DENMARK AND ONTARIO: 5-YEAR INCIDENCE RATES OF TKR (PER 1,000 PERSON-YEARS)

	Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
Denmark					
	Baseline Incidence Rate	5.38	4.76	6.01	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-0.15	-0.23	-0.07	0.001
Ontario, Canada					
	Baseline Incidence Rate	5.34	4.92	5.76	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-0.04	-0.09	0.02	0.19

TABLE 5.7: ESTIMATED CHANGES IN 5-YEAR TKR OUTCOMES FOLLOWING FIRST RA DIAGNOSIS, ASSOCIATED WITH THE INTRODUCTION OF BIOLOGICS IN DENMARK AND ONTARIO, CANADA

	'Post-intervention period' midpoint estimate		Absolute Change*		
	Without intervention	With intervention	estimate	lower 95% CI	upper 95% CI
Denmark					
Rheumatoid Arthritis (per 1,000 PYs)	8.07	6.37	-1.69	-3.54	0.15
Non-Rheumatoid Arthritis (per 1,000 PYs)	2.81	2.24	-0.57	-0.91	-0.23
Difference (RA - non-RA) (per 1,000 PYs)	5.38	4.14	-1.24	-1.89	-0.59
Ontario, Canada					
Rheumatoid Arthritis (per 1,000 PYs)	6.42	6.42	0.00	-	-
Non-Rheumatoid Arthritis (per 1,000 PYs)	1.30	1.48	0.19	0.01	0.37
Difference (RA - non-RA) (per 1,000 PYs)	5.38	5.05	-0.29	-0.71	0.13

* calculated by comparing estimated values for midpoint of post-intervention period to counterfactual values for the same timepoint (i.e. those estimated/expected taking into account only the pre-intervention level and trend). Estimated from final/parsimonious models specified using backward-stepwise selection (p-entry 0.049 and p-exit 0.20).

ONTARIO, CANADA

Baseline characteristics

Baseline characteristics of patient-level data prior to aggregation are reported in table 5.1. The THR/TKR time-series were based on a total of 99,974 RA patients and 499,870 non-RA controls. Mean age was 57 in both cohorts with identical sex ratios (70% female). Median follow-up was 5-years for both THR and TKR analyses in both the RA and non-RA cohorts. For the THR analysis there was a total of 400,729 PYs and 2,059,620 PYs of follow-up in RA and non-RA cohorts, respectively. During this time there were a total of 2,814 (median time-to-event equalling 1.24 years) and 4,478 (median time-to-event equalling 2.38 years) THR events in the RA and matched non-RA cohorts, respectively. Likewise, for the TKR analyses there was a total of 395,403 PYs and 2,056,522 PYs of follow-up in RA and non-RA cohorts, respectively. Over this time there was a total of

4,790 (median time-to-event equalling 1.2 years) and 5,344 (median time-to-event equalling 2.29 years) TKR events. Crude (i.e. unstandardized) incidence rates in the RA cohort were 7.02/1,000 PYs for THR and 12.11/1,000 PYs and TKR. These rates in the matched non-RA cohort were 2.17/1,000 PYs and 2.60/1,000 PYs, respectively.

THR time-series

Among the RA patient cohort, age and sex standardised 5-year incidence of THR (per 1,000 PYs) at the start of the study period was 5.29 [95% CI: 4.88 to 5.70], which remained unchanged during the pre-biologic period (Apr 1996 – Mar 2001) (figure 5.3A, table 5.2). Immediately following the intervention lag period there was a small but non-significant subsequent downward trend of -0.07 [95% CI: -0.14 to 0.01; p=0.093] per 6 months for the remainder of the study period (table 5.2). Among the matched non-RA cohort, the estimated standardized incidence of THR at the start of the study was 0.68 [95% CI: 0.57 to 0.79], but this increased by 0.04 [95% CI: 0.02 to 0.05; p<0.001] per timepoint throughout the pre-intervention period. Similar to the RA cohort, the intervention period was associated with a -0.02 [95% CI: -0.04 to 0.00; p=0.035] decrease in the prior upward trend in the non-RA controls (table 5.2). The modelled difference in THR between RA and non-RA patients at the start of the study was 4.30 [95% CI: 3.72 to 4.88] which remained constant throughout the pre-intervention period (table 5.3). However, after the introduction of biologics this delta steadily shrank by -0.10 [95% CI: -0.17 to -0.02; p=0.017] per timepoint throughout the post-intervention study period (figure 5.3b).

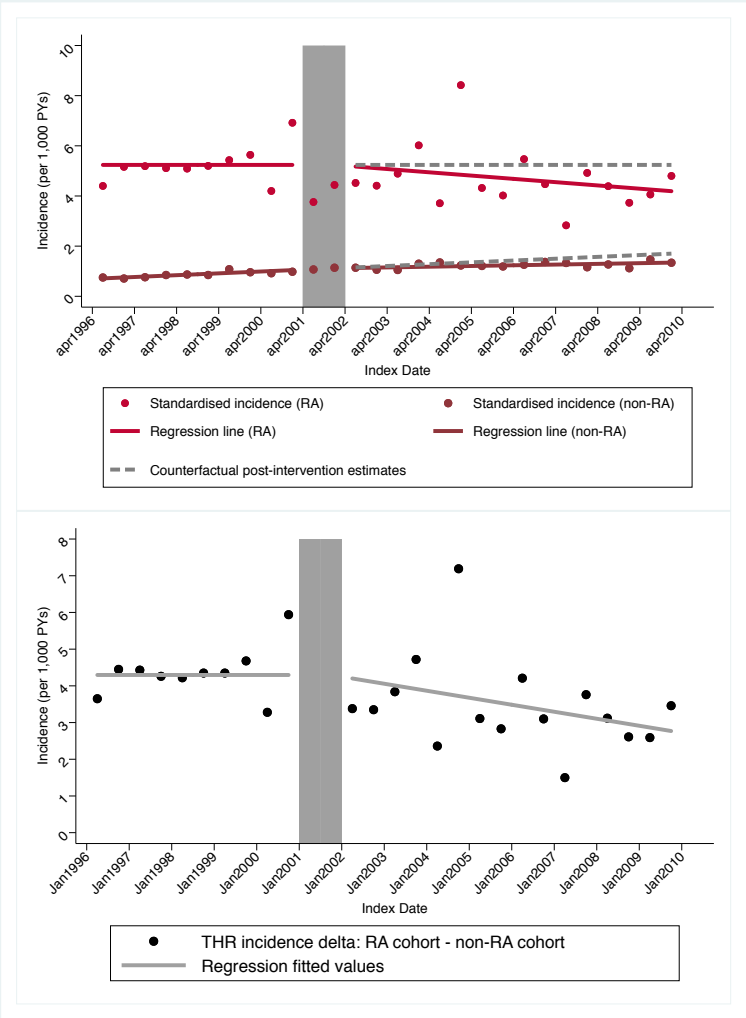


FIGURE 5.3: (A) AGE AND SEX STANDARDISED INCIDENCE RATES WITHIN ONTARIO OF THR WITHIN 5 YEARS OF RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS) BEFORE AND AFTER THE INTRODUCTION OF BIOLOGICS: STRATIFIED BY COHORT; (B) DIFFERENCE-IN-DIFFERENCES IN 5-YEAR INCIDENCE RATES OF THR BETWEEN COHORTS

Based on these coefficients, the estimated incidence of THR at the mid-point of the biologic era was 4.72 (table 5.4). The incidence at the same time point estimated solely by extrapolating the pre-biologic level was 5.24, therefore translating to no significant average change in rates (-0.52 [95% CI: -1.11 to 0.07]) (table 5.4). Over the biologic era, an absolute mean difference of -0.18 was observed in the modelled non-RA THR

incidence rate values compared to counterfactual estimates, i.e. what would have been observed for the non-RA individuals had their pre-intervention level/trend of outcome continued un-interrupted (figure 5.3A, table 5.4). In terms of the difference-in-differences, the coefficients yielded an average post-intervention reduction in differences (compared to counterfactual differences) of -0.76 [95% CI: -1.35 to -0.18] (figure 5.3B, table 5.4)

TKR time-series

Among RA patients, age and sex standardised 5-year incidence of TKR (per 1,000 PYs) at the start of the study period was 6.42 [95% CI: 6.12 to 6.72], which remained stable for the rest of the study period, with no significant changes observed following the introduction of biologics (figure 5.4A, table 5.5). However, among the matched non-RA cohort, the estimated incidence of TKR was 0.77 [95% CI: 0.68 to 0.86], and this increased by 0.03 [95% CI: 0.02 to 0.04; $p<0.001$] per timepoint throughout the pre-intervention period. The intervention period was associated with a further 'step change' increase in rates of 0.19 [95% CI: 0.00 to 0.37; $p=0.048$]. The modelled difference in TKR between the RA and non-RA cohorts was 5.34 [95% CI: 4.92 to 5.76] at the start of the study period which remained constant until the introduction of biologics, at which point it began to shrink by -0.04 [95% CI: -0.09 to 0.02] per timepoint although this was non-significant, $p=0.19$ (figure 5.4B, table 5.6).

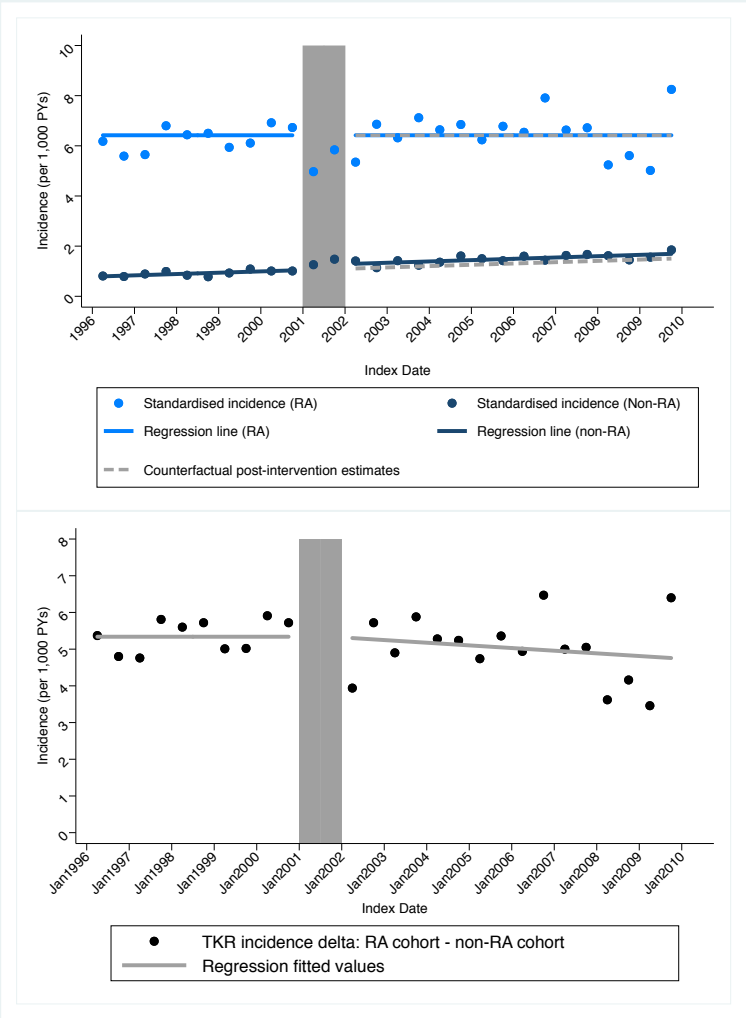


FIGURE 5.4: (A) AGE AND SEX STANDARDISED INCIDENCE RATES WITHIN ONTARIO OF TKR WITHIN 5 YEARS OF RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS) BEFORE AND AFTER THE INTRODUCTION OF BIOLOGICS: STRATIFIED BY COHORT; (B) DIFFERENCE-IN-DIFFERENCES IN 5-YEAR INCIDENCE RATES OF THR BETWEEN COHORTS

Over the biologic era, the coefficients for TKR rate changes yielded a significant mean absolute increase of 0.19 [95% CI: 0.01 to 0.37] in the modelled non-RA TKR incidence rates compared to counterfactual estimates, i.e. what would have been observed in the non-RA cohort had pre-intervention level/trend of outcome continued un-interrupted (figure 5.4A, table 5.7). This meant that overall there was a small and non-significant

reduction in average absolute difference-in-differences of -0.29 [95% CI: -0.71 to 0.13] over the post-intervention timepoints (versus counterfactual estimated differences) (table 5.7).

Prescriptions time-series (seniors only)

Among the senior patient sub-study, the incidence of csDMARD within the first year of RA diagnosis (amongst csDMARD naïve patients) at the start of the study period was 563.33/1,000 PYs [95% CI: 450.18 to 676.49] (table 5.8). This 1-year incidence rate increased dramatically throughout the study period (figure 5.5A), with an apparent (although not statistically significant) increase in both level and trend of prescribing following the introduction of biologics in 2001. The 5-year incidence of biologics was approximately 0 at the start of the study period and increased by 0.27/1,000 PYs throughout the study period, with an increase (although not statistically significant) of 1.59/1,000 PYs [95% CI: -0.74 to 3.93] following the 2001 phase-in period (figure 5.5B, table 5.8).

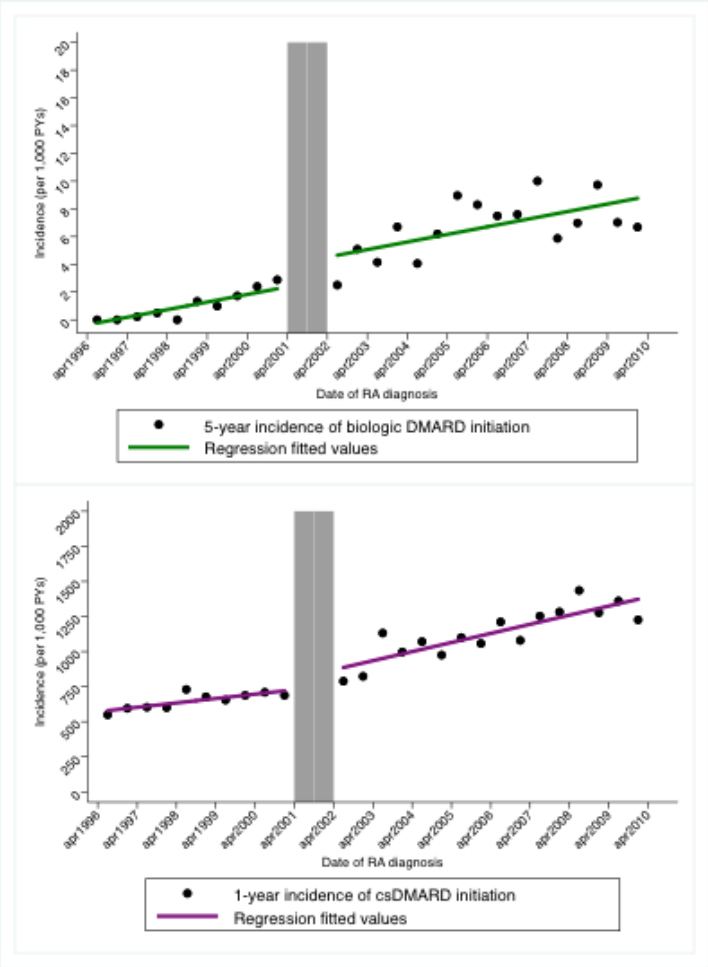


FIGURE 5.5: AGE AND SEX STANDARDISED INCIDENCE RATES OF PRESCRIPTIONS AMONG SENIORS IN ONTARIO FOLLOWING RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS). ESTIMATED BEFORE AND AFTER THE INTRODUCTION OF BIOLOGICS: (A) CSDMARD (1-YEAR) AND (B) BIOLOGIC (5-YEAR)

TABLE 5.8: ITS SEGMENTED LINEAR REGRESSION ANALYSIS FOR SENIOR RHEUMATOID ARTHRITIS PATIENTS WITHIN ONTARIO: 1-YEAR INCIDENCE RATES OF CSDMARD AND 5-YEAR INCIDENCE RATES OF BIOLOGICS (PER 1,000 PERSON-YEARS)

	Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
Ontario, Canada					
csDMARD (1-year)	Baseline Incidence Rate	563.33	450.18	676.49	<0.001
	Trend (per 6 months)	15.65	-2.59	33.88	0.089
	Level change after Biologics	102.34	-53.68	258.34	0.19
	Trend change after Biologics	16.80	-3.53	37.13	0.10
Biologic (5-year)	Baseline Incidence Rate	-0.50	-1.67	0.66	0.38
	Trend (per 6 months)	0.27	0.14	0.41	<0.001
	Level change after Biologics	1.59	-0.74	3.93	0.17
	Trend change after Biologics	-	-	-	-

In terms of the 1-year and 5-year cumulative incidence for csDMARD and biologics, these were approximately 34% and 0%, respectively at the start of the study period but were approximately 62% and 4% by the end of 2009 (appendix table 5.1, figure 5.6A). No changes in the cumulative incidence of csDMARDs was detected following the introduction of biologics in 2001, although there was a difference in cumulative incidence of biologics (figure 5.6).

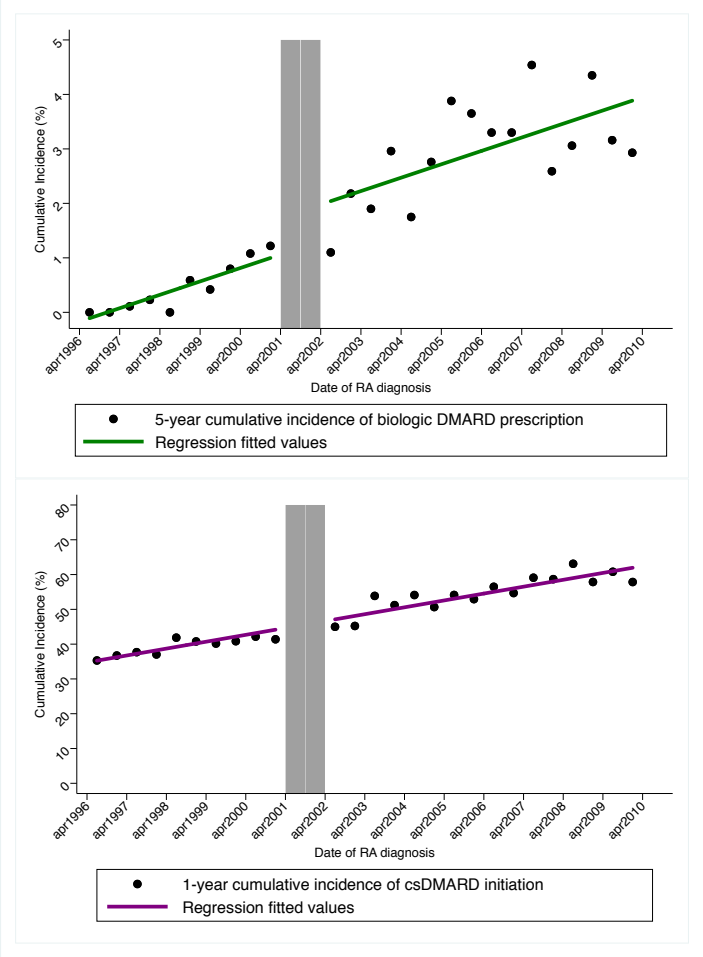


FIGURE 5.6: AGE AND SEX STANDARDISED CUMULATIVE INCIDENCE OF PRESCRIPTIONS AMONG SENIORS IN ONTARIO FOLLOWING RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS). ESTIMATED BEFORE AND AFTER THE INTRODUCTION OF BIOLOGICS: (A) CSDMARD (1-YEAR) AND (B) BIOLOGIC (5-YEAR)

Age stratified THR time-series

For the THR outcome, the incidence (per 1,000 PYs) at the start of the study period among the younger age category (<66 years old at index date), was 3.75 [95% CI: 3.28 to 4.23] among RA patients (figure 5.7A, table 5.9), which was considerably lower than the rate of 11.90 [95% CI: 9.30 to 14.49] for RA patients ≥66 years old (figure 5.8A, table 5.9). For younger RA patients, the introduction of biologics was associated with a subsequent

'step change' decrease in THR incidence of -0.89 [95% CI: -1.49 to -0.28, *p*=0.006], whilst a simultaneous increase (in the form of an upward trend change) was observed in the younger non-RA cohort: 0.37 [95% CI: 0.32 to 0.42, *p*<0.001]. This corresponded to a step-change difference-in-difference estimate of -1.02 [95% CI: -1.61 to -0.42; *p*=0.002] associated with the introduction of biologics (figure 5.7B, table 5.9).

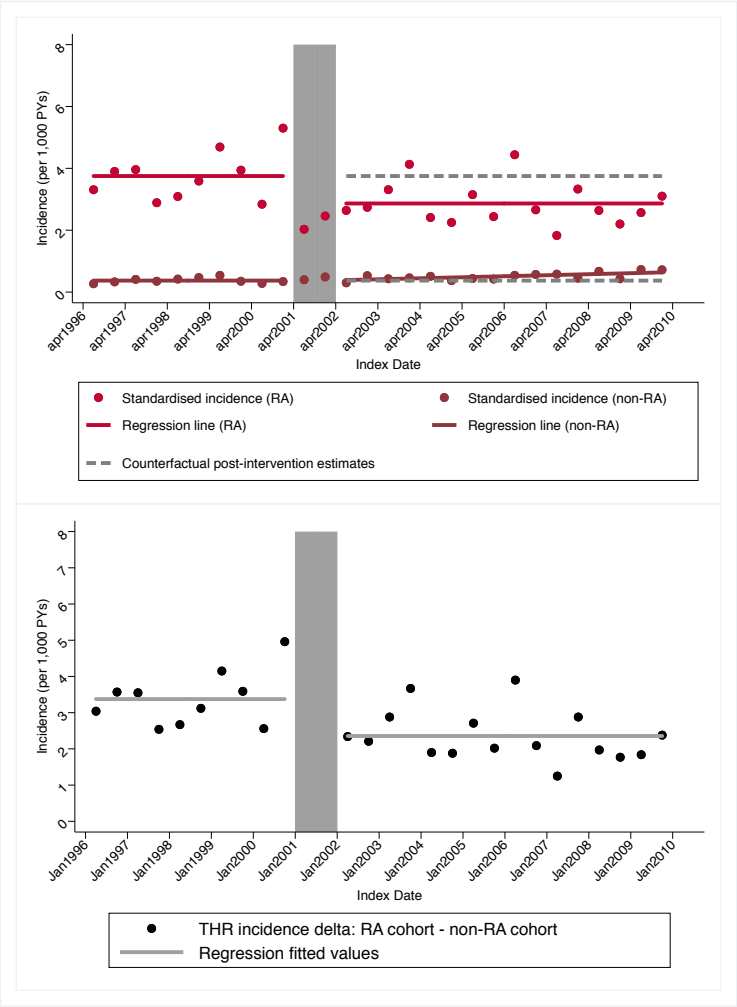


FIGURE 5.7: (A) AGE AND SEX STANDARDISED INCIDENCE RATES WITHIN ONTARIO (<66 YEARS OLD) OF THR WITHIN 5 YEARS OF RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS) BEFORE AND AFTER THE INTRODUCTION OF BIOLOGICS: STRATIFIED BY COHORT; (B) DIFFERENCE-IN-DIFFERENCES IN 5-YEAR INCIDENCE RATES OF THR BETWEEN COHORTS

TABLE 5.9: ITS SEGMENTED LINEAR REGRESSION ANALYSIS OF 5-YEAR INCIDENCE (PER 1,000 PERSON-YEARS) OF THR WITHIN ONTARIO AMONGST RHEUMATOID ARTHRITIS, CONTROL COHORTS AND DIFFERENCES BETWEEN THE TWO: STRATIFIED BY AGE

	Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
<66 years					
RA*	Baseline Incidence Rate	3.75	3.28	4.23	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-0.89	-1.49	-0.28	0.006
	Trend change after Biologics	-	-	-	-
non-RA*	Baseline Incidence Rate	0.37	0.32	0.42	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	0.37	0.32	0.42	<0.001
Diference (RA* - non-RA*)	Baseline Incidence Rate	3.38	2.9	3.85	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-1.02	-1.62	-0.42	0.002
	Trend change after Biologics	-	-	-	-
≥66 years					
RA*	Baseline Incidence Rate	11.90	9.30	14.49	<0.001
	Trend (per 6 months)	0.27	-0.05	0.59	0.091
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-0.54	-1.02	-0.06	0.03
non-RA*	Baseline Incidence Rate	2.28	1.51	3.05	<0.001
	Trend (per 6 months)	0.26	0.17	0.36	<0.001
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-0.25	-0.39	-0.11	0.001
Diference (RA - non-RA*)	Baseline Incidence Rate	9.67	8.36	10.98	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-0.28	-0.45	-0.11	0.003

*RA: rheumatoid arthritis

Among older RA patients, the incidence of THR increased prior to biologics being introduced, but a significant downward inflexion occurred after the phase-in period, equal to -0.54 [95% CI: -1.02 to -0.06, *p*=0.03] per timepoint (figure 5.8A, table 5.9). However, a similar pattern was also observed in the non-RA cohort, although the downward inflexion in the THR incidence trend was not so large for negative controls: -

0.25 [95% CI: -0.39 to -0.11, *p*=0.001] per timepoint, yielding a post-intervention difference-in-difference trend change estimate of -0.28 [95% CI: -0.45 to -0.11; *p*=0.00] per timepoint for the senior age group (figure 5.8B, table 5.9).

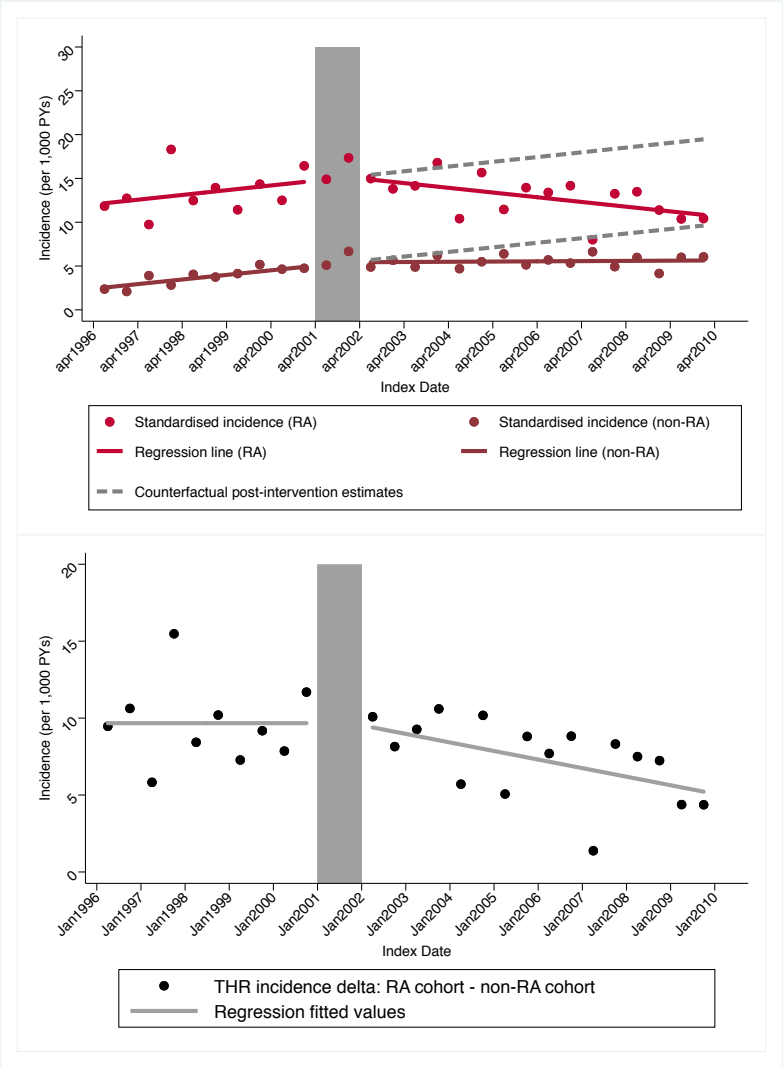


FIGURE 5.8: (A) AGE AND SEX STANDARDISED INCIDENCE RATES WITHIN ONTARIO (≥66 YEARS OLD) OF THR WITHIN 5 YEARS OF RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS) BEFORE AND AFTER THE INTRODUCTION OF BIOLOGICS: STRATIFIED BY COHORT; (B) DIFFERENCE-IN-DIFFERENCES IN 5-YEAR INCIDENCE RATES OF THR BETWEEN COHORTS

Age stratified TKR time-series

In terms of TKR, the incidence at the start of the study period among the younger RA patients was 3.96 [95% CI: 3.34 to 4.58] (figure 5.9A, table 5.10), which was considerably smaller than among those in the senior age category: 18.62 [95% CI: 17.42 to 19.82] (figure 5.10A, table 5.10). The incidence of TKR in RA patients remained approximately stable for both age categories during the study period, although a non-significant upward trend was observed in the younger age group. However, there was an almost identical upward trend in TKR incidence in younger non-RA patients, whilst in the older non-RA group the incidence of TKR increased both in terms of level and slope around the start of the biologic era. This meant that the difference in TKR incidence between the RA and non-RA cohort remained constant throughout the study period for the younger age group (figure 5.9B, table 5.10) but that a difference-in-difference estimate of -0.24 [95% CI: -0.47 to -0.01; p=0.04] was found in the senior age group (figure 5.10B, table 5.10).

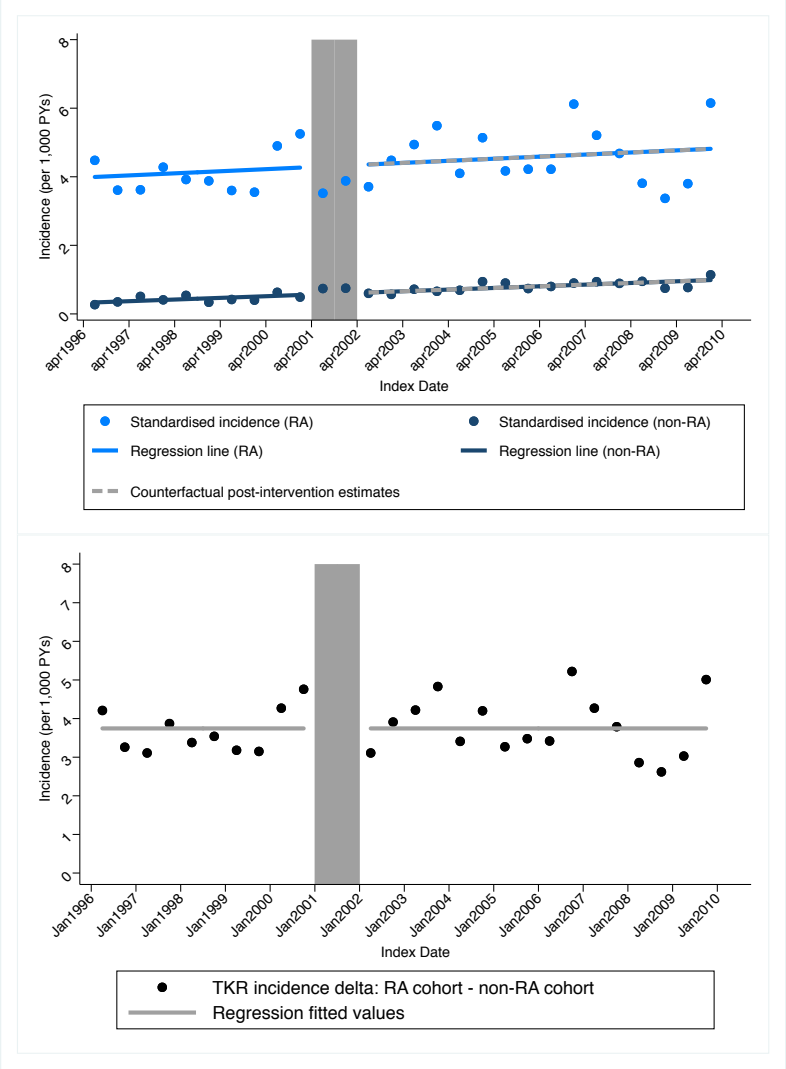


FIGURE 5.9: (A) AGE AND SEX STANDARDISED INCIDENCE RATES WITHIN ONTARIO (<66 YEARS OLD) OF TKR WITHIN 5 YEARS OF RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS) BEFORE AND AFTER THE INTRODUCTION OF BIOLOGICS: STRATIFIED BY COHORT; (B) DIFFERENCE-IN-DIFFERENCES IN 5-YEAR INCIDENCE RATES OF THR BETWEEN COHORTS

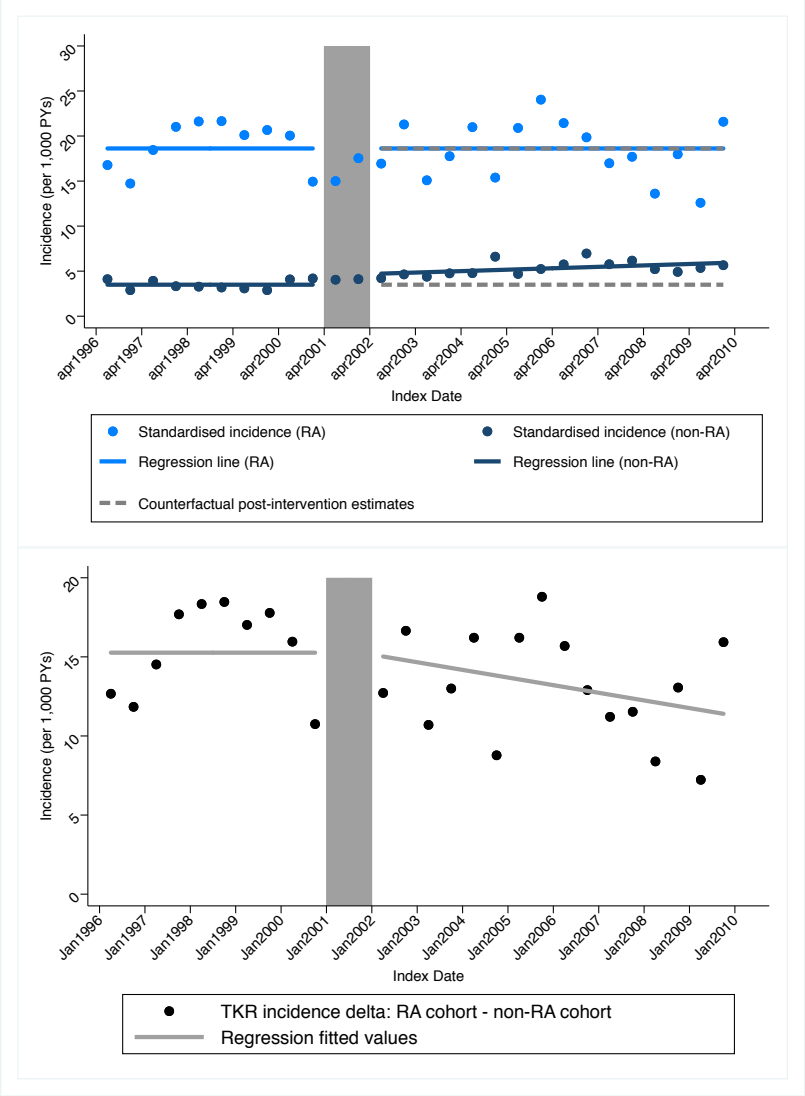


FIGURE 5.10: (A) AGE AND SEX STANDARDISED INCIDENCE RATES WITHIN ONTARIO (≥66 YEARS OLD) OF THR WITHIN 5 YEARS OF RA DIAGNOSIS (CALCULATED AT 6-MONTHLY INTERVALS) BEFORE AND AFTER THE INTRODUCTION OF BIOLOGICS: STRATIFIED BY COHORT; (B) DIFFERENCE-IN-DIFFERENCES IN 5-YEAR INCIDENCE RATES OF THR BETWEEN COHORTS

TABLE 5.10: ITS SEGMENTED LINEAR REGRESSION ANALYSIS OF 5-YEAR INCIDENCE (PER 1,000 PERSON-YEARS) OF TKR WITHIN ONTARIO AMONGST RHEUMATOID ARTHRITIS, CONTROL COHORTS AND DIFFERENCES BETWEEN THE TWO: STRATIFIED BY AGE

	Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
<66 years					
RA*	Baseline Incidence Rate	3.96	3.34	4.58	<0.001
	Trend (per 6 months)	0.03	-0.01	0.07	0.099
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-	-	-	-
non-RA*	Baseline Incidence Rate	0.31	0.23	0.39	<0.001
	Trend (per 6 months)	0.02	0.02	0.03	<0.001
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-	-	-	-
Diference (RA* - non-RA*)	Baseline Incidence Rate	3.75	3.46	4.03	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-	-	-	-
≥66 years					
RA*	Baseline Incidence Rate	18.62	17.42	19.82	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-	-	-	-
non-RA*	Baseline Incidence Rate	3.50	3.08	3.92	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	1.15	0.33	1.97	0.008
	Trend change after Biologics	0.08	0.01	0.15	0.034
Diference (RA - non-RA*)	Baseline Incidence Rate	15.27	13.53	17.00	<0.001
	Trend (per 6 months)	-	-	-	-
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-0.24	-0.47	-0.01	0.04

*RA: rheumatoid arthritis

COMBINED POPULATION-LEVEL FOREST PLOT

The absolute rate differences between average post-intervention fitted values compared to counterfactual estimates amongst the various RA cohorts are presented in figure 5.11. This indicates that overall rates of THR remained approximately stable post-introduction

of TNFi, although non-significantly elevated in Denmark, whilst there was an estimated average reduction in rates of TKR across the different populations. Difference-in-difference analyses revealed the delta (per 1,000 PYs) between RA and non-RA cohorts shrank for both the THR and TKR outcome in Denmark and Ontario (figure 5.12).

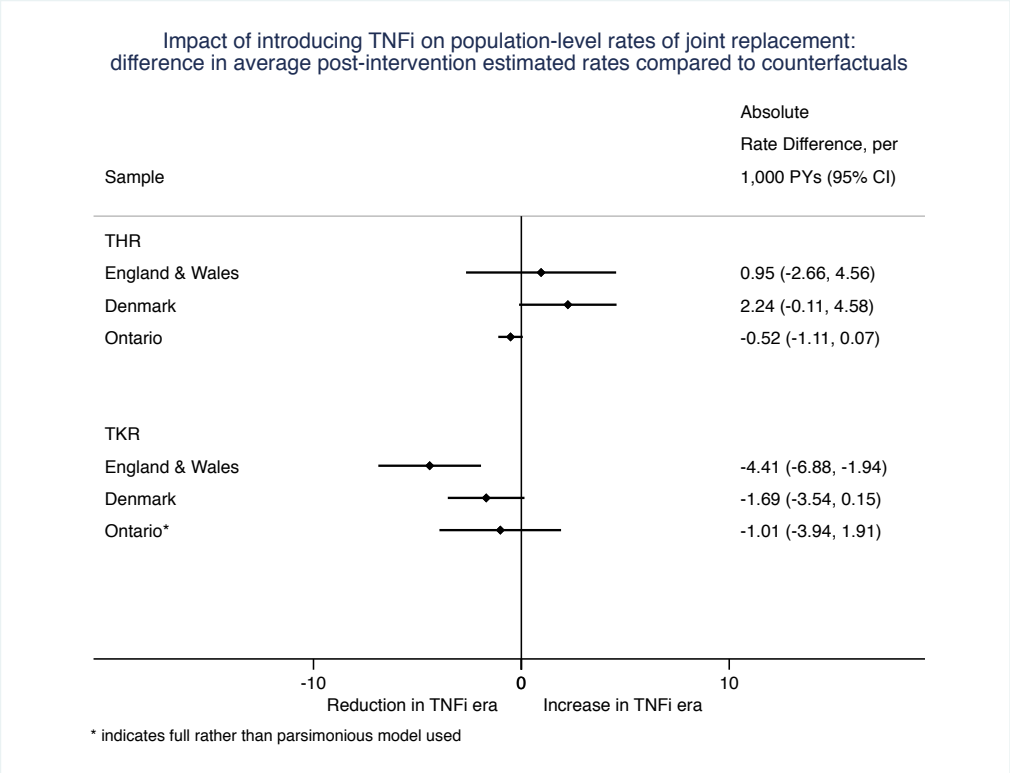


FIGURE 5.11: FOREST PLOT OF ESTIMATES OF AVERAGE ABSOLUTE RISK DIFFERENCE BETWEEN TNFI-ERA (FITTED VALUES FROM ITS REGRESSION MODELS) COMPARED TO COUNTERFACTUAL ESTIMATES, AMONGST RA COHORTS

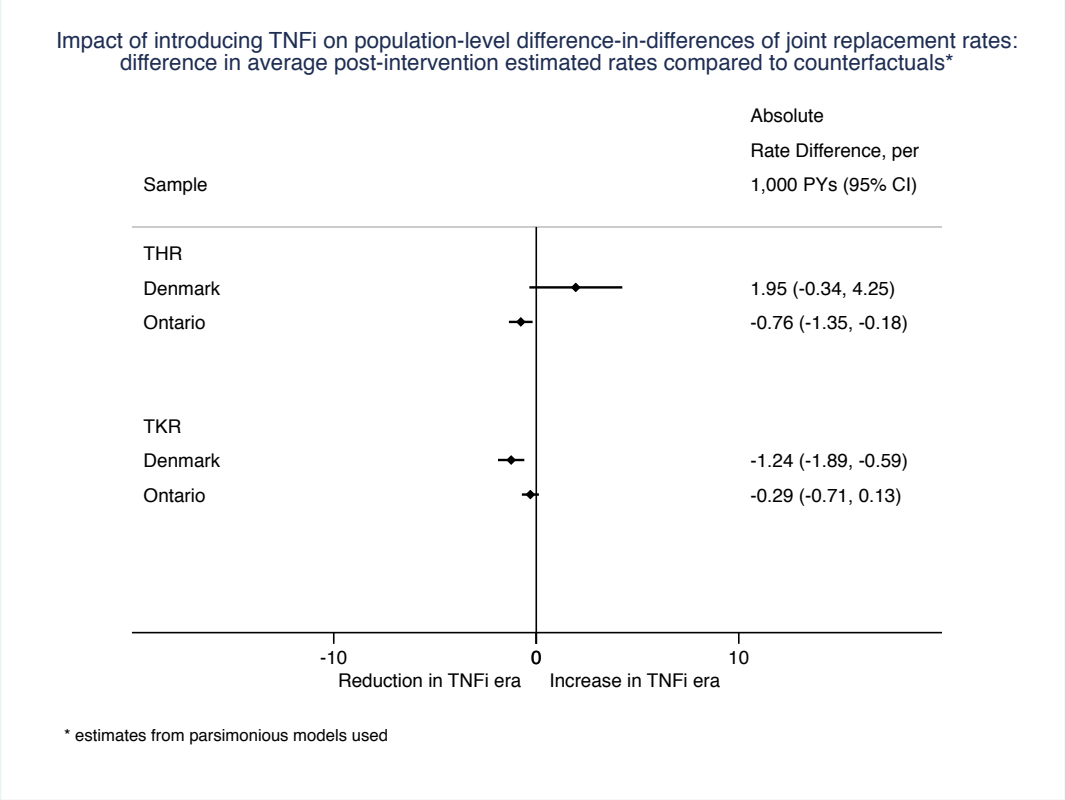


FIGURE 5.12: FOREST PLOT OF DIFFERENCE-IN-DIFFERENCES IN JOINT REPLACEMENT RATES: POST-INTERVENTION DIFFERENCE IN AVERAGE ESTIMATED DIFFERENCES (BETWEEN THE DISEASE COHORTS) COMPARED TO COUNTERFACTUAL DIFFERENCES

DISCUSSION

Main findings

The aim of this chapter was to use external datasets to replicate the investigation of the impact of introducing/approving biologic therapies for RA patients in England and Wales (chapter 4). This replication has been done using two large cohorts of RA patients from Denmark and Ontario, and additionally a control population of non-RA individuals from each region in order to allow more precise interpretations in regard to the specific effect

of biologics (203). The main findings were that whilst rates of joint replacement were markedly higher in RA patients compared to non-RA patients at the start of the study, the introduction of biologics was associated with the initiation of a significant closing of this incidence gap for TKR (Denmark) and THR (Ontario). This was clearly driven by downward inflexions in temporal trends of these outcomes in the RA patients which occurred following availability of biologics, and to an extent that was greater than any downward inflexions in the general population. However, the post-intervention decline in the 'excess' joint replacement rate for RA patients (versus non-RA individuals) in Ontario was not statistically significant for TKR, whilst this quantity for THR in Denmark was actually non-significantly larger during the biologic era despite an overall decline throughout the study period.

Findings in context

The Danish findings of reduced rates of TKR in RA patients following the introduction of biologics corresponds well with the findings from the previous analysis of CPRD data for England and Wales, as does the momentary increase in THR proceeded by a downward trend. The use of almost identical methodology, similar healthcare systems and geographically close region yielding consistent results in both the UK and Denmark is reassuring and strengthens the case for some causal link between the introduction of biologics and reduced TKR rates amongst RA patients. This is further underlined by the use of a non-RA control cohort in Denmark to compare the results of the RA patients to, which revealed that the reduction in RA patients was far greater than that in non-RA individuals (figure 5.2B). This also accords with a study from the Republic of Ireland which

reported on the descriptive trend in joint replacements for RA over a 15-year period and found a profound reduction in numbers of knee but not hip replacement within RA patients (94). However, in that analysis when the increase in hip and knee procedures for the general population was considered then a reduction in both hip and knee arthroplasties in RA relative to the general population was evident. Although the lack of a reduction in THR relative to the general population in the Danish data is not supported by this previous Irish investigation, the reduction in THR incidence in RA patients (relative to non-RA individuals) in Ontario is. The findings from Ontario are somewhat supported by a cohort study in Sweden which also found a reduction in the incidence of hip arthroplasty in RA patients during the biologic era (compared to pre-biologic), but not a reduction in incidence of knee arthroplasty (102). As was noted in chapter 4 though, this Swedish study should be interpreted with caution given no accounting was made for secular trends. Other previous studies have reported no temporal changes in the 5- or 10-year cumulative incidence of major joint surgery in RA (97, 98) or prevalence of RA amongst hip and knee arthroplasties performed in the US general population (38) (although a different investigation of the same study population used for this latter US study found significant reductions in numbers of both hip and knee replacement for RA when only 'primary' RA was considered (134)).

Possible mechanisms

The discussion in chapter 4 highlighted that a reduction in joint replacement rates following the introduction of biologics was a plausible finding given the wealth of prior RCT data indicating biologics for RA are effective in terms of both ACR response criteria,

quality of life and arrest of progression of structural damage (50). It seems reasonable, given this background, that a very valid explanation for the reduction in TKR rates is that the introduction of biologics ushered in greater suppression of inflammation and subsequent joint damage that expressed itself in an eventual reduced need for joint replacement procedures. The point was previously made that the knee maybe a more heavily and/or earlier effected joint in RA than the hip, e.g. the knee does not feature in either the 1987 ACR diagnostic criteria of DAS28 score (4, 191). This may explain a lack of reduction observed in THR rates within UK and Danish RA study populations, although this explanation does not fit the Ontario data. Here the reduction in joint replacement procedures within RA was observed at the hip (figure 5.3A) rather than knee (figure 5.4A). While the reasons for these differences in the Canadian data cannot be ascertained using the available data, an interesting phenomenon only observed in the Canadian data was that rates of knee replacement actually increased in the non-RA population during the biologic era. Factoring this in (with the difference-in-difference time-series) did suggest a small positive impact of biologics in RA patients (figure 5.4B, figure 5.12), which while being non-significant for the main analysis was far more pronounced for the senior sub-group (figure 5.8B).

While attributing a role to the introduction of biologics in the decrease in TKR (UK and Denmark) and THR (Ontario) is reasonable given the temporal association of changes and the fact that - on the whole - changes were greater in the RA cohorts than the non-RA cohorts, there is still reason to be cautious about this interpretation. Firstly, there is the lack of data on biologics available for the main analyses in either in the UK, Denmark or

Ontario. Not only so, but where prescription data were available (senior sub-group in Ontario), they indicated that the uptake of biologics was actually rather small (figure 5.5B & 5.6B), specifically that even at the end of the study period the 5-year cumulative incidence of biologic initiation was approximately 4%. Furthermore, the introduction of biologics was as much associated with an increase in the incidence of csDMARDs as it was biologics in the senior sub-group (figure 5.5A), and although no investigation was made concerning the dose of methotrexate prescribed, this may also have increased. This is understandable given that in Ontario, as in the UK, csDMARDs must have been tried and deemed to have been inadequate in controlling the disease before TNFis can be prescribed. This does however introduce a key confounder into the 'introduction of biologics' versus joint replacement relationship as here reported (106), although this was only observed for the senior sub-study and so should not automatically be assumed to necessarily be mirrored for the main analysis. Prior data also indicates that prescribing of csDMARDs in Ontario also increased markedly during the study period (132). Conversely, as mentioned in chapter 4, no sustained impact on methotrexate prescriptions in the UK coincided with the advent of biologics, as evident from a report of a prior investigation (194).

The second reason to be cautious about interpreting the reduction in joint replacement rates as caused by the introduction of biologics is that there were - in the main - concomitant reductions in the non-RA cohort (figures 5.1 – 5.3). As discussed above, these reductions were not as great as those in the RA patients, thereby suggesting some general factors might be at play in both (RA and non-RA) cohorts but also some specific

role of the introduction of biologics 'above and beyond' these general factors particular to the RA patient group. However, any reduction in the non-RA group is still cause for concern and weakens the case for a dogmatic interpretation in favour of biologics. One tentative possibility is that changes in the THR/TKR rates in the non-RA population during the post-intervention study period suggests a change in need and/or provision of THR/TKR for OA more generally. Despite a scarcity of data on the prevalence of OA amongst RA patients, there is some evidence to suggest this is substantially higher in RA patients than in the general population (211). This introduces a second confounder into the 'introduction of biologics' -> joint replacement relationship, i.e. the possibility of a greater concentration of OA in the RA cohort. Under this scenario one might naturally expect changes occurring in the non-RA group due to changes in prevalence/management of OA to be amplified in the RA group due to a greater concentration of OA therein. Despite this being a genuine concern, it is worth noting that certain aspects of the data still indicate this would be an inadequate explanation. For example, in the sub-group analysis in Ontario, TKR rates in the senior sub-group (figure 5.10) and THR rates in the younger sub-group increased markedly among non-RA individuals whilst the RA cohort experienced either a decrease or no change at all.

Other potential mechanisms are also possible, largely because of the impossibility of ruling out confounding factors. For example, the introduction of biologics may in itself have changed surgeon preference for operating rather than merely acting at a physiological level. That is, maybe with the availability of a new therapy and optimism surrounding the possibility of a much better patient prognosis that clinicians opted to

delay surgery in favour of pharmacotherapy. Neither can other factors be ruled out, such as a growing emphasis on earlier diagnosis/treatment (196), a natural evolution toward a milder form of RA or changes in the prevalence in other confounding characteristics such as smoking or obesity. Whilst these latter possibilities could readily influence general trends (for example the pre-biologic downward trend in THR incidence in Denmark), it still seems unlikely that they'd be responsible for some of the pronounced inflexions in joint replacement trends observed at the advent of biologics. One interesting option is the immediate increase observed for THR rates in Denmark in 2003 could have a link to the new government that formed at that time which anecdotally emphasized the importance of reducing operation waiting lists.

Limitations

There are numerous limitations to this investigation. As mentioned above, one cannot rule out the presence of confounding factors that may have changed concomitantly with the introduction of biologics and thereby introduced bias. The ITS approach assumes these factors to be either absent or inconsequential, which may not be the case here. Further work is evidently needed to explicitly estimate the impact of biologic use versus non-use at the patient level, and this will be further discussed and addressed in chapter 7. The use of population-based routinely collected "big data" as leveraged here means individual validation of either the exposure or outcome status of each individual was not carried out, although codes for the identification of RA patients in both Denmark and Ontario have previously been validated and shown to have a high degree of accuracy (130, 131, 206, 212). The use of NOMESCO and CCI codes to identify THR/TKR events are

standard and recognized coding structures although the individual validity of these for this outcome in these data sources has not been shown. The ITS approach is considered a strong quasi-experimental method, however the use of segmented linear regression as used here did impose certain assumptions onto the data, for example linearity of effect. This assumption was met for most of the models although there was some exception, for example the senior sub-group analysis of TKR as evident in figure 5.10B. Whilst the use of a control group was obviously a strength, the exact interpretation of results was limited owing to the nature of this comparator group. Specifically, in an ideal investigation the control group would greatly resemble the 'exposed' with the exception of exposure status. Here however, non-RA individuals had much lower baseline incidence of joint replacement and, as already discussed, may have been subject to confounding to a greater/lesser extent than the actual RA group. For this particular research question and data availability though, it is hard to conceive of a more appropriate control group than matched individuals without any inflammatory arthritis. However, recent developments on this subject are currently underway in the literature, with concepts such as propensity score weighted regression possible to account and correct for differences in characteristics between treated and control time-series (203, 213).

Despite efforts to harmonise methodology across the analyses of different study data, certain differences unfortunately remained mainly due to the sequential nature of the analytical process and the honing of processes along the way. For example, the UK data were provided without making exclusion of RA patients with a prior THR/TKR, whilst

individuals with a prior THR/TKR (at any time prior to index date) were excluded for the respective analysis of that outcome in the Danish analysis. In the Canadian data, given logistic/financial restrictions it was decided any individual with a prior THR or TKR in the 5 years prior to index date would be excluded for all analyses. Although this likely gave rise to only minor consequences in terms of exclusions, it was not ideal. Similarly, whilst it was possible to use the available CPRD data to explore changes in other patient characteristics over the study period, this data on patient characteristics was much more limited for the Danish and Canadian analysis.

Related to this point is that the ITS using CPRD data from England and Wales was originally meant to incorporate an OA comparator group with which to control for secular trends. This however proved unsuitable upon inspection of the data, where it became apparent that a large proportion of the OA cohort received a joint replacement in their first year of follow-up, and so indicated that very severe/late-stage OA was primarily being captured and was being coded as such just prior to a patient undergoing arthroplasty. For this reason, general-population controls were used for the replication analyses. Furthermore, the sample size of 30,000 RA patients in the DK analysis permitted 10 controls to be matched to each patient, although this was reduced to 5 controls in Ontario as the computational requirements and expense of matching 10 controls to such a large number of RA patients (~100,000) seemed unjustifiable given previous literature suggesting two controls may be sufficient (214). On the other hand, it was due to this much larger sample size in the Ontario investigation that an age-stratified approach was taken, whilst this was deemed unsuitable for the UK & DK

settings where it was known prior to ITS modelling that the sample size was smaller and for which the time series appeared relatively 'noisy' (large standard deviation around the regression line) without further stratification. As mentioned previously, this precise issue of selecting a format of aggregation that retains sufficient statistical power to detect changes is investigated in chapter 6.

Strengths

One of the most obvious strengths of this study is that it has investigated two independent populations in order to validate findings from a third population (England & Wales). This has facilitated the consistency of methodology and comparability of findings across populations (figures 5.11 & 5.12), which will be further discussed in chapter 8. Both the study samples used in this chapter were highly representative of their respective populations given they were nation-wide (Denmark) and region-wide (Ontario), and collectively were very large with >900,000 individuals included (considering those in the RA and non-RA cohorts). This not only increased the power available for detecting smaller intervention effect sizes, but allowed stratification by age. The use of multiple controls matched to each RA patient most likely more than compensated for the lower incidence rates in the non-RA groups, which otherwise would have caused a much noisier control time-series in a 1:1 matching. The use of a controlled/multiple group ITS design has permitting the control of not only secular trends in the RA group (which the single time-series in chapter 4 had already done) but also better control of other contemporaneous events impacting rates of outcome in both groups that may have confounded the estimate of biologic therapy availability in the

single (RA) group time-series (203). The availability of prescriptions data for the senior age sub-group in Ontario is another strength, allowing insight into the impact of the intervention on the most immediate outcome it was meant to target, i.e. use of biologics.

Conclusions

This chapter reports that the introduction of biologics in Denmark was associated with a reduction in TKR rates in RA which although non-significant in itself, represented a profound reduction when compared to rates in the general population (figure 5.2B). Consistent with the UK analysis, no overall impact on THR rates was found during the biologic era in Denmark. Within Ontario (Canada), the start of the biologic era was found to be associated with a (non-significant) reduction in rates of THR but not TKR. Compared to non-RA patients in Ontario, the reduction in THR represented a much more pronounced impact whilst a (non-significant) impact in TKR was also implied. Significant impact towards lower THR and TKR - as compared to the general population - were observed among the senior aged sub-group in Ontario. The introduction of biologics was followed by an increase in prescribing of both biologics and csDMARDs in the only population for which data were available (senior patients in Ontario). Given the probable element of ecological fallacy, i.e. confounding factors changing concomitantly to the introduction of biologics, notably increased usage of csDMARDs, further work is required to estimate the effect of biologics at the patient level, while addressing confounding factors. This will be addressed in chapter 7.

6. POWER AND SAMPLE SIZE CONSIDERATIONS IN INTERRUPTED TIME-SERIES ANALYSIS

INTRODUCTION

The previous two chapters have reported on analyses of population-level trends in THR and TKR amongst RA patients and non-RA controls. The impact of the introduction of biologic therapies on these trends has been estimated using an ITS approach, which is a method that is being increasingly used in epidemiology (173, 174, 215). It is an accessible and intuitive method that can be straight-forward to implement and has considerable strengths (170), described recently as "*the "next best" approach… …when randomisation is not possible*" (170). A common application is when population-level repeated measures of an outcome and/or exposure are available over time, both before and after some well-defined intervention such as a health policy change (173, 174, 216) or a naturally occurring event of interest (175, 176).

Despite the substantial growth in use of ITS methods, relatively little practical guidance has been developed in terms of methodological standards within the ITS framework (174, 215), including a scarcity of guidance on required sample size. Sample size planning is often a key component of designing a study and should be conducted prior to analysis (217), although this is an aspect very often overlooked in ITS studies, with many being underpowered (218).

In a traditional study involving frequentist hypothesis testing, the size of the sample being studied is one of the key determinants of the power of the test, i.e. the probability of correctly identifying an effect that truly exists in the population (or more formally, the probability of not failing to reject the null hypothesis of no effect when it should be rejected). A type 1 error is made when an observed effect size (e.g. between two groups) is concluded sufficient evidence to reject the null hypothesis, when in fact such an observed effect is merely the product of sampling variability and the null hypothesis should not actually be rejected. The type 1 error rate (i.e. alpha) is the rate at which null hypotheses would be incorrectly rejected in a hypothetical long-run of repeated sampling. When this error rate is very small, it is indicative of sufficient improbability of obtaining a given effect size from a particular study were the reality to be that the null hypothesis was in fact true. Traditionally the rate of 5% (i.e. 0.05) has been widely accepted in the medical research literature to indicate sufficient evidence for a finding to be considered "significant" rather than just chance, although the continuum nature of probability dictates the actual value of *p*-values be properly interpreted rather than judged merely by which side of 0.05 it falls. Conversely, a type 2 error (i.e. beta) is when a null hypothesis of no effect is not rejected when it should be, i.e. when there is a true effect/inequality in existence within the population but the sample data fails to adequately give evidence for it. The type 2 error determines a studies power (1-beta), meaning the power of a test can be interpreted as the probability of correctly rejecting the null hypothesis when in reality it should indeed be rejected due to a real difference in the population. Traditionally in the medical research literature a value of 80% (i.e. 0.8)

is usually required for a study to be considered to have acceptable power, although values of 90% would obviously be preferable (219).

Given these concepts, a studies sample size therefore needs to be sufficiently large in order to permit the studies primary hypothesis test to detect an effect of pre-specified size where it may truly exist in the population, with 80% power and at a <0.05 alpha. Assuming the 'conventional' values for type 1 and 2 errors, the real determinants of required sample size is the magnitude of a hypothesised effect size in conjunction with the variability (i.e. standard deviation) of the sample data.

However, in the analyses employed in previous chapters there was no formal assessment of statistical power carried out either before or after analysis. This doesn't have huge implications for how some of the specific findings of "no effect" should be interpreted as the uncertainty of estimates due to sampling variability is expressed in the 95% confidence intervals. Furthermore, post-hoc power calculation is not recommended (219). None-the-less, carrying out an underpowered analysis does in principle undermine the scientific credibility of a given investigation because it means the true effect size of relevant magnitude cannot be detected with sufficient confidence and therefore the study remains - to a greater or lesser extent - uninformative. Furthermore, it's been shown that in extreme cases of underpower, the product is an average vast over-estimation of the statistical significance of findings (220, 221). Although the ITS method has many strengths, if a given analysis is not adequately powered it may therefore lead to publication of not only weak but spurious findings (220, 221).

Information on the power associated with various numbers of repeated measures of an outcome (i.e. timepoints) has been previously reported for the ITS framework (222), with rules of thumb concerning the minimum number of pre- and post-intervention timepoints needed, such as three, (215) six, (223) eight, (182) and ≥10. (218) In this regard, all our analyses in previous chapters would be deemed to be adequately powered given that they all had ≥10 timepoints in final analyses, both before and after the introduction of biologics. However, these simple rules of thumb do not take into account the practical issue of considering a suitable underlying sample size of subjects/patients per aggregate timepoint (224). While longer time-series have been shown to have more power than short time-series, it seems reasonable to propose that ITS analyses (even those with many timepoints) with only a small number of subjects per timepoint may contain so much noise as to render it improbable of detecting a true impact of an intervention under study. Researchers seeking to aggregate patient-level data into a population-level timeseries in order to conduct an ITS are confronted with the trade-off between increasing the length of a time-series (at the expense of the number of subjects per timepoint) and increasing the number of subjects per timepoint (at the expense of the length of the time-series). The trade-off that optimises power, or at least ensures ≥80% power is clearly preferable, however guidance on assessing the available power to detect a specified "interruption" within the context of this trade-off is not provided in the literature.

Given this paucity of guidance on sample size calculation, the aim here was to use a simulation approach to estimate power in an ITS analysis, taking the UK analysis (as reported in chapter 4) as a case-study. The main objective was to quantify the power available in relation to the underlying sample size per timepoint, while varying a number of key parameters of interest, notably the number of total timepoints in the time-series. Furthermore, the intention was to develop Stata code to be readily usable by epidemiologists as a tool to generate estimates of required sample size for similar ITS applications.

METHODS

Study design

This was a Monte-Carlo simulation study, the strengths of which have been well described previously (225, 226). Pseudo-random data were generated with known characteristics/distributions defined by pre-specified input parameter values. Consequently, because the truth regarding these characteristics was known, it was possible to empirically evaluate the performance of a given statistical model when fitted to the simulated data (227, 228). The methods are here described according to the recently published guidelines by Morris *et al.* (227)

Aims

The aim was to describe the power associated with the mean sample size per timepoint to detect a change in: (i) level and (ii) trend in an outcome, following a defined

intervention in the ITS framework, using ordinary least-squares regression. The repeated outcome measure (timepoint) took the form of a proportion given that analyses in chapter 4 were initially performed using cumulative incidence (which was only subsequently updated to incidence rate). The potential impact of a range of values for various other factors such as total number of timepoints, effect size and location of intervention in the time-series were also investigated. The aim was primarily to develop a means to estimate power within the context of the specific case-study evaluating the impact of UK NICE TA36 on the cumulative incidence of TKR within the Clinical Practice Research Datalink (CPRD). (199). However, the intention was also to generalise the programming code developed in order to facilitate the estimation of power (and therefore the estimation of required sample size) when planning future ITS studies.

ITS scenarios

There are many factors within an ordinary least squares ITS framework that could conceivably influence the power to detect the impact of an intervention. Although the following is not an exhaustive list, it describes the main factors that were here investigated.

1. Nature of intervention impact (figure 6.1A)

The impact of an intervention can be modelled as a 'step' change in the level of outcome and/or a 'slope' change in the trend of outcome (170, 171). More complex realities can be incorporated such as multiple interventions, waning or delayed effects and non-linear

responses. (171, 173) However, for the purposes of the current work intervention effects were considered to be mediated through either: (i) a step change or (ii) a slope change

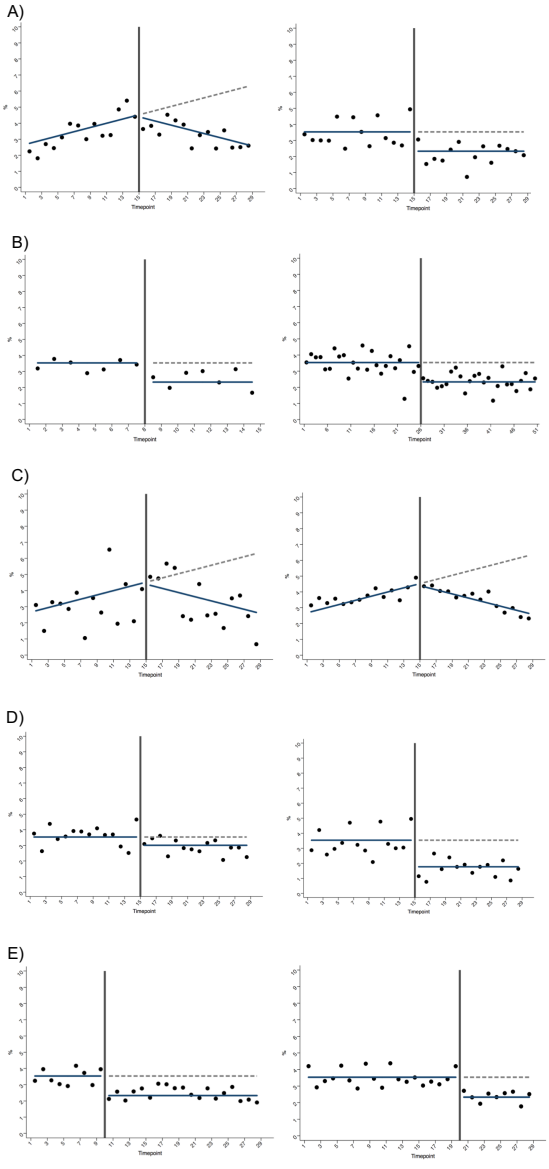


FIGURE 6.1: EXAMPLE SIMULATION SCENARIOS FOR: (A) SLOPE CHANGE (LEFT) VERSUS STEP CHANGE (RIGHT); (B) SHORTER TIME-SERIES (LEFT) VERSUS LONER TIME-SERIES (RIGHT); (C) SMALLER SAMPLE SIZE PER TIMEPOINT (LEFT) VERSUS LARGER SAMPLE SIZE PER TIMEPOINT (RIGHT); (D) SMALLER EFFECT SIZE (LEFT) VERSUS LARGER EFFECT SIZE (RIGHT); (E) EARLIER INTERVENTION (LEFT) VERSUS LATER INTERVENTION (RIGHT)

2. Total number of timepoints in the time-series, N (figure 6.1B)

The ITS approach requires repeated observations of an outcome event over time, usually at equally spaced intervals such as days, weeks, months, quarters, years, etc. As described in the introduction, the number of these timepoints has previously been shown to impact power. Nine values for total number of timepoints (N) were investigated, ranging from 6 to 50.

3. Number of subjects per timepoint, n (figure 6.1C)

The sample size per timepoint will impact the accuracy of outcome estimates and hence the dispersion of a given time-series. It is therefore an important factor influencing the power to detect an 'interruption'. To investigate this, 11 values for n were specified, ranging from approximately 150 to 5,700 patients per timepoint. The rationale for these values was that in the chapter 4 TKR time-series (which was the basis for this case-study), these values corresponded to a mean number of outcome events per timepoint that ranged from 5 to 200 (supplementary file 6.1).

4. Effect size – i.e. magnitude of intervention impact (figure 6.1D)

One of the assumptions of ITS analysis is that the pre-intervention level and trend of outcome can be used to predict post-intervention counterfactual estimates, i.e. what outcomes would be expected in the post-intervention period had the intervention not

occurred (171, 173). The impact of intervention can then be expressed as the difference between the estimated counterfactual outcome value for a given post-intervention timepoint versus the estimated modelled outcome value for the same timepoint using the observed data (179). In previous studies using segmented linear regression, this has often been done for the mid-point of the post-intervention period in order to yield an average post-intervention change. (199, 216, 229) The magnitude of this average post-intervention change expressed as a relative % was therefore used to express effect size, defined for mid-time-series interventions as the step or slope change resulting in a -15%, -34%, -50% and -75% reduction in the mean cumulative incidence.

5. Mean pre-intervention level and trend of outcome (figure 6.1E

The absolute pre-intervention level of outcome is an important factor. For example, a relative 50% reduction of a common outcome will be easier to detect than a relative 50% reduction of a rare outcome. Furthermore, a pre-intervention trend in outcome may exist, which may also have an effect on power. Two parameters were therefore specified: the mean pre-intervention outcome value (defined using the pre-intervention mid-point) in conjunction with a pre-intervention trend parameter. In main analyses only two scenarios were explored for this parameter (based on our prior CPRD study (199)), where mean pre-intervention cumulative incidence was 3.5% and where there was either: (i) no pre-intervention trend (for step change scenarios) or (ii) an upward trend (for slope change scenarios) (figure 1). Trend parameters were scaled according to N so that absolute pre-intervention values were constant across all mid-time-series intervention scenarios. Exact parameter values for these are provided in supplementary file 6.1.

6. Location of intervention in time-series (figure 6.1E)

Related to N, location of intervention in the time-series may also have an impact on power as this will affect the balance in number of pre-intervention and post-intervention timepoints to be modelled. Locations investigated were at: one-third, mid-way and two-thirds from the beginning of the time-series. For trend change scenarios, the same pre-intervention and post-intervention trends were used when investigating early/late interventions as per the corresponding mid-way intervention setting within each N scenario (supplementary file 6.1). This seemed the most realistic approach given that an intervention should have the same impact on the immediate level and/or trend irrespective of when that intervention occurs across the study period.

Data Generating Process

Data were generated in Stata v15.2, the general principles of which have been described elsewhere (230). Empty time-series datasets were created of length N (total number of timepoints). Three ITS variables were inserted: timepoint identifier (integer), post-intervention indicator (binary) and post-intervention timepoint identifier (integer) (171). The timepoint identifier was created first, then used in combination with the 'location of intervention' parameter to generate the other two ITS variables. The underlying sample size for each timepoint (n_t) was simulated from a normal distribution with mean n (a key parameter of interest; 11 values investigated) and standard deviation of n/3. The number of outcome events occurring per timepoint were drawn as a binomial random variate (n_t, p_t), where n_t represents the sample size and p_t the probability of outcome. p_t was a linear

function defined using the ITS variables in combination with other scenario-specific parameter values (equation included in table 6.1). The number of events per timepoint and n_t were used to derive the cumulative incidence time-series. A total of 1,000 Monte-Carlo repetitions were carried out for each unique scenario.

Methods of analysis

A segmented linear regression model was fitted to each created dataset. This took the form of model (1) for step change scenarios and model (2) for slope change scenarios:

(1) $Y_t = \beta_0 + \beta_1 * \text{timepoint}_t + \beta_2 * \text{intervention\_indicator}_t + e$.

(2) $Y_t = \beta_0 + \beta_1 * \text{timepoint}_t + \beta_3 * \text{post\_intervention\_timepoint}_t + e$.

Here, Y_t is the value of outcome at timepoint t. β_0 estimates the level of the outcome just before the beginning of the time-series. β_1 estimates the pre-intervention trend, β_2 the change in level between the time point immediately before vs. after the intervention and β_3 the change in trend occurring immediately after the intervention. e is the error term.

TABLE 6.1: SIMULATED VALUES FOR CALCULATING PT^ FOR MID-TIME-SERIES INTERVENTION SCENARIOS RESULTING IN A RELATIVE REDUCTION (EFFECT SIZE) FROM A MEAN PRE-INTERVENTION INCIDENCE OF 3.5%: BY 'TYPE' OF IMPACT AND TOTAL NUMBER OF TIMEPOINTS IN TIME-SERIES ±

						Post-intervention Relative Reduction							
						-15%		-34%		-50%		-75%	
Type of intervention impact	Total no. timepoints	Pre-intervention mid-point value	Intercept	Pre-intervention slope *	Expected post-intervention mid-point value**	observed post-intervention mid-point value	derived post-intervention slope	observed post-intervention mid-point value	derived post-intervention slope	observed post-intervention mid-point value	derived post-intervention slope	observed post intervention mid-point value	derived post-intervention slope
slope	12	3.534	2.612	0.307	5.378	4.571	-0.269	3.549	-0.609	2.689	-0.896	1.344	-1.344
	16	3.534	2.612	0.230	5.378	4.571	-0.202	3.549	-0.457	2.689	-0.672	1.344	-1.008
	20	3.534	2.612	0.184	5.378	4.571	-0.161	3.549	-0.366	2.689	-0.538	1.344	-0.807
	24	3.534	2.612	0.154	5.378	4.571	-0.134	3.549	-0.305	2.689	-0.448	1.344	-0.672
	28	3.534	2.612	0.132	5.378	4.571	-0.115	3.549	-0.261	2.689	-0.384	1.344	-0.576
	32	3.534	2.612	0.115	5.378	4.571	-0.101	3.549	-0.229	2.689	-0.336	1.344	-0.504
	40	3.534	2.612	0.092	5.378	4.571	-0.081	3.549	-0.183	2.689	-0.269	1.344	-0.403
	50	3.534	2.612	0.074	5.378	4.571	-0.065	3.549	-0.146	2.689	-0.215	1.344	-0.323
step	12	3.534	3.534	0	3.534	3.004	0	2.332	0	1.767	0	0.883	0
	16	3.534	3.534	0	3.534	3.004	0	2.332	0	1.767	0	0.883	0
	20	3.534	3.534	0	3.534	3.004	0	2.332	0	1.767	0	0.883	0
	24	3.534	3.534	0	3.534	3.004	0	2.332	0	1.767	0	0.883	0
	28	3.534	3.534	0	3.534	3.004	0	2.332	0	1.767	0	0.883	0
	32	3.534	3.534	0	3.534	3.004	0	2.332	0	1.767	0	0.883	0
	40	3.534	3.534	0	3.534	3.004	0	2.332	0	1.767	0	0.883	0
	50	3.534	3.534	0	3.534	3.004	0	2.332	0	1.767	0	0.883	0

^ p_t =(Intercept + Pre-intervention_slope*timepoint$_t$ + Post-intervention_step_change*intervention_indicator + Post-intervention_slope_change*Post-intervention_timepoint$_t$)/100. Here

± The sample size per timepoint was drawn from a normal distribution with mean n and S.D. n/3. Values for n were: 141, 283, 424, 566, 707, 849, 1132, 1415, 2264, 2830 and 5659.

* This value was derived from the pre-intervention trend in the TKR CRPD study (chapter 4). In this study, the pre-intervention trend was B = 0.1317 over 14 pre-intervention time points

** Assuming the pre-intervention trend would have continued uninterrupted without the intervention

Estimands

The target of inference was change in outcome following an intervention, specifically testing the null hypothesis of no change (i.e. β_2=zero [model 1] or β_3=zero [model 2]). The outcome at each timepoint was a proportion, which in this particular case-study was the 5-year cumulative incidence of TKR (199).

Performance

The coefficients, standard error and *p*-values from these models were stored and the empirical power to reject the null hypothesis of no post-intervention change was calculated as the proportion of simulations where the *p*-value for the intervention variable coefficient (step/slope change) was <0.05. (228, 230, 231) This was represented graphically as contour plots across scenarios according to N and n. For convenience of comparison, additional presentation was made for power according to different effect size and location scenarios while keeping N constant (N=28). Also calculated for midway step and slope change scenarios (while keeping N constant) was the percentage bias (228) of the regression coefficients, defined as:

$$\% \text{ Bias} = \left[\frac{\text{average estimate across simulations} - \text{true parameter value}}{\text{true parameter value}} \right] * 100$$

Sensitivity analysis

The case-study was focussed on a scenario where the mean pre-intervention level of outcome was 3.5%. This corresponds to a relatively 'rare' outcome, and whilst not

uncommon to encounter such incidences, other more common disease incidences are also likely to be routinely encountered and this will have an effect on power given the relative effect size is a function of the absolute pre-intervention level. This is therefore a key parameter to recognise and in order to demonstrate/explore the impact of pre-intervention level of outcome, the main analyses were repeated to investigate power for slope and step changes while keeping N constant (N=28) but varying pre-intervention level from 3.5% to (i) 8% and (ii) 20%.

Stata programme

The present analysis was primarily based on the study as reported in chapter 4 and therefore explored parameter value ranges as adapted from the CPRD study. For this, Stata code was developed by myself following attendance at a training course on simulation studies using Stata (227). However, this was subsequently adapted into a Stata programme with associated documentation (appendix file 5 & 6) in order to provide a ready-to-use means for assessing power associated with any valid list of (nine) input parameter values, as described in appendix file 5. The substantive programming code was authored by myself, however this consisted of a very large number of separate do.files. A more efficient/generalised coded loop allowing all parameter values to be simultaneously inputted was subsequently developed in collaboration with another DPhil student (K. Berencsi) possessing an advanced level of programming expertise.

RESULTS

Results are presented below describing the impact of N and n on power within the various ITS scenarios. It is important to interpret these while recognising the main results pertain to the case-study based on the UK cumulative incidence TKR time-series (chapter 4), where the mean pre-intervention level of outcome for mid-time-series interventions was 3.5%, the Stata programme developed can be used to explore any valid combination of alternative input parameter values (appendix file 5 & 6).

Slope change

As would be expected, power increased as N and/or n increased (figure 6.2) and as effect sizes became larger (figure 6.2, figure 6.3A). Figure 6.2 depicts power for different N and n combinations for each effect size investigated. These indicated that nearly all mid-time-series intervention scenarios with a large effect size (-75%) had at least 80% power when there were >24 total timepoints, even when there was a very small sample size per timepoint (~150 subjects, corresponding here to only 5 outcome events per timepoint). However, when the effect size was small (-15%) then to achieve 80% power an analysis had to either contain a large N or very large n (figure 6.2). While keeping other factors constant (effect size =-34% and N=28), power was greater in scenarios with mid-time-series interventions, with comparably less power in scenarios with earlier/later interventions (figure 6.4A). The % bias in model coefficients was small throughout all simulations and this trended towards zero as sample size increased (Figure 6.5A)

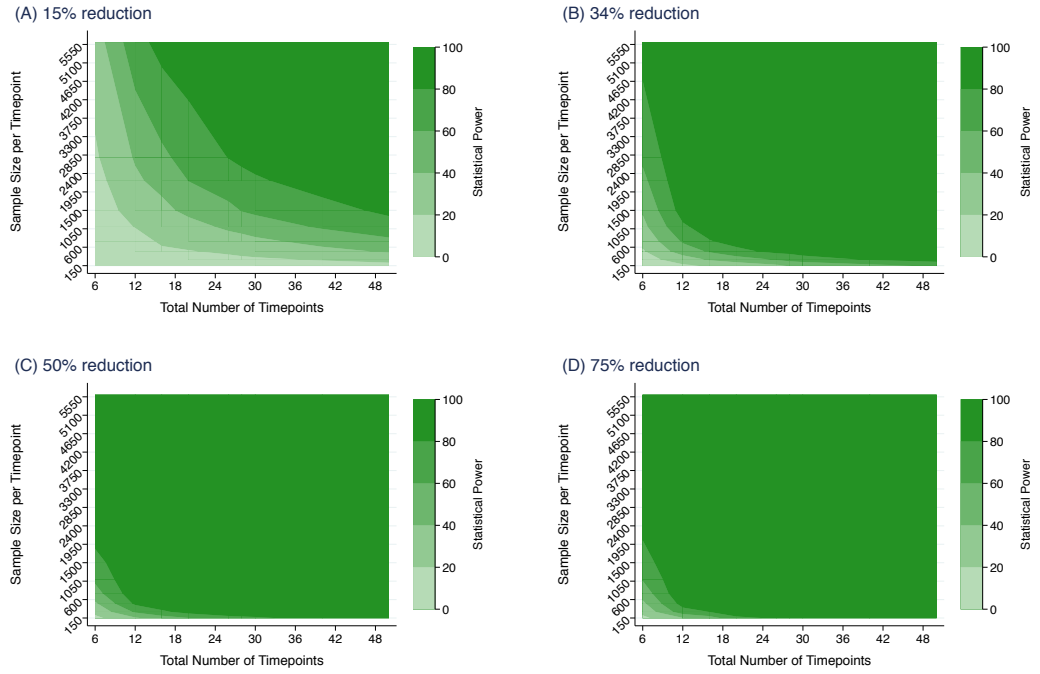


FIGURE 6.2: EMPIRICAL POWER TO DETECT A RELATIVE REDUCTION IN OUTCOME (OF VARIOUS MAGNITUDES) MEDIATED VIA A SLOPE CHANGE, WHERE MEAN PRE-INTERVENTION INCIDENCE IS 3.5%: BY NUMBER OF TIMEPOINTS AND MEAN SAMPLE SIZE PER TIMEPOINT

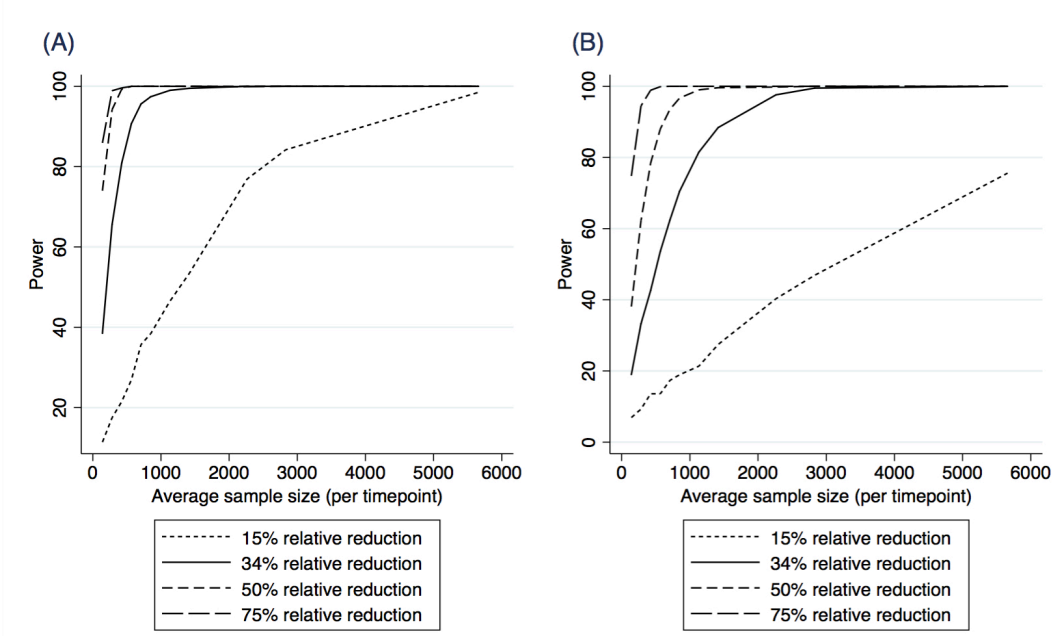


FIGURE 6.3: EMPIRICAL POWER IN CASE-STUDY (STRATIFIED BY EFFECT SIZE) TO DETECT AN INTERVENTION RESULTING IN: (A) A SLOPE CHANGE OR (B) A STEP CHANGE.

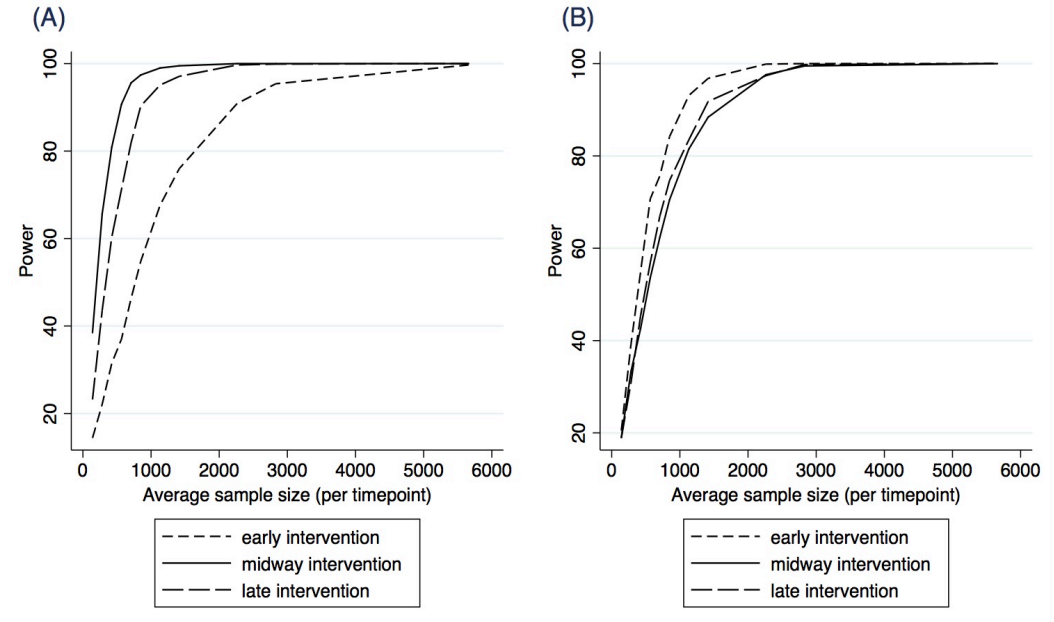


FIGURE 6.4: EMPIRICAL POWER IN CASE-STUDY (STRATIFIED BY INTERVENTION LOCATION) TO DETECT AN INTERVENTION RESULTING IN: (A) A SLOPE CHANGE OR (B) A STEP CHANGE

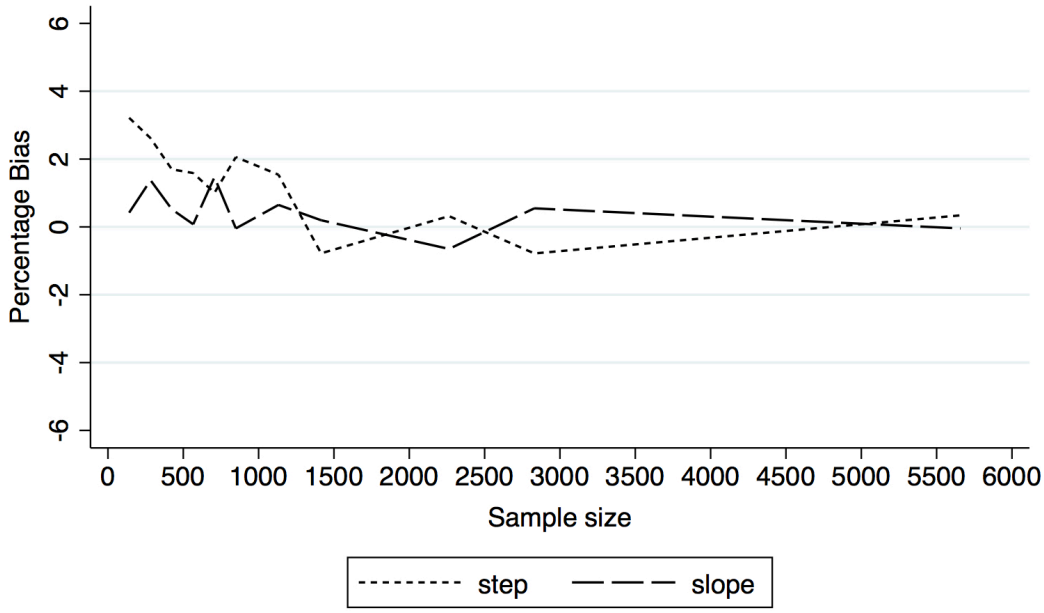


FIGURE 6.5: PERCENTAGE BIAS IN ESTIMATES OF INTERVENTION IMPACT IN CASE-STUDY: STRATIFIED BY STEP OR SLOPE CHANGE

Step change

Similar to slope change scenarios, power increased as N and n became larger (figure 6.6) or as the effect size was larger (figure 6.6, figure 6.3B). Generally, there was less power in step change scenarios than in corresponding slope change scenarios (figure 6.6 versus 6.4), with nearly all mid-time-series intervention scenarios being inadequately powered when the effect-size was only -15% (figure 6.6, figure 6.3B). Even when effect sizes were large and number of timepoints was moderate (14 pre-intervention and 14 post-intervention timepoints), analyses were underpowered if sample size per timepoint was low (figure 6.3B). Interestingly, little difference was found in power following an early or late intervention as compared to when the intervention occurred mid-way through

(figure 6.4). The % bias in model coefficients was likewise small throughout all simulation and this trended towards zero as sample size increased (Figure 6.5B)

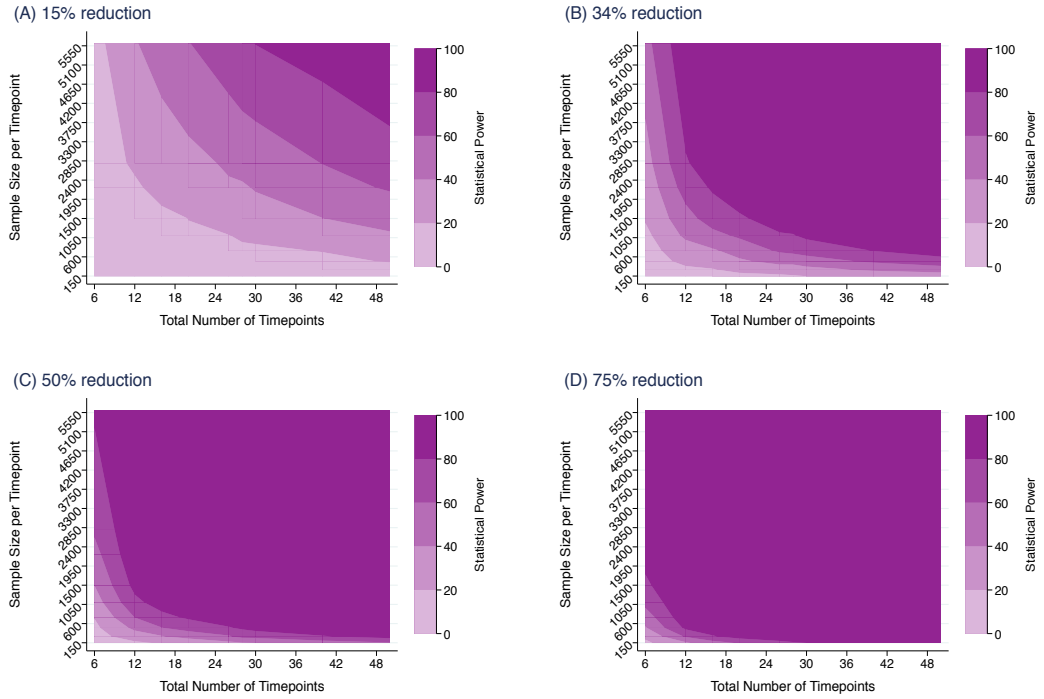


FIGURE 6.6: EMPIRICAL POWER TO DETECT A RELATIVE REDUCTION IN OUTCOME (OF VARIOUS MAGNITUDES) MEDIATED VIA A STEP CHANGE, WHERE MEAN PRE-INTERVENTION INCIDENCE IS 3.5%: BY NUMBER OF TIMEPOINTS AND MEAN SAMPLE SIZE PER TIMEPOINT

Sensitivity analyses

In scenarios exploring the impact of higher values for the mean pre-intervention cumulative incidence, there was greater power to detect the same relative effect sizes compared to the setting of 3.5% as used in the main analysis. This was the case for where mean pre-intervention cumulative incidence was 8% (figure 6.7) and 20% (figure 6.8). In

cases of a 20% pre-intervention outcome, most scenarios (except for a significant proportion of those with only a -15% relative effect size or with larger effect sizes but very small sample sizes) had 100% power.

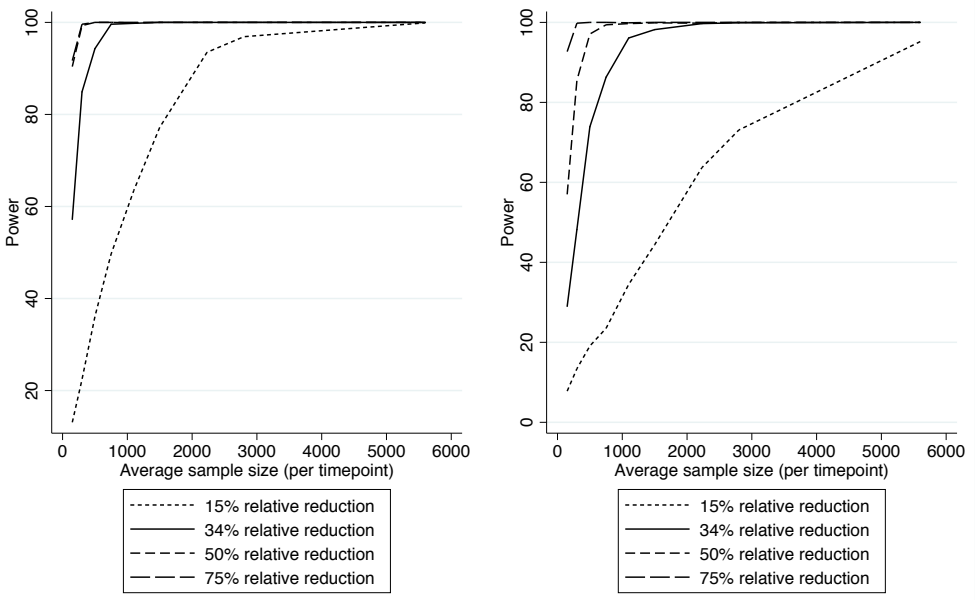


FIGURE 6.7: EMPIRICAL POWER (STRATIFIED BY EFFECT SIZE) TO DETECT IN SCENARIOS OF MEAN PRE-INTERVENTION INCIDENCE OF 8.0% AN INTERVENTION RESULTING IN: (A) A SLOPE CHANGE OR (B) A STEP CHANGE

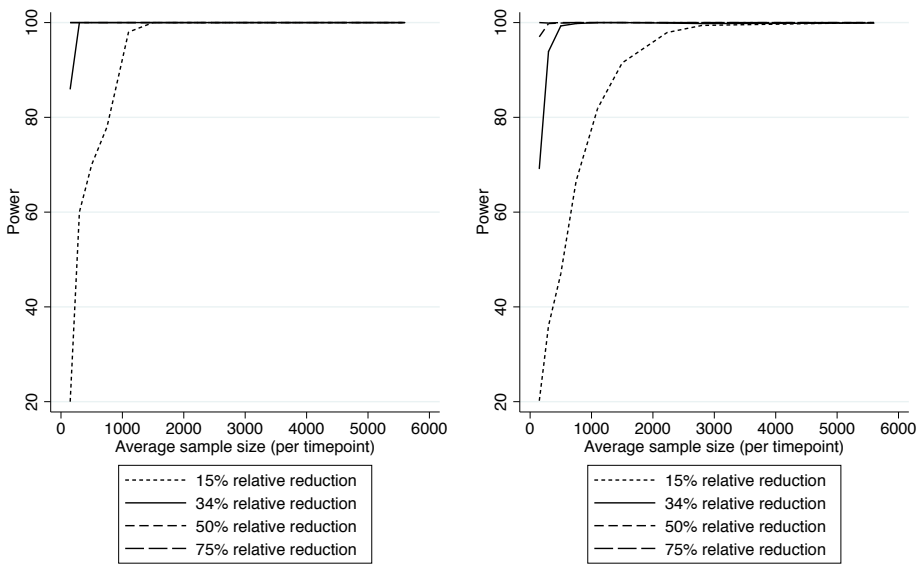


FIGURE 6.8: EMPIRICAL POWER (STRATIFIED BY EFFECT SIZE) TO DETECT IN SCENARIOS OF MEAN PRE-INTERVENTION INCIDENCE OF 20.0% AN INTERVENTION RESULTING IN: (A) A SLOPE CHANGE OR (B) A STEP CHANGE.

DISCUSSION

This Monte-Carlo simulation study has sought to demonstrate a means of estimating power in ordinary least squares ITS within specific scenarios according to effect size, number of timepoints, number of individuals per timepoint, location of intervention in the time-series and intervention impact mediated via slope vs. step changes. A binomial distribution has been used to simulate the outcome at each timepoint within each time-series for each specific scenario. Model parameters were informed and adapted from the analysis of cumulative incidence of TKR in the analysis of CPRD data in chapter 4.

Main findings

This study demonstrates that simple rules regarding the number of timepoints are not adequate by themselves to denote an ITS analysis as sufficiently powered. Other factors such as the sample size per timepoint, expected effect size, location of intervention in the time-series and pre-intervention trends need to be collectively considered. For example, this simulation analysis can be used to provide the power available in the TKR time series analysis in chapter 4. That is, with 14 pre- and 14 post-intervention timepoints with 500 patients per timepoint and where the pre-intervention incidence started at 2.61% and finished at 4.45%, then the simulation programme estimates 86% power to detect a 34% average slope change reduction. This indicates that the TKR time-series in chapter 4, for which the average sample size per timepoint was just over 600, was sufficiently powered to detect a relative reduction of at least a third (figure 6.2B). These results underline the importance of robust pre-study sample size planning, which here would entail hypothesising about the magnitude of any expected effect size, the average pre-intervention rate of outcome and the possible combination of number of timepoints throughout the study period and the average number of individuals underlying each timepoint.

Findings in context

That power increases as N increases is an expected finding and has previously been shown for fixed ratios of effect size to the standard deviation of the timeseries (222, 232). However, here addressing the previously unknown trade-off between N and n is an important consideration and a helpful development. Firstly, because the standard

deviation of a given number of population-level timepoints may likely be difficult for applied researchers to estimate in advance of a proposed ITS study. Secondly, because exploring this trade-off between N and n informs to what extent it may be beneficial (in terms of power) when generating an aggregate ITS dataset to sacrifice sample size per timepoint in order to increase the number of timepoints (or vice-versa). It allows a combination of N and n to be selected to optimise power. Although the exact nuances of this unique trade-off were scenario specific, in most cases only very little gain in power was achieved when a time-series was lengthened at the expense of timepoint sample size. This though was not the case for where a very short time-series was lengthened, where gains in power were more noticeable. For example, to achieve at least 80% power in the mid-way -50% step change scenario (figure 6.3C), approximately 5,100 individuals were required per timepoint in a six timepoint time-series although this was less than 1,500 for a 12 timepoint time-series. Yet when this was extended to a 24 timepoint time-series then there was still a requirement of approximately 600 individuals per timepoint to achieve 80% power.

A differential power according to whether an intervention impact is mediated via a slope or step change does not seem to have previously been investigated. This study has found that power was greater in slope change scenarios, a likely explanation being that the effect size was the average difference between post-intervention values and counterfactuals, which in the case of slope change scenarios continued to increase as per the pre-intervention slope and therefore made detection of a change more probable. Within scenarios with a slope change, power was found to be greater in settings with a

balanced number of pre-intervention and post-intervention timepoints (as opposed to earlier/later interventions). On the other hand, the location of the intervention had little impact on power to detect step changes and was even marginally greater when the intervention occurred early. Although this was unexpected, it is not without some support from previous work (222) which reported a similar finding for heteroskedastic models.

There is some discussion in the literature as to whether, rather than calculating the estimated sample size required to achieve 80% power to detect a pre-specified effect at a given type 1 error rate, that instead it may be more appropriate to estimate the sample size required to achieve a certain level of precision, i.e. 95% confidence interval width (219, 233). Although this is not mainstream, it is very much mathematically linked to the concept of power and does have certain advantages. Determining the required sample size for achieving a desired level of precision for the 95% confidence interval of an effect size is more intuitive than the concept of power, tangibly expressing the impact of being underpowered on the informativeness (i.e. precision) of results. It is also more consistent with the continuous nature of strength of evidence rather that the dichotomising of results that is involved in powering for a significance test to reject the null hypothesis. That said, the concept of power is now very well established in epidemiology and is often what is required by funding and ethical committees in their assessment of the validity of a proposed investigation.

Limitations

The analysis is subject to various limitations. Each timepoint was treated to be independent of other timepoints as these were cumulative incidence measures, which as such meant that a given subject/patient could only be included in a single timepoint. The impact of autocorrelation on estimates has not therefore been explored, although this remains a subject for further investigation. Despite the availability of ITS approaches that explicitly model autocorrelation, such as autoregressive integrated moving average (ARIMA) models (234), it would seem that where the assumptions of OLS regression are met then this is preferable for epidemiological studies where the goal is likely to be causal inference rather than future prediction. Indeed, while autocorrelation needs to be addressed where present, it has been noted that in epidemiological studies it can often be accounted for by controlling for other variables (173), and interestingly of a recent review of over 200 drug utilization studies implementing ITS analysis, 50% were found to use segmented linear regression (174). Specification of ARIMA models are frequently cited to require a minimum of 50 timepoints (235), with >100 being preferable (234), yet it is common to have less than this minimum available in epidemiology contexts using routinely collected data (171, 198, 222, 229). For these reasons, the focus here was on 'short' time-series where 50 timepoints has been considered a maximum, with the use of Durbin Watson statistics to confirm first-order autocorrelation was not present. Previous work has investigated the relationship between the number of timepoints and power in the presence of autocorrelation (222, 236), where positive autocorrelation has been shown to reduce power and negative autocorrelation to increase power (222). Similarly, seasonality has not been considered nor situations

where there may be a delay or waning intervention effect. These remain the topics further extensions/complexities to the simulation programme.

Another limitation is that the definition of effect size as the difference between post-intervention timepoints and counterfactual timepoints (i.e. what would have been observed had pre-intervention level/slope continued uninterrupted) involves extrapolation and therefore uncertainty. While this is often done in practice, with uncertainty of model estimates expressed using confidence intervals (179), there is still the assumption that pre-intervention trends would have continued unchanged. The advantage of using this approach also means that the effect size can be expressed as a relative proportion of the pre-intervention mean, rather than a very scenario specific coefficient. This prioritises practicality given that this is an intuitive effect size in comparison to a coefficient for a slope change, which is very dependent upon the number of timepoints in the time-series.

Similarly, the basis of this work was the setting of the time-series analysis in chapter 4, where mean pre-intervention cumulative incidence of TKR was 3.5%. The impact of other parameters here explored may be different in contexts with different values for mean pre-intervention incidence, and although this wasn't exhaustively explored here, the sensitivity analysis showed that power was indeed greater when pre-intervention incidence was higher within otherwise like-for-like scenarios. The focus of the work was also on the case-study where reductions in outcome followed the intervention, however the Stata programme (as described in appendix file 5) could just as well be used to

estimate sample size requirements for scenarios where outcome is expected to increase post-intervention, for example in drug utilisation following a change in prescribing guidance.

Furthermore, only scenarios where the repeated outcome measure is a cumulative incidence (i.e. a proportion) has been investigated. The reason for this is because the ITS analyses as reported in chapters 4 & 5 had initially been planned with cumulative incidence as the main outcome measure. The work as reported in this chapter was thereafter carried out in conjunction with this initial analysis plan. It was after the results of chapter 4 had already been presented at an international conference that it was decided incidence rate was a more preferable measure given that it accounts for censoring. However, cumulative incidence is a very common epidemiological measure and therefore should prove useful, but incorporating other common measures such as person-year rates, means (for example length of hospital stay or drug doses prescribed) and frequencies is a logical next step and remains the subject for imminent further investigation. It is worth mentioning that results for the ITS analyses in chapter 4 which use incidence rates are almost identical to those using cumulative incidence (there suggesting censoring remained approximately stable over the timeseries).

Strengths

The disentangling of N and n is a key strength and novel aspect of the current study, as is the separate consideration of post-intervention step and slope changes. The development and inclusion of a Stata programme is an important feature of the

investigation, facilitating researchers to estimate sample size requirements for future ITS studies and thereby promoting the avoidance of carrying out underpowered and therefore under-informative analyses. Although the programme is currently confined to Stata, plans for imminent further work includes using this tool as the basis for a readily available online calculator. Another strength of using the Stata tool is that it not only produces the heat map graphs as included here, but summary output in tabular format which can then be sorted by various parameters. This allows one to identify which scenario (in terms of N and n) provides maximum power for a given total sample size. Alternatively, it can be used to determine the precise total sample size required to achieve 80% power for a given scenario. It is also worth mentioning that the parameter values for the UK TKR time-series (chapter 4) formed the basis of this case-study, providing a "real world" clinical scenario (199) in order to increase the applicability of the findings, rather than starting from arbitrary parameter values.

Conclusions

Multiple factors influence the power of OLS ITS analysis and these should be collectively taken into account when considering the feasibility of a proposed ITS study. This chapter demonstrates how a simulation approach can be used to estimate the power available within specific ITS scenarios. The Stata code developed can be used to facilitate pre-analysis sample size planning of ITS studies within similar applications in order to reduce future wasteful, inefficient and possibly misleading research. Further work includes extended these simulations to applications using other outcome measures rather than simply cumulative incidence, and translating the Stata programme into an online tool.

7. PATIENT-LEVEL ANALYSIS OF IMPACT OF TNFI VERSUS CSDMARD ON SUBSEQUENT NEED FOR JOINT REPLACEMENT AMONG RA PATIENTS IN THE UK

INTRODUCTION

Chapter 1 described the emergence of biologic therapies such as TNFi and how this has revolutionised the management of RA over recent decades (50, 237). Analyses as reported in chapters 4 & 5 were carried out in order to explicitly estimate the impact of introducing TNFi and suggested some role in subsequent decreasing trends in joint replacement rates, although there were reasons to be cautious about a causal relationship. In particular, although declining rates of TKR (UK and Denmark) and THR (Ontario) were observed for RA patients (compared to counterfactuals), there was also a reduction in joint replacement rates in the non-RA control populations (chapter 5). Furthermore, there is prior evidence that there was an increasing usage of csDAMARDs throughout the timeframe of the study (132, 194). This latter observation regarding csDMARDs is understandable given that patients were only eligible (at least in the UK) for TNFi after failing to adequately respond to two 6-month episodes of conventional therapy. This dual increase in both biologics and csDMARDs renders the population-level analyses of chapters 4 & 5 difficult to interpret with precision given that the exact causes for any change in THR/TKR rates remain unclear. Patient-level data on this topic is required to disentangle these issues, yet such data remain scarce (91).

One previous study has attempted to estimate the patient-level impact of biologics versus csDMARD (108), however that study failed to adjust for many relevant confounding characteristics, such as disease severity, prior medications and comorbidities, which they conclude likely biased their reported associations. Indeed, as described in chapter 1, one of the fundamental challenges being faced within the field of pharmacoepidemiology is that of the issue of confounding by indication (109). In the case of TNFi use and subsequent need for joint replacement, it is likely that users of biologics are - on average - at significantly higher baseline risk of joint replacement than non-users (34). Failing to account for these systematic differences between treated and untreated patients in an observational study would render results unreliable, particularly where the aim is to estimate comparative effectiveness (110). There has been increasing development and use of novel statistical methodologies within the field of pharmacoepidemiology to better minimise the bias resulting from confounding by indication (117), including various methods that are quasi-experimental in nature and which offer the opportunity for causal inference where an RCT is likely to be unfeasible (112, 116). Important examples include ITS, instrumental variable (IV) and regression discontinuity design (RDD) analyses (122), although applications of these methods to-date within the field of musculoskeletal epidemiology appears to be limited.

The British Society for Rheumatology Biologics Register: Rheumatoid Arthritis (BSRBR-RA) was set up in 2001, primarily in order to monitor the safety of TNFi (71, 238). It is now one of the largest biologics registers in the world and is a rich dataset containing

many clinical, diagnostic and demographic variables. The breadth of data included is one of the resources great strengths and permits control for many factors that may be involved in the confounding by indication problem that would likely bias cruder approaches to estimating the effectiveness of TNFi. Despite the use of BSRBR-RA for many studies on both treatment safety and effectiveness (238), it has not been used yet to investigate the potential impact of treatment on need for joint replacement.

The current aim was therefore to leverage the BSRBR:RA data in conjunction with quasi-experimental methodologies, in order to estimate the comparative effectiveness of TNFi vs. csDMARDs on subsequent rates of joint replacement among UK RA patients. The primary outcomes for this investigation, as per previous chapters, were THR and TKR (analysed separately) although other (non-hip/non-knee) joint replacement (OJR) was also investigated as a secondary outcome.

METHODS

Data sources and exposures

Data were obtained from the British Society for Rheumatology Biologics Register for RA (BSRBR-RA), as detailed in chapter 2. This register contains prospectively collected observational data on over 20,000 RA patients recruited from 2001 onwards, primarily to evaluate the real-world safety of TNFi. In addition to recruiting initiators of biologic therapies, the BSRBR-RA contains a comparator cohort of non-biologic treated RA

patients on csDMARDs, entry into which was dependent on having active disease (guide DAS28 >4.2).

Outcomes

The primary outcomes of interest were first occurrence of THR or TKR, analysed separately. Patients were followed up from date of registration into BSRBR-RA until the earliest date of either outcome event, follow-up form date indicating switching exposure status (stopping among TNFi users or starting biologics among csDMARD users), last follow-up form or death. OJR (a composite outcome consisting of elbow, shoulder, hand or other small joint replacement) was a secondary outcome.

Study population

The study sample (Figure 7.1) consisted of all biologic-naïve RA patients, either in the control cohort or those initiating a TNFi (etanercept, infliximab or adalimumab) no more than 6 months prior to registration within BSRBR-RA. All patients however were permitted exposure to csDMARD therapy. Patients with a THR or TKR recorded prior to registration were excluded, as were patients with less than 6 months of follow-up (i.e. those who did not return at least one follow-up questionnaire). In analyses of OJR, further exclusion was made of patients who had undergone an OJR prior to baseline.

Statistical analysis

Statistical analyses were carried out in Stata 15.1 and R. As described in the introduction, given the probable confounding by indication, i.e. TNFi users vs. csDMARD users likely having a different baseline risk of THR/TKR, it was decided a-priori to match TNFi users to comparable csDMARD users. This was done based on their propensity for receiving treatment, which is a method now commonly carried out in the pharmacoepidemiology literature (239, 240). Propensity scores (PS), i.e. the probability of receiving TNFi conditional on observed baseline characteristics (113) (including those predictive of outcome) were estimated for all patients using logistic regression. Here, treatment status was the outcome in the model and predictors (fully listed in appendix table 7.2) consisted of: age, gender, ethnicity, index of multiple deprivation (IMD), body mass index (BMI), smoking status, year of registration (quintiles), time since RA diagnosis, DAS28, health assessment questionnaire (HAQ) score, quality of life (SF36 domains), 1987 ACR criteria, systemic involvement, co-morbidities and co-medications. Each TNFi patient was matched to the csDMARD patient with the most similar PS within a caliper distance of 0.2 standard deviations of the logit of the PS, as this has previously been shown to greatly reduce bias in estimation of effectiveness (113, 241). Patients failing to be matched within this caliper width were excluded from further analysis. Matching with replacement (242) was used which entailed matching each TNFi user to its most similar csDMARD user but then 'returning' the csDMARD user back into the pool of potential matches for further TNFi users. This choice of matching is commonly used in situations of fewer available untreated patients than treated patients, as was the case here with fewer in the csDMARD than TNFi cohort (242, 243). The reason for this is that it maximises efficiency in that it potentially finds a match (and therefore makes use of) all

subjects in the treated cohort, although it does result in a pseudo-population for the untreated group (i.e. a matched csDMARD cohort not necessarily representative of the 'actual' csDMARD population). TNFi users were assigned weights of one and csDMARD users a weight proportional to the number of times matched. The weights amongst the csDMARD users totalled the number of unique csDMARD users included in the matched sample, which was accounted for in estimation of standard errors in subsequent analyses.

Baseline characteristics of the TNFi user and csDMARD cohorts were summarised and differences between the cohorts assessed by way of absolute standardised mean differences (SMD) (244, 245). SMDs are preferable to *p*-values for assessing balance as *p*-values are impacted by a combination of balance and sample size. That is, all other factors being equal, *p*-values comparing the matched cohorts may be different from those comparing the unmatched cohorts simply due to changes in sample size alone. SMDs on the other hand are scaled/standardised according to sample size, with smaller SMD values indicative of greater balance between cohorts. SMDs express the difference in means in units of standard deviation, which for continuous variables is the difference in sample means between the matched cohorts and for dichotomous variables is the difference in prevalence (245). Incidence rates of each outcome event with 95% confidence intervals (CI) were calculated (accounting for csDMARD weighting) among matched TNFi users and csDMARD users. Cox regression was used to compare THR, TKR and OJR rates, with weights added to account for the matching with replacement procedure. The proportional hazards assumption was checked in final models using the

Schoenfeld and scaled Schoenfeld residuals (using the stphtest Stata command) which indicated no significant breach of proportionality ($P \geq 0.1$). All patients were censored at 12 years due to the small sample size of the csDMARD cohort after this time, which made the model unstable (most likely due to changes in the diminished sample size of unique csDMARD users being amplified in the weighted model). Baseline characteristics were adjusted for in the final models if they were not sufficiently similar post-matching, defined as SMD values >0.1 (246).

Propensity score analysis was chosen as the primary method a-priori given that this approach has been shown to be capable of eliminating a large majority of the bias due to measured confounding (113, 241) and is an intuitive method aiming to impose quasi-randomisation upon observational data. Although there is some methodological literature supporting the superiority of matching over other propensity-score based methods (241, 247), the choice to use matching was primarily due to the relative ease with which the method can be seen to have "worked", i.e. balanced baseline characteristics between matched cohorts. The use of other methods to explore the impact of unmeasured confounding was always proposed only as a sensitivity approach given that it was suspected (as was confirmed in analyses) that the more stringent assumptions of these latter methods may be breached and therefore undermine the validity of such investigations as primary analyses.

Missing data

Missingness in baseline data was identified and this was imputed using multiple imputation (MI) using chained equations (248-251), with all the variables used in the statistical analysis (described above) included into the imputation model, in addition to the time-to-event variables. Missingness was assumed to be "missing at random", i.e. dependent on other observed variables but randomly distributed once these dependencies are accounted for (248, 251, 252). This was not assumed where a variable was only collected for England (i.e. not Scotland or Wales), as was the case for IMD, which was not imputed. The R package mice was implemented, with partial mean matching (PMM) used for all numeric variables given that this approach has been shown to be valid in the case of non-normally distributed or only 'semi-continuous' variables (253). Ten imputed datasets were created as used elsewhere (254) and owing to the fact that the missingness was small (appendix table 7.1). Although consensus in the literature is lacking on the issue of choosing the number of imputations, it has been previously demonstrated that m=10 is suitable for situations where missingness is not large (250). Distributions of variables with imputed data were visually compared to those in the original dataset. Balance assessment was carried out for the cohort prior to matching in only one of the imputed datasets, but final models were run for all 10 imputed datasets (255) and hazard ratios (HRs) were pooled using Rubin's rules (250, 256).

Age and disease severity were a-priori specified as potential effect modifiers of the association between TNFi use and subsequent need for joint replacement. The reason for specifying age as a potential effect modifier was that it strongly predicted joint

replacement (chapter 3) and it seemed a foreseeable possibility that age may modify the TNFi versus joint replacement relationship (257). It also seemed plausible that TNFi might provide more or less benefit to those with severe disease activity (previous data indicate patients with high DAS achieve greater DAS reductions on TNFi but are less likely to achieve remission (34, 258)). Following the main analysis, these potential interactions were tested by including interaction terms for approximately median age (</≥ 60 years old) and DAS28 (≤/>5.1 NICE cut-off) in the weighted Cox model in a single imputed dataset. The likelihood ratio test was used to assess model fit and in the event of a significant interaction (p<0.1), matching and survival models were re-run and pooled within strata of the significant effect modifier.

Sensitivity analysis of PS matching

Given the fact that BSRBR-RA stopped recruitment to either cohort in 2008 but that the TNFi cohort reopened in 2010 allowing further enrolment (chapter 2), it was a-priori decided to further explore this issue. This should have been largely addressed in the main analysis given that there was vast overlap in calendar time between recruitment of the TNFi group and the csDMARD group (table 7.1) and the inclusion of calendar time in the PS model. However, in a sensitivity analysis the main PS matching and survival models were repeated following exclusion of the TNFi users recruited into the registry after the csDMARD cohort had closed. Secondly, in the main survival analysis models, patients were censored at time of changing exposure status (starting biologics amongst the csDMARD cohort or stopping amongst the TNFi cohort) as this seemed the 'cleanest' approach, i.e. isolating the effect of treatment from that of the exposure status switched

to. However, this may have introduced selection bias in the form of informative censoring (259). For this reason, survival analysis models were repeated with up to 180 days added after the follow-up form date indicating change in exposure status, although patients dying, lost to follow-up or reaching the end of study period in this timeframe were censored on these dates instead.

Instrumental variable sensitivity analysis

An instrumental variable (IV) approach was carried out as a sensitivity analysis, which sought to address the issue of unmeasured confounding given the non-random allocation of treatment in the BSRBR-RA. IVs can be leveraged if fulfilling certain key assumptions in order to deal with this issue and yield unbiased results (122, 260). Physician preference was used as the instrument, as has been done previously in other contexts (260, 261). The time-specific preference for prescribing biologics was estimated among each of the consultants who recruited patients into the BSRBR-RA. This variable was called 'physician preference', as estimated and applied elsewhere (261). This preference variable was calculated in a time-dependent manner by sorting BSRBR-RA patients according to recruiting consultant and registration date, and then by calculating the relative frequency of biologics prescribed over the previous 12 patients recruited into the registry by each consultant ID. If this relative frequency for a given patient was >=0.5 then the consultant was defined (at that point in time) as "preferring biologics", if it was <0.5 then they were considered as "preferring csDMARDs". In such manner the first 12 prescriptions per consultant ID were excluded from the effect estimate as they were only used to generate the preference IV variable. The strength of the IV was

assessed by comparing the proportion of TNFi users between the different IV status groups. Given that an uncensored proportion (cumulative incidence) for THR/TKR was used as the outcome in this approach, a 5-year outcome was considered more suitable as this would contain less attrition due to mortality and loss-to-follow-up. However, a 12-year outcome was also explored. Crude associations between the IV and THR/TKR (analysed separately) were carried out using simple cross tabulations and chi-square tests in order to provide an initial demonstration of crude relationships.

The IV was then used in a 2-stage least squares linear regression, as has previously been carried out (261) and described (262-264). Treatment exposure was the dependent variable in the first model containing the IV (consultant preference) and measured covariates as independent variables. This first model provided the estimated expected value of exposure to TNFi dependent on IV status and (for the adjusted analysis) measured confounders. These predicted values were then included along with (for the adjusted analysis) measured confounders as independent variables into the second stage which had THR/TKR (analysed separately) as the outcome. The coefficient for the predicted value of being treated (given the IV and measured covariates) in this second regression was the result of intrinsic interest, and can be interpreted as the risk difference between treated and untreated patients who were neither "always" going to take TNFi nor "never" going to take TNFi. That is, it's the complier average causal effect estimate (CACE), meaning the effect of treatment among the sub-population induced to take treatment solely due to the physicians' preference (122), and would have hypothetically taken the alternative treatment had they been seen by a different

physician with the alternative preference. The risk difference was multiplied by 100 so that the interpretation was the risk difference per 100 patients treated with TNFi versus csDMARDs (261).

Regression discontinuity sensitivity analysis

Regression discontinuity design (RDD) analyses, although a mainstream method in econometrics and social sciences, has been little used in the field of epidemiology (265, 266). As per other quasi-experimental approaches, the essence of the method is to mimic an RCT environment where 'allocation' of treatment in the final analytical dataset is quasi-random due to careful selection of who/when to measure (rather than who/when to expose). It is usually applied to situations where treatment in the real-world is allocated according to a threshold rule regarding a cut-off value of a continuous variable, such as CD4 count (267) or cardiovascular risk score (265). It assumes exchangeability in that patients should be continuously comparable within a sufficiently small bandwidth around the cut-off value of the threshold rule. This can be achieved given that such cut-off values are usually arbitrary and incorporate some measurement error, thus giving rise to a situation where patients "just below" the threshold are almost identical (in terms of measured and unmeasured confounders) to those "just above" except for the fact one group is eligible for treatment whilst the other is not. In the context of the present investigation, NICE guidance only recommends TNFi therapy for individuals with an elevated DAS28 score of over 5.1. This value of baseline DAS28 was therefore considered as a cut-off value in an RDD analysis, the impact of which was estimated on rates of THR/TKR (analysed separately), in a similar fashion as has previously been done

to estimate the impact of HIV treatment eligibility (267). Visual inspection of the density of DAS values was made in order to assess whether there was bunching of values around the cut-off, which has been previously suggested as a good assessment of whether some manipulation of the score had been performed, assuming that in the real world this would be continuous across the cut-off. Baseline characteristics were described among those "just below" versus "just above" the cut-off, and SMDs used to assess exchangeability within numerous bandwidths: +/- 0.2 units, +/- 0.3 units, +/- 0.5 units and +/- 0.7 units. The incidence (per 1,000 PYs) of THR/TKR was estimated per unit of the DAS28 score, and segmented linear regression models fitted to the data within the numerous bandwidths investigated. These models estimated the difference in THR/TKR incidence at the cut-off for treatment eligibility, taking into account trends within the various band-widths. These were also depicted graphically. The importance of exploring various band-widths has been previously described (267, 268), given that the choice of bandwidth can have a significant impact on the difference at the cut-off. Furthermore, it has previously been noted that RDD analyses within epidemiology will often be "fuzzy" rather than "sharp" (265), meaning that allocation of treatment is influenced by the cut-off in a probabilistic rather than deterministic fashion. In other words, there may be a significant number of 'non-compliers' who may receive treatment despite being below the threshold or may be withheld treatment despite being above the threshold. In such fuzzy scenarios, it is important to scale the impact of treatment eligibility (as defined by the threshold) on the outcome, by the actual impact on uptake of therapy (267), as was similarly done in the IV analysis of physician preference. This was here done using a 2-SLS approach, with the threshold defined as the IV, yielding estimation of the CACE around the threshold.

RESULTS

Baseline Characteristics

Baseline characteristics before matching are presented in table 7.1. Of 13,126 eligible RA patients identified in BSRBR-RA, 97% (9,558) of the TNFi users and 51% (1,644) of the csDMARD users were retained following PS matching (Figure 1). Given the 1:1 matching with replacement, a total of 19,116 patient records were used in subsequent analyses, with each csDMARD user being used a median of 3 (IQR: 1-6) times.

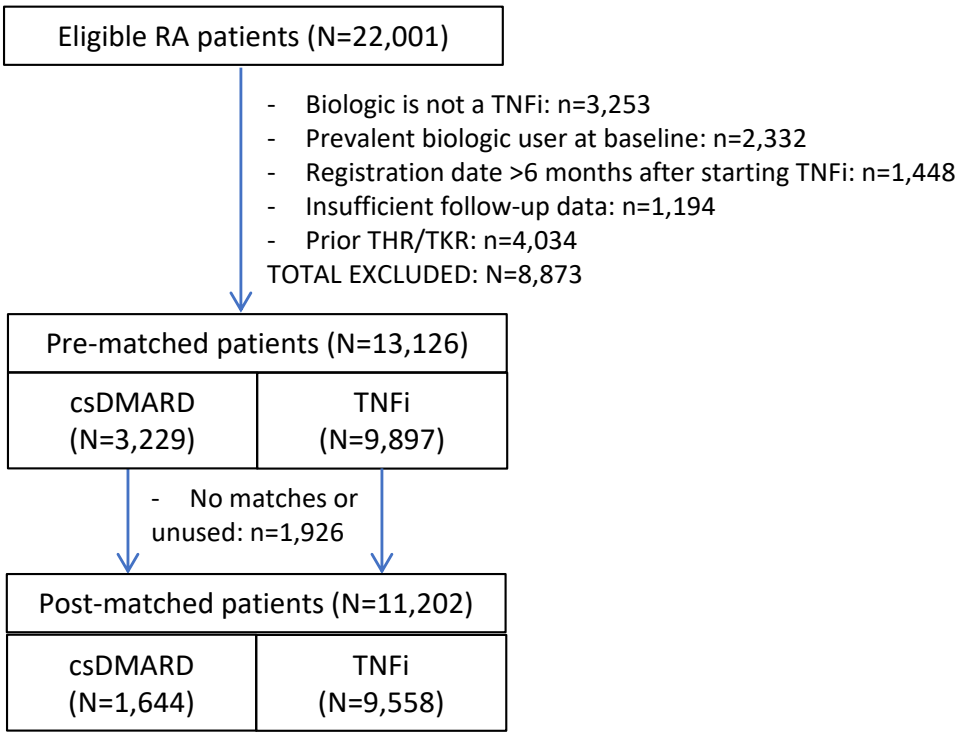


FIGURE 7.1: POPULATION FLOW DIAGRAM OF RA PATIENTS REGISTERED WITHIN BSRBR-RA

TABLE 7.1: BASELINE CHARACTERISTICS OF PROPENSITY SCORE MATCHED COHORTS *BEFORE* MATCHING: STRATIFIED BY USE OF TNFI VS. CONVENTIONAL SYNTHETIC DMARDS

	CS-DMARD (N=3,229)		Biologic (N=9,897)		SMD
Characteristic	**N**	**%**	**N**	**%**	
Age: Mean (S.D.)	59.0 (12.4)		55.1 (12.3)		-0.32
Gender: Female	2,338	72.4	7,522	76	0.08
ethnicity: White/Caucasian	3,150	97.6	9,428	95.3	0.12
Index Multiple Deprivation					
Quintile 1	609	18.9	1339	13.5	-0.15
Quintile 2	561	17.4	1524	15.4	-0.05
Quintile 3	481	14.9	1719	17.4	0.07
Quintile 4	525	16.3	1789	18.1	0.05
Quintile 5	421	13	1743	17.6	0.13
Unknown	632	19.6	1785	18	-0.04
BMI	27.3 (6.1)		27.1 (6.3)		-0.03
Smoking?					
Current	815	25.2	2306	23.3	-0.05
Ex	1,250	38.7	3,696	37.3	-0.03
Calendar period of registration					
Oct 01 - Oct 03	251	7.8	2517	25.4	0.49
Nov 03 - Aug 04	538	16.7	2083	21	0.11
Sept 04 - Aug 05	764	23.7	1820	18.4	-0.13
Aug 05 - May 07	1108	34.3	1466	14.8	-0.47
May 07 - May 16	568	17.6	2013	20.3	0.07
Years since diagnosis: Median (IQR)	8.6 (9.5)		11.1 (8.9)		0.27
DAS 28: Mean (S.D.)	5.10 (1.30)		6.48 (1.00)		1.18
Overall HAQ score	1.46 (0.75)		1.92 (0.62)		0.66
ACR: Ever rheumatoid positive	1,853	57.4	6,306	63.7	0.13
ACR: Deformity of ≥3 joint areas	2,325	72	8,452	85.4	0.33
ACR: Erosions on hands/feet	1,436	44.5	5,543	56	0.23
ACR: Ever had nodules	961	29.8	4,203	42.5	0.27
ACR: Symmetry	2,107	65.3	8,211	82.9	0.41
ACR: Deformity of hand joint	2,280	70.6	7,929	80.1	0.22
ACR: Morning stiffness >1 hour	2,839	87.9	9,305	94	0.21
Non-major prior joint replacement^	460	14.2	2094	21.1	0.18

Results shown are for the 10th imputed dataset. Full covariate table in appendix

^ Composite variable consisting of: shoulder, elbow, neck or other small joint replacement (e.g. hand)

SMD: standardised mean difference (smaller values indicative of better balance)

TABLE 7.2: BASELINE CHARACTERISTICS OF PROPENSITY SCORE MATCHED COHORTS *AFTER* MATCHING*: STRATIFIED BY USE OF TNFI VS. CONVENTIONAL SYNTHETIC DMARDS

	CS-DMARD (N=9,558) (1,644 unique patients)		TNFi (N=9,558)		SMD
Characteristic	**N**	**%**	**N**	**%**	
Age: Mean (S.D.)	55.2 (12.1)		55.2 (12.3)		0
Gender: % Female	7,289	76.3	7,259	75.9	-0.01
ethnicity: % White/Caucasian	9,114	95.4	9,118	95.4	0.00
Index Multiple Deprivation					
Quintile 1	1282	13.4	1322	13.8	0.01
Quintile 2	1420	14.9	1485	15.5	0.02
Quintile 3	1413	14.8	1640	17.2	0.07
Quintile 4	1544	16.2	1710	17.9	0.05
Quintile 5	1284	13.4	1650	17.3	0.11
Unknown	2615	27.4	1751	18.3	-0.22
BMI	26.8 (5.9)		27.1 (6.3)		0.06
Smoking?					
% Current	2448	25.6	2259	23.6	-0.05
% Ex	3272	34.2	3577	37.4	0.07
Calendar period of registration					
Oct 01 - Oct 03	1481	15.5	2262	23.7	0.21
Nov 03 - Aug 04	2037	21.3	2037	21.3	0
Sept 04 - Aug 05	1823	19.1	1802	18.9	-0.01
Aug 05 - May 07	1664	17.4	1462	15.3	-0.06
May 07 - May 16	2553	26.7	1995	20.9	-0.14
Years since diagnosis: Median (IQR)	10.8 (10.7)		11.0 (8.8)		0.02
DAS 28: Mean (S.D.)	6.47 (1.09)		6.43 (0.98)		-0.04
Overall HAQ score	1.91 (0.63)		1.91 (0.62)		-0.01
ACR: Ever rheumatoid positive: %	5,532	57.9	6,055	63.4	0.11
ACR: Deformity of ≥3 joint areas?: %	7,425	77.7	8,118	84.9	0.19
ACR: Erosions on hands/feet: %	4,651	48.7	5,282	55.3	0.13
ACR: Ever had nodules %	4,034	42.2	3,988	41.7	-0.01
ACR: Symmetry %	7,468	78.1	7,883	82.5	0.11
ACR: Deformity of hand joint %	6,771	70.8	7,602	79.5	0.20
ACR: Morning stiffness >1 hour: %	8,901	93.1	8,966	93.8	0.03
Non-major prior joint replacement^	1742	18.2	1989	20.8	0.07

Results shown are for the 10th imputed dataset. Matching was performed using replacement of the cs-DMARD users. 9,558 biologic users were each matched to one of the 3,229 cs-DMARD users (with replacement). Number of csDMARD patients represented in final matched sample was 1,644. Full covariate table in appendix

^ Composite variable consisting of: shoulder, elbow, neck or other small joint replacement (e.g. hand)

SMD: standardised mean difference (smaller values indicative of better balance)

Missingness was identified in 16 variables (appendix table 7.1). These, along with the % of missing data were: age (3%), ethnicity (15%), IMD (18%), BMI (12%), smoking status (1%), disease duration (4%), DAS-28 (1%), HAQ score (7%) and SF36 (17% for all eight domains). The distribution of values for these variables, both within the original dataset and imputed datasets were inspected and examples of these distributions are included in appendix figure 7.1. The distributions for variables following imputation were very similar to those in the complete case in the original dataset (appendix figure 7.1).

Baseline characteristics of TNFi users were markedly different in the unmatched study sample compared to the csDMARD cohort (table 7.1, appendix table 7.2), especially in aspects of disease severity. Specifically, the TNFi cohort had on average higher DAS28, HAQ score, proportion fulfilling the 1987 ACR RA criteria, lower health-related quality of life (as per SF36), longer disease duration and a higher prevalence of prior non-major joint replacement. Conversely, baseline characteristics between exposure cohorts were much more similar post-matching (table 7.2, appendix table 7.2). The only persisting differences (SMD>0.1) between the matched cohorts were calendar period of registration, low deprivation and the proportion of patients fulfilling ACR criteria (which were hence added as covariates into the final models).

Comparative effectiveness on THR

A total of 589 THRs were reported during follow-up (median = 4.94 years [IQR:1.52 to 10.04] for TNFi and 5.97 years [IQR: 2.05 to 9.55] for csDMARD) of the propensity-matched cohorts. Incidence rates (per 1,000 PYs) were 5.22 [95% CI: 4.66 to 5.88] and

6.30 [95% CI: 4.24 to 9.76] among TNFi users and csDMARD users, respectively (table 7.3). Comparing TNFi to csDMARDs yielded a pooled HR = 0.91 [95% CI: 0.64 to 1.31; *p*=0.62], which when adjusted for any remaining post-matching imbalance in baseline covariates was 0.86 [95% CI: 0.60 to 1.22; *p*=0.39] (figure 7.2, figure 7.3).

Comparative effectiveness on TKR

Among the matched sample, a total of 864 TKRs were reported during followup (median = 4.85 years [IQR:1.50 to 10.01] for TNFi and 5.98 years [IQR: 2.03 to 9.55] for csDMARD) of the propensity-matched cohorts. Incidence rates (per 1,000 PYs) were 8.89 [95% CI: 8.13 to 9.72] and 8.09 [95% CI: 5.32 to 12.89] among TNFi users and csDMARD users, respectively (table 7.3). This yielded a pooled HR = 1.18 [95% CI: 0.90 to 1.56; *p*=0.24], which when adjusted for any remaining post-matching imbalance in baseline covariates was 1.11 [95% CI: 0.84 to 1.47; *p*=0.46] (figure 7.2, figure 7.3).

Comparative effectiveness on OJR

Among the matched sample, a total of 336 OJRs occurred during follow-up (median = 4.93 years [IQR:1.52 to 10.02] for TNFi and 5.98 years [IQR: 2.05 to 9.12] for csDMARD) among the propensity-matched cohorts. Incidence rates (per 1,000 PYs) were 4.34 [95% CI: 3.76 to 5.02] and 3.87 [95% CI: 1.97 to 8.73] among TNFi and csDMARD users, respectively (table 7.3). There was no significant difference in OJR rates between the exposure cohorts (figure 7.2, figure 7.3)

TABLE 7.3: CRUDE INCIDENCE OF JOINT REPLACEMENT AMONG MATCHED TNFI AND CSDMARD COHORTS*

	csDMARD						TNFi			
	No. patients (crude)	No. outcome events (crude)	No. patients (matched with replacement)	No. outcome events (matched with replacement)	median follow up (matched with replacement), years (IQR)	rate (matched with replacement) per 1000 PYs (95% CI)	No. patients	No. outcome events	median follow up, years (IQR)	rate per 1000 PYs (95% CI)
THR	1,644	50	9,558	302	5.97 (2.05-9.55)	6.30 (4.24-9.76)	9,558	287	4.94 (1.52-10.04)	5.22 (4.66-5.88)
TKR	1,644	51	9,558	384	5.98 (2.03-9.55)	8.09 (5.32-12.89)	9,558	480	4.85 (1.50-10.01)	8.89 (8.13-9.72)
OJR ±	1,393	23	9,558	149	5.98 (2.00-9.13)	3.87 (1.97-8.73)	7,537	187	4.93 (1.52-10.02)	4.34 (3.76-5.02)

Results shown are for the 10th imputed dataset. Each TNFi-user matched 1:1 to a csDMARD user, with replacement. Standard errors were corrected for number of times a csDMARD user was included

± Anlysis of OJR was restricted to the subset of patients with no prior OJR

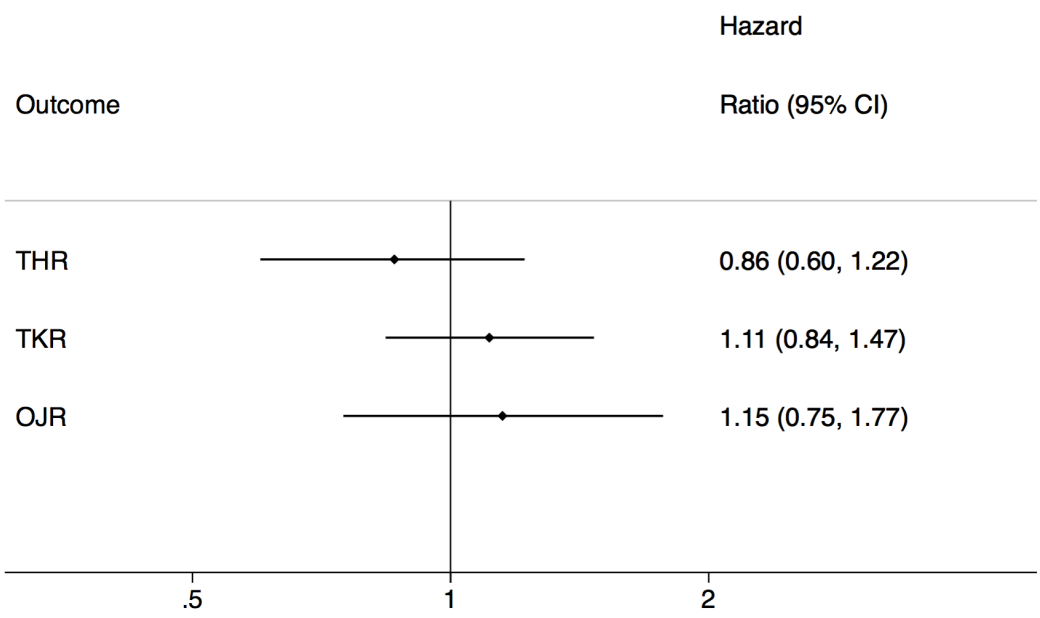
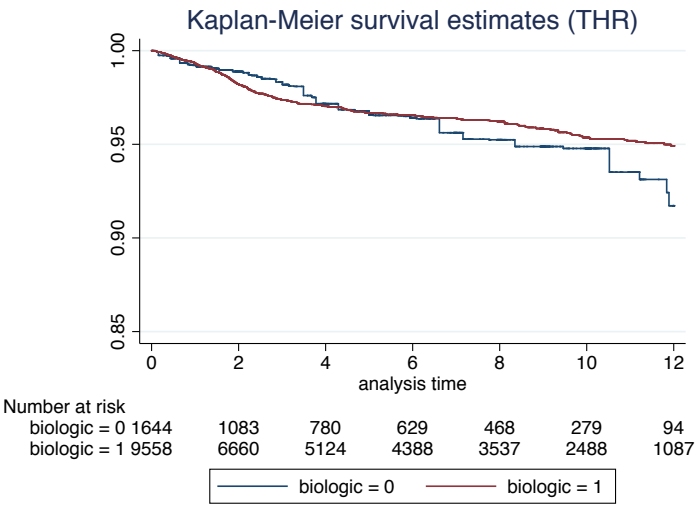
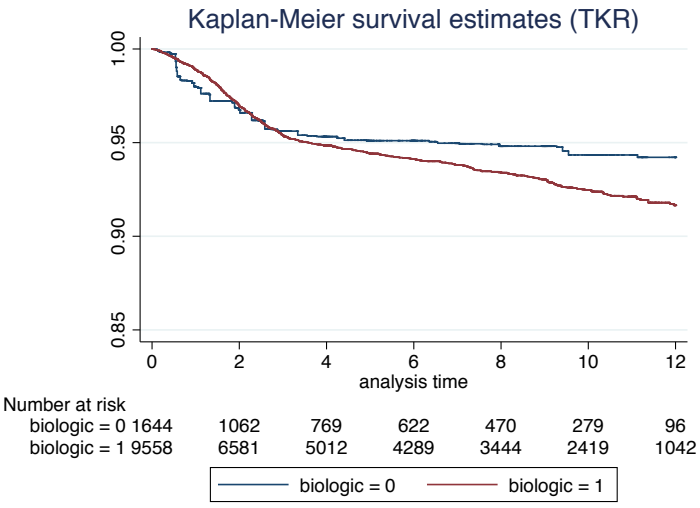


FIGURE 7.2: POOLED HAZARD RATIOS COMPARING TNFI TO CSDMARD ON SUBSEQUENT RATES OF JOINT REPLACEMENT





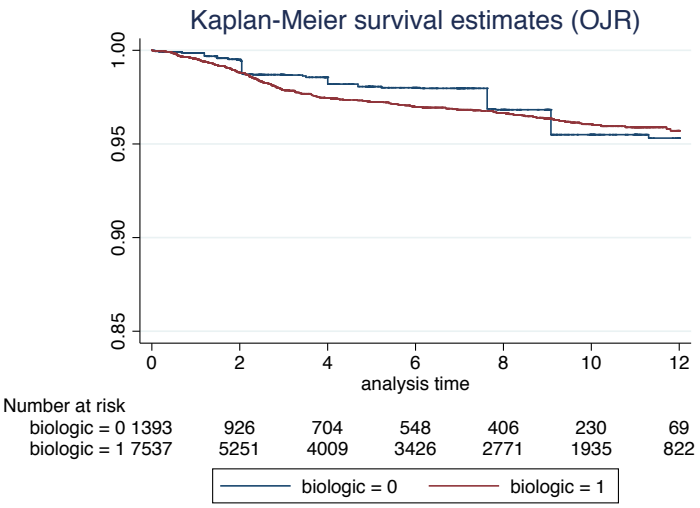


FIGURE 7.3: KAPLAN-MEIER SURVIVAL PLOTS STRATIFIED BY TNFI VERSUS CSDMARD STATUS FOR (A) THR, (B) TKR AND (C) OJR

Interactions

Age was found to be a significant (p<0.001) effect modifier for both THR and TKR outcomes, although disease severity was not (p>0.1). In subsequent stratified analyses (table 7.4), TNFi was associated with an estimated 40% reduction in incidence of THR among older patients (HR = 0.60 [95% CI: 0.41 to 0.87; p=0.008]) (figure 7.3). Differences in THR or TKR incidence rates between TNFi and csDMARD cohorts among younger patients were non-significant (figure 7.4).

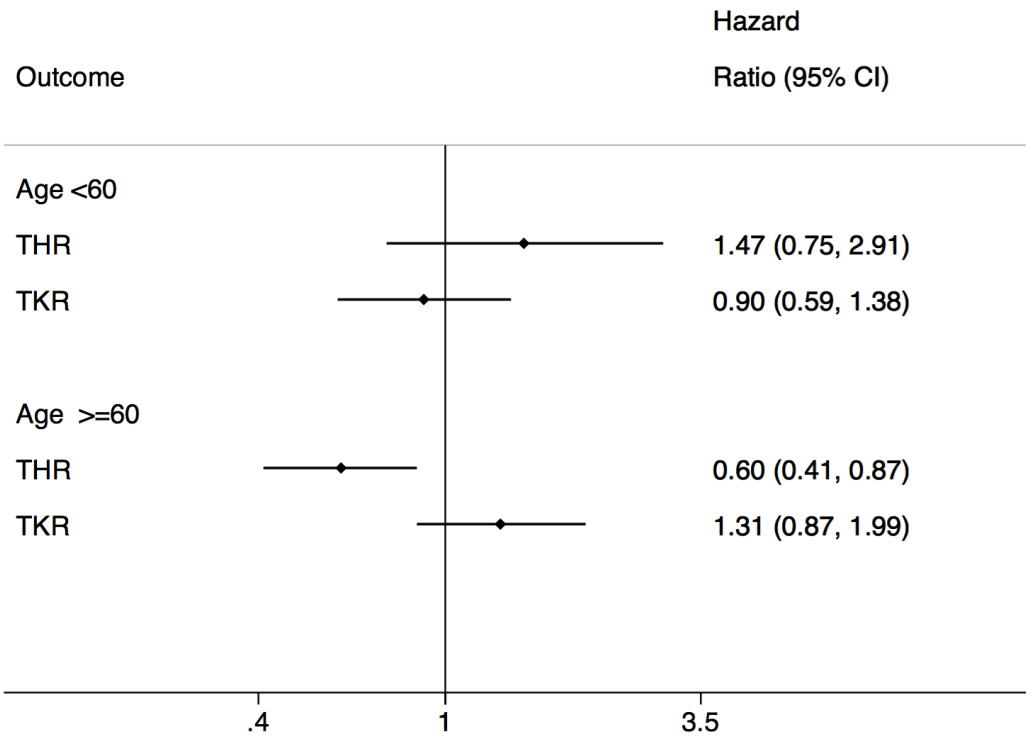


FIGURE 7.4: POOLED HAZARD RATIOS COMPARING TNFI TO CSDMARD ON SUBSEQUENT RATES OF JOINT REPLACEMENT, STRATIFIED BY AGE

TABLE 7.4 CRUDE INCIDENCE OF JOINT REPLACEMENT AMONG MATCHED TNFI AND CSDMARD COHORTS*: STRATIFIED BY AGE

	csDMARD					TNFi			
	No. patients (crude)	No. outcome events (crude)	No. patients (matched with replacement)	No. outcome events (matched with replacement)	median follow up (matched with replacement), years (IQR)	rate (matched with replacement) per 1000 PYs (95% CI)	No. patients	No. outcome events	median follow up, years (IQR)	rate per 1000 PYs (95% CI)
<u><60 years</u>										
THR	884	13	5,912	66	5.99 (1.99 - 9.99)	2.36 (0.93 - 7.72)	5,912	159	6.10 (1.84 - 10.98)	4.33 (3.71 - 5.08)
TKR	884	24	5,912	251	5.98 (1.97 - 9.98)	9.13 (5.08 - 18.06)	5,912	267	5.96 (1.73 - 10.97)	7.40 (6.56 - 8.37)
<u>≥60 years</u>										
THR	730	34	3,518	218	5.96 (2.32 - 9.05)	11.36 (7.23 - 18.82)	3,518	124	3.91 (1.39 - 8.96)	7.03 (5.90 - 8.43)
TKR	730	29	3,518	165	5.96 (2.28 - 9.47)	8.69 (7.46 - 10.13)	3,518	205	3.19 (1.27 - 8.67)	11.85 (10.33 - 13.58)

* Results shown are for the 10th imputed dataset. Each TNFi-user matched 1:1 to a csDMARD user, with replacement. PS matching and analyses performed seperately for each age strata. Each TNFi-user matched 1:1 to a csDMARD user, with replacement. Standard errors were corrected for number of times a csDMARD user was included.

Sensitivity PS matching results

Results were very similar when PS matching and subsequent survival analysis was repeated following exclusion of (n=1,213) patients recruited into the TNFi cohort after the csDMARD cohort had closed. Pooled comparative effectiveness estimates were adjusted HR=0.81 [95% CI: 0.55 to 1.18] for THR and 1.11 [95% CI: 0.83 to 1.50] for TKR. Results were also similar for the analysis adding up to 180 days (vital status permitting) to the follow-up of patients after switching exposure status, i.e. stopping (amongst the TNFi cohort) or starting biologics (amongst the csDMARD cohort). The adjusted HRs from this analysis (appendix table 7.3) were 0.85 (0.63 to 1.15); p=0.29 for THR and 1.06 (0.81 to 1.37); p=0.68 for TKR.

Sensitivity IV results

The IV proved to be a strong predictor of treatment, with 4,466 (87%) of patients cared for under a physician 'preferring' biologics actually given TNFi, which was only 660 (13%) amongst patients cared for under a physician 'preferring' csDMARDs. The sample size for these analyses were lower than in the main analysis due to the first 12 prescriptions for each physician being discarded from analysis owing to their use to estimate the IV. In terms of crude associations between the 5-year THR/TKR outcomes and actual exposure, there was higher risk of THR (p<0.001) and TKR (p<0.001) amongst TNFi versus csDMARD treated individuals (table 7.5). These simple cross tabulations for the IV status showed that among those with a physician preferring TNFi there was no difference in THR risk (p=0.68) but a higher risk of TKR (p=0.006) (table 7.5). However, there were key differences in levels of observed confounders between the two groups (table 7.6), with

the instrument being associated with age, disease duration, DAS score, HAQ score, prevalence of positive ACR criteria, number of comedications, steroid use and prior joint replacement.

TABLE 7.5: CRUDE ASSOCIATIONS BETWEEN (1) TREATMENT RECEIVED AND 5-YEAR OUTCOME AND (2) INSTRUMENTAL VARIABLE AND 5-YEAR OUTCOME

	Treatment received				
	Biologic (N=9,897)		csDMARD (N=3,229)		p-value
	n	%	n	%	
THR	308	3.1	59	1.8	<0.001
TKR	508	5.1	76	2.4	<0.001
	physician preference (IV) *				
	Biologic (N=5,018)		csDMARD (N=2,901)		
	n	%	n	%	
THR	129	2.6	79	2.7	0.68
TKR	214	4.3	88	3	0.006

* Physician preference for biologics of csDMARD determined using </>= 0.5 cutoff of relative frequency of biologics within physicians' prior 12 prescriptions. Sample size is lower than full population oweing to 'burn in' for creating IV for each physician

TABLE 7.6: IV PHYSICIAN PREFERENCE (USING LAST 12 PRESCRIPTIONS)

Characteristic	IV=0	IV=1	SMD
Age	58.9	56.4	-0.21
Gender (% Female)	74	75	0.041
BMI	27.2	27.0	-0.041
Time since RA diagnosis	9.52	11.34	0.19
DAS score	5.34	6.36	0.83
HAQ mean	1.54	1.89	0.50
ACR: % >5 criteria	53	69	0.33
no. comorbidities (%)			
0	36	38	0.05
1	33	34	0.02
2	17	18	0.02
>=3	14	10	-0.13
number comedications (median)	6	7	0.27
number of prior csDMARDs (median)	4	4	0.36
Current glucocorticoid use (%)	25	37	0.28
non-major prior surgery (%)	16	22	0.14

Results of the 2-stage least squares analysis indicated no significant difference in the absolute risk of THR between patients treated with TNFi as opposed to csDMARDs (as a result of physician preference), but a significant increase in TKR risk (tables 7.7). Specifically, treatment was associated with -0.23 [95% CI: -1.34 to 0.88; *p*=0.68] less THR per 100 patients treated, but 1.86 [95% CI: 0.60 to 3.12; *p*=0.004] for TKR. Upon adjustment for the variables that were imbalanced between the two groups, there was a stronger reduction in THR (absolute risk reduction of -1.94 per 100 patients [95% CI: -3.82 to -0.06; *p*=0.043]) but no association between biologics use and TKR (0.50 per 100 patients [95% CI: -1.57 to 2.56; *p*=0.64]). Results were similar when the 12-year time-frame was used, as similar to the main analysis (appendix table 7.4 & 7.5)

TABLE 7.7: IV (PHYSICIAN PREFERENCE) ANALYSIS OF BIOLOGICS IMPACT ON NEED FOR 5-YEAR THR/TKR

	Absolute risk difference between patients treated by physicians preferring biologics versus csDMARDs (per 100 patients)			
	Estimate	lower 95% C.I.	upper 95% C.I.	P-value
IV: crude				
THR	-0.23	-1.34	0.88	0.68
TKR	1.86	0.60	3.12	0.004
IV: covariate adjusted				
THR	-1.94	-3.82	-0.06	0.043
TKR	0.50	-1.57	2.56	0.64

Sensitivity RDD results

There were 411 patients with a DAS28 of 5.0≤5.1 ("just below" the cut-off) and 682 with a DAS28 of 5.2≤5.3 ("just above" the cut-off). Patient characteristics were very similar between these two groups, although there were differences in age, disability (as per HAQ score) and pain (table 7.8).

TABLE 7.8: BASELINE CHARACTERISTICS OF RA PATIENTS "JUST BELOW" (5.00≤5.19) AND "JUST ABOVE" (5.20≤5.39) THE NICE DAS THRESHOLD FOR TNFI TREATMENT ELIGIBILITY

	Below DAS Threshold: 5.0≤5.1 (n=411)		Above DAS Threshold: 5.2≤5.3 (n=682)	
	n	%	n	%
Age: Mean (S.D.)	56.9 (13.7)		55.2 (12.9)	
Gender: Female	305	74.2%	482	70.7%
ethnicity: White/Caucasian	334	96.8%	541	96.8%
Index Multiple Deprivation				
Quintile 1	78	19.0%	108	15.8%
Quintile 2	62	15.1%	107	15.7%
Quintile 3	69	16.8%	123	18.0%
Quintile 4	72	17.5%	135	19.8%
Quintile 5	70	17.0%	106	15.5%
Unknown	60	14.6%	103	15.1%
BMI	26.7 (5.7)		27.4 (6.4)	
Smoking?				
Current/Ex	244	59.4%	413	60.6%
Calendar period of registration				
Quintile 1	52	12.7%	91	13.3%
Quintile 2	55	13.4%	114	16.7%
Quintile 3	66	16.1%	109	16.0%
Quintile 4	111	27.0%	159	23.3%
Quintile 5	127	30.9%	209	30.6%
Years since diagnosis: Median (IQR)	9.6 (9.3)		9.9 (8.8)	
Overall HAQ score	1.52 (0.67)		1.60 (0.68)	
ACR: Ever rheumatoid positive	249	60.6%	398	58.4%
ACR: Deformity of >=3 joint areas	317	77.1%	544	79.8%
ACR: Erosions on hands/feet	190	46.2%	337	49.4%
ACR: Ever had nodules	142	34.5%	242	35.5%
ACR: Symmetry	292	71.0%	507	74.3%
ACR: Deformity of hand joint	301	73.2%	531	77.9%
ACR: Morning stiffness >1 hour	370	90.0%	624	91.5%
SF36: mean (/100) (S.D.)				
physical function	34.4 (24.1)		30.7 (23.3)	
limited physically	33.7 (27.7)		31.1 (28.0)	
limited emotionally	53.6 (34.6)		50.8 (35.6)	
energy	32.7 (20.0)		32.0 (20.1)	
emotional health	59.0 (20.2)		58.8 (21.0)	
social	42.4 (20.8)		42.3 (20.8)	
pain	39.6 (21.3)		35.6 (22.0)	
general	43.6 (14.8)		43.4 (15.8)	
Comorbidities: Ever prior				
0	174	42.3%	282	41.3%
1	121	29.4%	218	32.0%
2	76	18.5%	108	15.8%
≥3	40	9.7%	74	10.9%
Current medication: Median (IQR)	7 (5-8)		7 (5-8)	
Any systemic involvement	89	21.7%	131	19.2%
Non-major prior TJR	69	16.8%	126	18.5%

continued on next page...

…continued from last page

	SMD comparing +/- different bandwidths			
	+/- 0.2	+/- 0.3	+/- 0.5	+/- 0.7
Age: Mean (S.D.)	-0.18	-0.16	-0.20	-0.24
Gender: Female	-0.07	-0.08	-0.06	-0.02
ethnicity: White/Caucasian	-0.02	-0.03	0.05	0.02
Index Multiple Deprivation				
Quintile 1	-0.10	-0.11	-0.17	-0.16
Quintile 2	0.06	0.07	0.04	-0.01
Quintile 3	0.04	0.02	0.02	0.01
Quintile 4	0.05	-0.02	0.07	0.08
Quintile 5	-0.09	-0.02	-0.01	0.01
Unknown	0.04	0.06	0.04	0.05
BMI	0.09	0.06	0.03	0.00
Smoking?				
Current/Ex	0.02	-0.02	-0.03	-0.07
Calendar period of registration				
Quintile 1	0.05	0.08	0.20	0.24
Quintile 2	0.04	0.02	-0.01	0.00
Quintile 3	-0.02	0.00	0.01	0.00
Quintile 4	-0.05	-0.06	-0.11	-0.12
Quintile 5	0.00	-0.04	-0.06	-0.08
Years since diagnosis: Median (IQR)	0.07	0.04	0.03	0.10
Overall HAQ score	0.14	0.23	0.24	0.34
ACR: Ever rheumatoid positive	0.03	0.06	0.06	0.06
ACR: Deformity of >=3 joint areas	0.10	0.06	0.06	0.12
ACR: Erosions on hands/feet	0.09	0.13	0.12	0.18
ACR: Ever had nodules	-0.06	0.04	0.04	0.03
ACR: Symmetry	0.04	0.02	0.17	0.26
ACR: Deformity of hand joint	0.04	0.00	0.03	0.08
ACR: Morning stiffness >1 hour	-0.04	-0.07	0.05	0.11
SF36: mean (/100) (S.D.)				
physical function	-0.10	-0.22	-0.22	-0.29
limited physically	-0.09	-0.21	-0.24	-0.31
limited emotionally	-0.05	-0.12	-0.15	-0.19
energy	0.00	-0.09	-0.13	-0.21
emotional health	0.05	-0.01	-0.09	-0.16
social	0.10	-0.07	-0.09	-0.19
pain	-0.16	-0.25	-0.27	-0.37
general	0.03	0.01	-0.08	-0.12
Comorbidities: Ever prior				
0	0.03	0.03	0.08	0.11
1	-0.08	0.02	-0.06	-0.06
2	0.02	-0.06	-0.03	-0.06
≥3		0.00	0.00	-0.01
Current medication: Median (IQR)	0.02	0.04	0.11	0.17
Any systemic involvement	-0.10	-0.07	-0.05	0.00
Non-major prior TJR	0.06	0.04	0.08	0.09

The differences between patients below versus above the cut-off became more pronounced and widespread as the bandwidths became wider (table 7.8). The segmented linear regression models included a total of 1,593 patients when the +/- 0.3 bandwidth was used, 2,643 patients when the +/- 0.5 bandwidth was used, and 3,721 patients when the bandwidth of +/= 0.7 was used (table 7.9). For the THR models, the impact of treatment eligibility (according to the DAS28 threshold) ranged from -6.88/1,000 PYs (p=0.16) for the narrowest bandwidth to -4.37/1,000 PYs (p=0.026) for the widest. Estimated impact on TKR rates under the narrow bandwidth scenario was -1.66 (p=0.79), which was similar for the wider bandwidths (table 7.9). Scaled results according to uptake in probability of treatment allocation are presented in table 7.10, which consisted of point estimates indicating large but non-significant reductions in THR and TKR associated with actual uptake in TNFi therapy.

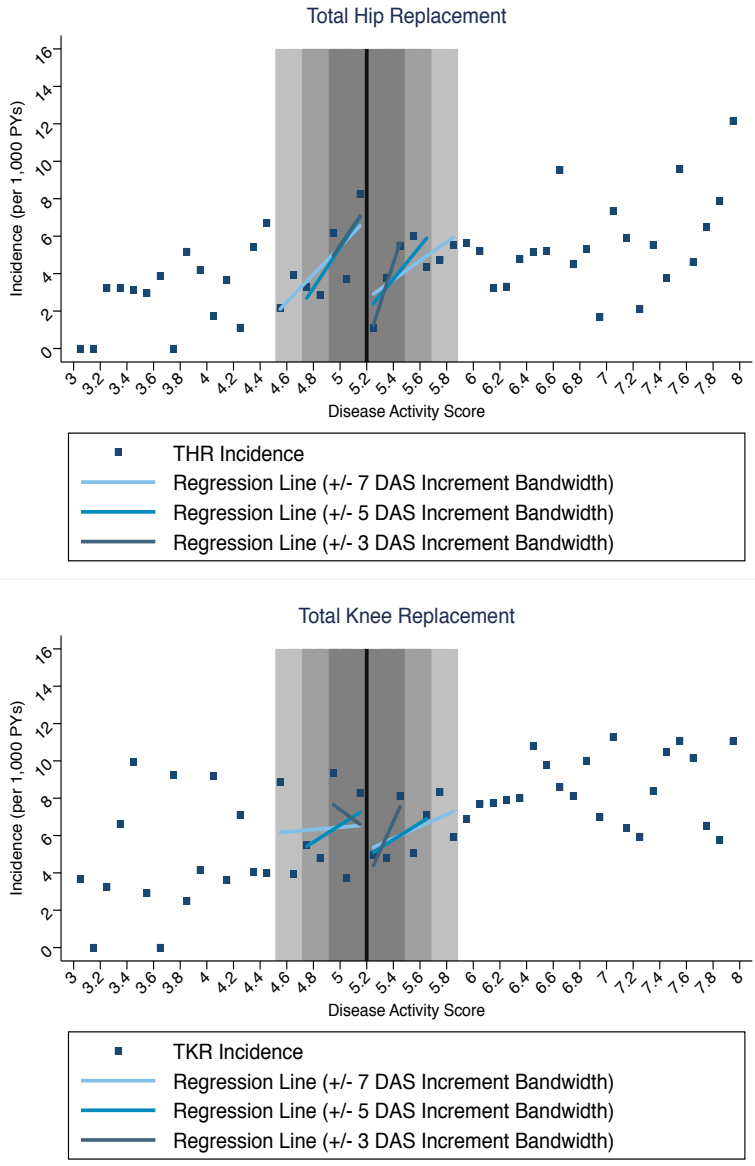


FIGURE 7.5 INCIDENCE OF JOINT REPLACEMENT ACCORDING TO DAS-28, WITH REGRESSION FITTED VALUES DRAWN EITHER SIDE OF NICE TREATMENT ELIGIBILITY CUT-OFF (>5.1) AROUND VARIOUS BANDWIDTHS

TABLE 7.9: EFFECT OF TNFI TREATMENT ELIGIBILITY (NICE 5.2 DAS THRESHOLD) ON INCIDENCE OF TOTAL HIP (THR) AND TOTAL KNEE REPLACEMENT (TKR): STRATIFIED BY BANDWIDTH AROUND THE TREATMENT THRESHOLD

	Number of included patients	Number of events of interest	Absolute difference at threshold (per 1,000 PYs)	Lower 95% C.I.	Upper 95% C.I.	p-value
THR						
+/- 0.3 DAS bandwidth	1,593	42	-6.875883	-22.47595	8.724183	0.16
+/- 0.5 DAS bandwidth	2,643	70	-5.686833	-10.98349	-0.3901766	0.044
+/- 0.7 DAS bandwidth	3,721	100	-4.3740774	-7.908102	-0.8400533	0.026
TKR						
+/- 0.3 DAS bandwidth	1,593	58	-1.667887	-25.58156	22.24579	0.79
+/- 0.5 DAS bandwidth	2,643	94	-2.583645	-9.483033	4.315743	0.40
+/- 0.7 DAS bandwidth	3,721	139	-1.226838	-6.4165	3.962825	0.61

TABLE 7.10: EFFECT OF TNFI TREATMENT ELIGIBILITY (AS PER NICE 5.2 DAS THRESHOLD) ON INCIDENCE OF TOTAL HIP (THR) AND TOTAL KNEE REPLACEMENT (TKR): STRATIFIED BY BANDWIDTH AROUND THE TREATMENT THRESHOLD: COMPLIER AVERAGE CAUSAL EFFECT (CACE)*

	CACE* (per 1,000 PYs)	Lower 95% C.I.	Upper 95% C.I.	p-value
THR				
+/- 0.3 DAS bandwidth	-94.98411	-394.5163	204.5481	0.31
+/- 0.5 DAS bandwidth	-44.90892	-105.8737	16.05589	0.12
+/- 0.7 DAS bandwidth	-30.29051	-62.53822	1.957212	0.06
TKR				
+/- 0.3 DAS bandwidth	-44.61958	-243.6128	154.3737	0.44
+/- 0.5 DAS bandwidth	-21.09271	-70.84501	28.65959	0.34
+/- 0.7 DAS bandwidth	-10.8776	-47.15181	25.39661	0.52

* complier average causal effect (CACE) is the impact of treatment eligibility, scaled by actual uptake of treatment. It can be interpreted as the effect of biologics only among those induced to initiate biologics due to the DAS eligibility criteria, rather than among those who would of taken it / not taken it anyway (always takers/never takers). Estimates obtained by using the threshold as an IV for probability of treatment assignment, one of the assumptions of which is that the threshold effect is only mediated through uptake of biologics (breach of this assumption likely explains the unrealistically large point estimates here obtained)

DISCUSSION

Main findings

This chapter addresses the scarcity of patient-level data comparing use of TNFi versus csDMARDs on rates of joint replacement in RA. Using a large UK-based RA biologics register, overall no difference in subsequent rates of joint replacement was found between PS matched TNFi and csDMARD users (figure 7.2). When stratified by age, TNFi was associated with a significant 40% reduction (p=0.008) in THR incidence among patients ≥60 years old, although non-significant increases were observed in TKR for the same age group and in THR for those <60 years old. Sensitivity analyses which sought to address the possibility of bias due to unmeasured confounding provided some support for a reduction in rates of THR but not TKR, although these analyses were largely inconclusive due to being compromised by persisting confounding and other violations of modelling assumptions.

Findings in context

The overall incidence rates of THR and TKR (results not shown) of 4.95/1,000 PYs and 7.84/1,000 PYs, respectively, align well with previous estimates of joint replacement among RA patients within the UK (chapter 3) and elsewhere (101, 102, 224).

As noted in the background, a registry-based study on this topic has been previously conducted by Aaltonen and colleagues using the Register of Biologic Treatment in Finland (ROB-FIN) (108). That study found increased rates of major joint replacement

associated with use of biologic therapy, although the authors concluded residual confounding was an issue given the small number of confounding factors for which the analysis was adjusted. Specifically, only age, sex, disease duration, HAQ, visual analogue score and number of prior joint replacements were included. This therefore is most likely still biased by residual confounding by indication, given that biologic users in the present analysis had much greater DAS28 scores, worse quality of life measures and greater rates of steroid use, amongst other differences. Furthermore, whilst HAQ was there matched on, their results clearly show that even after matching, HAQ was significantly higher in the biologics group yet this was not further adjusted for. Interestingly, whilst significantly higher rates of knee and other joint replacement were there found in the biologic group compared to csDMARD group, hip replacement rates were found to be similar (0.94 [0.72 to 1.22] versus 0.89 [0.73 to 1.08] per 100 PYs).

Potential Mechanisms

As has been previously described, emerging observational data indicate that the number and/or incidence of RA related joint surgery has been in decline across many developed countries (91, 92, 94, 95, 98, 100, 102, 134, 135, 224, 269-271). Many of these prior studies have inferred a role of biologics in this decline and the analyses performed within chapters 4 & 5 specifically demonstrated this within England & Wales, Denmark and Ontario. Given this context, it seems plausible to have here found a 40% reduction in THR rates associated with TNFi use among a more elderly subgroup of patients, and the lack of translation of positive findings on joint erosion from prior RCTs into a "real world" reduction in rates of joint replacement within the main study sample is initially

surprising. For instance, one meta-analysis of 70 RCT studies found annual radiographic progression was 0.6% less in patients treated with biologics compared to those on a single DMARD (85). Similarly, another meta-analysis has demonstrated that patients on initial combination therapy (methotrexate plus a biologic agent) are 30% more likely to experience non-progression at 1-year than those on methotrexate alone (88).

However, there is heterogeneity in existent RCT data and this suggests that an expectation of widespread reduction in joint replacement associated with TNFi use is potentially unwarranted. For instance, whilst RCT studies have shown reduced radiographic progression among biologic users versus csDMARD monotherapy, an almost equal reduction has been achieved among combination csDMARD users relative to csDMARD monotherapy (85). Similarly, whilst TNFi has been shown to confer early benefits over combination csDMARD therapy, these benefits have been reported to disappear during the second year of follow-up (86), possibly due to time-to-efficacy and time to achieve maximal dose of csDMARDs. The use of etanercept (vs. oral triple therapy) resulted in only small radiographic benefits in another trial (272), with another showing (csDMARD) triple therapy to be non-inferior to biologics in terms of change in DAS28 at 48 weeks (273). There is a large degree of variation in the nature of comparator groups used in prior studies (89, 90, 257), and this could be an important factor in considering the lack of effect as found for the main comparisons here carried out. It maybe the lack of overall effect reflects the potential non-inferiority of high intensity csDMARD versus TNFi on radiographic outcomes seen over the longer-term in RCTs. It's worth noting that the median number of prior/current csDMARDs among the matched

groups was 4 (IQR: 3-6), suggesting a reasonably aggressive use of conventional therapy in both groups. Furthermore, a recent investigation of the prevalence of sustained remission and/or low disease activity amongst TNFi users within the BSRBR-RA reported that overall this was infrequent: 15% for remission and 26% for low disease activity. These figures were not investigated amongst the csDMARD cohort, and a robust BSRBR-RA comparative effectiveness analysis of TNFi versus csDMARDs on longer-term DAS28 and HAQ is a potential subject for further research.

The reduction in THR incidence among an older patient subgroup is interesting given no corresponding reduction in TKR incidence was found, yet one might expect any effect to be expressed approximately equally at hip and knee if not more so at the knee given data from previous chapters suggested RA involvement to be greater at the knee than hip. One tentative reason could be that the long disease duration at baseline (~10 years) meant there was already more destruction of knees within the cohort, leading to relatively greater potential for prevention of further destruction at the hip over knee. Although a possibility, details of differential natural history of RA disease at these two joints are not well established despite some data reporting greater synovitis at the knee versus hip (187-189). It's also very difficult to disentangle the impact of TNFi on improved function and overall quality of life and how this may have mediated effects on longer-term progression of joint damage, potentially differentially at the knee and hip. Although somewhat speculative, it is interesting that the unmatched TNFi group were younger yet had higher rates of anti-osteoporosis medication and calcium supplementation, providing a possible pathway in reduced trauma-related THRs via TNFi associated

improvement in bone quality, especially among the older subgroup. It should be emphasised however that the positive impact on THR incidence was only observed in a subgroup of patients that have not been well studied in this regard previously, although previous analysis of published RCTs has shown a trend towards greater effectiveness of biologics with age (257).

Strengths and Limitations: PS matching

Some of the study's key strengths are related to the use of the BSRBR-RA. This is one of the largest RA registers in the world which made it possible to adjust for many potential confounders and stratify analyses where there was significant effect modification. It was also possible to accurately censor follow-up given the linkage to NHS Digital (formerly HSCIC) to access mortality data. The use of PS matching is a strong method for dealing with bias arising from the likely confounding by indication (274) and generally speaking good post-matching balance of characteristics between the two exposure groups was achieved across most variables, particularly those most predictive of joint replacement (34, 137). The PS analytical approach taken, in which comparable csDMARD matches were found for each TNFi user - while much improving the internal validity - does mean our findings are not average treatment effects generalizable to the entire RA population but an estimate of the "average treatment effect on the treated" (ATT). That is, it estimates the treatment effect among those on TNFi compared to had they not been exposed to TNFi (275). This is evident given that the csDMARD users included in the matched sample had more established and severe disease (thereby making them comparable to the TNFi sample) than the unmatched csDMARD sample (Appendix Table

7.2A – 7.2B). The study findings should also be interpreted in the context of a study sample with relatively long median disease duration at baseline (table 7.2), which means findings are not necessarily generalisable to the context of early RA. Furthermore, the study outcome was composed of a combination of physician-reported and self-reported incidence of THR, TKR and OJR as per BSRBR-RA follow-up questionnaires. This may have introduced misclassification bias if events were under-reported, although this would likely act non-differentially in regard to TNFi status and minimally during the early years of follow-up, during which time study participants were sent questionnaires every six months.

However, despite the richness of the data resource, the potential for residual confounding by indication is a key limitation of the current study given that the number of variables for which data were collected was still limited (and that treatment was not randomly allocated). For example, prior to matching there was a much higher disease severity among the TNFi group (Table 7.1) thus giving rise to the possibility that the overall lack of reduction in joint replacement rates among TNFi users may be due in part to a greater prevalence of unmeasured aspects of disease severity and unresponsiveness to therapy in this group. That is, a possible baseline 'disadvantage' persisted even after PS matching (34). On the other hand, estimates of impact of TNFi exposure may be subject to a general healthy user bias in that a clinician perceives sufficient patient ability to tolerate and benefit from more intensive therapy regimens, which may have here contributed to the reduced rate of THR in the older TNFi cohort.

Furthermore, while residual baseline confounding may have impacted on treatment estimates, it's also quite likely time-varying confounding will have affected the results too (276). That is, whilst covariate balance was achieved at baseline, these patient groups may well have subsequently deviated substantially over time in terms of variables that confounded the exposure -> outcome relationship. This may have been compounded further were these variables to partially mediate the impact of therapy on the outcome, given that this would rule out a conventional time-varying confounder adjusted analysis (276, 277). Similarly, there may have been some "healthy adherence" bias where sicker patients (possibly at higher risk of joint replacement) were differentially censored (259). These aspects remain the subject for future investigation.

Strengths and Limitations: IV sensitivity analysis

The use of an IV approach as a sensitivity analysis sought to address the issue of unmeasured confounding, assuming physician preference to be a strong predictor of exposure but unrelated/weakly-related to confounders of the TNFi versus joint replacement relationship. However, the findings of this IV approach should be interpreted with caution given the instrument was here associated with several measured confounders (table 7.6) which undermines its validity as a means to obtain unbiased estimates in the presence of unmeasured confounding. Although this constituted failure to meet one of the key assumptions of the IV approach, there was good evidence indicating physician preference was a strong predictor of TNFi treatment ('relevance condition' assumption) (278, 279). However, the exclusion restriction assumption that physician preference was independent of outcome except via

prescriptions of the actual treatment (261, 279) is clearly not valid. Reasons for the lack of independence between the IV and other confounders could be explained if physicians apparently 'preferring' TNFi were also the physicians caring for patients with relatively more/less severe RA, or even possibly physicians temporally clustering their consultations according to disease severity. There is still the entangling issue that physicians more likely to prescribe biologics may also be prescribing csDMARDs more aggressively and earlier. Despite this lack of independence, it's reassuring that the adjustment of these known confounders acted to move the IV estimate for THR away from the null and towards a protective effect, suggesting that adjustment for further (measured and unmeasured) characteristics might only act to confirm/strengthen the protective effect on THR. However, this cannot be known for sure and whilst the GRADE criteria (280) explicitly recommend upgrading the rating of the quality of evidence produced by an observational study when it is considered that residual confounding would act to further the confidence of an observed effect, it would be more prudent to be cautious regarding the IV results.

Strengths and Limitations: RDD sensitivity analysis

The availability of a known treatment threshold for initiation of TNFi provided a rare opportunity for carrying out an RDD approach sensitivity analysis, previously described as one of the strongest possible study designs when an RCT is not possible (266). While meeting the >5.1 DAS28 threshold cut-off was here associated with a reduction in THR incidence, this should not be interpreted as the effect of treatment per se but rather the effect of meeting treatment eligibility criteria concerning DAS28. This estimation of

meeting the DAS treatment eligibility criteria is of interest in and of itself, but unless treatment assignment is completely deterministic at the threshold then scaling of the effect estimates is recommended in order to generate the CACE, i.e. the estimated causal effect of treatment among those who initiated it as a consequence of meeting the threshold (with the fuzziness of "always takers" and "never takers" removed).

A 2-SLS IV approach was carried out to estimate the CACE at the threshold (table 7.10), which showed the actual impact of uptake in TNFi at the threshold to be non-significant. However, there are strong reasons to suspect these results are invalid given failure to satisfy multiple assumptions. Whilst appendix figure 7.2A suggests there was indeed a discontinuity in the numbers allocated TNFi around the threshold, albeit fuzzy, once these were translated into probabilities (relative frequency of TNFi compared to all patients within each DAS score unit) there was very little discontinuity to be observed (appendix figure 7.2B). That is, the threshold hardly predicted treatment probability and therefore a causal effect on THR/TKR could not be reliably identified because as an IV the threshold is evidently excessively weak (267). Furthermore, this undertaking highlighted a limitation of attempting to use an RDD approach in the context of data from a drug registry that primarily sought to recruit users of treatment, as there was not a generalisable sample of untreated patients (appendix figure 7.2C). This is apparent from appendix figure 7.3 which indicates there was significant discontinuity in the density of the assignment variable (DAS) at the threshold which would be very unlikely to naturally occur in a representative sample of all RA patients in the real world (122). The consequence of this heavier emphasis on sampling treated patients is that the

impact of eligibility on the *probability* of treatment cannot be reliably ascertained, yet estimation of the CACE relies upon the derivation of this probability of treatment, below and above the threshold, with which to scale estimates of the threshold effect on the outcome.

There is also a clinical problem in CACE estimation in that this uses the >5.1 threshold as an IV for exposure to TNFi. This assumes any of the impact of the IV on outcome (THR/TKR) is mediated solely via TNFi, yet in the situation of very little discontinuity in the probability of treatment (i.e. a very weak IV) then this is inappropriate and likely explains the unrealistically large/biased CACE estimates (278). Given that a DAS of >5.1 places a patient within the "severe RA" category this is likely to in and of itself put them at modified risk of THR/TKR in a way that is not mediated solely by TNFi exposure. For example, through increased monitoring, access to services and more intensive usage of conventional therapies. Linked to this is the fact there were other aspects of treatment eligibility - other than the cut-off - which contributed to overall NICE eligibility, such as failure to adequately respond to two csDMARD regimens or that the elevated DAS of >5.1 had to be "continuing". While the exclusion restriction assumption (that patients are very similar at the cut-off) appeared to be met within the smallest DAS bandwidth here inspected (+/- 0.2) (table 7.8), the "just below" and "just above" groups became increasingly dissimilar as the bandwidths widened. This is a limitation of the naïve analysis of threshold impact on THR/TKR incidence, given that only the wider bandwidth analyses produced a significant reduction in THR although these were not adjusted for confounding, which was demonstrably existent in the wider bandwidths (table 7.8). As

has been previously observed (267), it's likely the smaller bandwidths are susceptible to less bias but also less precision in estimation of eligibility impact, and vice-versa for wider bandwidths. Results from narrower bandwidths are also limited in their interpretation as they are restricted to a relatively very small proportion of the total RA population.

Further work

It would be informative to investigate whether these findings are replicated in healthcare settings elsewhere in the world. Although this process has begun, using the Danish data as used in chapter 5 linked to DANBIO, this investigation is in its infancy. Exploring the impact of TNFi among incident RA cases would be interesting given that the BSRBR-RA contains unusually established RA (median disease duration at baseline approximately 10 years). There is the potential to improve the PS matching as performed here, possibly with the use of machine learning algorithms to hone the PS estimation model in order to improve/maximise balance. Although some exploration of this issue was carried out in this project (results not shown), the available Stata package (eltmle) used produced no better balance between matched cohorts than that achieved with the standard logistic regression approach already used. This is likely because eltmle optimises model fit rather than balance between groups, while for PS matching it is balance rather than model fit that is the ideal target. Further work also includes using more computationally intensive methods to address time-varying confounding.

Conclusions

This chapter has reported on the patient-level impact of TNFi versus csDMARD therapy within a large prospective registry of RA patients. Overall, no difference in subsequent incidence of joint replacement was found between matched TNFi users and csDMARD users, although a 40% relative reduction in THR rates was found among older patients. Sensitivity analyses seeking to address unmeasured confounding suggested TNFi may be associated with lower rates of THR, although these analyses should be interpreted with caution given their strong assumptions. Future studies are needed to confirm these findings and/or further elucidate the relationship between TNFi use and joint replacement.

CHAPTER 8: DISCUSSION AND CONCLUSION

Summary of main findings

This thesis set out to describe the relationship between TNFi treatment and subsequent need for hip and knee replacement amongst RA patients, primarily within the UK. Joint damage is a central feature of RA and its progressive and often permanent nature, in conjunction with pain, loss of function and failure to adequately respond to therapeutic options are strong indications for eventual joint replacement surgery (33, 34).

However, little data exists on basic epidemiological descriptive parameters concerning hip and knee replacement amongst RA patients in the UK, which is regrettable given that these are major and expensive operations. More information on these RA-related procedures would likely be of interest to patients, clinicians and those allocating health-care resources. This scarcity of data was addressed, as described in chapter 3, using population-based routinely collected health data to bring to light various features of this long-term but less-studied outcome. Cumulative incidence function plots (figure 3.1) indicated that the approximately linear progression of joint erosion (as known from prior literature (25)) is reflected in an approximately constant accumulation of THR/TKR risk from the time of RA diagnosis. This rose to approximately 5% and 7% at 10 years for THR and TKR, respectively. These estimates in conjunction with previously reported estimates

for the general population (147) suggest a greater need for both THR and TKR among RA patients than for non-RA individuals. Average time-to-event, from RA diagnosis, was 3.2 [IQR: 1.3, 6.5] and 3.5 [IQR: 1.5, 6.7] years for THR and TKR, respectively. There was significant variation across several demographic characteristics. Both THR and TKR rates increased with age at RA diagnosis, although these rates declined among patients ≥75 years old. THR rates were higher among females and TKR rates increased with increasing BMI. There were lower rates of THR and TKR among current smokers and lower rates of both these procedures according to increasing deprivation (IMD). Geographic variation in TKR rates was an interesting finding, suggesting inequalities in either need and/or provision of these procedures in the UK.

The issue of temporal variation in THR/TKR across the study period was not included in the descriptive epidemiology study as this was addressed in detail in the context of testing the hypothesis of whether there were changes in the level and/or trend of these procedures following NICE approval of TNFi for RA in 2002 (chapter 4). This was a relatively novel approach (ITS) in that whilst numerous previous studies from elsewhere in the world had reported on descriptive temporal trends in joint replacement and anti-rheumatic prescriptions amongst RA patients, none had formally investigated whether the rates of joint replacement within RA changed following the introduction of biologics. The results of this population-level analysis indicated that the introduction of TNFi therapy in 2002 was associated with a subsequent reduction in rates of TKR but not THR within England & Wales, as compared to what would have been expected based on a continuation of pre-TNFi era data. Specifically, while controlling for secular trend, there

was a -4.41 [95% CI: -6.88 to -1.94] per 1,000 PY reduction in incidence of TKR associated with introduction of TNFi therapy.

The consistency of findings from the UK analysis of temporal association between introduction of TNFi and joint replacement rates in RA was then investigated by repeating the ITS analyses using data from two different countries. These 'external validation' studies made use of large routinely collected health data from Denmark and Ontario (Canada). The Danish data on the impact of TNFi introduction on joint replacement rates in RA corroborated findings from England & Wales, in that an overall reduction was observed in rates of TKR but not THR (figure 5.11). Conversely, the data from Ontario suggested rates of THR but not TKR were reduced among RA patients following the introduction of TNFi, and this was particularly evident when this latter analysis was stratified by age (figures 5.7 to 5.10).

In addition to the age stratified analysis in Ontario, several other extensions were made to the basic time-series method within these 'validation' cohort analyses. The first of these consisted of analysing the temporal association between the introduction of TNFi in 2002 with subsequent incidence of prescriptions for biologics and csDMARDs within the senior subset of the RA cohort from Ontario. This clearly showed that an increase in the incidence rates of both biologics and csDMARDs prescriptions occurred concomitantly with the introduction of TNFi (figure 5.5).

The second of these extensions was to incorporate a non-RA control cohort to both validation analyses in order to test whether there was a difference in the differences (in THR/TKR between the RA and non-RA cohorts) post-introduction of TNFi. That is, did the intervention impact the delta between the two exposure cohorts. These analyses indicated that while a reduction in joint replacement was observed in the overall non-RA control cohorts (except for TKR in Ontario, where rates were stable), these were smaller than that observed in the RA cohorts yielding an overall shrinking of the deltas between the two exposure cohorts during the biologic era (figure 5.12). Furthermore, among the younger strata in Ontario, the introduction of TNFi was associated with divergent effects on THR between the exposure cohorts: an increase among the non-RA cohort but a simultaneous reduction in the RA cohort (figure 5.7), strengthening the case for some role of biologics.

The key clinical question of the causal impact of TNFi therapy on need for joint replacement was addressed in chapter 7, using BSRBR-RA drug registry data. The results of this PS-matched patient-level analysis, comparing TNFi users to users of csDMARDs (not on biologics) revealed no significant differences in joint replacement rates between the two exposure groups. This was the case for the THR, TKR and OJR outcomes (figure 7.2). Pre-specified stratified analyses according to age (*p*-interaction <0.001) highlighted a stronger reduction in THR rates among TNFi users ≥60 years old compared to csDMARD users <60 years old (HR=0.60 [95% CI: 0.41 – 0.87]), although rates among younger patients were not significantly different between the exposure cohorts nor were rates of TKR in either age group (figure 7.3). Two sensitivity analyses were conducted using

methods that sought to account for the possibility of unmeasured confounding. Neither of these provided firm conclusions given a failure to meet key modelling assumptions but were nonetheless overall supportive of the main PS findings of no impact on THR/TKR, although they could be perceived as providing weak evidence of a reduction in THR rates associated with TNFi (table 7.7).

The aims of this DPhil project were not merely clinical but also methodological. Specifically, to advance the relatively novel quasi-experimental approaches in ways that were feasible within the time constraints of a 3-year project. To this end, a simulation study was conducted to provide a means of estimating power within ordinary least-squares ITS analysis, in order to facilitate researchers in achieving an adequate sample size for future studies in similar contexts. This undertaking revealed that numerous parameters need to be collectively considered when seeking to estimate an adequate sample size for an ITS study. All of the following were found in at least some scenarios to be influential: pre-intervention value of outcome, magnitude of intervention effect, whether intervention impact is mediated via a step or slope change, location of intervention in time-series, number of time-series timepoints and sample size per timepoint. A Stata programme was developed to loop through the simulations pertaining to numerous scenario-specific parameters. An illustrative example of this was where a mean pre-intervention level of outcome was 3.5%, to achieve 80% power to detect a relative 34% mid-timeseries post-intervention step change reduction, with 14 pre- and 14 post-intervention timepoints, one needed over 1,000 subjects per timepoint (i.e. >28,000 total subjects). Furthermore, several novel methodologies in the area of

epidemiology (e.g. the IV and RDD approach) were implemented. Although in this particular application the assumptions were not met and these analyses remained largely invalid, they none-the-less provided demonstration of the potential within the field of musculoskeletal pharmacoepidemiology for addressing the issue of confounding by indication.

Strengths and limitations

In assessing the strengths of the project, the clinical and financial relevance of the question being addressed should be briefly mentioned. This is evidenced by the fact the literature states some form of causal relationship between TNFi and joint replacement is suspected (44, 108, 281), the inclusion of which into economic modelling is already being explored (281), yet the nature of which remains undetermined.

One of the first and foremost strengths is probably the high-quality methodological approaches applied. Given that carrying out an RCT on the subject would likely be unfeasible (given ethical, financial and follow-up requirements), the "next best" approach would seem to be methods described as quasi-experimental in nature (112). That is, the methodologies employed within chapters 4-7 are broadly considered some of the best available alternatives which attempt to mimic a randomised experiment (113, 170, 266). Their high quality is largely due to the combination of high internal validity (when modelling assumptions are met), especially for approaches that deal with unmeasured confounding (116, 282), whilst combining high external generalisability compared to the narrowly defined population of most RCTs (115). For example, although

the PS matched cohorts in chapter 7 were not representative of all RA patients, as a conventional multivariable regression-adjustment approach might be, yet nonetheless they yielded estimates of the causal treatment effect amongst patients with characteristics reflecting "real world" users of TNFi. Likewise, the robust difference-in-difference approach to modelling the differential changes following introduction of TNFi among RA patients over and above non-RA patients provided a means to control for many biases simultaneously, while remaining in a study context that was highly generalizable.

Furthermore, the breadth of methods applied is comprehensive. The "big 5" quasi-experimental designs have elsewhere been listed as: *"(1) Instrumental Variables, (2) Regression Discontinuity, (3) Interrupted Time-Series, (4) Difference-in-Difference and (5) Fixed Effects Designs*" (283), four of which have been used here. What is reassuring is the consistency of findings amongst (although admittedly, not between) the population-level and patient-level analyses performed, building a more solid body of evidence. For example, the main findings from PS-matched, IV and RDD were all indicative of no causal association between TNFi and joint replacement, although these latter two approaches were to some degree inconclusive due to the invalidation caused by breaches of assumptions.

Following on from the breadth of the methodologies is the breadth of the data utilised. The availability of multiple data sources permitted the investigation of the impact of TNFi introduction at the population-level in three different healthcare systems from at least

three different countries, which again indicated consistency. The population-based ITS analyses collectively showed a reduction in average TKR rates in RA patients compared to counterfactuals during the biologic era (figure 5.11), although non-significantly in Ontario. Similarly, all these ITS analyses showed no overall significant impact on THR rates in RA patients (compared to counterfactuals) (figure 5.11), although in Ontario weak evidence was found for a reduction which became a lot stronger in age stratified approaches (table 5.9). Another strength of the data, in addition to the number of data sources used, was their size. The Danish dataset covered the entirety of Denmark and the size of the even larger Ontario sample permitted age-stratification. Although the size of the BSRBR-RA is also noteworthy, its key strength is likely it's richness in terms of variables included, here enabling a robust PS matching procedure that has clinical credibility. A previous attempt to estimate the impact of biologics on joint replacement using PS methodology was severely limited given a lack of variables used to generate comparable groups (108), rendering such prior estimates lacking in basic reliability. Here there were many demographic, quality of life, co-medication and clinical variables used to generate and demonstrate baseline balance.

However, as noted in chapter 5 in the context of the difference-in-difference ITS approach, not all "strong" study designs are equally robust. There is often a gradient. This principle applied more generally leads to the issue of the comparative strengths and limitations of the various components of this project, which are now discussed.

The difference-in-differences ITS approach was clearly a useful addition to a single-group ITS approach, making the modelling in chapter 5 superior to that in chapter 4. There were though several limitations that applied to all these population-level ITS analyses. Firstly,

a 5-year incidence rate outcome was used as this was deemed necessary to ensure comparability of measures throughout the time-series, rather than individuals at the beginning of the study period having more follow-up to those at the end. Although restricting analyses to incident rather than prevalent RA is arguably a strength as this reduces the confounding of previous exposure and disease duration, overall this does however limit the scope of these studies to the situation of early RA. What all these population-level estimates failed to rule out was confounding at the same time as the introduction of TNFi in the RA group, as they were essentially ecological. A particular concern is the increase in csDMARDs that was observed to occur concomitantly to introduction of TNFi among the Ontarian seniors, suggesting this may have also been an issue in the UK and Danish analyses. Were this to be the case, it is probable that the observed reduction in TKR rates (figure 5.11) at the population-level were influenced to a greater or lesser extent by csDMARDs rather than solely biologics. More fundamentally, these analyses have to be interpreted as the impact of TNFi availability rather than the causal impact of TNFi use versus non-use. That is, even if confounding factors were to be inconsequential or even non-existent, there is almost certainly non-compliance in that there will be some "always takers" (before the guideline) and many "never takers" (after the guideline). This means the impact estimates are at best the evaluation of health policy impact on an RA population rather than the effect of exposure per se on an individual patient's risk.

On the other hand, the patient-level analyses of chapter 7 were much stronger at addressing the very issues that the population-level analyses of chapters 4-5 had been relatively weak to resolve. These patient-level analyses sought to disentangle the issue

of temporal confounding (e.g. csDMARD intensity) concurrent to the introduction of TNFi, whilst meeting the research need for a clinically meaningful causal estimate of TNFi exposure. A-priori, the collective use of PS, IV and RDD analyses was an exceedingly powerful set of methods for the attainment of unbiased estimates in the presence of both measured and unmeasured confounding. Had the assumptions of the IV and RDD been fully met then these would theoretically have been more robust than the PS matching as the latter demonstrably works to balance measured covariates but by design cannot provide balance on those unmeasured (except via proxy). However, given the failure of the IV and RDD approaches to meet their key assumptions the weight of confidence in the conclusions from chapter 7 should rest mainly on the PS findings as these were less stringent in the demands they made about mechanism of effect, yet still achieved good balance across a host of relevant confounders.

Whilst the strengths of the robust analysis of rich, patient-level data from the BSRBR-RA should be cause for reassurance, there are nonetheless various important limitations. Although balance was mainly established in the PS matching, the issue of whether the matched groups were really comparable as per a well conducted RCT is doubtful. Some degree of residual confounding by indication is likely given the exposure cohorts were so different before matching and that the number of variables included in the PS were finite. This is a well-recognised challenge of using observational data for comparative effectiveness research (as opposed to a comparative safety study) (110) and is a key limitation of the patient-level approach, which could conceivably bias results in either direction. It may be that no overall effect of TNFi was identified in the main analysis

owing to aspects of disease severity and/or unresponsiveness to conventional therapy within the TNFi group that persisted after matching and put this group at a baseline 'disadvantage'. On the other hand, it could be that the reduction in THR amongst elderly TNFi users was driven by a healthy-user bias where a prescribing clinician perceived greater ability to tolerate and benefit from more intensive therapy regimens which was itself a possible driver of subsequently lower joint replacement rates.

There are other data-related limitations arising from the BSRBR-RA. It initially recruited all users of TNFi nationally, which given the novelty of the therapy meant a disproportionately large number of severe RA cases with long disease duration entered the study sample. This preponderance of established disease means the results are limited in their generalisability. Although the availability of ongoing longitudinal follow-up data is a clear strength, is does provide greater opportunity for the exposure cohorts to become increasingly confounded over time in ways that mean joint replacement rates diverge for reasons other than TNFi exposure status. Similarly, adherence bias is a potential issue given that patients remaining on therapy maybe different in their risk of outcome relative to those stopping and being censored. For instance, if csDMARD users adhering in this exposure group were in the main to be doing so due to responding well and having their disease under control, then this progressively makes this group less at risk of joint replacement than the TNFi group where patients who respond well and have their disease under control may be more likely to stop therapy and become censored. However, these issues would likely still be encountered in a hypothetical long follow-up RCT on the subject and they are not easily resolved.

These varying strengths and limitations to the analyses utilised (many of which are the consequence of underlying differences in study design) are possibly the explanation for differences in the results obtained between the population-level (chapter 4 & 5) versus patient-level (chapter 7) studies. While the population-level analyses supported a reduction in TKR rates amongst RA patients, and a reduction in both TKR and THR rates when changes in the non-RA population was accounted for, neither of these reductions were reflected in the main analysis of patient level BSRBR-RA data. As described above, this maybe because the population-level analyses are confounded by an increase in csDMARDs (which the PS analysis accounted for) or it could be the patient-level analysis are confounded by residual imbalance in disease severity/responsiveness to therapy (which the ITS analyses accounted for). Besides this, the established, severe and prevalent nature of baseline RA with subsequent long-term follow-up in the BSRBR-RA is a very different context to the incident (and most likely relatively mild) nature of RA with only 5-years follow-up in the ITS approach. Essentially, chapters 4, 5 & 7 are answering variations around the central clinical question of the thesis which, although not inconsistent with each other, are not necessarily expected to be identical in terms of findings.

Future work

Although the findings here reported move the field on in terms of a better understanding of the relationship between TNFi and need for joint replacement in RA, there are many potential avenues for future research.

The descriptive epidemiology of hip and knee replacement (reported in chapter 3) should motivate further research into the inequalities in rates of THR/TKR according to deprivation and inequalities in rates of TKR across geographic region for patients with RA. It maybe that new strategies are needed to rectify inequality in provision of services, although this is by no means adequately demonstrated from this DPhil as there are other possible explanations related to patient need and willingness which are all subjects deserving of further investigation.

The ITS population-trend analyses have been carried out in three countries, and while further large study samples would add further to this body of evidence, probably more informative going forward would be analyses on the impact of TNFi introduction on rates of smaller joint surgeries such as hand, wrist, elbow, shoulder and feet. Given their smaller size it maybe these joints suffer destruction within a faster and/or more consequential disease process, into which the arrest of inflammation via TNFi intervention maybe more pronounced. Another interesting extension of the current ITS analyses would be to gain access to a similar population-based study sample that includes data on TNFi and csDMARD therapy (as per the senior sub-group in Ontario) over the study period in order to get a better understanding of the relative effects of TNFi availability on actual utilization.

The issue of power in ITS analyses has already been explored in some detail in chapter 6, however there are many extensions that could be made to this piece of methodological work. The growing utilization of ITS analysis in epidemiology presents a need for more

guidance on this topic, and of particular benefit would be an interactive tool which could be freely available to researchers planning future ITS studies in order to gain insight into sample size requirements to achieve adequate power in relation to the various parameters investigated. This process has already been initiated with the ongoing development of an R programme but this is yet to be completed. Furthermore, a pressing limitation of the current work is that it only considers an outcome measurement in the form of cumulative incidence, so more work is required to modify the code already developed to simulate scenarios for incidence rates and raw count data. Also, whilst the issue of auto-correlation is difficult to satisfactorily address in situations of less than 50 timepoints, this is nonetheless another factor to be incorporated into such a sample size calculator tool.

Although there was the valuable opportunity to carry out the ITS analyses evaluating trends in joint replacement from three different countries, there is the outstanding issue of replication of the patient-level PS matched analysis of chapter 7 using data from elsewhere in the world. This is an important research need as there are a number of peculiarities specific to BSRBR-RA that likely act to limit the generalisability of the main patient-level analyses here carried out. For example, the long disease duration at baseline and the fact those recruited reflect UK policy of restricting TNFi to those with continuing severe disease (DAS >5.1) who have already failed conventional therapy for over 6 months (237). To truly answer the research question under study, further research such as an international RCT of incident RA patients with a very large sample size and follow-up (given the nature of the outcome) would be an ideal study design to aspire to.

This however is very unlikely to be feasible given the ethical implications and costs, which are likely to be prohibitive (as discussed in chapter 1). Combining observational data from other registries, for example RABBIT (biologics registry of 17,000 RA patients in Germany), may be the next best alternative. Combining a number of biologics registries in this way would allow stratification on more variables and potentially improve generalisability in terms of disease duration and DAS values of patients at baseline. It would also allow estimation of TNFi effect in healthcare systems without the >5.1 DAS threshold rule as per UK guidance. This testing of the same study hypothesis using more data and in different healthcare settings from elsewhere in the world would be prudent before any firm conclusions are drawn.

In terms of specific advice to future researchers seeking to replicate or adapt the methodological approaches used in this thesis, I would comment that the use of an ITS is certainly worth considering given the strengths of this design and relative ease of conducting such a study. However, some of the strengths of this approach are lost in contexts of exclusively long-term outcomes or where there is no well-defined intervention period. The use of a 5-year outcome and 1-year/2-year intervention period in this project addressed these limitations to an extent but demonstrates the method should not be applied irrespective of these considerations. Also, if serious confounding events occur concomitantly to a proposed 'intervention' event then the strength and therefore usefulness of the findings from the ITS approach are likely to be reduced. In terms of using an ITS to investigate similar research questions in the future, it would seem sensible to prioritise access to prescriptions data (exposure) rather than just clinical

outcomes. Comparable control populations to inspect the difference-in-differences in outcome should also be made available if at all possible. Furthermore, this thesis demonstrates that patient-level approaches are likely to be essential to properly evaluate questions of comparative-effectiveness given the inability of population-level analyses to truly rule out confounding events and ecological fallacies in general. PS matching is a good option to deal with confounding by indication at the person-level but only if lots of measured covariates are available to balance on, including at the very least those most predictive of outcome. One thing to consider is the ratio of treated and untreated patients available in prospective data sources and to ensure there are sufficient numbers in each group for an analysis to be feasible. It is worth trying to do an IV or RDD approach (if the situation permits) to account for unmeasured confounding and testing assumptions but I would not recommend using these as primary analyses given their strong underlying assumptions that underpin their validity. Also worth mentioning is the need to use fairly advanced epidemiological methods simultaneously in a single analysis, e.g. the main analysis in chapter 7 used weighted survival analysis on propensity score matched cohorts in the context of multiple imputed datasets to account for missing data. It is therefore essential to develop and maintain well documented programming code within a capable statistics package such as R or Stata.

Clinical implications

Despite the arising need for further research, there are still numerous outputs from the current project that could have potential clinical implications. Firstly, the basic descriptive statistics of the cumulative risk of THR/TKR during the course of RA (chapter

3) and how the delta rates between RA and non-RA patients are still in existence yet are shrinking (chapter 5) is deserving of clinicians' attention. That is, major joint replacement is an important long-term outcome of RA that affects a non-trivial proportion of patients but seems to be a modifiable risk. While this may not noticeably change clinical practice, it should function to increase awareness of this important long-term disease outcome among RA patients and those managing their care.

Outcomes for hip and knee replacement in RA are generally considered good (37), with marked improvement in patient quality of life (as measured using SF36) following surgery (284). However, these patients are known to be at increased risk of various adverse events compared to patients undergoing these procedures for osteoarthritis including: infection, dislocation, myocardial infarction and revision (285, 286). Furthermore, these are expensive operations (138), and even when successfully performed still put the patient at a lifelong risk of requiring revision surgery. Given this background, a reduction in need for joint replacement in RA is clearly to be sought after and the growing belief over recent years, both anecdotally and in the literature (237, 281) is that TNFi therapy has a positive impact on joint replacement rates in RA. New data on the possible protective role of TNFi should be of genuine clinical interest, for example given that it could be incorporated into a health economic model of TNFi. Were such data to be of sufficient quality to be used in this way then this could potentially change clinical practice in that TNFi would probably be deemed more cost effective and restrictions on use could potentially be made more lenient.

However, no overall effect was observed in the PS matched cohort study and (as was discussed in chapter 7) the RCT data is heterogeneous and not unanimously supportive of biologics being superior to combination csDMARD therapy, especially over longer time-frames (86, 287). In the light of this, but also in the light of the limitations as discussed in previous paragraphs relating to methodologies and data used, findings here reported of no overall impact of TNFi would not in all likelihood constitute sufficient evidence to change clinical practice in favour of csDMARDS. Although a 40 % reduction in THR rates was observed amongst older TNFi users, this would most likely need to be validated in other contexts in order to overcome doubts regarding residual confounding (e.g. possible healthy user bias). Therefore, while further research is certainly required in the interim, an immediate clinical implication of this project might well be that the growing belief and conviction among clinicians that biologics reduce the need for joint replacement in RA may need to be tempered with more caution, especially in light of research elsewhere (108). On the whole, the findings lend support to the statement by Scott and colleagues in their Lancet seminar in relation to the improvement in various features and outcomes of RA over recent decades, *"better treatment seems the dominant factor. …better conventional treatment seems especially important"* (1).

Conclusions

This DPhil has set out to determine the relationship between TNFi use and subsequent need for joint replacement in RA. Estimates of the population-level impact of introducing TNFi within three different countries indicate that on the whole, rates of TKR but not THR decreased during the biologic era (compared to estimated counterfactuals). Relative to

trends of joint replacement among matched non-RA patients, a favourable impact on both TKR and THR among RA patients following the introduction of TNFi was implied. However, confounding exposures cannot be ruled out in these ecological investigations. In patient-level analyses of UK drug registry data, a 40% reduction in THR rates was observed among older PS matched TNFi users, offering some support for biologics playing a role in reducing need for joint replacement. On the other hand, main PS findings indicated no significant difference in joint replacement rates between TNFi and csDMARD users, suggesting other factors apart from TNFi are likely to be involved in the aforementioned downward population trends in joint replacement rates in RA. Overall, this DPhil suggests that a firm belief that biologics are superior to csDMARDs in reducing the need for joint replacement is currently unwarranted. It highlights the importance of further patient-level research from other countries and healthcare systems, investigating the impact of TNFi use on need for both large and small joint replacement in RA in order to confirm and/or further elucidate the relationship.

REFERENCES

1. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. Lancet. 2010;376(9746):1094-108.
2. Feldmann M, Brennan FM, Maini RN. Rheumatoid arthritis. Cell. 1996;85(3):307-10.
3. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, 3rd, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis. 2010;69(9):1580-8.
4. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis and rheumatism. 1988;31(3):315-24.
5. Dougados M, Soubrier M, Antunez A, Balint P, Balsa A, Buch MH, et al. Prevalence of comorbidities in rheumatoid arthritis and evaluation of their monitoring: results of an international, cross-sectional study (COMORA). Ann Rheum Dis. 2014;73(1):62-8.
6. Ollier WE, Harrison B, Symmons D. What is the natural history of rheumatoid arthritis? Best Pract Res Clin Rheumatol. 2001;15(1):27-48.
7. Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. Lancet. 2009;373(9664):659-72.
8. Verstappen SM. Outcomes of early rheumatoid arthritis--the WHO ICF framework. Best Pract Res Clin Rheumatol. 2013;27(4):555-70.
9. Matcham F, Scott IC, Rayner L, Hotopf M, Kingsley GH, Norton S, et al. The impact of rheumatoid arthritis on quality-of-life assessed using the SF-36: a systematic review and meta-analysis. Semin Arthritis Rheum. 2014;44(2):123-30.
10. Barrett EM, Scott DG, Wiles NJ, Symmons DP. The impact of rheumatoid arthritis on employment status in the early years of disease: a UK community-based study. Rheumatology (Oxford). 2000;39(12):1403-9.
11. Fex E, Larsson BM, Nived K, Eberhardt K. Effect of rheumatoid arthritis on work status and social and leisure time activities in patients followed 8 years from onset. J Rheumatol. 1998;25(1):44-50.
12. Cross M, Smith E, Hoy D, Carmona L, Wolfe F, Vos T, et al. The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis. 2014;73(7):1316-22.
13. Symmons D, Turner G, Webb R, Asten P, Barrett E, Lunt M, et al. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology (Oxford). 2002;41(7):793-800.
14. Humphreys JH, Verstappen SM, Hyrich KL, Chipping JR, Marshall T, Symmons DP. The incidence of rheumatoid arthritis in the UK: comparisons using the 2010

ACR/EULAR classification criteria and the 1987 ACR classification criteria. Results from the Norfolk Arthritis Register. Ann Rheum Dis. 2013;72(8):1315-20.
15. Choy E. Understanding the dynamics: pathways involved in the pathogenesis of rheumatoid arthritis. Rheumatology (Oxford). 2012;51 Suppl 5:v3-11.
16. Smolen JS, Aletaha D, Koeller M, Weisman MH, Emery P. New therapies for treatment of rheumatoid arthritis. Lancet. 2007;370(9602):1861-74.
17. Breedveld FC, Kalden JR. Appropriate and effective management of rheumatoid arthritis. Ann Rheum Dis. 2004;63(6):627-33.
18. Symmons DP, Chakravarty K. Can immunisation trigger rheumatoid arthritis? Ann Rheum Dis. 1993;52(12):843-4.
19. van der Woude D, Houwing-Duistermaat JJ, Toes RE, Huizinga TW, Thomson W, Worthington J, et al. Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. Arthritis and rheumatism. 2009;60(4):916-23.
20. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis and rheumatism. 2000;43(1):30-7.
21. Silman AJ, Pearson JE. Epidemiology and genetics of rheumatoid arthritis. Arthritis Res. 2002;4 Suppl 3:S265-72.
22. Carty SM, Snowden N, Silman AJ. Should infection still be considered as the most likely triggering factor for rheumatoid arthritis? (Reprinted from The Journal of Rheumatology, vol 30, pg 425-429, 2003). Annals of the Rheumatic Diseases. 2004;63:46-9.
23. Schett G, Stach C, Zwerina J, Voll R, Manger B. How antirheumatic drugs protect joints from damage in rheumatoid arthritis. Arthritis and rheumatism. 2008;58(10):2936-48.
24. Schett G, Gravallese E. Bone erosion in rheumatoid arthritis: mechanisms, diagnosis and treatment. Nat Rev Rheumatol. 2012;8(11):656-64.
25. Scott DL, Pugner K, Kaarela K, Doyle DV, Woolf A, Holmes J, et al. The links between joint damage and disability in rheumatoid arthritis. Rheumatology (Oxford). 2000;39(2):122-32.
26. Tamas MM, Filippucci E, Becciolini A, Gutierrez M, Di Geso L, Bonfiglioli K, et al. Bone erosions in rheumatoid arthritis: ultrasound findings in the early stage of the disease. Rheumatology (Oxford). 2014;53(6):1100-7.
27. Boini S, Guillemin F. Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. Ann Rheum Dis. 2001;60(9):817-27.
28. Ory PA. Interpreting radiographic data in rheumatoid arthritis. Ann Rheum Dis. 2003;62(7):597-604.
29. Wolfe F, Sharp JT. Radiographic outcome of recent-onset rheumatoid arthritis: a 19-year study of radiographic progression. Arthritis and rheumatism. 1998;41(9):1571-82.
30. Scott DL, Steer S. The course of established rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2007;21(5):943-67.

31. Plant MJ, Jones PW, Saklatvala J, Ollier WE, Dawes PT. Patterns of radiological progression in early rheumatoid arthritis: results of an 8 year prospective study. J Rheumatol. 1998;25(3):417-26.
32. Lindqvist E, Jonsson K, Saxne T, Eberhardt K. Course of radiographic damage over 10 years in a cohort with early rheumatoid arthritis. Ann Rheum Dis. 2003;62(7):611-6.
33. Gademan MG, Hofstede SN, Vliet Vlieland TP, Nelissen RG, Marang-van de Mheen PJ. Indication criteria for total hip or knee arthroplasty in osteoarthritis: a state-of-the-science overview. BMC musculoskeletal disorders. 2016;17(1):463.
34. Nikiphorou E, Norton S, Young A, Carpenter L, Dixey J, Walsh DA, et al. Association between rheumatoid arthritis disease activity, progression of functional limitation and long-term risk of orthopaedic surgery: combined analysis of two prospective cohorts supports EULAR treat to target DAS thresholds. Ann Rheum Dis. 2016;75(12):2080-6.
35. Wolfe F, Zwillich SH. The long-term outcomes of rheumatoid arthritis: a 23-year prospective, longitudinal study of total joint replacement and its predictors in 1,600 patients with rheumatoid arthritis. Arthritis and rheumatism. 1998;41(6):1072-82.
36. Hulsmans HM, Jacobs JW, van der Heijde DM, van Albada-Kuipers GA, Schenk Y, Bijlsma JW. The course of radiologic damage during the first six years of rheumatoid arthritis. Arthritis and rheumatism. 2000;43(9):1927-40.
37. Nelissen RG. The impact of total joint replacement in rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2003;17(5):831-46.
38. Young BL, Watson SL, Perez JL, McGwin G, Singh JA, Ponce BA. Trends in Joint Replacement Surgery in Patients with Rheumatoid Arthritis. J Rheumatol. 2018;45(2):158-64.
39. National Joint Registry for England and Wales: 10th Annual Report. 2013

.
40. National Joint Registry for England and Wales: 14th Annual Report. 2017.
41. Guidelines for the management of rheumatoid arthritis. American College of Rheumatology Ad Hoc Committee on Clinical Guidelines. Arthritis and rheumatism. 1996;39(5):713-22.
42.
American_College_of_Rheumatology_Subcommittee_on_Rheumatoid_Arthritis_Guidelines. Guidelines for the Management of Rheumatoid Arthritis: 2002 Update. Arthritis & Rheumatism. 2002;46(2):328 - 46.
43. Scott D. Guidelines for arthritis: ten years on. Clin Med (Lond). 2001;1(5):389-91.
44. Luqmani R, Hennell S, Estrach C, Birrell F, Bosworth A, Davenport G, et al. British Society for Rheumatology and british health professionals in Rheumatology guideline for the management of rheumatoid arthritis (the first two years). Rheumatology (Oxford). 2006;45(9):1167-9.
45. Luqmani R, Hennell S, Estrach C, Basher D, Birrell F, Bosworth A, et al. British Society for Rheumatology and British Health Professionals in Rheumatology guideline

for the management of rheumatoid arthritis (after the first 2 years). Rheumatology (Oxford). 2009;48(4):436-9.
46. NICE. Rheumatoid arthritis in adults: management (CG79). 2009.
47. Short CLM, B. Rheumatoid Arthritis: Historical Aspects. J Chron Dis. 1959;10(5):368-87.
48. Upchurch KS, Kay J. Evolution of treatment for rheumatoid arthritis. Rheumatology (Oxford). 2012;51 Suppl 6:vi28-36.
49. Wilske KR. Inverting the therapeutic pyramid: observations and recommendations on new directions in rheumatoid arthritis therapy based on the author's experience. Semin Arthritis Rheum. 1993;23(2 Suppl 1):11-8.
50. Scott DL. Biologics-based therapy for the treatment of rheumatoid arthritis. Clin Pharmacol Ther. 2012;91(1):30-43.
51. Caldwell JR, Furst DE. The efficacy and safety of low-dose corticosteroids for rheumatoid arthritis. Semin Arthritis Rheum. 1991;21(1):1-11.
52. Mazzantini M, Talarico R, Doveri M, Consensi A, Cazzato M, Bazzichi L, et al. Incident comorbidity among patients with rheumatoid arthritis treated or not with low-dose glucocorticoids: a retrospective study. J Rheumatol. 2010;37(11):2232-6.
53. Donahue KE, Gartlehner G, Jonas DE, Lux LJ, Thieda P, Jonas BL, et al. Systematic review: comparative effectiveness and harms of disease-modifying medications for rheumatoid arthritis. Ann Intern Med. 2008;148(2):124-34.
54. Guo Q, Wang Y, Xu D, Nossent J, Pavlos NJ, Xu J. Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies. Bone Res. 2018;6:15.
55. Pincus T, Yazici Y, Sokka T, Aletaha D, Smolen JS. Methotrexate as the "anchor drug" for the treatment of early rheumatoid arthritis. Clin Exp Rheumatol. 2003;21(5 Suppl 31):S179-85.
56. Weinblatt ME, Coblyn JS, Fox DA, Fraser PA, Holdsworth DE, Glass DN, et al. Efficacy of low-dose methotrexate in rheumatoid arthritis. The New England journal of medicine. 1985;312(13):818-22.
57. Singh JA, Saag KG, Bridges SL, Jr., Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Rheumatol. 2016;68(1):1-26.
58. Smolen JS, Landewe R, Bijlsma J, Burmester G, Chatzidionysiou K, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. Ann Rheum Dis. 2017;76(6):960-77.
59. Hannonen P, Mottonen T, Hakola M, Oka M. Sulfasalazine in early rheumatoid arthritis. A 48-week double-blind, prospective, placebo-controlled study. Arthritis and rheumatism. 1993;36(11):1501-9.
60. Plosker GL, Croom KF. Sulfasalazine: a review of its use in the management of rheumatoid arthritis. Drugs. 2005;65(13):1825-49.
61. Strand V, Cohen S, Schiff M, Weaver A, Fleischmann R, Cannon G, et al. Treatment of active rheumatoid arthritis with leflunomide compared with placebo and methotrexate. Leflunomide Rheumatoid Arthritis Investigators Group. Arch Intern Med. 1999;159(21):2542-50.

62. Katz SJ, Russell AS. Re-evaluation of antimalarials in treating rheumatic diseases: re-appreciation and insights into new mechanisms of action. Curr Opin Rheumatol. 2011;23(3):278-81.
63. Monaco C, Nanchahal J, Taylor P, Feldmann M. Anti-TNF therapy: past, present and future. Int Immunol. 2015;27(1):55-62.
64. Emery P. Treatment of rheumatoid arthritis. Bmj. 2006;332(7534):152-5.
65. Elliott MJ, Maini RN, Feldmann M, Long-Fox A, Charles P, Katsikis P, et al. Treatment of rheumatoid arthritis with chimeric monoclonal antibodies to tumor necrosis factor alpha. Arthritis and rheumatism. 1993;36(12):1681-90.
66. Elliott MJ, Maini RN, Feldmann M, Long-Fox A, Charles P, Bijl H, et al. Repeated therapy with monoclonal antibody to tumour necrosis factor alpha (cA2) in patients with rheumatoid arthritis. Lancet. 1994;344(8930):1125-7.
67. Elliott MJ, Maini RN, Feldmann M, Kalden JR, Antoni C, Smolen JS, et al. Randomised double-blind comparison of chimeric monoclonal antibody to tumour necrosis factor alpha (cA2) versus placebo in rheumatoid arthritis. Lancet. 1994;344(8930):1105-10.
68. Lipsky PE, van der Heijde DM, St Clair EW, Furst DE, Breedveld FC, Kalden JR, et al. Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. The New England journal of medicine. 2000;343(22):1594-602.
69. GUIDELINES FOR PRESCRIBING TNF-α BLOCKERS IN ADULTS WITH RHEUMATOID

ARTHRITIS (British Society for Rheumatology), (April 2001).
70. Dillon A. Technological Appraisal Guidance - No. 36: Guidance on the use of etanercept and infliximab for the treatment of rheumatoid arthritis. National Institute for Clinical Excellence. March 2002.
71. Watson K, Symmons D, Griffiths I, Silman A. The British Society for Rheumatology biologics register. Ann Rheum Dis. 2005;64 Suppl 4:iv42-3.
72. Kiely PD, Deighton C, Dixey J, Ostor AJ, British Society for Rheumatology Standards G, Audit Working G. Biologic agents for rheumatoid arthritis--negotiating the NICE technology appraisals. Rheumatology (Oxford). 2012;51(1):24-31.
73. Ledingham J, Gullick N, Irving K, Gorodkin R, Aris M, Burke J, et al. BSR and BHPR guideline for the prescription and monitoring of non-biologic disease-modifying anti-rheumatic drugs. Rheumatology (Oxford). 2017;56(12):2257.
74. Deighton C, Hyrich K, Ding T, Ledingham J, Lunt M, Luqmani R, et al. BSR and BHPR rheumatoid arthritis guidelines on eligibility criteria for the first biological therapy. Rheumatology (Oxford). 2010;49(6):1197-9.
75. Finzel S, Rech J, Schmidt S, Engelke K, Englbrecht M, Stach C, et al. Repair of bone erosions in rheumatoid arthritis treated with tumour necrosis factor inhibitors is based on bone apposition at the base of the erosion. Ann Rheum Dis. 2011;70(9):1587-93.
76. Moller Dohn U, Boonen A, Hetland ML, Hansen MS, Knudsen LS, Hansen A, et al. Erosive progression is minimal, but erosion healing rare, in patients with rheumatoid arthritis treated with adalimumab. A 1 year investigator-initiated follow-

up study using high-resolution computed tomography as the primary outcome measure. Ann Rheum Dis. 2009;68(10):1585-90.
77. Lukas C, van der Heijde D, Fatenajad S, Landewe R. Repair of erosions occurs almost exclusively in damaged joints without swelling. Ann Rheum Dis. 2010;69(5):851-5.
78. Smolen JS, Aletaha D, Bijlsma JW, Breedveld FC, Boumpas D, Burmester G, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis. 2010;69(4):631-7.
79. Nam JL, Ramiro S, Gaujoux-Viala C, Takase K, Leon-Garcia M, Emery P, et al. Efficacy of biological disease-modifying antirheumatic drugs: a systematic literature review informing the 2013 update of the EULAR recommendations for the management of rheumatoid arthritis. Ann Rheum Dis. 2014;73(3):516-28.
80. Orme ME, Macgilchrist KS, Mitchell S, Spurden D, Bird A. Systematic review and network meta-analysis of combination and monotherapy treatments in disease-modifying antirheumatic drug-experienced patients with rheumatoid arthritis: analysis of American College of Rheumatology criteria scores 20, 50, and 70. Biologics. 2012;6:429-64.
81. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis and rheumatism. 1995;38(6):727-35.
82. Felson DT, LaValley MP. The ACR20 and defining a threshold for response in rheumatic diseases: too much of a good thing. Arthritis Res Ther. 2014;16(1):101.
83. Lloyd S, Bujkiewicz S, Wailoo AJ, Sutton AJ, Scott D. The effectiveness of anti-TNF-alpha therapies when used sequentially in rheumatoid arthritis patients: a systematic review and meta-analysis. Rheumatology (Oxford). 2010;49(12):2313-21.
84. Hazlewood GS, Barnabe C, Tomlinson G, Marshall D, Devoe D, Bombardier C. Methotrexate monotherapy and methotrexate combination therapy with traditional and biologic disease modifying antirheumatic drugs for rheumatoid arthritis: abridged Cochrane systematic review and network meta-analysis. Bmj. 2016;353:i1777.
85. Graudal N, Jurgens G. Similar effects of disease-modifying antirheumatic drugs, glucocorticoids, and biologic agents on radiographic progression in rheumatoid arthritis: meta-analysis of 70 randomized placebo-controlled or drug-controlled studies, including 112 comparisons. Arthritis and rheumatism. 2010;62(10):2852-63.
86. Graudal N, Hubeck-Graudal T, Faurschou M, Baslund B, Jurgens G. Combination Therapy With and Without Tumor Necrosis Factor Inhibitors in Rheumatoid Arthritis: A Meta-Analysis of Randomized Trials. Arthritis care & research. 2015;67(11):1487-95.
87. Murray E, Butylkova Y, Ellis A, Skup M, Kalabic J, Garg V. A Systematic Review and Meta-Analysis of Comparative Efficacy of Biologics in Treating Patients with Rheumatoid Arthritis: Assessment of Long-Term Radiographic Progression from Published Clinical Trials. Arthritis Rheumatol. 2016;68.
88. Kuriya B, Arkema EV, Bykerk VP, Keystone EC. Efficacy of initial methotrexate monotherapy versus combination therapy with a biological agent in early rheumatoid

arthritis: a meta-analysis of clinical and radiographic remission. Ann Rheum Dis. 2010;69(7):1298-304.
89. Thorlund K, Druyts E, Avina-Zubieta JA, Wu P, Mills EJ. Why the findings of published multiple treatment comparison meta-analyses of biologic treatments for rheumatoid arthritis are different: an overview of recurrent methodological shortcomings. Ann Rheum Dis. 2013;72(9):1524-35.
90. Estellat C, Ravaud P. Lack of head-to-head trials and fair control arms: randomized controlled trials of biologic treatment for rheumatoid arthritis. Arch Intern Med. 2012;172(3):237-44.
91. Jamsen E, Virta LJ, Hakala M, Kauppi MJ, Malmivaara A, Lehto MU. The decline in joint replacement surgery in rheumatoid arthritis is associated with a concomitant increase in the intensity of anti-rheumatic therapy: a nationwide register-based study from 1995 through 2010. Acta Orthop. 2013;84(4):331-7.
92. Mertelsmann-Voss C, Lyman S, Pan TJ, Goodman SM, Figgie MP, Mandl LA. US trends in rates of arthroplasty for inflammatory arthritis including rheumatoid arthritis, juvenile idiopathic arthritis, and spondyloarthritis. Arthritis Rheumatol. 2014;66(6):1432-9.
93. Leon L, Abasolo L, Carmona L, Rodriguez-Rodriguez L, Lamas JR, Hernandez-Garcia C, et al. Orthopedic surgery in rheumatoid arthritis in the era of biologic therapy. J Rheumatol. 2013;40(11):1850-5.
94. Harty L, O'Toole G, FitzGerald O. Profound reduction in hospital admissions and musculoskeletal surgical procedures for rheumatoid arthritis with concurrent changes in clinical practice (1995-2010). Rheumatology (Oxford). 2015;54(4):666-71.
95. National Joint Registry for England and Wales: 1st Annual Report, (2004).
96. National Joint Registry for England and Wales: 4th Annual Report, (2007).
97. Nikiphorou E, Carpenter L, Morris S, MacGregor AJ, Dixey J, Kiely P, et al. Hand and Foot Surgery Rates in Rheumatoid Arthritis Have Declined From 1986 to 2011, but Large-Joint Replacement Rates Remain Unchanged Results From Two UK Inception Cohorts. Arthritis Rheumatol. 2014;66(5):1081-9.
98. Shourt CA, Crowson CS, Gabriel SE, Matteson EL. Orthopedic surgery among patients with rheumatoid arthritis 1980-2007: a population-based study focused on surgery rates, sex, and mortality. J Rheumatol. 2012;39(3):481-5.
99. Momohara S, Inoue E, Ikari K, Ochi K, Ishida O, Yano K, et al. Recent trends in orthopedic surgery aiming to improve quality of life for those with rheumatoid arthritis: data from a large observational cohort. J Rheumatol. 2014;41(5):862-6.
100. Louie GH, Ward MM. Changes in the rates of joint surgery among patients with rheumatoid arthritis in California, 1983-2007. Ann Rheum Dis. 2010;69(5):868-71.
101. Rodriguez-Rodriguez L, Ivorra-Cortes J, Leon L, Fernandez-Gutierrez B, Abasolo L, Gonzalez-Alvaro I, editors. Temporal Trends of Total Joint Replacement in Rheumatoid Arthritis Patients: A Survival Study2016: Arthritis Rheumatol.
102. Hekmat K, Jacobsson L, Nilsson JA, Petersson IF, Robertsson O, Garellick G, et al. Decrease in the incidence of total hip arthroplasties in patients with rheumatoid arthritis--results from a well defined population in south Sweden. Arthritis Res Ther. 2011;13(2):R67.

103. Yasui T, Nishino J, Kadono Y, Matsui T, Nakamura K, Tanaka S, et al. Impact of biologics on the prevalence of orthopedic surgery in the National Database of Rheumatic Diseases in Japan. Mod Rheumatol. 2010;20(3):233-7.
104. Koenig AS, Mardekian J, Kotak S. Impact of Biologics On Total Knee Replacement and Total Hip Replacement Rates in Rheumatoid Arthritis Patients: Results From US Marketscan Database. Arthritis and rheumatism. 2012;64(10):S786-S.
105. Ollendorf DA, Peterson AN, Doyle J, Huse DM. Impact of leflunomide versus biologic agents on the costs of care for rheumatoid arthritis in a managed care population. Am J Manag Care. 2002;8(7 Suppl):S203-13.
106. Moura CS, Abrahamowicz M, Beauchamp ME, Lacaille D, Wang YS, Boire G, et al. Early medication use in new-onset rheumatoid arthritis may delay joint replacement: results of a large population-based study. Arthritis Research & Therapy. 2015;17.
107. Widdifield J, Moura CS, Wang Y, Abrahamowicz M, Paterson JM, Huang A, et al. The Longterm Effect of Early Intensive Treatment of Seniors with Rheumatoid Arthritis: A Comparison of 2 Population-based Cohort Studies on Time to Joint Replacement Surgery. J Rheumatol. 2016;43(5):861-8.
108. Aaltonen KJ, Virkki LM, Jamsen E, Sokka T, Konttinen YT, Peltomaa R, et al. Do biologic drugs affect the need for and outcome of joint replacements in patients with rheumatoid arthritis? A register-based study. Semin Arthritis Rheum. 2013;43(1):55-62.
109. Joseph KS, Mehrabadi A, Lisonkova S. Confounding by Indication and Related Concepts. Current Epidemiology Reports. 2014;1(1):1-8.
110. Miettinen OS. The need for randomization in the study of intended effects. Statistics in medicine. 1983;2(2):267-71.
111. Barton S. Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. Bmj. 2000;321(7256):255-6.
112. Barnighausen T, Tugwell P, Rottingen JA, Shemilt I, Rockers P, Geldsetzer P, et al. Quasi-experimental study designs series-paper 4: uses and value. J Clin Epidemiol. 2017;89:21-9.
113. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.
114. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. Br J Cancer. 2014;110(3):551-5.
115. Reyes C, Pottegard A, Schwarz P, Javaid MK, Van Staa TP, Cooper C, et al. Real-Life and RCT Participants: Alendronate Users Versus FITs' Trial Eligibility Criterion. Calcified tissue international. 2016;99(3):243-9.
116. Geldsetzer P, Fawzi W. Quasi-experimental study designs series-paper 2: complementary approaches to advancing global health knowledge. J Clin Epidemiol. 2017;89:12-6.
117. Lao KS, Chui CS, Man KK, Lau WC, Chan EW, Wong IC. Medication safety research by observational study design. Int J Clin Pharm. 2016;38(3):676-84.

118. Osborne J. Best Practices in Quantitative Methods2008.
119. Shadish WR, Cook TD, Campell DT. Experimental and Quasi-Experimental Designs for Generalised Causal Inference. Belmont, USA: Wadsworth Cangage Learning; 2002.
120. King G, Keohane RO, Verba S. The Importance of Research Design in Political-Science. Am Polit Sci Rev. 1995;89(2):475-81.
121. Gunasekara FI, Carter K, Blakely T. Glossary for econometrics and epidemiology. J Epidemiol Community Health. 2008;62(10):858-61.
122. Barnighausen T, Oldenburg C, Tugwell P, Bommer C, Ebert C, Barreto M, et al. Quasi-experimental study designs series-paper 7: assessing the assumptions. J Clin Epidemiol. 2017;89:53-66.
123. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). International journal of epidemiology. 2015.
124. Office_for_National_Statistics. Key Health Statistics from General Practice 1998. London; 2000.
125. Thomas SL, Edwards CJ, Smeeth L, Cooper C, Hall AJ. How accurate are diagnoses for rheumatoid arthritis and juvenile idiopathic arthritis in the general practice research database? Arthritis and rheumatism. 2008;59(9):1314-21.
126. Culliford D, Maskell J, Judge A, Cooper C, Prieto-Alhambra D, Arden NK, et al. Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink. Osteoarthritis Cartilage. 2015;23(4):594-600.
127. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scandinavian journal of public health. 2011;39(7 Suppl):30-3.
128. Schmidt M, Pedersen L, Sorensen HT. The Danish Civil Registration System as a tool in epidemiology. European journal of epidemiology. 2014;29(8):541-9.
129. Widdifield J, Paterson JM, Bernatsky S, Tu K, Tomlinson G, Kuriya B, et al. The epidemiology of rheumatoid arthritis in Ontario, Canada. Arthritis Rheumatol. 2014;66(4):786-93.
130. Widdifield J, Bombardier C, Bernatsky S, Paterson JM, Green D, Young J, et al. An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. BMC musculoskeletal disorders. 2014;15:216.
131. Widdifield J, Bernatsky S, Paterson JM, Tu K, Ng R, Thorne JC, et al. Accuracy of Canadian health administrative databases in identifying patients with rheumatoid arthritis: a validation study using the medical records of rheumatologists. Arthritis care & research. 2013;65(10):1582-91.
132. Tatangelo M, Paterson M, Tomlinson GA, Bansback N, Widdifield J, Gomes T, et al. Temporal Trends in Drug Prescription, Utilization and Costs Among Rheumatoid Arthritis (RA) Patients Show Wide Regional Variation Despite Universal Drug Coverage. Arthritis Rheumatol. 2015;67.
133. MedDRA. Welcome to MedDRA https://www.meddra.org [
134. David G, Tandon N, Waters H, Gunnarsson C, Kavanaugh A. Rheumatoid Arthritis and Joint Replacement: Impact of Biologics. Americal Journal of Pharmaceutical Benefits. 2014;6(6):256-64.

135. Nystad TW, Fenstad AM, Furnes O, Havelin LI, Skredderstuen AK, Fevang BT. Reduction in orthopaedic surgery in patients with rheumatoid arthritis: a Norwegian register-based study. Scand J Rheumatol. 2015:1-7.
136. James D, Young A, Kulinskaya E, Knight E, Thompson W, Ollier W, et al. Orthopaedic intervention in early rheumatoid arthritis. Occurrence and predictive factors in an inception cohort of 1064 patients followed for 5 years. Rheumatology. 2004;43(3):369-76.
137. Gwinnutt JM, Symmons DPM, MacGregor AJ, Chipping JR, Lapraik C, Marshall T, et al. Predictors of and outcomes following orthopaedic joint surgery in patients with early rheumatoid arthritis followed for 20 years. Rheumatology (Oxford). 2017;56(9):1510-7.
138. Burn E, Edwards CJ, Murray DW, Silman A, Cooper C, Arden NK, et al. Trends and determinants of length of stay and hospital reimbursement following knee and hip replacement: evidence from linked primary care and NHS hospital records from 1997 to 2014. BMJ Open. 2018;8(1):e019146.
139. Prieto-Alhambra D, Javaid MK, Judge A, Maskell J, Cooper C, Arden NK, et al. Hormone replacement therapy and mid-term implant survival following knee or hip arthroplasty for osteoarthritis: a population-based cohort study. Ann Rheum Dis. 2015;74(3):557-63.
140. Hennekens CH, Buring SL. EPIDEMIOLOGY IN MEDICINE: Lippincott Williams & Wilkins; 1987. 383 p.
141. Supporting Data Quality. National Joint Registry for England and Wales; 2014/16.
142. National Joint Registry for England and Wales: 11th Annual Report. 2014

.
143. National Joint Registry for England and Wales: 9th Annual Report. 2012

.
144. Berry SD, Ngo L, Samelson EJ, Kiel DP. Competing risk of death: an important consideration in studies of older adults. Journal of the American Geriatrics Society. 2010;58(4):783-7.
145. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. International journal of epidemiology. 2012;41(3):861-70.
146. Ford E, Carroll J, Smith H, Davies K, Koeling R, Petersen I, et al. What evidence is there for a delay in diagnostic coding of RA in UK general practice records? An observational study of free text. BMJ Open. 2016;6(6):e010393.
147. Culliford DJ, Maskell J, Kiran A, Judge A, Javaid MK, Cooper C, et al. The lifetime risk of total hip and knee arthroplasty: results from the UK general practice research database. Osteoarthritis Cartilage. 2012;20(6):519-24.
148. Litwic A, Edwards MH, Dennison EM, Cooper C. Epidemiology and burden of osteoarthritis. Br Med Bull. 2013;105:185-99.
149. Richter M, Crowson CS, Matteson EL, Makol A. Orthopedic Surgery among Patients with Rheumatoid Arthritis: A Population-based study to Identify Risk factors, Sex differences, and Time trends. Arthritis care & research. 2017.

150. Kuperman EF, Schweizer M, Joy P, Gu X, Fang MM. The effects of advanced age on primary total knee arthroplasty: a meta-analysis and systematic review. BMC Geriatr. 2016;16:41.
151. de Vries LM, Sturkenboom MC, Verhaar JA, Kingma JH, Stricker BH. Complications after hip arthroplasty and the association with hospital procedure volume. Acta Orthop. 2011;82(5):545-52.
152. Hunt LP, Ben-Shlomo Y, Clark EM, Dieppe P, Judge A, MacGregor AJ, et al. 45-day mortality after 467,779 knee replacements for osteoarthritis from the National Joint Registry for England and Wales: an observational study. Lancet. 2014;384(9952):1429-36.
153. Hunt LP, Ben-Shlomo Y, Clark EM, Dieppe P, Judge A, MacGregor AJ, et al. 90-day mortality after 409,096 total hip replacements for osteoarthritis, from the National Joint Registry for England and Wales: a retrospective analysis. Lancet. 2013;382(9898):1097-104.
154. Judge A, Arden NK, Cooper C, Kassim Javaid M, Carr AJ, Field RE, et al. Predictors of outcomes of total knee replacement surgery. Rheumatology (Oxford). 2012;51(10):1804-13.
155. Hofstede SN, Gademan MG, Vliet Vlieland TP, Nelissen RG, Marang-van de Mheen PJ. Preoperative predictors for outcomes after total hip replacement in patients with osteoarthritis: a systematic review. BMC musculoskeletal disorders. 2016;17:212.
156. Gordon M, Greene M, Frumento P, Rolfson O, Garellick G, Stark A. Age- and health-related quality of life after total hip replacement: decreasing gains in patients above 70 years of age. Acta Orthop. 2014;85(3):244-9.
157. Leyland KM, Judge A, Javaid MK, Diez-Perez A, Carr A, Cooper C, et al. Obesity and the Relative Risk of Knee Replacement Surgery in Patients With Knee Osteoarthritis: A Prospective Cohort Study. Arthritis Rheumatol. 2016;68(4):817-25.
158. Reyes C, Leyland KM, Peat G, Cooper C, Arden NK, Prieto-Alhambra D. Association Between Overweight and Obesity and Risk of Clinically Diagnosed Knee, Hip, and Hand Osteoarthritis: A Population-Based Cohort Study. Arthritis Rheumatol. 2016;68(8):1869-75.
159. Clement ND, Jenkins PJ, Brenkel IJ, Walmsley P. Predictors of mortality after total knee replacement: a ten-year survivorship analysis. The Journal of bone and joint surgery British volume. 2012;94(2):200-4.
160. Judge A, Arden NK, Batra RN, Thomas G, Beard D, Javaid MK, et al. The association of patient characteristics and surgical variables on symptoms of pain and function over 5 years following primary hip-replacement surgery: a prospective cohort study. BMJ Open. 2013;3(3).
161. Davis AM, Wood AM, Keenan AC, Brenkel IJ, Ballantyne JA. Does body mass index affect clinical outcome post-operatively and at five years after primary unilateral total hip replacement performed for osteoarthritis? A multivariate analysis of prospective data. The Journal of bone and joint surgery British volume. 2011;93(9):1178-82.

162. Berstock JR, Beswick AD, Lenguerrand E, Whitehouse MR, Blom AW. Mortality after total hip replacement surgery: A systematic review. Bone Joint Res. 2014;3(6):175-82.
163. Singh JA. Smoking and outcomes after knee and hip arthroplasty: a systematic review. J Rheumatol. 2011;38(9):1824-34.
164. Dixon T, Shaw M, Ebrahim S, Dieppe P. Trends in hip and knee joint replacement: socioeconomic inequalities and projections of need. Ann Rheum Dis. 2004;63(7):825-30.
165. Judge A, Welton NJ, Sandhu J, Ben-Shlomo Y. Equity in access to total joint replacement of the hip and knee in England: cross sectional study. Bmj. 2010;341:c4092.
166. Chaturvedi N, Ben-Shlomo Y. From the surgery to the surgeon: does deprivation influence consultation and operation rates? Br J Gen Pract. 1995;45(392):127-31.
167. Hawker GA, Wright JG, Glazier RH, Coyte PC, Harvey B, Williams JI, et al. The effect of education and income on need and willingness to undergo total joint arthroplasty. Arthritis and rheumatism. 2002;46(12):3331-9.
168. Juni P, Dieppe P, Donovan J, Peters T, Eachus J, Pearson N, et al. Population requirement for primary knee replacement surgery: a cross-sectional study. Rheumatology (Oxford). 2003;42(4):516-21.
169. Frankel S, Eachus J, Pearson N, Greenwood R, Chan P, Peters TJ, et al. Population requirement for primary hip-replacement surgery: a cross-sectional study. Lancet. 1999;353(9161):1304-9.
170. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. Bmj. 2015;350:h2750.
171. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther. 2002;27(4):299-309.
172. Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc. 2006;13(1):16-23.
173. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. International journal of epidemiology. 2016.
174. Jandoc R, Burden AM, Mamdani M, Levesque LE, Cadarette SM. Interrupted time series analysis in drug utilization research is increasing: systematic review and recommendations. J Clin Epidemiol. 2015;68(8):950-6.
175. Laliotis I, Ioannidis JPA, Stavropoulou C. Total and cause-specific mortality before and after the onset of the Greek economic crisis: an interrupted time-series analysis. Lancet Public Health. 2016;1(2):e56-e65.
176. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health. 2012;66(12):1182-6.

177. NICE. NICE technological appraisal guidance: NICE; 2017 [01/02/2017]. Available from: https://www.nice.org.uk/About/What-we-do/Our-Programmes/NICE-guidance/NICE-technology-appraisal-guidance.
178. Critical values for Durbin-Watson test [Available from: https://web.stanford.edu/~clint/bench/dwcrit.htm.
179. Zhang F, Wagner AK, Soumerai SB, Ross-Degnan D. Methods for estimating confidence intervals in interrupted time series analyses of health interventions. J Clin Epidemiol. 2009;62(2):143-8.
180. Zhang FW, A.K.; Soumerai, S.B.; Ross-Degnan, D.;. Estimating confidence intervals around relative changes in outcome in segmented linear regression analysis of time series data. NESUG 15th Annual Conference; Buffalo: NY; 2002.
181. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Statistics in medicine. 2000;19(3):335-51.
182. Penfold RB, Zhang F. Use of Interrupted Time Series Analysis in Evaluating Health Care Quality Improvements. Acad Pediatr. 2013;13(6):S38-S44.
183. Maini RN, Breedveld FC, Kalden JR, Smolen JS, Furst D, Weisman MH, et al. Sustained improvement over two years in physical function, structural damage, and signs and symptoms among patients with rheumatoid arthritis treated with infliximab and methotrexate. Arthritis and rheumatism. 2004;50(4):1051-65.
184. Emery P, Breedveld FC, Hall S, Durez P, Chang DJ, Robertson D, et al. Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial. Lancet. 2008;372(9636):375-82.
185. Gwinnutt JM, Symmons DPM, MacGregor AJ, Chipping JR, Marshall T, Lunt M, et al. Have the 10-year outcomes of patients with early inflammatory arthritis improved in the new millennium compared with the decade before? Results from the Norfolk Arthritis Register. Ann Rheum Dis. 2018;77(6):848-54.
186. Lehtimaki MY, Kaarela K, Hamalainen MM. Incidence of hip involvement and need for total hip replacement in rheumatoid arthritis. An eight-year follow-up study. Scand J Rheumatol. 1986;15(4):387-91.
187. Duthie RB, Harris CM. A radiographic and clinical survey of the hip joint in sero-positive rheumatoid arthritis. Acta Orthop Scand. 1969;40(3):346-64.
188. Eberhardt K, Fex E, Johnsson K, Geborek P. Hip involvement in early rheumatoid arthritis. Ann Rheum Dis. 1995;54(1):45-8.
189. Kelly IG. Surgical treatment of the rheumatoid hip. Ann Rheum Dis. 1990;49 Suppl 2:858-62.
190. Smolen JS, Breedveld FC, Eberl G, Jones I, Leeming M, Wylie GL, et al. Validity and reliability of the twenty-eight-joint count for the assessment of rheumatoid arthritis activity. Arthritis and rheumatism. 1995;38(1):38-43.
191. Fransen JS, G; Piet, L. C. M. van Riel. Rheumatoid Arthritis Measures. Arthritis & Rheumatism. 2003;49(5S):214 - 24.
192. National Joint Registry for England and Wales: 8th Annual Report, (2011).
193. National Joint Registry for England and Wales: 12th Annual Report. 2015.

194. Judge A, Wallace G, Prieto-Alhambra D, Arden NK, Edwards CJ. Can the publication of guidelines change the management of early rheumatoid arthritis? An interrupted time series analysis from the United Kingdom. Rheumatology (Oxford). 2015;54(12):2244-8.
195. Pearce N. The ecological fallacy strikes back. J Epidemiol Community Health. 2000;54(5):326-7.
196. Luqmani R, Hennell S, Estrach C, Birrell F, Bosworth A, Davenport G, et al. British society for rheumatology and british health professionals in rheumatology guideline for the management of rheumatoid arthritis (the first two years). Rheumatology. 2006;45(9):1167-9.
197. Dayer MJ, Jones S, Prendergast B, Baddour LM, Lockhart PB, Thornhill MH. Incidence of infective endocarditis in England, 2000-13: a secular trend, interrupted time-series analysis. Lancet. 2015;385(9974):1219-28.
198. Hawley S, Leal J, Delmestri A, Prieto-Alhambra D, Arden NK, Cooper C, et al. Anti-Osteoporosis Medication Prescriptions and Incidence of Subsequent Fracture Among Primary Hip Fracture Patients in England and Wales: An Interrupted Time-Series Analysis. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2016;31(11):2008-15.
199. Hawley S, Cordtz R, Lene D, Edwards CJ, Arden NK, Delmestri A, et al. Association between NICE guidance on biologic therapies with rates of hip and knee replacement among rheumatoid arthritis patients in England and Wales: An interrupted time-series analysis. American College of Rheumatology; Washington DC: Arthritis Rheumatol.; 2016.
200. Somers M-AZ, P.; Jacob, R.; Bloom, H;. The Validity and Precision of the Comparative Interrupted Time Series Design and the Difference-in-Difference Design in Educational Evaluation. 2013.
201. Tang Y, Liu C, Zhang Z, Zhang X. Effects of prescription restrictive interventions on antibiotic procurement in primary care settings: a controlled interrupted time series study in China. Cost Eff Resour Alloc. 2018;16:1.
202. Wing C, Simon K, Bello-Gomez RA. Designing Difference in Difference Studies: Best Practices for Public Health Policy Research. Annu Rev Public Health. 2018;39:453-69.
203. Lopez Bernal J, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. International journal of epidemiology. 2018.
204. Linden A. Conducting interrupted time-series analysis for single- and multiple-group comparisons. The Stata Journal. 2015;15:480-500.
205. Law MR, Majumdar SR, Soumerai SB. Effect of illicit direct to consumer advertising on use of etanercept, mometasone, and tegaserod in Canada: controlled longitudinal study. Bmj. 2008;337:a1055.
206. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sorensen HT. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. BMC Med Res Methodol. 2011;11:83.

207. Institute_for_Rational_Pharmacotherpay. Treatment with TNF-inhibitors guidelines 2002 [Available from: http://www. irf. dk/ dk/ publikationer/ vejledninger/ tnf- alfa_ haemmende_ behandling_

02. htm.
208. Hetland ML. DANBIO--powerful research database and electronic patient record. Rheumatology (Oxford). 2011;50(1):69-77.
209. Drugs for the management of rheumatoid arthritis: clinical evaluation, no. 146 (2018).
210. Ministry_of_Health_and_Long-term_Care_Exceptional_Access_Program_(EAP). EAP Reimbursement Criteria for Frequently Requested Drugs. 2015.
211. Starodubtseva IA. [Prevalence of secondary osteoarthritis in patients with rheumatoid arthritis and risk factors for its progression]. Adv Gerontol. 2014;27(4):693-8.
212. Ibfelt EH, Sorensen J, Jensen DV, Dreyer L, Schiottz-Christensen B, Thygesen PH, et al. Validity and completeness of rheumatoid arthritis diagnoses in the nationwide DANBIO clinical register and the Danish National Patient Registry. Clin Epidemiol. 2017;9:627-32.
213. Linden A, Adams JL. Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation. J Eval Clin Pract. 2011;17(6):1231-8.
214. Austin PC. Statistical criteria for selecting the optimal number of untreated subjects matched to each treated subject when using many-to-one matching on the propensity score. American journal of epidemiology. 2010;172(9):1092-7.
215. Ewusie JE, Blondal E, Soobiah C, Beyene J, Thabane L, Straus SE, et al. Methods, applications, interpretations and challenges of interrupted time series (ITS) data: protocol for a scoping review. BMJ Open. 2017;7(6):e016018.
216. Hawton K, Bergen H, Simkin S, Dodd S, Pocock P, Bernal W, et al. Long term effect of reduced pack sizes of paracetamol on poisoning deaths and liver transplant activity in England and Wales: interrupted time series analyses. Bmj. 2013;346:f403.
217. Lenth RV. Some practical guidelines for effective sample size determination. Am Stat. 2001;55(3):187-93.
218. Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. Int J Technol Assess Health Care. 2003;19(4):613-23.
219. Goodman SN, Berlin JA. The Use of Predicted Confidence Intervals When Planning Experiments and the Misuse of Power When Interpreting Results. Ann Intern Med. 1994;122(3):200-6.
220. Van Calster B, Steyerberg EW, Collins GS, Smits T. Consequences of relying on statistical significance: Some illustrations. Eur J Clin Invest. 2018;48(5):e12912.
221. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14(5):365-76.

222. Zhang F, Wagner AK, Ross-Degnan D. Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. J Clin Epidemiol. 2011;64(11):1252-61.
223. Fretheim A, Zhang F, Ross-Degnan D, Oxman AD, Cheyne H, Foy R, et al. A reanalysis of cluster randomized trials showed interrupted time-series studies were valuable in health system evaluation. J Clin Epidemiol. 2015;68(3):324-33.
224. Cordtz RL, Hawley S, Prieto-Alhambra D, Hojgaard P, Zobbe K, Overgaard S, et al. Incidence of hip and knee replacement in patients with rheumatoid arthritis following the introduction of biological DMARDs: an interrupted time-series analysis using nationwide Danish healthcare registers. Ann Rheum Dis. 2018;77(5):684-9.
225. Chang M. Monte Carlo Simulation for the Pharmaceutical Industry: Concepts, Algorithms, and Case Studies: CRC Press; 2010.
226. Ripley BD. Stochastic Simulation. New York, United States: John Wiley and Sons Ltd; 2006.
227. Morris T, White I, Crowther M. Using simulation studies to evaluate statistical methods. Tutorial in Biostatistics. 2017.
228. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. Statistics in medicine. 2006;25(24):4279-92.
229. Hawton K, Bergen H, Simkin S, Brock A, Griffiths C, Romeri E, et al. Effect of withdrawal of co-proxamol on prescribing and deaths from drug poisoning in England and Wales: time series analysis. Bmj. 2009;338:b2270.
230. Feiveson AH. Power by simulation. The Stata Journal. 2002;2(2):107-24.
231. Sayers A, Crowther MJ, Judge A, Whitehouse MR, Blom AW. Determining the sample size required to establish whether a medical device is non-inferior to an external benchmark. BMJ Open. 2017;7(8):e015397.
232. McLeod AI, Vingilis ER. Power computations in time series analyses for traffic safety interventions. Accid Anal Prev. 2008;40(3):1244-8.
233. Rothman KJ, Greenland S. Planning Study Size Based on Precision Rather Than Power. Epidemiology. 2018;29(5):599-603.
234. Box GEPJ, G.M.; Reinsel, G.C.; Ljung, G.M. Time Series Analysis: Forecasting and Control. 5th ed: WILEY; 2015 29/06/2015.
235. Chatfield C. The Analysis of Time Series: an introduction. Chatfield CT, M.; Zidek, J., editor. Taylor & Francis e-Library: CHAPMAN & HALL/CRC; 2009.
236. McLeod AI, Vingilis ER. Power computations for intervention analysis. Technometrics. 2005;47(2):174-81.
237. Curtis JR, Singh JA. Use of biologics in rheumatoid arthritis: current and emerging paradigms of care. Clin Ther. 2011;33(6):679-707.
238. Dennison EM, Packham J, Hyrich K. The BSRBR-RA at 15 years. Rheumatology (Oxford). 2016;55(12):2093-5.
239. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006;98(3):253-9.
240. Jackson JW, Schmid I, Stuart EA. Propensity Scores in Pharmacoepidemiology: Beyond the Horizon. Curr Epidemiol Rep. 2017;4(4):271-80.

241. Austin PC. The performance of different propensity-score methods for estimating relative risks. J Clin Epidemiol. 2008;61(6):537-45.
242. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1-21.
243. Lunt M, Solomon D, Rothman K, Glynn R, Hyrich K, Symmons DP, et al. Different methods of balancing covariates leading to different effect estimates in the presence of effect modification. American journal of epidemiology. 2009;169(7):909-17.
244. Austin PC. Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research. Communications in Statistics - Simulation and Computation. 2009;38(6).
245. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in medicine. 2009;28(25):3083-107.
246. Nguyen TL, Collins GS, Spence J, Daures JP, Devereaux PJ, Landais P, et al. Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. BMC Med Res Methodol. 2017;17(1):78.
247. Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. Statistics in medicine. 2006;25(12):2084-106.
248. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017;9:157-66.
249. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Bmj. 2009;338:b2393.
250. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc. 1996;91(434):473-89.
251. Carpenter J, Kenward MG. Missing data in randomised controlled trials - a practical guide. 2007.
252. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods. 2002;7(2):147-77.
253. Vink G, Frank LE, Pannekoek J, van Buuren S. Predictive mean matching imputation of semicontinuous variables. Stat Neerl. 2014;68(1):61-90.
254. Eulenburg C, Suling A, Neuser P, Reuss A, Canzler U, Fehm T, et al. Propensity Scoring after Multiple Imputation in a Retrospective Study on Adjuvant Radiation Therapy in Lymph-Node Positive Vulvar Cancer. PLoS One. 2016;11(11):e0165705.
255. Mitra R, Reiter JP. A comparison of two methods of estimating propensity scores after multiple imputation. Statistical methods in medical research. 2016;25(1):188-204.
256. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York, Chichester, Brisbane, Toronto, Singapore: JOHN WILEY & SONS; 1987.
257. Kanters S, Druyts E, Mills EJ, Thorlund K. What drives the comparative effectiveness of biologics vs. methotrexate in rheumatoid arthritis? Meta-regression

and graphical inspection of suspected clinical factors. Rheumatology (Oxford). 2014;53(7):1264-73.
258. Daien CI, Morel J. Predictive factors of response to biological disease modifying antirheumatic drugs: towards personalized medicine. Mediators Inflamm. 2014;2014:386148.
259. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15(5):615-25.
260. Ertefaie A, Small DS, Flory JH, Hennessy S. A tutorial on the use of instrumental variables in pharmacoepidemiology. Pharmacoepidemiol Drug Saf. 2017;26(4):357-67.
261. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiology. 2006;17(3):268-75.
262. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Statistics in medicine. 2014;33(13):2297-340.
263. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. J Bus Econ Stat. 2001;19(1):2-16.
264. Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. American journal of epidemiology. 2009;169(3):273-84.
265. O'Keeffe AG, Geneletti S, Baio G, Sharples LD, Nazareth I, Petersen I. Regression discontinuity designs: an approach to the evaluation of treatment efficacy in primary care using observational data. Bmj. 2014;349:g5293.
266. Vandenbroucke JP, le Cessie S. Commentary: regression discontinuity design: let's give it a try to evaluate medical and public health interventions. Epidemiology. 2014;25(5):738-41.
267. Bor J, Moscoe E, Mutevedzi P, Newell ML, Barnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology. 2014;25(5):729-37.
268. Moscoe E, Bor J, Barnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. J Clin Epidemiol. 2015;68(2):122-33.
269. National Joint Registry for England and Wales: 5th Annual Report, (2008).
270. Singh JY, B; Watson, S; Perez, J; McGwin, G; Ponce, B. Trends in Joint Replacements Surgery in Patients with Rheumatoid Arthritis. Arthritis Rheumatol. 2016;68 (suppl 10).
271. Hawley S, Cordtz R, Lene D, Edwards CJ, Arden NK, Delmestri A, et al. Association between NICE guidance on biologic therapies with rates of hip and knee replacement among rheumatoid arthritis patients in England and Wales: An interrupted time-series analysis. Semin Arthritis Rheu. 2018;47(5):605-10.
272. Moreland LW, O'Dell JR, Paulus HE, Curtis JR, Bathon JM, St Clair EW, et al. A randomized comparative effectiveness study of oral triple therapy versus etanercept plus methotrexate in early aggressive rheumatoid arthritis: the treatment of Early

Aggressive Rheumatoid Arthritis Trial. Arthritis and rheumatism. 2012;64(9):2824-35.
273. O'Dell JR, Mikuls TR, Taylor TH, Ahluwalia V, Brophy M, Warren SR, et al. Therapies for active rheumatoid arthritis after methotrexate failure. The New England journal of medicine. 2013;369(4):307-18.
274. Austin PC. The performance of different propensity-score methods for estimating relative risks. J Clin Epidemiol. 2008;61(6):537-45.
275. Pirracchio R, Carone M, Rigon MR, Caruana E, Mebazaa A, Chevret S. Propensity score estimators for the average treatment effect and the average treatment effect on the treated may yield very different estimates. Statistical methods in medical research. 2016;25(5):1938-54.
276. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5):550-60.
277. Williamson T, Ravani P. Marginal structural models in clinical research: when and how to use them? Nephrol Dial Transpl. 2017;32:84-90.
278. Lousdal ML. An introduction to instrumental variable assumptions, validation and estimation. Emerg Themes Epidemiol. 2018;15:1.
279. Greenland S. An introduction To instrumental variables for epidemiologists. International journal of epidemiology. 2000;29(6):1102.
280. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol. 2011;64(12):1311-6.
281. Barton P. Development of the Birmingham Rheumatoid Arthritis Model: past, present and future plans. Rheumatology (Oxford). 2011;50 Suppl 4:iv32-iv8.
282. Rockers PC, Rottingen JA, Shemilt I, Tugwell P, Barnighausen T. Inclusion of quasi-experimental studies in systematic reviews of health systems research. Health Policy. 2015;119(4):511-21.
283. Tugwell P, Knottnerus JA, McGowan J, Tricco A. Big-5 Quasi-Experimental designs. J Clin Epidemiol. 2017;89:1-3.
284. March LM, Barcenilla AL, Cross MJ, Lapsley HM, Parker D, Brooks PM. Costs and outcomes of total hip and knee joint replacement for rheumatoid arthritis. Clin Rheumatol. 2008;27(10):1235-42.
285. Ravi B, Escott B, Shah PS, Jenkinson R, Chahal J, Bogoch E, et al. A systematic review and meta-analysis comparing complications following total joint arthroplasty for rheumatoid arthritis versus for osteoarthritis. Arthritis and rheumatism. 2012;64(12):3839-49.
286. Burn E, Edwards CJ, Murray DW, Silman A, Cooper C, Arden NK, et al. The impact of rheumatoid arthritis on the risk of adverse events following joint replacement: a real-world cohort study. Clin Epidemiol. 2018;10:697-704.
287. Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, van Zeben D, Kerstens PJ, Hazes JM, et al. Comparison of treatment strategies in early rheumatoid arthritis: a randomized trial. Ann Intern Med. 2007;146(6):406-15.

APPENDIX

APPENDIX FILE 1: CPRD READ CODES USED TO IDENTIFY RA PATIENTS

medcode	readcode	readterm
844	N040.00	Rheumatoid arthritis
6639	14G1.00	H/O: rheumatoid arthritis
6916	N040P00	Seronegative rheumatoid arthritis
8350	N040T00	Flare of rheumatoid arthritis
12019	N04X.00	Seropositive rheumatoid arthritis, unspecified
5723	N042200	Rheumatoid nodule
27603	N04..00	Rheumatoid arthritis and other inflammatory polyarthropathy
9707	N047.00	Seropositive errosive rheumatoid arthritis
8583	N042000	Rheumatic carditis
100187	38DZ.00	Disease activity score in rheumatoid arthritis
23552	N041.00	Felty's syndrome
28853	N04y012	Fibrosing alveolitis associated with rheumatoid arthritis
31054	N040S00	Rheumatoid arthritis - multiple joint
30548	N040N00	Rheumatoid vasculitis
50863	N040D00	Rheumatoid arthritis of knee
105507	66HB000	Rheumatoid arthritis annual review
31724	N04y000	Rheumatoid lung
53621	N040R00	Rheumatoid nodule
48832	N040700	Rheumatoid arthritis of wrist
42299	N040800	Rheumatoid arthritis of MCP joint
18155	N040Q00	Rheumatoid bursitis
51239	N040F00	Rheumatoid arthritis of ankle
103829	38DZ000	Disease activity score 28 joint in rheumatoid arthritis
44203	N040100	Other rheumatoid arthritis of spine
41941	N040900	Rheumatoid arthritis of PIP joint of finger
51238	N040K00	Rheumatoid arthritis of 1st MTP joint
37431	N042z00	Rheumatoid arthropathy + visceral/systemic involvement NOS
21358	N040200	Rheumatoid arthritis of shoulder
59738	N040500	Rheumatoid arthritis of elbow
49067	N040B00	Rheumatoid arthritis of hip
46436	N042100	Rheumatoid lung disease
49227	N042.00	Other rheumatoid arthropathy + visceral/systemic involvement
56202	Nyu1G00	[X]Seropositive rheumatoid arthritis, unspecified
31209	F396400	Myopathy due to rheumatoid arthritis
62401	F371200	Polyneuropathy in rheumatoid arthritis
44743	N040000	Rheumatoid arthritis of cervical spine
93715	Nyu1100	[X]Other seropositive rheumatoid arthritis
63198	N040A00	Rheumatoid arthritis of DIP joint of finger
56838	N04y011	Caplan's syndrome
70658	N040H00	Rheumatoid arthritis of talonavicular joint
70221	Nyu1200	[X]Other specified rheumatoid arthritis
102088	7P20300	Delivery of rehabilitation for rheumatoid arthritis

APPENDIX FILE 2: CRPD READ/OXMIS CODES USED TO IDENTIFY THR/TKR

OUTCOME EVENTS

THR

Read/OXMIS codes	Read/OXMIS Terms
XE2n7	Primary total prosthetic replacement of hip joint NEC
XaF7k	Primary hybrid total replacement of hip joint NEC
XE08o	Other total prosthetic replacement of hip joint
XE08j	Total prosthetic replacement of hip joint using cement
XaF7j	Primary hybrid total replacement of hip joint
XE08k	Primary cemented total hip replacement
7K22y	Other specified total prosthetic replacement of hip joint
XaF7l	Prosthetic hybrid total replacement of hip joint
X606J	Total hip replacement
XE08m	Total prosthetic replacement of hip joint not using cement
7K20.16	Freeman total replacement of hip joint using cement
7K20000	Primary cemented total hip replacement
7K20.1G	THR - Total prosthetic replacement of hip joint using cement
7K21.13	Lord total replacement of hip joint not using cement
7K21.15	Monk total replacement of hip joint not using cement
7K20.1C	Muller total replacement of hip joint using cement
7K22.12	THR - Other total prosthetic replacement of hip joint
7K22000	Primary total prosthetic replacement of hip joint NEC
7K20.17	Furlong total replacement of hip joint using cement
7K21.16	Ring total replacement of hip joint not using cement
7K20.11	Arthroplasty of hip joint using cement
7K21.12	Furlong total replacement of hip joint not using cement
7K21.11	Freeman total replacement of hip joint not using cement
7K20.1F	Turner total replacement of hip joint using cement
7K20.1A	McKee total replacement of hip joint using cement
7K20.12	Aufranc total replacement of hip joint using cement
7K21.17	THR - Total prosthetic replacement hip joint without cement
7K21.00	Total prosthetic replacement of hip joint not using cement
7K20.14	Exeter total replacement of hip joint using cement
7K20.1D	Pretoria total replacement of hip joint using cement

7K20.19	Ilch total replacement of hip joint using cement
7K21.14	Madreporique total replacement of hip joint not using cement
7K20.18	Howse total replacement of hip joint using cement
7K20.1E	Stanmore total replacement of hip joint using cement
7K22.00	Other total prosthetic replacement of hip joint
7K20300	Primary hybrid total replacement of hip joint NEC
7K20.1B	Monk total replacement of hip joint using cement
7K20.13	Charnley total replacement of hip joint using cement
7K20.00	Total prosthetic replacement of hip joint using cement
7K20.15	Farrer total replacement of hip joint using cement
7K20011	Charnley cemented total hip replacement
7K21y00	Other specified total prosthetic replacement of hip joint not using cement
7K20y00	Other specified total prosthetic replacement of hip joint using cement
7K21000	Primary uncemented total hip replacement
7K21z00	Total prosthetic replacement of hip joint not using cement NOS
7K20z00	Total prosthetic replacement of hip joint using cement NOS
7K22z00	Total prosthetic replacement of hip joint NOS

TKR

Read/OXMIS codes	**Read/OXMIS Terms**
X6060	Total knee replacement
XE08w	Total prosthetic replacement of knee using cement
XE08y	Total prosthetic replacement of knee joint not using cement
XE090	Other total prosthetic replacement of knee joint
XE091	Primary hybrid total knee replacement NEC
7K30.00	Total prosthetic replacement of knee joint using cement
7K30.11	Anametric total replacement of knee joint using cement
7K30.13	Attenborough total replacement of knee joint using cement
7K30.15	Cavendish total replacement of knee joint using cement
7K30.16	Charnley total replacement of knee joint using cement
7K30.17	Deane total replacement of knee joint using cement
7K30.18	Denham total replacement of knee joint using cement
7K30.19	Freeman total replacement of knee joint using cement
7K30.1A	Geomedic total replacement of knee joint using cement
7K30.1B	Geometric total replacement of knee joint using cement
7K30.1C	Guepar hinge replacement of knee joint using cement
7K30.1D	Gunston total replacement of knee joint using cement
7K30.1E	Herbert total replacement of knee joint using cement

7K30.1F	Ilch total replacement of knee joint using cement
7K30.1G	Irving total replacement of knee joint using cement
7K30.1H	Liverpool total replacement of knee joint using cement
7K30.1I	Manchester total replacement of knee joint using cement
7K30.1J	Marmor total replacement of knee joint using cement
7K30.1L	Melbourne total replacement of knee joint using cement
7K30.1N	Polycentric total replacement of knee joint using cement
7K30.1P	Sheehan total replacement of knee joint using cement
7K30.1Q	Shiers total replacement of knee joint using cement
7K30.1R	Stanmore total replacement of knee joint using cement
7K30.1S	Swanson total replacement of knee joint using cement
7K30.1T	Uci total replacement of knee joint using cement
7K30.1V	TKR -Total prosthetic replacement of knee joint using cement
7K31.00	Total prosthetic replacement of knee joint not using cement
7K31.12	TKR - Total prosthetic replacement knee joint without cement
7K32.00	Other total prosthetic replacement of knee joint
7K32.12	TKR - Other total prosthetic replacement of knee joint
7K32000	Primary total knee replacement NEC
7K32011	Primary hybrid total knee replacement NEC
K812	TOTAL KNEE REPLACEMENT
7K30000	Primary cemented total knee replacement
7K30y00	Total prosthetic replacement of knee joint using cement OS
7K30z00	Total prosthetic replacement of knee joint using cement NOS
7K31000	Primary uncemented total knee replacement
7K31y00	Total prosthetic replacement knee joint not using cement OS
7K31z00	Total prosthetic replacement knee joint not using cement NOS
7K32y00	Other total prosthetic replacement of knee joint OS
7K32z00	Other total prosthetic replacement of knee joint NOS

APPENDIX FILE 3: NDORMS "BIG DATA" CPDR STANDARD OPERATING PROCEDURES

CPRD: Data SOPs

Reading Data: General Rules

1) Check if the patients respect the inclusion/eligibility criteria

   If there are patients who do not respect the inclusion criteria and their number is >= 1% of the total number of the patients expected to reflect them, contact CPRD. If their number is < 1%, flag them for exclusion. If the study is a case-control study, flag for exclusion also their controls.

2) When applicable, check if the control patients are "real" controls.

   If there are control patients who are not "real" and their number is >= 1% of the total number of the controls, contact CPRD. If their number is < 1%, flag them for exclusion.

3) When applicable and if requested, drop cases without controls

4) When applicable, check if the proportion between the total number of case and control patients is correct. The right proportion is: case = 1/3, controls = 2/3. If this is not correct, contact CPRD.

5) If ONS DEATH data are provided:

   a) Check the ONS "dod" and the "death_matchrank".

      We consider valid only links where:

      - ONS "dod" between "Start" and "End" of ONS coverage
      - ONS "dod" = CPRD "deathdate" (regardless the "death_matchrank") OR "death_matchrank" equals 1 or 2.

   Therefore, we exclude all ONS patient links where ONS dod <> CPRD deathdate AND death_matchrank >= 3, but we keep the CPRD corresponding patients.

   b. Check the ONS death records for duplications

      If there are duplications, implement the following rules for each patient:

      i. If CPRD death date is available, compare it to the ONS dod: if one coincides, keep this ONS death record, bin the other ONS records. If more than one coincides, take the most likely ONS correct date based on the lower death_matchrank. If it is not possible to decide as the death_matchrank is the same, bin all the ONS links.

      ii. If CPRD death date is available and no ONS dod coincides with it, check if the difference between the ONS dates and the CPRD date is within 1 month: if there is one ONS date

within the washout window, keep this ONS death record and bin the other ONS records. If more than one dod is within 1 month, take the one which date is closest to the CPRD death date.

iii. If CPRD death date is available and no ONS dod is within 1-month washout window, take the most likely ONS dod date based on the lower death_matchrank.

iv. If CPRD death date is not available, take the most likely ONS dod date based on the lower death_matchrank. Bin the ONS links with same death_matchrank.

c. Calculate the date of death considering both CPRD and ONS as follows:
If ONS dod is present, use it regardless the presence of CPRD deathdate
If ONS dod is not present and CPRD deathdate is present, use CPRD deathdate

6) For CPRD data:
   a. Exclude all types of records were the dates are marked as invalid by CPRD (i.e. '01/01/2500')
   b. Exclude all types of records were it is not possible to determine an eventdate or a reliable approximation of it.
   c. Exclude all types of records where events occur BEFORE the maximum of the followings:
      - First Registration Date at the current practice (patient.frd)
      - Up-To-Standard for research for the practice (practice.uts) + a washout window of INTERVAL 1 YEAR
   d. Exclude all types of records where events occur AFTER any of the following, i.e. after the minimum of the followings:
      - Date when the data were extracted by CPRD or, if this information is not available, when the data were received by us
      - Practice last Collection Date (practice.lcd)
      - Transfer Out Date of the patient (patient.tod)
      - Date of death of the patient as calculated above

7) If HES data are provided:
   a) Check the hes_matchrank.
   We consider valid only links where hes_matchrank equals 1 or 2. Therefore, we flag for exclusion all HES patient links where hes_matchrank >= 3, but we keep the CPRD corresponding patients.
   b) Flag for exclusion all the CPRD patids that represent an individual who moved to a subsequent practice, keeping the patid that the person had in the last practice. This will allow to keep possible death date information.

c) Exclude all types of records were it is not possible to determine an eventdate or a reliable approximation of it. For example, in table HES_PROCEDURE_EPI, we use the field evdate when not null and the field epistart otherwise: in this way we did not exclude any records.

d) Exclude all types of HES records where events occur BEFORE the maximum of the followings:
   - "Start" of HES coverage

e) Exclude all types of HES records where events occur AFTER any of the following, i.e. after the minimum of the followings:
   - Date of death of the patient as calculated above
   - "End" of HES coverage

Note: Applying consistent rules to the CPRD and HES data allows the exclusion of all those events that may represent possible duplications.

8) Bin all the Therapy records where the prodcode and the bnfcode are incompatible, i.e. they do not refer to the same product

9) If BMI is used in the study, consider valid numbers only between 10 and 60.

10) If BMI, smoking and drinking alcohol are used in the study, bin all the record that report duplicated and inconsistent information for the same patient in the same eventdate.

11) If BMI, smoking and drinking alcohol are used in the study, if a validity time window is provided around the index date, bin all the data reported to have happened before or after that valid time window.

12) To calculate DDDs in prescriptions apply the following algorithm to the THERAPY(t) table:

```
If (t.numdays >= ddd_lower_limit) and (t.numdays <= ddd_upper_limit)
    Then ddds = t.numdays
    Else
        if (t.qty >0) and (ndd > 0)
            and (qty/ndd >= ddd_lower_limit) and (qty/ndd <= ddd_upper_limit)
                Then ddds = t.qty/t.ndd
            Else ddds = 28
```

where ddd_lower_limit and ddd_upper_limit can be set differently for different drugs in different studies. Defaults are holds in MySQL database "cprd__library"

APPENDIX FILE 4: CODES USED TO IDENTIFY THR/TKR IN ORAD/OHIP

THA

Primary

Prior to 2002 (ICD9/CCP) [variable PRCODE1-10].

• 93.51, total hip replacement with methyl methacrylate; and 93.59, other total hip replacement.

April 1, 2002 onward (ICD10/CCI) [variable INCODE1-20].

• 1.VA.53, implantation of internal device, hip joint -- 1.VA.53.LA-PN (open approach) and 1.VA.53.PN-PN (robotics-assisted approach)

TKA

Primary

Prior to 2002 (ICD9/CCP)

• 93.41, geomedic and polycentric total knee replacement

April 1, 2002 onward (ICD10/CCI)

• 1.VG.53, implantation of internal device, knee joint (includes both TKAs and partial knee replacements (single component prosthetic devices and cement spacers to maintain consistency with above)

Documentation for Stata programme: ITSSimulationLoop

Description

ITSSimulationLoop produces estimates of power for detection of changes in level or trend of outcome following an intervention of interest in the interrupted time-series (ITS) framework. These estimates can be used to explore underlying sample size requirements (per timepoint and in total) for a proposed ITS analysis. The current version of the programme only allows repeated measures over time to take the form of cumulative incidence. It allows specification of general parameters as described below.

Specification of parameters

General parameters

Seed: intializes the random number generator. Possible values: any integer.

Simulation Number: specifies how many times each simulation scenario is repeated. Possible values: any positive integer. Methods to calculate the number of simulations required are available elsewhere. We would not recommend using less than 1,000.

ITS scenario specific parameters

Pre-intervention Setting: specifies the location of the known value of the pre-intervention line. A knowledge of the pre-intervention level of outcome is required and this parameter acts as an indication of where in the pre-intervention section of the time-series the level to be inputted is referring to. Possible values: Beginning or End.

Beginning: specifies that the absolute value of the pre-intervention line is known at the beginning of the pre-intervention period

End: specifies that the absolute value of the pre-intervention line is known at the end of the pre-intervention period

Event Intercept: specifies the absolute value of the pre-intervention line either at the beginning or at the end of the pre-intervention period (see previous setting). Possible values: any real number between 0 and 100. For example, a cumulative incidence of 1% should be typed in as 1.

Pre-intervention Change: where there is some pre-intervention slope in existence then this parameter specifies the total measure change over the pre-intervention period based on which the pre-intervention slope will be calculated. Possible values: any real number between -100 and 100. An example: 1% should be typed in as 1. This parameter is only relevant in the case of slope change models. For step change models the MeasureChange should be specified as 0 and will be ignored during execution.

Timepoint Parameter 1:specifies the proportion of time points in the pre-intervention period relative to the total number of time points. Possible values are any real number between 0 and 1 not including 0 or 1. This value is held constant over all simulation scenarios.

Timepoint Parameter 2 (List): total number of time-points in the time-series. Numerous values can be explored with this parameter and entered in the form of a list. Possible values are any list of positive integers enclosed in double quotation marks.

Average Sample Size (List): specifies a list of sample sizes per timepoint to test. (Possible values: any list of positive integers enclosed in double quotation marks. Note: a sufficiently large sample size should be used in order for proportions to approximate normality)

Sample Size Variance: given that the sample size per timepoint will likely be subject to some variation across the time-series, this parameter specifies the value of this variance as a percentage of the average population size per scenario. Possible values are any real number between 0 and 100 not including 0.

Model Type: specifies whether the ITS model should model a step change or a slope change following the intervention. Possible values: Step or Slope.

Mean Change (List): specifies the effect size of the intervention we seek to detect, expressed as a relative percentage change in average post-intervention counterfactuals (as extrapolated from the pre-intervention level and trend) versus post-intervention modelled values (as based on what we seek to observe given the intervention). Possible values are any list of real numbers between -100 and 100 enclosed in double quotation marks. The values should represent percentages. An example: 1% should be typed in as 1.

Step change model: The MeanChangeList values will be translated to step-changes.

Slope change model: The MeanChangeList values will be translated to average mean changes

Storing parameters

Path: specifies the path to the folder to save the simulation results in. Possible values: any valid path enclosed in double quotation marks.

Postfix: specifies the postfix that is added to the end of each file that is a result of the current simulation. Date and time are automatically added to the filenames by the module. Possible Values: any string no longer than 10 characters.

Example

An example application would be to estimate the power associated with various underlying sample sizes (in terms of number of timepoints and sample size per timepoint) for the time-series described and depicted in Hawley S *et al*. Arthritis Rheumatol. 2016 68 (suppl 10). To do this, after running the code to define the programme, one could then execute the following code (noting that 1000 simulations may take several minutes to complete):

ITSSimulationLoop 20180408 1000 Beginning 2.61209 1.8437 0.5 "12 16 20 26 30 40 50" "150 300 500 750 1100 1500 2230 2800 5600" 30 Slope "-34" “[enter valid file pathway]" ITSfilename

Appendix file 6: Stata programme for ITS power estimation


```
**Title: ITSSimulationLoop
**Version: v0.2
**Date: 8th May 2018
**Authors: Samuel Hawley and Klara Berencsi
**Institution: University of Oxford

capture noisily {
    program drop GenerateITSDataSet
}

program GenerateITSDataSet
    args                                    ///
        /* Simulation Settings */           ///
            SimulationCounter               ///
        /* TimePoint Distribution */        ///
            TimePointParameterI             ///
            TimePointParameterII            ///
        /* Model Settings */                ///
            ModelType                       ///
            MeanChange                      ///
        /* PreIntervention Settings */      ///
            PreIntMeasureChange             ///
            PreInterventionSetting          ///
                EventIntercept              ///
        /* Population Settings */           ///
            AveragePopulationSize           ///
            PopulationVariance              ///

    clear

    display "`SimulationCounter'"


    /* Translating Arguments into Simulation Scenario Parameters
*/
        /* TimePoints */
            global TotTP    =`TimePointParameterII'
```

```
            global TotTP    =`TimePointParameterII'
            global PreIntTP =round(`TimePointParameterI'*
`TimePointParameterII')

        /* PreIntervention Line */
            /* PreIntervention Setting */
                if upper("`PreInterventionSetting'")=="BEGINNING"
 {
                    local BegEventIntercept = `EventIntercept'
                }
                if upper("`PreInterventionSetting'")=="END" {
                    local BegEventIntercept = `EventIntercept'-
`PreIntMeasureChange'
                }

            /* Definition of pre-intervention line parameters */
                local  TotChange=`PreIntMeasureChange'
                local  SlopeTP  =`PreIntMeasureChange'/$PreIntTP

        /* PostIntervention Line */
            if upper("`ModelType'")=="STEP" {
                global InterventionJump =(`BegEventIntercept')*
`MeanChange'/100
                global PostIntSlopeTP   =0
            }
            if upper("`ModelType'")=="SLOPE" {
                global InterventionJump =0
                local  MidPostIntTP     =($TotTP-$PreIntTP)/2
                global PostIntSlopeTP   =((`BegEventIntercept'+
`TotChange'+`SlopeTP'*`MidPostIntTP')*`MeanChange'/100)/
`MidPostIntTP'
                display $PostIntSlopeTP
            }

            }

    /* Set up first 4 columns of the database */
        *allocate memory for database
            set     obs $TotTP
        *assign observations
            gen     TimePoint = _n
            gen     Dots = $TotTP
        *create intervention timepoints
            gen     Intervention = 0 if TimePoint<=$PreIntTP
            replace Intervention = 1 if TimePoint >$PreIntTP
        *create post-intervention timepoint
            gen     IntTP = 0               if _n<=$PreIntTP
            replace IntTP =(_n-$PreIntTP) if _n >$PreIntTP

    /* Calculating the pre-intervention line */

    /* Calculating the expected values of Population size in
each timepoint */
        gen PopSize=round(rnormal(`AveragePopulationSize', ((
`AveragePopulationSize')*`PopulationVariance'/100)))

        display "Normal Distribution used for Population Size
Generation"

    /* Calculating the expected values of Events in each
timepoint */
            gen     EventSize=`BegEventIntercept'/100+TimePoint*
`SlopeTP'/100                                                 if
TimePoint<=$PreIntTP
            replace EventSize=`BegEventIntercept'/100+TimePoint*
`SlopeTP'/100+$InterventionJump/100+IntTP*$PostIntSlopeTP/100 if
TimePoint >$PreIntTP
```

```
86              replace EventSize=`BegEventIntercept'/100+TimePoint*
    `SlopeTP'/100+$InterventionJump/100+IntTP*$PostIntSlopeTP/100 if
    TimePoint >$PreIntTP
87
88          /* Assuming Binomial Distribution */
89              gen AbsoluteOutcome=rbinomial(PopSize,EventSize)
90              gen OutcomeMeasure =AbsoluteOutcome/PopSize*100
91              display "Binomial Distribution used for Event
    Generation"
92
93
94      /* Showing relevant macro variables for comparison with
    analysis results */
95          display "`SlopeTP' $PostIntSlopeTP $InterventionJump"
96      /* Saving simulation counter */
97          global CurrentSimulation `SimulationCounter'
98
99  end
100
101 capture noisily {
102     program drop AnalyseITSDataSet
103 }
104
105 program AnalyseITSDataSet
106     args                                        ///
107         /* Model Settings */                    ///
108             ModelType                           ///
109             MeanChange                          ///
110         /* Population Settings */               ///
111             AveragePopulationSize               ///
112
113     if upper("`ModelType'")=="STEP" {
114         *carry out regression
115             regress OutcomeMeasure TimePoint Intervention

115             regress OutcomeMeasure TimePoint Intervention
116
117         *extract p-value
118             matrix  Results = r(table)
119             scalar  p_Intervention = Results[4,2]
120             display p_Intervention
121
122         *post results to file of estimates
123             post ests ($TotTP) ((`AveragePopulationSize')) (
    $TotTP*`AveragePopulationSize') ((`MeanChange')) ((
    $InterventionJump)) ///
124                         ($CurrentSimulation) (_b[_cons]) (_b[
    TimePoint]) (_b[Intervention]) (_se[Intervention]) (
    p_Intervention)
125     }
126
127     if upper("`ModelType'")=="SLOPE" {
128         *carry out regression
129             regress OutcomeMeasure TimePoint IntTP
130
131         *extract p-value
132             matrix  Results = r(table)
133             scalar  p_IntTP = Results[4,2]
134             display p_IntTP
135
136         *post results to file of estimates
137             post ests ($TotTP) ((`AveragePopulationSize')) ((
    $TotTP*(`AveragePopulationSize'))) ((`MeanChange')) ((
    $PostIntSlopeTP)) ///
138                         ($CurrentSimulation) (_b[_cons]) (_b[
    TimePoint]) (_b[IntTP]) (_se[IntTP]) (p_IntTP)
139     }
140
141     *reset global macro variables
142         macro drop TotTP
```

```
142         macro drop TotTP
143         macro drop PreIntTP
144         macro drop PostIntSlopeTP
145         macro drop InterventionJump
146         macro drop CurrentSimulation
147 end
148
149 capture noisily {
150     program drop OutputITSResults
151 }
152
153 program OutputITSResults
154     args                                    ///
155         /* Simulation Settings */           ///
156             Seed                            ///
157             SimulationNumber                ///
158         /* TimePoint Distribution */        ///
159             TimePointParameterI             ///
160         /* Model Settings */                ///
161             ModelType                       ///
162         /* PreIntervention Settings */      ///
163             PreIntMeasureChange             ///
164             PreInterventionSetting          ///
165                 EventIntercept              ///
166         /* Population Settings */           ///
167             PopulationVariance              ///
168         /* Path to Results */               ///
169             Path                            ///
170             Postfix                         ///
171
172
173     clear
174     display "`Path'"
175     use "`Path'/Simulation_`Postfix'.dta"
176
177     /* Saving Settings of Simulation as MetaData Attached to the
    Database of Simulation Results */
178         notes:  Seed                        ="`Seed'"
                        ///


178         notes:  Seed                        ="`Seed'"
                        ///
179                 SimulationNumber            =
    "`SimulationNumber'"            ///
180                 TimePointParameterI         =
    "`TimePointParameterI'"         ///
181                 ModelType                   ="`ModelType'"
                      ///
182                 PreIntMeasureChange         =
    "`PreIntMeasureChange'"         ///
183                 PreInterventionSetting      =
    "`PreInterventionSetting'"      ///
184                 EventIntercept              ="`EventIntercept'"
               ///
185                 PopulationVariance          =
    "`PopulationVariance'"          ///
186
187
188     /* Calculating Measures by Every Each Parameter Combination */
189         /* Power */
190             bys MeanChange TotTimePoints PopulationSize: egen
    PowerNumerator   = total(p_EstPostIntMeasure < 0.05)
191             bys MeanChange TotTimePoints PopulationSize:  gen
    PowerDenominator = _N
192                                                           gen
    PowerScenario    = (PowerNumerator/PowerDenominator)*100
193

194         /* Absolute Bias */
195             bys MeanChange TotTimePoints PopulationSize: egen
    MeanEstPostIntMeasure = mean(EstPostIntMeasure)
196             gen AbsoluteBias=MeanEstPostIntMeasure-PostIntMeasure
197
198         /* Percentage Bias */
```

```
198         /* Percentage Bias */
199             gen PercentageBias=((MeanEstPostIntMeasure-
    PostIntMeasure)/PostIntMeasure)*100
200
201
202     /* Final DataSet With Results */
203         keep MeanChange TotTimePoints PopulationSize
    TotSampleSize ///
204                 PowerScenario AbsoluteBias PercentageBias
205         duplicates drop
206
207     /* Output Results */
208         putdocx begin,landscape font(Calibri, 9)
209
210             putdocx paragraph
211             putdocx text ("Simulation Results with the Following
    Settings:")
212             putdocx paragraph
213             putdocx text ("Seed=`Seed'")
214             putdocx paragraph
215             putdocx text ("SimulationNumber=`SimulationNumber'")
216             putdocx paragraph
217             putdocx text (
    "PreInterventionSetting=`PreInterventionSetting'")
218             putdocx paragraph
219             putdocx text ("EventIntercept=`EventIntercept'")
220             putdocx paragraph
221             putdocx text (
    "PreIntMeasureChange=`PreIntMeasureChange'")
222             putdocx paragraph
223             putdocx text (
    "TimePointParameterI=`TimePointParameterI'")
224             putdocx paragraph
225             putdocx text (
    "PopulationVariance=`PopulationVariance'%")


225             putdocx text (
    "PopulationVariance=`PopulationVariance'%")
226             putdocx paragraph
227             putdocx text ("ModelType=`ModelType'")
228
229             putdocx table tbl1 = data("MeanChange TotTimePoints
    PopulationSize TotSampleSize PowerScenario AbsoluteBias
    PercentageBias RootMeanSquareError"), varnames ///
230                     border(start, nil) layout(autofitw) border(
    end, nil)
231
232         putdocx save "`Path'/Estimates_`Postfix'.docx", replace
233
234     /* Save Database with Estimates */
235         save "`Path'/Estimates_`Postfix'.dta", replace
236
237 end
238
239 capture noisily {
240     program drop GraphGeneration
241 }
242
243 program GraphGeneration
244     args                                    ///
245         /* PreIntervention Settings */      ///
246             PreInterventionSetting          ///
247                 EventIntercept              ///
248             PreIntMeasureChange             ///
249         /* TimePoint Distribution */        ///
250             TimePointParameterIIList        ///
251         /* Population Settings */           ///
252             AveragePopulationSizeList       ///
253         /* Model Settings */                ///
254             ModelType                       ///
```

```
254             ModelType                       ///
255             MeanChange                      ///
256         /* Path to Results */               ///
257             Path                            ///
258             Postfix                         ///
259
260     clear
261     use "`Path'/Estimates_`Postfix'.dta"
262     drop if MeanChange!=`MeanChange'
263
264     /* PreIntervention Setting */
265                 if upper("`PreInterventionSetting'")=="BEGINNING"
 {
266                     local BegEventIntercept = `EventIntercept'
267                 }
268                 if upper("`PreInterventionSetting'")=="END" {
269                     local BegEventIntercept = `EventIntercept'-
`PreIntMeasureChange'
270                 }
271
272     /* List of axis tick marks */
273         local PreMinTPList =subinstr("`TimePointParameterIIList'"
," ","","",.)
274         local MinTPList    =     min(`PreMinTPList')
275         local MaxTPList    =     max(`PreMinTPList')
276         local DistTP       =   round((`MaxTPList'-`MinTPList')/7)
277         local MaxTPList    =     `MinTPList'+8*`DistTP'
278         local PreMinPSList =subinstr(
"`AveragePopulationSizeList'"," ","","",.)
279         local MinPSList    =     min(`PreMinPSList')
280         local MaxPSList    =     max(`PreMinPSList')


280          local MaxPSList    =     max(`PreMinPSList')
281          local DistPS       =   round((`MaxPSList'-`MinPSList')/7)
282          local MaxPSList    =`MinPSList'+8*`DistPS'
283          local AvgPreMeasure=round(`BegEventIntercept'+0.5*
`PreIntMeasureChange',0.01)
284
285     twoway contour PowerScenario PopulationSize TotTimePoints,
ccuts(0(20)100)                                 ///
286         xlabel(`MinTPList'(`DistTP')`MaxTPList') ylabel(
`MinPSList'(`DistPS')`MaxPSList',angle(forty_five)) ///
287         xtitle("Total Number of Timepoints" ,size(medsmall)
margin(medsmall))                               ///
288         ytitle("Sample Size per Timepoint"  ,size(medsmall)
margin(large))                                  ///
289         ztitle("Statistical Power"          ,size(medsmall)
margin(small))                                  ///
290         title("Statistical Power to Detect `ModelType' Change
resulting in an Average"                        /*
291             */"`MeanChange'% Relative Change in Outcome Where
Mean Pre-Intervention"                          /*
292             */"Incidence is `AvgPreMeasure'% (by Number of
Timepoints and Mean"                            /*
293             */"Sample Size per Timepoint)"  ,size(small) margin(
medsmall))                                      ///
294         plotregion(fcolor(white)) graphregion(fcolor(white))
crule(intensity) ecolor(green)
295
```

```
295
296     graph export "`Path'/Figures/Power_MCh`MeanChange'.png", as(
    png) replace
297
298 end
299
300 capture noisily {
301     program drop ITSSimulationLoop
302 }
303
304 program ITSSimulationLoop
305     args                                    ///
306         /* Simulation Settings */           ///
307             Seed                            ///
308             SimulationNumber                ///
309         /* PreIntervention Settings */      ///
310             PreInterventionSetting          ///
311                 EventIntercept              ///
312             PreIntMeasureChange             ///
313         /* TimePoint Distribution */        ///
314             TimePointParameterI             ///
315             TimePointParameterIIList        ///
316         /* Population Settings */           ///
317             AveragePopulationSizeList       ///
318             PopulationVariance              ///
319         /* Model Settings */                ///
320             ModelType                       ///
321             MeanChangeList                  ///
322         /* Path to Results */               ///
323             Path                            ///
324             Postfix                         ///
325
326     clear
327
328     /* Checking the Values Specified as Arguments of the
    ITSSimulationLoop */
329         /* Seed */
330             if int(`Seed') != `Seed' {
```

Appendix Table 3.1: Validation measures of outcome events during follow-up according to CPRD versus HES (allowing 30-day or 90-day time window for agreement): for THR and TKR

	Sensitivity	Specificity	PPV	NPV
30 day				
HIP	77.9	98.6	67	99.2
KNEE	82.2	98.4	71.6	99.1
180 day				
HIP	81	98.6	69.2	99.1
KNEE	84.4	98.5	73.3	99.2

*using HES as reference standard, agreement only counted if date of surgery within 30/180 days of CPRD THR/TKR date

Appendix Table 4.1: THR time-series: 5-year cumulative incidence outcome

timepoint	year	No. RA diagnoses	No. with subsequent THR (5-year)	Crude 5-year cumulative incidence of THR	Standardised 5-year cumulative incidence of THR	Standardised 5-year cumulative incidence of THR" lower 95% C.I.	Standardised 5-year cumulative incidence of THR" upper 95% C.I.
1	1995	314	6	1.91	1.95	0.40	3.50
2	1995	287	6	2.09	2.17	0.44	3.89
3	1996	331	9	2.72	2.59	0.92	4.27
4	1996	330	11	3.33	3.19	1.33	5.05
5	1997	355	11	3.10	3.00	1.25	4.74
6	1997	455	13	2.86	2.86	1.34	4.39
7	1998	374	10	2.67	2.38	0.91	3.85
8	1998	398	9	2.26	2.23	0.79	3.68
9	1999	418	8	1.91	1.96	0.63	3.30
10	1999	551	10	1.81	1.75	0.67	2.83
11	2000	498	14	2.81	2.77	1.33	4.20
12	2000	662	19	2.87	2.90	1.62	4.18
13	2001	711	15	2.11	2.06	1.03	3.09
14	2001	780	23	2.95	2.93	1.75	4.10
15	2002	750	25	3.33	3.36	2.07	4.64
16	2002	774	13	1.68	1.66	0.77	2.55
17	2003	753	31	4.12	4.12	2.70	5.53
18	2003	823	41	4.98	5.01	3.53	6.48
19	2004	798	25	3.13	3.03	1.86	4.20
20	2004	836	18	2.15	2.12	1.15	3.08
21	2005	784	20	2.55	2.50	1.42	3.59
22	2005	694	23	3.31	3.33	2.00	4.66
23	2006	772	14	1.81	1.75	0.84	2.66
24	2006	660	23	3.48	3.47	2.09	4.85
25	2007	720	19	2.64	2.84	1.60	4.08
26	2007	649	21	3.24	3.19	1.87	4.52
27	2008	670	15	2.24	2.44	1.20	3.67
28	2008	697	6	0.86	0.85	0.16	1.54
29	2009	661	7	1.06	1.03	0.27	1.78

Appendix Table 4.2: TKR time-series: 5-year cumulative incidence outcome

Appendix Table 4.2: TKR time-series: 5-year cumulative incidence outcome							
timepoint	year	No. RA diagnoses	No. with subsequent TKR (5-year)	Crude 5-year cumulative incidence of TKR	Stansardised 5-year cumulative incidence of TKR	Stansardised 5-year cumulative incidence of TKR lower 95% C.I.	Stansardised 5-year cumulative incidence of TKR upper 95% C.I.
1	1995	314	9	2.87	2.98	1.06	4.90
2	1995	287	5	1.74	2.20	0.31	4.09
3	1996	331	12	3.63	3.63	1.62	5.64
4	1996	330	6	1.82	1.88	0.39	3.37
5	1997	355	15	4.23	4.27	2.16	6.38
6	1997	455	19	4.18	4.28	2.42	6.13
7	1998	374	10	2.67	2.36	0.91	3.82
8	1998	398	13	3.27	3.29	1.54	5.04
9	1999	418	20	4.78	4.72	2.71	6.74
10	1999	551	24	4.36	4.39	2.67	6.12
11	2000	498	20	4.02	3.93	2.24	5.62
12	2000	662	21	3.17	3.27	1.91	4.63
13	2001	711	38	5.34	5.31	3.66	6.95
14	2001	780	30	3.85	3.89	2.54	5.24
15	2002	750	25	3.33	3.31	2.04	4.59
16	2002	774	29	3.75	3.78	2.43	5.14
17	2003	753	31	4.12	3.95	2.58	5.31
18	2003	823	47	5.71	5.67	4.11	7.22
19	2004	798	23	2.88	2.76	1.65	3.88
20	2004	836	36	4.31	4.19	2.85	5.53
21	2005	784	44	5.61	5.46	3.90	7.02
22	2005	694	24	3.46	3.48	2.12	4.84
23	2006	772	26	3.37	3.33	2.07	4.58
24	2006	660	18	2.73	2.68	1.46	3.90
25	2007	720	28	3.89	3.98	2.54	5.42
26	2007	649	22	3.39	3.31	1.96	4.67
27	2008	670	15	2.24	2.14	1.06	3.21
28	2008	697	22	3.16	3.02	1.77	4.26
29	2009	661	18	2.72	2.77	1.51	4.03

Appendix Table 4.3: Temporal Trends in 5-Year THR and TKR Incidence Rates Among 17,505 Incident Rheumatoid Arthritis Patients Diagnosed From 1995 to 2009 (per 1,000 Person-Years): SENSITIVITY ANALYSIS using full models (i.e. no selection)

Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
Total Hip Replacement				
Intercept	5.43	3.48	7.37	<0.001
Trend [1]	0.03	-0.20	0.25	0.81
Level change after NICE TA 36	3.74	0.79	6.69	0.015
Trend change after NICE TA 36	-0.49	-0.83	-0.15	0.007
Total Knee Replacement				
Apr-Oct 1995 Incidence Rate	5.83	3.50	8.16	<0.001
Trend [1]	0.32	0.05	0.60	0.023
Level change after NICE TA 36	-0.20	-3.74	3.34	0.91
Trend change after NICE TA 36	-0.68	-1.09	-0.27	0.002

[1] per 6 months

Appendix Table 4.4: Temporal Trends in 5-Year Cumulative Incidence (%) of THR and TKR Among 17,505 Incident Rheumatoid Arthritis Patients Diagnosed From 1995 to 2009: SENSITIVITY ANALYSIS using cumulative incidence

Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
Total Hip Replacement				
Intercept	2.48	2.10	2.86	<0.001
Trend [1]	-	-	-	-
Level change after NICE TA 36	1.71	0.80	2.62	0.001
Trend change after NICE TA 36	-0.21	-0.31	-0.10	<0.001
Total Knee Replacement				
Apr-Oct 1995 Incidence Rate	2.61	1.71	3.52	<0.001
Trend [1]	0.13	0.04	0.22	0.005
Level change after NICE TA 36	-	-	-	-
Trend change after NICE TA 36	-0.29	-0.47	-0.12	0.002

- = P≥0.2

[1] per 6 months

Appendix Table 5.1: Interrupted Time-Series Linear Regression Analysis: CUMULATIVE INCIDENCE of csDMARDs (1-year) and biologics (5-year) (%) among SENIOR RA patients in Ontario, Canada

	Parameter	Coefficient	Lower 95% C.I.	Upper 95% C.I.	*P*-Value
Ontario, Canada					
csDMARD (1-year)	Baseline Incidence Rate	34.30	32.40	36.10	<0.001
	Trend (per 6 months)	0.99	0.88	1.10	<0.001
	Level change after Biologics	-	-	-	-
	Trend change after Biologics	-	-	-	-
Biologic (5-year)	Baseline Incidence Rate	-0.23	-0.75	0.29	0.37
	Trend (per 6 months)	0.12	0.06	0.18	<0.001
	Level change after Biologics	0.67	-0.37	1.72	0.19
	Trend change after Biologics	-	-	-	-

Appendix table 7.1: baseline characteristics of RA patients: variables with missing data (stratified by incident use of Biologic vs. Conventional Synthetic DMARDs)

	All eligible RA patients (N=13,128)			
	CS-DMARD (N=3,229)		**Biologic (N=9,897)**	
Characteristic	**N**	**%**	**N**	**%**
Age: Mean (S.D.)	59.1 (12.4)		55.1 (12.3)	
Missing	39	1.2	402	4.1
ethnicity: White/Caucasian	2,583	97.5	8,151	95.7
Missing	581	18	1,383	14.0
Index Multiple Deprivation				
Quintile 1	609	18.9	1339	13.5
Quintile 2	561	17.4	1524	15.4
Quintile 3	481	14.9	1719	17.4
Quintile 4	525	16.3	1789	18.1
Quintile 5	421	13	1743	17.6
Missing	632	19.6	1785	18.0
BMI	27.3 (6.0)		27.1 (6.4)	
Missing	184	5.7	1,326	13.4
Smoking?				
Current	812	25.3	2290	23.4
Ex	1,242	38.7	3,673	37.5
Missing	17	0.5	95	1.0
Years since diagnosis: Median (IQR)	8.6 (9.5)		11.1 (8.9)	
Missing	61	1.9	506	5.1
DAS 28: Mean (S.D.)	5.11 (1.30)		6.48 (1.00)	
Missing	45	1.4	67	0.7
Overall HAQ score	1.58 (1.24)		2.11 (1.28)	
Missing	645	20	616	6.2
SF36				
physical function: mean (/100) (S.D.)	36.9 (27.2)		21.9 (21.1)	
Missing	627	19.4	1574	15.9
limited physically: mean (/100) (S.D.)	38.5 (28.9)		19.9 (23.2)	
Missing	625	19.4	1567	15.8
limited emotionally: mean (/100) (S.D.)	56.9 (34.6)		42.2 (36.2)	
Missing	625	19.4	1562	15.8
energy: mean (/100) (S.D.)	35.3 (20.5)		25.3 (19.9)	
Missing	626	19.4	1573	15.9
emotional health: mean (/100) (S.D.)	61.8 (21.6)		53.1 (21.5)	
Missing	626	19.4	1568	15.8
social: mean (/100) (S.D.)	44.6 (21.5)		33.3 (21.6)	
Missing	624	19.3	1565	15.8
pain: mean (/100) (S.D.)	42.4 (23.3)		25.2 (19.6)	
Missing	626	19.4	1560	15.8
general: mean (/100) (S.D.)	44.7 (15.2)		40.1 (14.2)	
Missing	624	19.3	1569	15.9

Appendix table 7.2A: Baseline characteristics of RA patients: stratified by incident use of Biologic vs. Conventional Synthetic DMARDs: *BEFORE MATCHING*

	All eligible RA patients (N=13,126)				
	CS-DMARD (N=3,229)		**Biologic (N=9,897)**		**SMD**
Characteristic	**N**	**%**	**N**	**%**	
Age: Mean (S.D.)	59.0 (12.4)		55.1 (12.3)		-0.32
Gender: Female	2,338	72.4%	7,522	76.0%	0.08
ethnicity: White/Caucasian	3,150	97.6%	9,428	95.3%	0.12
Index Multiple Deprivation					
Quintile 1	609	18.9%	1339	13.5%	-0.15
Quintile 2	561	17.4%	1524	15.4%	-0.05
Quintile 3	481	14.9%	1719	17.4%	0.07
Quintile 4	525	16.3%	1789	18.1%	0.05
Quintile 5	421	13.0%	1743	17.6%	0.13
Unknown	632	19.6%	1785	18.0%	-0.04
BMI	27.3 (6.1)		27.1 (6.3)		-0.03
Smoking?					
Current	815	25.2%	2306	23.3%	-0.05
Ex	1,250	38.7%	3,696	37.3%	-0.03
Calendar period of registration					
Quintile 1	251	7.8%	2517	25.4%	0.49
Quintile 2	538	16.7%	2083	21.0%	0.11
Quintile 3	764	23.7%	1820	18.4%	-0.13
Quintile 4	1108	34.3%	1466	14.8%	-0.47
Quintile 5	568	17.6%	2013	20.3%	0.07
Years since diagnosis: Median (IQR)	8.6 (9.5)		11.1 (8.9)		0.27
DAS 28: Mean (S.D.)	5.10 (1.30)		6.48 (1.00)		1.18
Overall HAQ score	1.46 (0.75)		1.92 (0.62)		0.66
ACR: Ever rheumatoid positive	1,853	57.4%	6,306	63.7%	0.13
ACR: Deformity of >=3 joint areas	2,325	72.0%	8,452	85.4%	0.33
ACR: Erosions on hands/feet	1,436	44.5%	5,543	56.0%	0.23
ACR: Ever had nodules	961	29.8%	4,203	42.5%	0.27
ACR: Symmetry	2,107	65.3%	8,211	82.9%	0.41
ACR: Deformity of hand joint	2,280	70.6%	7,929	80.1%	0.22
ACR: Morning stiffness >1 hour	2,839	87.9%	9,305	94.0%	0.21
SF36: mean (/100) (S.D.)					
physical function	36.4 (27.0)		22.6 (21.6)		-0.56
limited physically	38.2 (29.0)		20.6 (23.6)		-0.66
limited emotionally	56.3 (34.8)		43.2 (36.4)		-0.37
energy	35.2 (20.7)		25.5 (19.9)		-0.48
emotional health	61.4 (21.6)		53.4 (21.4)		-0.37
social	44.4 (21.4)		34.1 (21.6)		-0.48
pain	41.9 (23.3)		25.6 (19.7)		-0.76
general	44.6 (15.4)		40.3 (14.3)		-0.29
Comorbidities: Ever Prior					
Hypertension	956	29.6%	2777	28.1%	-0.03
Angina	240	7.4%	406	4.1%	-0.14
MI	143	4.4%	259	2.6%	-0.10
Stroke	101	3.1%	183	1.8%	-0.08

Continued on next page...

...Continued from last page

Epilesy	43	1.3%	112	1.1%	-0.02
COPD	267	8.3%	512	5.2%	-0.12
Peptic ulcer	210	6.5%	697	7.0%	0.02
TB	70	2.2%	218	2.2%	0.00
liver disease	62	1.9%	203	2.1%	0.01
renal disease	84	2.6%	202	2.0%	-0.04
diabetes	215	6.7%	576	5.8%	-0.04
hyperthyroidism	137	4.2%	327	3.3%	-0.05
depression	547	16.9%	1936	19.6%	0.07
cancer	208	6.4%	291	2.9%	-0.17
demyelin	9	0.3%	15	0.2%	-0.03
asthma	451	14.0%	1059	10.7%	-0.1
Current medication					
antibiotic	38	1.2%	98	1.0%	-0.02
anticancer	8	0.2%	6	0.1%	-0.05
anticoagulant	103	3.2%	310	3.1%	0.00
antidepressant	335	10.4%	1218	12.3%	0.06
antidiabetic	127	3.9%	343	3.5%	-0.03
antihypertensive	815	25.2%	2116	21.4%	-0.09
antiosteoporosis	426	13.2%	1615	16.3%	0.09
antipsychotic	2	0.1%	18	0.2%	0.03
antithrombotic	43	1.3%	94	0.9%	-0.04
aspirin	370	11.5%	629	6.4%	-0.18
asthma inhaler	250	7.7%	448	4.5%	-0.14
betablocker	342	10.6%	946	9.6%	-0.03
calcium	387	12.0%	1786	18.0%	0.17
copd_inhaler	34	1.1%	59	0.6%	-0.05
digoxin	32	1.0%	53	0.5%	-0.05
diuretic	51	1.6%	96	1.0%	-0.05
glucocorticoid (current)	701	21.7%	3935	39.8%	-0.05
glucocorticoid (prior use only)	1036	32.1%	3278	33.1%	0.02
hormone_contraceptive	8	0.2%	94	0.9%	0.09
Hormone replaement therapy	116	3.6%	446	4.5%	0.05
Hypnotic	84	2.6%	309	3.1%	0.03
insulin	62	1.9%	173	1.7%	-0.01
lithium	1	0.0%	12	0.1%	0.03
csDMARD (current/prior): median used (IQR)	3 (2-5)		4 (4-6)		0.65
nitrate	38	1.2%	72	0.7%	-0.05
NSAID	1217	37.7%	4032	40.7%	0.06
opiod	1334	41.3%	4680	47.3%	0.12
PPI	688	21.3%	2715	27.4%	0.14
statin	440	13.6%	833	8.4%	-0.17
thyroid hormone	293	9.1%	812	8.2%	-0.03
Systemic involvement ever (any)	573	17.7%	2679	27.1%	0.23
eye involvement	196	6.1%	882	8.9%	0.11
systemic vasculitis	21	0.7%	147	1.5%	0.08
nailfold vasulitis	33	1.0%	174	1.8%	0.06
pulminory involvement	58	1.8%	255	2.6%	0.05
other systemic symptoms	79	2.4%	222	2.2%	-0.01
Non-major prior total joint replacement (any)	460	14.2%	2094	21.2%	0.18

Appendix table 7.2B: Baseline characteristics of RA patients: stratified by incident use of Biologic vs. Conventional Synthetic DMARDs: *AFTER MATCHING*

Characteristic	CS-DMARD (N=9,558) (1,644 unique patients)		Users of Biologic (N=9,558)		SMD
	N	%	N	%	
Age: Mean (S.D.)	55.2 (12.1)		55.2 (12.3)		0
Gender: Female	7,289	76.3%	7,259	75.9%	-0.01
ethnicity: White/Caucasian	9,114	95.4%	9,118	95.4%	0.00
Index Multiple Deprivation					
Quintile 1	1282	13.4%	1322	13.8%	0.01
Quintile 2	1420	14.9%	1485	15.5%	0.02
Quintile 3	1413	14.8%	1640	17.2%	0.07
Quintile 4	1544	16.2%	1710	17.9%	0.05
Quintile 5	1284	13.4%	1650	17.3%	0.11
Unknown	2615	27.4%	1751	18.3%	-0.22
BMI	26.8 (5.9)		27.1 (6.3)		0.06
Smoking?					
Current	2448	25.6%	2259	23.6%	-0.05
Ex	3272	34.2%	3577	37.4%	0.07
Calendar period of registration					
Quintile 1	1481	15.5%	2262	23.7%	0.21
Quintile 2	2037	21.3%	2037	21.3%	0.00
Quintile 3	1823	19.1%	1802	18.9%	-0.01
Quintile 4	1664	17.4%	1462	15.3%	-0.06
Quintile 5	2553	26.7%	1995	20.9%	-0.14
Years since diagnosis: Median (IQR)	10.8 (10.7)		11.0 (8.8)		0.02
DAS 28: Mean (S.D.)	6.47 (1.09)		6.43 (0.98)		-0.04
Overall HAQ score	1.91 (0.63)		1.91 (0.62)		-0.01
ACR: Ever rheumatoid positive	5,532	57.9%	6,055	63.4%	0.11
ACR: Deformity of >=3 joint areas	7,425	77.7%	8,118	84.9%	0.19
ACR: Erosions on hands/feet	4,651	48.7%	5,282	55.3%	0.13
ACR: Ever had nodules	4,034	42.2%	3,988	41.7%	-0.01
ACR: Symmetry	7,468	78.1%	7,883	82.5%	0.11
ACR: Deformity of hand joint	6,771	70.8%	7,602	79.5%	0.20
ACR: Morning stiffness >1 hour	8,901	93.1%	8,966	93.8%	0.03
SF36: mean (/100) (S.D.)					
physical function	23.3 (22.4)		23.0 (21.7)		-0.01
limited physically	22.6 (21.8)		21.0 (23.8)		-0.02
limited emotionally	42.5 (34.3)		43.1 (36.0)		0.02
energy	26.2 (18.5)		25.8 (20.0)		-0.02
emotional health	53.5 (21.5)		53.6 (21.5)		0.01
social	33.3 (22.5)		34.3 (21.6)		0.05
pain	26.7 (19.6)		26.0 (19.7)		-0.03
general	39.0 (13.9)		40.3 (14.3)		0.09
Comorbidities: Ever Prior					
Hypertension	2581	27.0%	2676	28.0%	0.02
Angina	421	4.4%	403	4.2%	-0.01
MI	230	2.4%	257	2.7%	0.02
Stroke	139	1.5%	179	1.9%	0.03

Continued on next page...

...Continued from last page

Epilesy	117	1.2%	110	1.2%	-0.01
COPD	713	7.5%	504	5.3%	-0.09
Peptic ulcer	642	6.7%	666	7.0%	0.01
TB	188	2.0%	204	2.1%	0.01
liver disease	132	1.4%	195	2.0%	0.05
renal disease	172	1.8%	198	2.1%	0.02
diabetes	690	7.2%	551	5.8%	-0.06
hyperthyroidism	359	3.8%	320	3.3%	-0.02
depression	1726	18.1%	1862	19.5%	0.04
cancer	316	3.3%	287	3.0%	-0.02
demyelin	10	0.1%	14	0.1%	0.01
asthma	949	9.9%	1035	10.8%	0.03
Current medication					
antibiotic	156	1.6%	97	1.0%	-0.05
anticancer	4	0.0%	6	0.1%	0.01
anticoagulant	306	3.2%	298	3.1%	-0.01
antidepressant	1281	13.4%	1170	12.2%	-0.04
antidiabetic	413	4.3%	332	3.5%	-0.04
antihypertensive	2112	22.1%	2047	21.4%	-0.02
antiosteoporosis	1590	16.6%	1546	16.2%	-0.01
antipsychotic	4	0.0%	18	0.2%	0.04
antithrombotic	109	1.1%	92	1.0%	-0.02
aspirin	588	6.2%	622	6.5%	0.02
asthma inhaler	542	5.7%	444	4.6%	-0.05
betablocker	829	8.7%	910	9.5%	0.03
calcium	1674	17.5%	1667	17.4%	0.00
copd_inhaler	53	0.6%	56	0.6%	0.00
digoxin	47	0.5%	52	0.5%	0.01
diuretic	111	1.2%	96	1.0%	-0.02
glucocorticoid (current)	3227	33.8%	3664	38.3%	0.10
glucocorticoid (prior use only)	3181	33.3%	3234	33.8%	0.01
hormone_contraceptive	77	0.8%	76	0.8%	0.00
Hormone replaement therapy	402	4.2%	423	4.4%	0.01
Hypnotic	341	3.6%	300	3.1%	-0.02
insulin	237	2.5%	164	1.7%	-0.05
lithium	3	0.0%	9	0.1%	0.03
csDMARD (current/prior): median used (IQR)	4 (3-6)		4 (3-6)		0.04
nitrate	50	0.5%	72	0.8%	0.03
NSAID	3713	38.8%	3882	40.6%	0.04
opiod	4784	50.1%	4518	47.3%	-0.06
PPI	2714	28.4%	2607	27.3%	-0.03
statin	808	8.5%	823	8.6%	0.01
thyroid hormone	973	10.2%	792	8.3%	-0.07
Systemic involvement ever (any)	2245	23.5%	2546	26.6%	0.07
eye involvement	714	7.5%	836	8.7%	0.05
systemic vasculitis	97	1.0%	134	1.4%	0.04
nailfold vasulitis	157	1.6%	168	1.8%	0.01
pulminory involvement	239	2.5%	242	2.5%	0.00
other systemic symptoms	363	3.8%	215	2.2%	-0.09
Non-major prior total joint replacement (any)	1742	18.2%	1989	20.8%	0.07

Appendix table 7.3: Crude incidence of joint replacement among matched TNFi and csDMARD cohorts, with adjusted hazard ratios. Sensitivity analysis using 6 month persistence of exposure status after switching

	csDMARD						TNFi				
	No. patients (crude)	No. outcome events (crude)	No. patients (matched with replacement)	No. outcome events (matched with replacement)	median follow up (matched with replacement), years (IQR)	rate (matched with replacement) per 1000 PYs (95% CI)	No. patients	No. outcome events	median follow up, years (IQR)	rate per 1000 PYs (95% CI)	adj HR
THR	1644	50	9558	302	5.98 (2.42-9.55)	6.16 (4.14-9.54)	9558	304	5.05 (1.94-10.07)	5.28 (4.72-5.92)	0.85 (0.63 - 1.15); p=0.29
TKR	1644	51	9558	384	6.00 (2.39-9.55)	8.92 (5.20-12.60)	9558	489	4.92 (1.91-10.02)	8.65 (7.91-9.47)	1.06 (0.81 - 1.37); p=0.68

*Results shown are for the 10th imputed dataset. Each TNFi-user matched 1:1 to a csDMARD user, with replacement. Standard errors were corrected for ±Anlysis of OJR was restricted to the subset of patients with no prior OJR

Appendix table 7.4: Crude associations between (1) treatment received and outcome and (2) instrumental variable and outcome (12 year outcome)

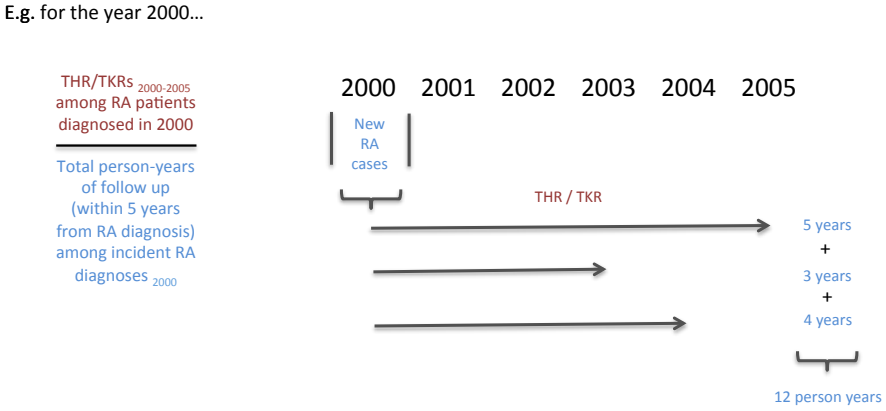
	Treatement received				
	Biologic (N=9,897)		csDAMRD (N=3,229)		p-value
	n	%	n	%	
THR	427	4.3	84	2.6	<0.001
TKR	671	6.8	95	2.9	<0.001
	physician preference (IV) *				
	Biologic (N=5,018)		csDAMRD (N=2,901)		
	n	%	n	%	
THR	182	3.6	106	3.7	0.95
TKR	281	5.6	114	3.9	0.001

* physician preference for biologics of csDMARD determined using </>= 0.5 cutoff of relative frequency of biologics within physicians' prior 12 prescriptions. Sample size is lower than full population oweing to 'burn in' for creating IV for each physician

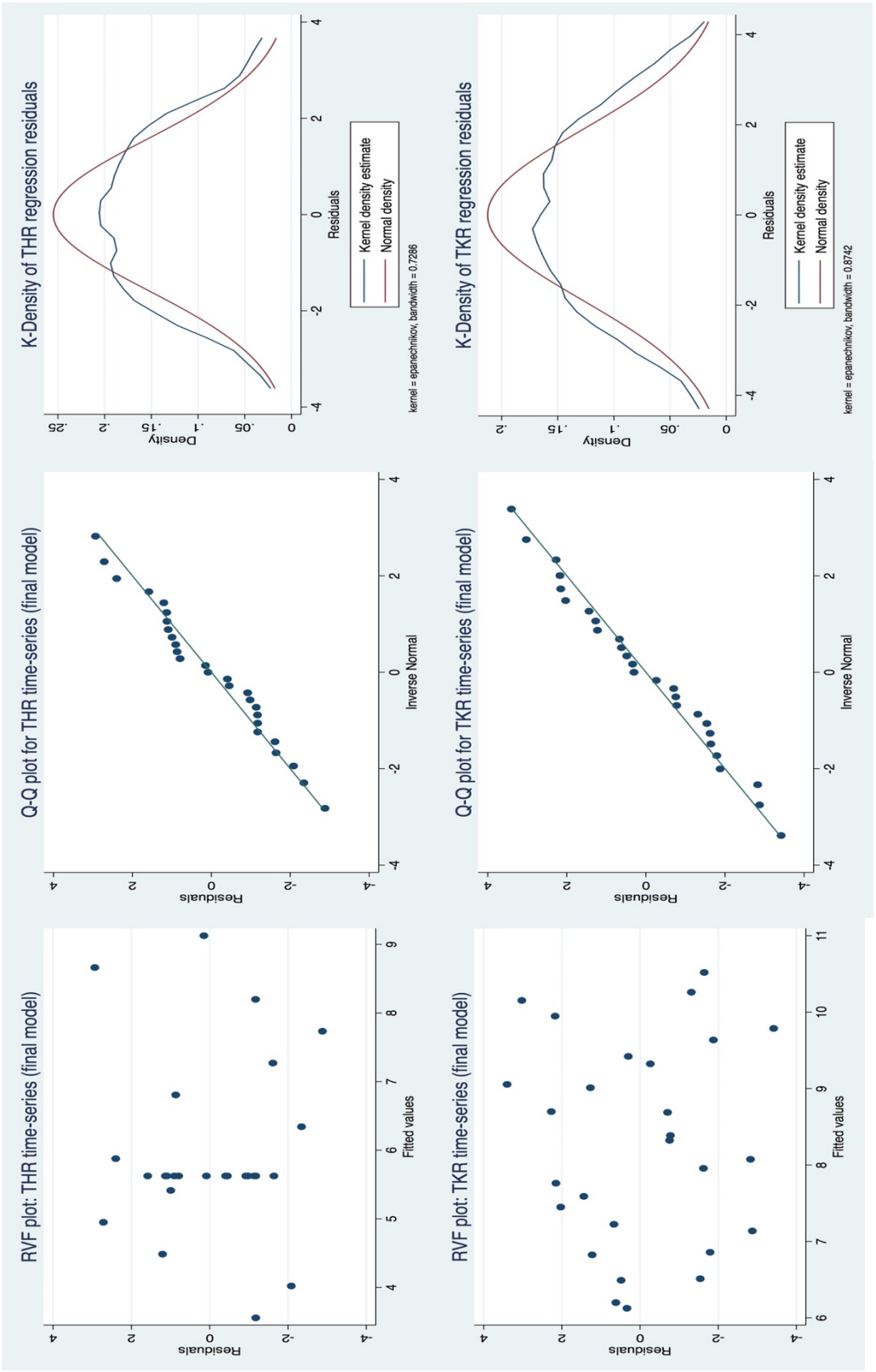
Appendix table 7.5: IV (physician preference) analysis of biologics impact on need for THR/TKR: 12 year outcome

	Absolute risk difference between patients treated by physicians preferring biologics versus csDMARDs (per 100 patients)			
	Estimate	lower 95% C.I.	upper 95% C.I.	P-value
IV: crude				
THR	0.00	-1.33	1.25	0.95
TKR	2.55	2.34	4.32	<0.001
IV: covariate adjusted				
THR	-2.03	-4.21	0.16	0.069
TKR	0.89	-1.43	3.21	0.45

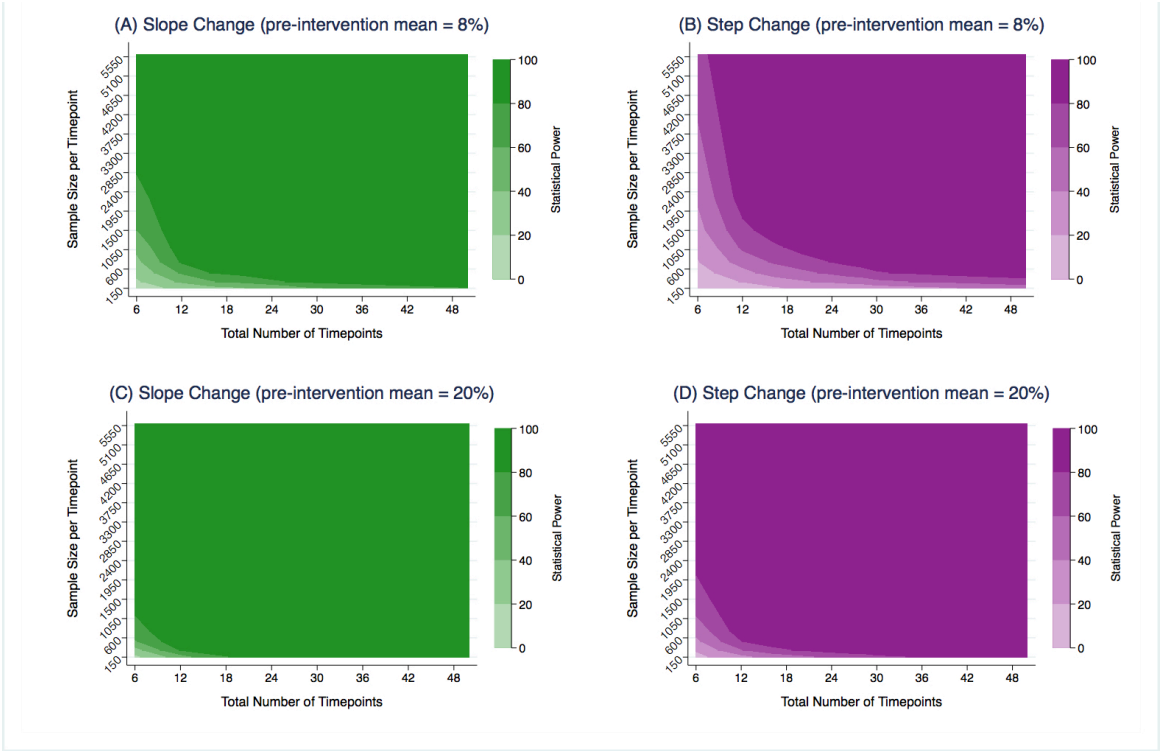
APPENDIX FIGURES



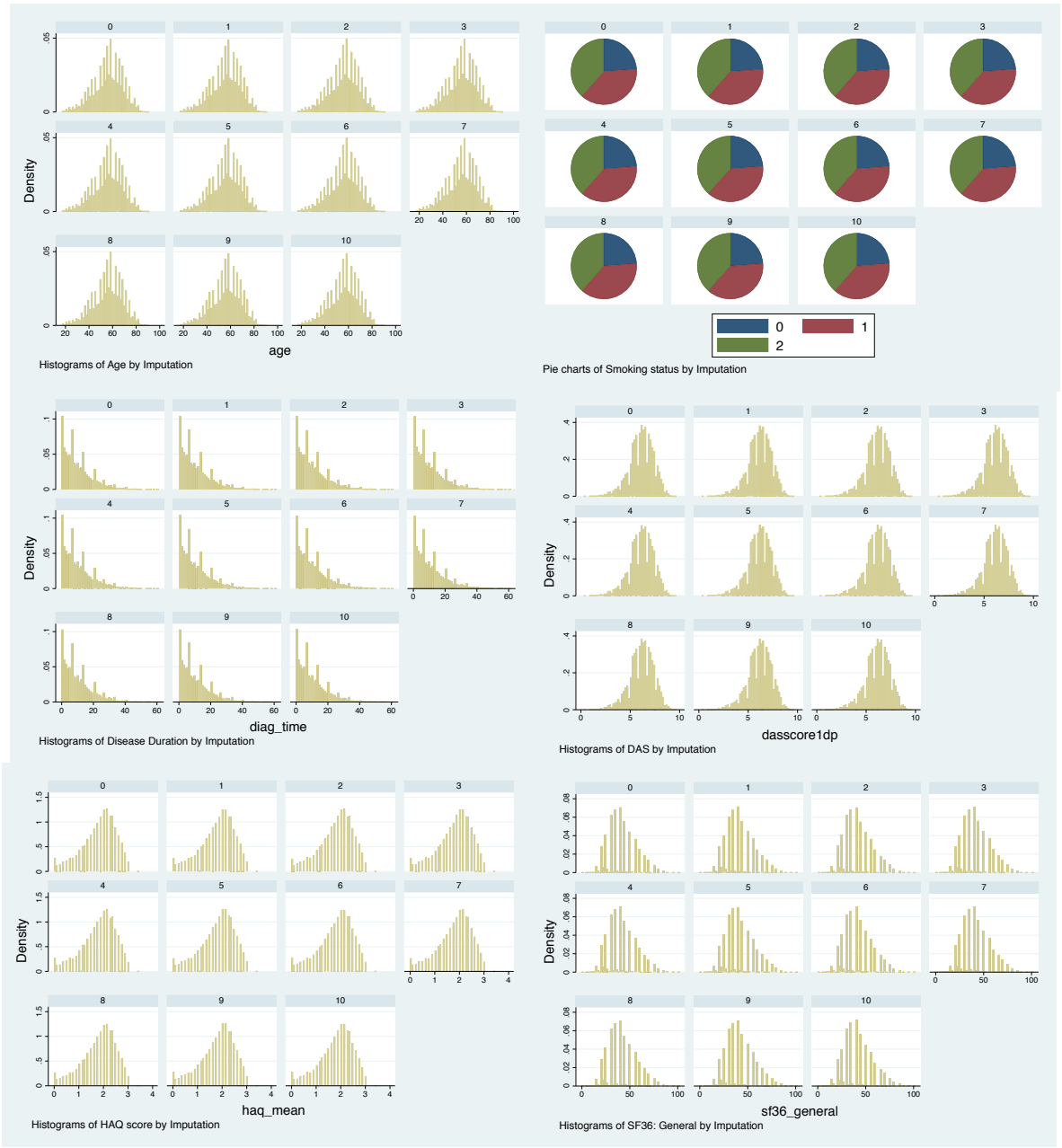
Appendix Figure 4.1: Example of timepoint measurement (note: here each 'timepoint' is a year, however in analysis each timepoint was a six month period)



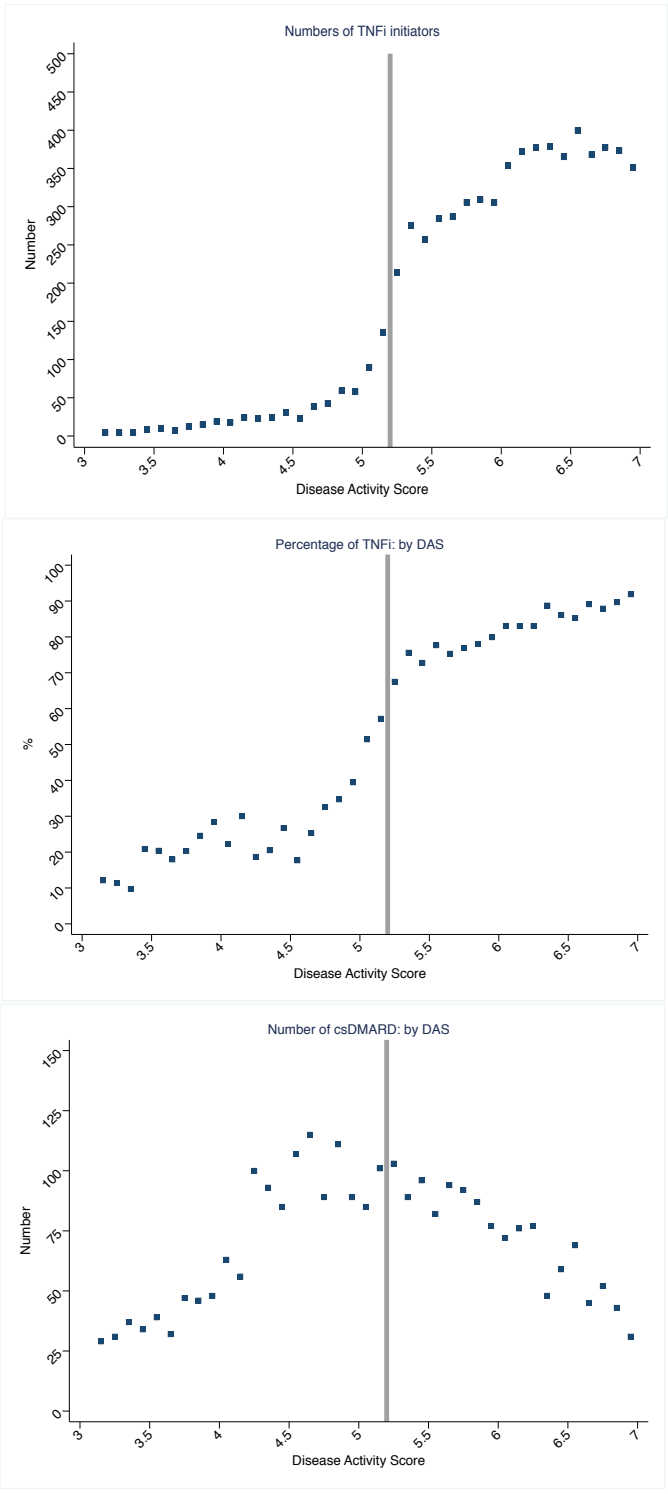
Appendix Figure 4.2: Regression diagnostic plots: RVF plot for THR time-series (top left); RVF plot for TKR time-series (bottom left); QQ plot for THR time-series (top middle); QQ plot for TKR time-series (bottom middle); Kernel density plot for THR time-series (top right); kernel density plot for TKR time-series (bottom right)



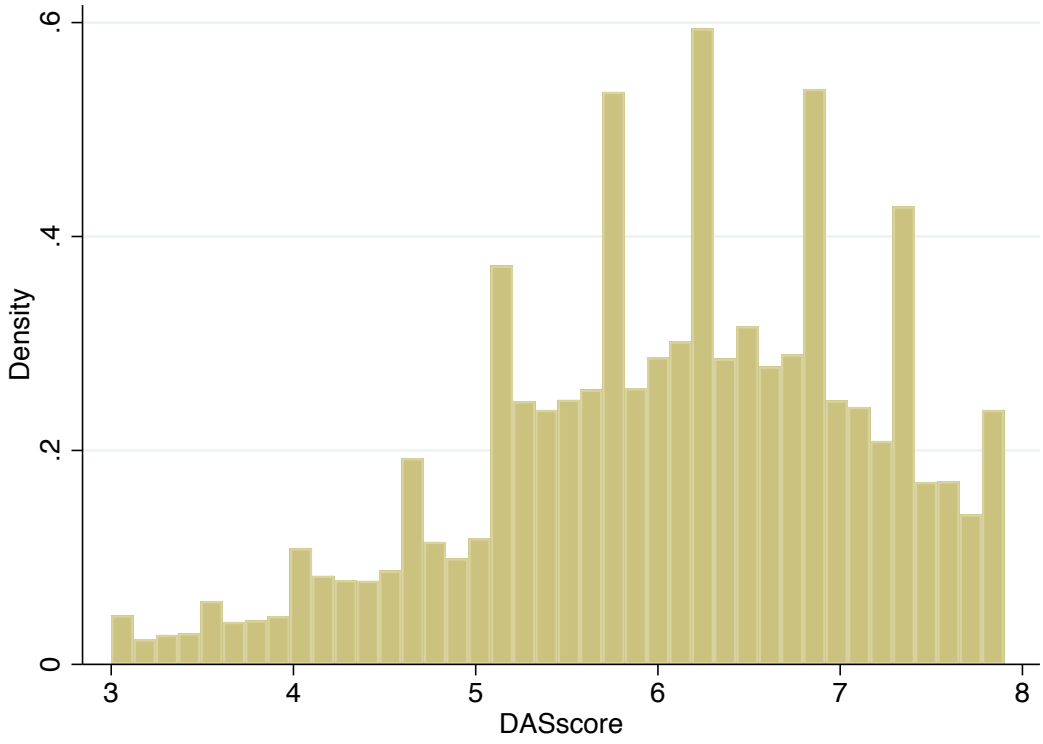
Appendix Figure 6.1: empirical power to detect a relative 34% reduction in outcome, where mean pre-intervention incidence is either 8% or 20%: by number of timepoints and mean sample size per timepoint



Appendix Figure 7.1: Examples of distributions of variables with missing data, by imputation (0 = original data): (A) Age, (B) Smoking, (C) Disease Duration, (D) DAS28, (E) HAQ score, (F) SF36: general



Appendix figure 7.2: TNFi >5.1 treatment threshold and (A) number of TNFi prescriptions per unit of DAS28, (B) probability (relative frequency) of TNFi prescriptions per unit of DAS28, (C) number of csDMARD prescriptions per unit of DAS28)



Appendix figure 7.3: Histogram for DAS28 amongst both TNFi and csDMARD patients