Published in final edited form as:

Curr Opin Rheumatol. 2013 May ; 25(3): . doi:10.1097/BOR.0b013e32835fd8c0.

New Methods for Determining Comparative Effectiveness in RA

Huifeng Yun, PhD¹ and Jeffrey R Curtis, MD MS MPH¹

¹University of Alabama at Birmingham; Birmingham, AL

Abstract

Purpose of review—To provide an overview of recently-published articles describing or applying newer methods for evaluating comparative effectiveness research (CER) in Rheumatoid Arthritis (RA).

Recent findings—Historically, clinical trials in RA have compared newer therapies against placebo. Newer trials designs designed to increase the relevance of trial results to real-world settings include head-to-head comparisons, some that incorporate non-inferiority, factorial and crossover designs. Extensions of traditional meta-analysis through network meta-analysis can combine direct and indirect evidence together and can compare multiple treatments to each other.

Observational data used to support CER include disease registries, administrative claims data, and electronic medical records. Pooling and linking across these data sources and applying newer epidemiologic methods to analyze such data can provide more valid inferences regarding optimal treatment regimens for RA.

Summary—CER methods in RA include head-to-head clinical trials, advanced techniques to summarize and aggregate data across studies, enrich the data available in observational settings, and enhance the methods used for analysis. Efforts to continue to apply and improve these methodologies will address key needs of clinicians, patients, and health policy decision-makers to generate evidence regarding real-world risks and benefits.

Keywords

Comparative effectiveness research; rheumatoid arthritis; clinical trials; meta-analysis; observational studies

Introduction

Comparative effectiveness research (CER) seeks to inform health-care decisions by providing evidence of the effectiveness, benefits, and harms of different treatment options (1). The management of RA has progressed dramatically in recent years through the development of multiple efficacious biologic and non-biologic agents. However, few comparative data are available to allow clinicians, patients, and other stakeholders to assess the risks and benefits of various therapies to guide optimal decisions for individual patients (2). Further research to examine the comparative effectiveness of different biologic therapies for RA was listed as one of the top 25 priority research topics by the Institute of Medicine (IOM) (3).

A number of methodologies are available to examine CER questions in RA and other chronic diseases(4) These include (1) head-to-head clinical trials, (2) synthesis of clinical

trials through meta-analyses or related techniques; and (3) studies using observational data. In this review, methodologies that have been or might be used for CER in RA will be described, drawing whenever possible on examples from recent arthritis clinical trials, meta-analyses and observational studies.

Clinical trials in RA

Traditional RCTs typically have used placebos as comparison groups and excluded clinically relevant patient subgroups like older patients and those with high comorbidities (5), so most of them are not adequate to answer needed real-world questions posed by the IOM. Head-to-head RCTs play a central role in CER and are generally the most well-accepted method of producing evidence for CER (6). However, a recent literature review showed that only 5 trials involved a head-to-head comparison of biologic DMARDs among 91 randomized controlled trials of biologic DMARDs ended after 2002 (7). The updated Agency for Healthcare Research and Quality (AHRQ) report summarized the findings of head-to-head studies on disease activity and remission for biologic DMARDs (4). (Table 1)

New head-to-head clinical trials in RA

The AMPLE (Abatacept versus Adalimumab comparison in Biologic Naïve Subjects with Background Methotrexate) study is one of the first head-to-head, non-inferiority studies of biologics in RA patients. It compared the efficacy and safety of subcutaneous abatacept vs adalimumab, both with background methotrexate. A key design element is that it was powered to demonstrate non-inferiority at a 12% non-inferiority margin, meaning that the lower bound of the 95% confidence interval had to exclude 12% in order for abatacept to be considered non-inferior to adalimumab. This study found that the ACR20 response rates were 64.8% for abatacept and 63.4% for adalimumab, a difference estimate of 1.8% (95% CI -5.6 – 9.2) (8), indicating that abatacept was non-inferior to adalimumab. This design must establish a non-inferiority margin (a difference) that must be defended on clinical grounds, and guidelines for choosing this difference are not consistent. Non-inferiority margins in the 10–12% range have been used in some past studies and have been accepted by the FDA and EMEA (9). Moreover, non-inferiority trials are generally large (like AMPLE, with a size of 646 patients), usually larger than a trial designed to show superiority, because they must exclude a small degree of inferiority compared to an active comparator drug.

In contrast, the ADACTA (Adalimumab and ACTAmra) trial was designed to show the superiority of mono-therapy with tocilizumab vs adalimumab for RA. As a superiority study, its sample size was substantially smaller ($n=325$) than it in AMPLE trial. The results of this study showed that TCZ was superior to ADA in the change in DAS28 (-3.3 vs. -1.8 , $p < 0.001$). Because manufacturers typically do not make available to investigators matching placebo versions of their products (10), blinding the administration of parenteral TCZ and subcutaneous ADA was one particular challenge for ADACTA trial (11). In addition to a blinded clinical assessor, patients in this study were blinded by placing a physical barrier (e.g. towel, pillow) between the patient's field of vision and the site of injection; investigators were further masked to the lab value of C-reactive protein, which was felt to have the potential to unblind the trial. These requirements for blinding pose particular challenges in future head-to-head studies of subcutaneous products where manufacturers generally do not make available blinded placebo in identical appearing syringes.

Used a two-by-two factorial design Treatment of Early Aggressive Rheumatoid Arthritis (TEAR) trial (12) randomly assign patients to 1 of 4 treatment arms: immediate MTX plus etanercept, immediate oral triple therapy (MTX plus sulfasalazine plus hydroxychloroquine), or step-up from MTX monotherapy to one of the two combination

therapies at week 24. The factorial design tests the marginal effects of each treatment strategy and requires a smaller sample size than a typical 4-arm parallel group design(11). Based on DAS28-ESR change from week 48 to week 102, this study found no clinical differences between the 2 treatment approaches. Similarly, RACAT (**R**heumatoid **A**rthritis: **C**omparison of **A**ctive **T**herapies) trial also found the strategy of MTX plus etanercept has comparable effectiveness with MTX plus sulfasalazine plus hydroxychloroquine . RACAT trial used a cross-over design to switch patients for those who had failed their initial treatments at 24 weeks (13). Factorial designs (TEAR), as well as crossover designs (RACAT) allow patients to receive alternative treatment options rather than simply drop out of the study, but at the cost of added complexity and some challenges in interpretation of results.

N-of-1 trials have been suggested to help decide on optimal treatments for patient when the treatments effects are rapid. N-of-1 trials are single-patient experiments that incorporate multiple crossover periods where each patient receives a randomized sequence of treatments. At least two studies have applied this design to an RA population (14, 15). Kravitz et suggested that this study design may offer “a better way to balance clinical judgment, consumer choice, and cost” after comparing different access models to expensive biologic therapies for the treatment of RA (15). Acceptability to patients and clinicians, as well as assumptions about rapidity of onset and washout between treatment periods, may pose some limitations in applying this trial design to RA. Nevertheless, despite a paucity of real-world applied examples in RA, N-of-1 trials remain a potentially promising study design option for some treatment regimens.

Cluster controlled trials

Cluster-controlled trials (CCT) are designed to evaluate interventions that operate at multiple levels and are used increasingly in public health related research, such as hospital- and physician-based studies. The rationale for a cluster-based trial is that an intervention is either directed at the physician or could influence physician behavior to such a degree that it would compromise the study’s results. Several CER related CCT designs have been applied among patients with RA including randomized cluster trials, non-randomized but controlled cluster trials and multi-level cluster trials. Examples of these types of interventions in RA include strategy trials (e.g. a treat-to-target trial), and a quality improvement trial.

The Consortium of Rheumatology Researchers of North America, Inc (CORRONA) has initiated a treat to target (T2T) trial, a multi-center cluster randomized trial in adults with RA(16). This study is among the first in U.S. to compare the effectiveness of the T2T approach for the care of RA versus usual care. T2T requires frequent measurement of disease activity and prescribed treatment changes. The requirement for a cluster trial is that intervention physicians likely would become accustomed to treating aggressively for intervention patients and control patients alike. Therefore, participating sites (i.e. clusters) rather than individual patients were randomized to the two treatments arms. Because measures collected from individuals of the same group tend to be more similar than individuals from different groups, the outcomes of patients from the same participating site are likely to be correlated, often measured by a non-zero intra-class correlation (ICC). This correlation will require a larger sample size and must be considered in designing such studies .

Desai, et al conducted a cluster controlled, non-randomized trial that allocated physicians to intervention or control groups based on the volume of patients in their practices (17). Fourteen rheumatologists were selected to receive a point-of-care paper reminder form for patients who did not have pneumococcal vaccination, whereas the other 21 rheumatologists were selected not to receive the reminder form and served as controls. Although the non-

random design was a limitation of this study, intervention physicians had improvement in their patients being up-to-date with respect to pneumococcal vaccination (68 to 80%, $p = 0.006$) whereas control physicians had no appreciable change in the same time frame (52% in both periods, $p = 0.90$)(17).

Some study designs may deploy interventions at the patient level but need to account for the practices of their treating physician. Warriner and colleagues conducted a multi-level randomized trial to identify an intervention to improve osteoporosis screening with dual energy x-ray absorptiometry (DXA) by offering patients the opportunity to self-schedule a DXA test. The trial included but was not limited to RA patients. Patients were randomized within physician practices and the intervention patients were compared to control patients within each practice. The results showed that mailing a simple educational osteoporosis brochure and allowing self-scheduling of a DXA scan significantly improve osteoporosis screening rates (18). This study design that stratifies randomization within physician practices is useful when strong provider-level effects are expected and may be used in future studies, especially quality improvement studies, where the focus of interventions is not only on patients but also on physicians, allied health care providers (e.g. pharmacists), and health system-related factors and barriers to care.

Large Pragmatic trials

Compared to traditional clinical trials, pragmatic clinical trials (PCTs) are conducted in “real-world” situations to compare the effectiveness of several treatments or treatment strategies. Strengths include randomization, efficiency because of modest primary data collection, minimal monitoring burden, and suitability for safety assessment (19). The efficiency of conducting PCTs in the U.S. was recently summarized (20) and key design elements included 1) using community practices and engaging Practice Based Research Networks (PBRNs) to improve efficiency, effectiveness and generalizability of patient and provider site recruitment; 2) simplifying enrollment procedures and data collection requirements, possibly through Internet technology using tablet PCs; and 3) employing administrative data linkages to identify selected outcomes as key endpoints of the trial (especially safety outcomes), thus reducing cost and measurement burden for long-term follow-up. Administrative data include claims for medical services and prescription drugs, and including these linkages greatly simplifies collection of selected outcomes, important comorbidities, medication usage and drug adherence.

Application of these new technologies was formative in the development of the Varicella Zoster Vaccine (VERVE), a large pragmatic safety trial designed with support of the American College of Rheumatology Within Our Reach Program. A tablet PC is used to guide screening and facilitate patient-based data collection. Multiple health plan and registry databases were linked together to demonstrate trial feasibility and to assist in assessing long-term outcomes. RA patients were identified on the basis of physician diagnoses and receipt of anti-TNF therapy. Although the primary safety outcomes are clinical and immunologic, follow-up for longer-term effectiveness outcomes to reduce herpes zoster risk is facilitated through a linkage with administrative claims data (21). Trials conducted with safety outcomes in mind, as well as pharmacovigilance efforts, may be especially benefitted by such linkages between a clinical trial (or registry) and health plan or administrative data.

Although not an RA-focused trial, the Post-Myocardial Infarction Free Rx Event and Economic Evaluation (MI FREEE) trial is another example of a large pragmatic trial. After discharge for myocardial infarction, participants were randomly assigned to full prescription coverage or usual prescription coverage. Patients in the full-coverage group had no cost sharing for any branded or generic statin, beta blocker, angiotensin-converting-enzyme (ACE) inhibitor, or angiotensin-receptor blocker (ARB) for every prescription. As the

randomization happened at the level of plan sponsor, all eligible beneficiaries in a plan received the same coverage. Using linked claims for filled prescriptions, procedures, physician encounters, hospitalizations and inpatient deaths, the investigators assessed outcomes (e.g. MI) using validated algorithms and medication adherence by using pharmacy data (22). As adherence to RA medications may be in part related to medication costs, a trial designed to vary out-of-pocket patient costs of RA medications to improve adherence and outcomes could be used in future RA-related CER as a potential mechanism to overcome disparities in use of non-biologic and biologic DMARDs.

Synthesis of existing clinical information

Traditional systematic reviews and meta-analyses are often used to summarize the effects of an intervention against a single comparator group, usual placebo. Since direct evidence from head-to-head comparison trials for RA medications is often unavailable, indirect comparison and network comparison were developed to provide more sophisticated methods for comparisons across several interventions (23, 24). Indirect comparisons allows the synthesis of data using the relative effects of the several treatments versus the common comparator. For example, a trial comparing treatment A versus placebo and another trial comparing treatment B versus placebo allow indirect comparison of treatment A versus B. Network meta-analysis, sometimes used interchangeably with 'mixed treatment comparisons' when both direct and indirect evidence is available, is an extension of conventional meta-analysis and can incorporate multiple pairwise comparisons across a range of interventions and provide estimates of relative treatment effect (25). Further variations on this method include mixed treatment comparisons using Bayesian approaches that seek to combine evidence from a-priori information and enhance confidence in the process of decision making in CER analysis; published examples are available in RA (26).

Observational Research to Address CER in RA

Observational studies offer the ability to study outcomes of RA treatments in routine medical practice and fill critical gaps in RA CER.(2). Observational data sources to answer CER questions in RA include cohorts/registries, administrative claims and electronic medical records, each with strengths and limitations (Table 2). CORRONA, the National Data Bank(NDB) for Rheumatic Diseases, the American College of Rheumatology's electronic health record-based Rheumatology Informatics System for Effectiveness (RISE) (27) and several large European registries in the U.K., Sweden, Germany and Spain and other countries are examples of RA cohorts and registries that could be used for CER. Each of these data sources, as well as many others, are well-positioned to contribute important comparative effectiveness data to the rheumatology community. Several examples from these registries directly comparing biologics to one another and identifying predictors of response have been recently published (28–30).

Nested Cohorts and Learning Networks

As a variation on a large disease registry, a nested cohort inside that registry can incorporate features of RCTs to study CER. For example, the recently launched CORRONA Comparative Effectiveness Registry to study Therapies for Arthritis and Inflammatory Conditions (CERTAIN) study uses the existing CORRONA network of participating sites to recruit new biologic users among RA patients. CERTAIN explicitly recruits patients based upon pre-defined explicit inclusion/exclusion criteria and establishes a new time axis for data collection based upon patients starting or switching therapies. Other design features include systematic collection of RA disease activity measures at standardized intervals that are different from the parent cohort, a biospecimen repository, centrally run laboratory

results, linkages with multiple external data sources, and on-site monitoring for quality control.

As another example of a CER effort nested inside a disease registry, the Childhood Arthritis Rheumatology and Research Alliance (CARRA) network has developed consensus-based treatment protocols to understand the comparative effectiveness of diverse therapeutic options available for systemic juvenile idiopathic arthritis (JIA) and other conditions(31). These protocols were designed to use best-available evidence to minimize unwarranted heterogeneity in how patients are treated in order to generate comparative effectiveness data. Although currently deployed in an observational setting, randomization to one of these treatment protocols for specific patient types would be possible if needed to minimize confounding.

Administrative health plan and electronic health record data

Complementary to data from a disease registry, administrative claims data provided by governmental sources (e.g. the U.S. Center for Medicare and Medicaid Services) and commercial insurance companies includes information on diagnoses, treatments, prescription drug use, and outcome events for millions of beneficiaries. They are useful data sources for evaluation the CER of RA medications, especially with respect to studying rare safety events where large numbers of patients are needed.

Electronic medical records (EMR) provide phenotypically richer information and include clinical details, medical history, patients' symptoms, the results of physical examinations and diagnostic tests (32). The U.S. Veterans Health Administration health care system and Kaiser Permanente (a large U.S. managed care organization) are the best examples of large health care systems and provider network sharing the same electronic medical record system that could be used for research.

Combining data for CER analyses

Although the largest registries and health plan datasets contain huge amount of information, each of them by themselves still may be too small to address key questions regarding rare exposures or outcomes. Aggregating data across different data systems is therefore potentially useful. However, sharing of raw person-level data between data systems may be impractical or even impossible. As one example of a mechanism that pooled person-level data, projects can adopt common, standardized data formats across datasets to allow for pooling of masked, person-level data. A recently published analysis examining biologic safety across a range of outcomes pooled four independent datasets together using such a distributed data model; individual identifiers were masked using propensity scores (33, 34). The U.S. Mini-Sentinel project uses a similar distributed data model (35)

Pooling and linking data sources together

Because administrative data typically lack detailed clinical information, linking a RA registry or cohort with other data sources may allow for more valid assessment of exposures, confounders, and outcomes. Standing linkages in some European countries between an RA cohort and a separate data source (e.g. cancer registry) provide such examples. While these data sources are most easily linked through having unique identifier, they may also be linked using multiple non-unique identifiers. More recently, several novel linkages between an RA registry and health plan data have been described. For example, Veterans Affairs (VA) RA registry (VARA) data have been collected by rheumatologists at 11 VHA facilities throughout the United States since 2003. With unique identifiers and patient consent, participants in VARA have been linked to both Veterans Health Administration's administrative databases (e.g. hospitalizations, pharmacy refills) as well as VHA EHR data.

This type of linkage can facilitate CER in multiple ways; one recent example derived and validated a claims-based algorithm to determine the effectiveness of RA medications and compared it to a clinical gold standard (DAS28) available in the registry. (36) This approach has been subsequently applied in very large datasets containing tens of thousands of RA patients to study the comparative effectiveness of various biologics for RA (37). In addition, using a deterministic linkage on multiple, non-unique identifiers, the CORRONA registry has been linked with U.S. Medicare data to collect information regarding medication refills, hospitalizations, treatment-associated costs, and other health services utilization (38).

New methodologic approaches to analyzing CER data

Channeling of patients with more severe or active disease is an important issue in validly analysis of observational data (39). Propensity scores (PS) may be useful to control confounder and have been frequently used for observational studies of RA medications, especially when studying rare outcomes. On expectations, PS balance observed covariates between exposed and comparator-treated patients, although PS do not balance unmeasured confounders (40). Variations on propensity scores may prove useful to extend this methodology. **High Dimensional Propensity Scores** (hd-PS) allow for semi-automated variable selection using administrative data and may offer equal or better control of confounding compared to investigator-driven covariate selection (41). When exposures are relatively rare (e.g. introduction of a new medication) and when outcomes are more common, or when there are multiple treatment groups to be compared, disease risk scores may serve a similar function to propensity scores and balance confounders by collapsing them into a single score, interpreted as the predicted risk for an outcome in the absence of the exposure of interest (42). RA-specific disease risk scores have been derived and validated to estimate the risk of serious infections using clinical and administrative data (43, 44). When time-depending confounding is particularly problematic, **Marginal Structural Models (MSMs)** can be used to estimate the causal effect of a time-dependent exposure in the presence of time-dependent confounders affected by previous treatments. This approach has been applied to study the safety of RA medication in a longitudinal cohort where complex relationships in changes in glucocorticoid dose, disease activity, functional status, and medication discontinuation can be disentangled (45). Finally, **Instrumental Variable (IV)** analysis is a method borrowed from health economics intended to reduce bias introduced by measured and unmeasured confounding. When unmeasured characteristics potentially bias results, instrumental variable analysis may provide a less biased estimate of treatment effects. While instrumental variables (IV) has traditionally been limited to comparing only two treatments/interventions concurrently, a new approach has been developed to estimate multiple treatment effects (46, 47).

Conclusion

This article reviews recent methods likely to prove useful for addressing comparative effectiveness research questions in RA. These methods include head-to-head trials using active comparison groups, cluster controlled trials comparing the effectiveness of interventions that operate at a group level when the allocation of RA individual participants may not be desirable, large pragmatic trial using internet methods, network meta-analysis and related techniques to aggregate and summarize results. Variations on traditional disease-based registries include nesting subcohorts with more rigorous and standardized assessment and interventions, pooling and linking data sources together, and application of new analytic methods to control for confounding in observational studies. With these new methods, comparative effectiveness research can help clinicians, patients, and other stakeholders to assess the risks and benefits of various RA therapies to guide optimal decisions for individual patients.

References and recommended reading

1. Agency for Healthcare Research and Quality. [assessed on 12/14/12] Effective Health Care Program. <http://effectivehealthcare.ahrq.gov/index.cfm/what-is-comparative-effectiveness-research1/>.
2. Schneeweiss S, Gagne JJ, Glynn RJ, Ruhl M, Rassen JA. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. *Clinical pharmacology and therapeutics*. 2011; 90(6):777–790. Epub 2011/11/04. [PubMed: 22048230]
3. Institute of Medicine. Initial National Priorities for Comparative Effectiveness Research. Washington, DC: Institute of Medicine; 2009.
4. Donahue, KE.; Jonas, DE.; Hansen, RA.; Roubey, R.; Jonas, B.; Lux, LJ., et al. Drug Therapy for Rheumatoid Arthritis in Adults: An Update. Rockville (MD): 2012.
5. Greenberg JD, Kishimoto M, Strand V, Cohen SB, Olenginski TP, Harrington T, et al. Tumor necrosis factor antagonist responsiveness in a United States rheumatoid arthritis cohort. *The American journal of medicine*. 2008; 121(6):532–538. Epub 2008/05/27. [PubMed: 18501236]
6. Luce BR, Kramer JM, Goodman SN, Connor JT, Tunis S, Whicher D, et al. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Ann Intern Med*. 2009; 151(3):206–209. Epub 2009/07/02. [PubMed: 19567619]
7. Estellat C, Ravaud P. Lack of head-to-head trials and fair control arms: randomized controlled trials of biologic treatment for rheumatoid arthritis. *Archives of internal medicine*. 2012; 172(3):237–244. Epub 2012/02/15. [PubMed: 22332155]
8. Weinblatt ME, Schiff M, Valente R, van der Heijde D, Citera G, Zhao C, et al. Head-to-head comparison of subcutaneous abatacept versus adalimumab for Rheumatoid Arthritis. *Arthritis Rheum*. 2012 Epub 2012/11/22. . ** This study is one of the first head-to-head, non-inferiority studies of biologics in RA patients. It compared the efficacy and safety of subcutaneous abatacept vs adalimumab, in combination with background methotrexate.
9. D'Agostino RB Sr, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med*. 2003; 22(2):169–186. Epub 2003/01/10. [PubMed: 12520555]
10. Kavanaugh A, Emery P, van Vollenhoven RF, Dikranian AH, Alten R, Klearman M, et al. Tocilizumab Monotherapy Compared with Adalimumab Monotherapy in Patients with Rheumatoid Arthritis: Results of a 24-Week Study. *Arthritis & Rheumatism*. 2012; 64(10 (Suppl)):S333–S334.
11. Peduzzi P, Kyriakides T, O'Connor TZ, Guarino P, Warren SR, Huang GD. Methodological issues in comparative effectiveness research: clinical trials. *The American journal of medicine*. 2010; 123(12 Suppl 1):e8–e15. Epub 2011/01/12. [PubMed: 21184867]
12. 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 Rheum*. 2012; 64(9):2824–2835. Epub 2012/04/18. [PubMed: 22508468] . * A two year follow up study used a two- by- two factorial design to randomly assign subjects to 1 of 4 treatment arms: immediate treatment with MTX plus etanercept, immediate oral triple therapy (MTX plus sulfasalazine plus hydroxychloroquine), or step-up from MTX monotherapy to one of the combination therapies at week 24.
13. O'Dell JR, Mikuls TR, Taylor T, Ahlumaia V, Brophy M, Warren S, et al. Rheumatoid Arthritis Comparison of Active Therapies in Methotrexate Suboptimal Responders: Validation of the Strategy of Conventional Disease Modifying Anti-Rheumatic Drugs Before Biologicals. *Arthritis & Rheumatism*. 2012; 64(10 (supplement)):S551.
14. Pincus T, Huizinga TW, Yazici Y. N-of-1 trial of low-dose methotrexate and/or prednisolone in lieu of anti-CCP, MRI, or ultrasound, as first option in suspected rheumatoid arthritis? *J Rheumatol*. 2007; 34(2):250–252. Epub 2007/02/17. [PubMed: 17304647]
15. Kravitz RL, Duan N, White RH. N-of-1 trials of expensive biological therapies: a third way? *Archives of internal medicine*. 2008; 168(10):1030–1033. Epub 2008/05/28. [PubMed: 18504330]

16. Agency for Healthcare Research and Quality. Registries for Evaluating patient Outcomes: A User's Guide. 2012
17. Desai SP, Lu B, Szent-Gyorgyi LE, Bogdanova AA, Turchin A, Weinblatt M, et al. Increasing pneumococcal vaccination for immunosuppressed patients: A cluster quality improvement trial. *Arthritis Rheum.* 2012 Epub 2012/10/10. . * A cluster controlled, non-randomized trial that allocated rheumatologists to intervention or control groups based on the volume of patients in their practices.
18. Warriner AH, Outman RC, Kitchin E, Chen L, Morgan S, Saag KG, et al. A randomized trial of a mailed intervention and self-scheduling to improve osteoporosis screening in postmenopausal women. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research.* 2012; 27(12):2603–2610. Epub 2012/07/28. [PubMed: 22836812]
19. Chalkidou K, Tunis S, Whicher D, Fowler R, Zwarenstein M. The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clinical trials.* 2012; 9(4):436–446. Epub 2012/07/04. [PubMed: 22752634]
20. Saag KG, Mohr PE, Esmail L, Mudano AS, Wright N, Beukelman T, et al. Improving the efficiency and effectiveness of pragmatic clinical trials in older adults in the United States. *Contemporary clinical trials.* 2012; 33(6):1211–1216. Epub 2012/07/17. [PubMed: 22796098] . * This study summarized novel approaches to the design of pragmatic trials that can improve the efficiency of conducting comparative effectiveness research.
21. Curtis JR, Chen L, Xie F, Zhang J, Saag KG, Cofield S, et al. Use of Health Plan Data to Assess Feasibility of Large Pragmatic Clinical Trials in Rheumatoid Arthritis. *Arthritis & Rheumatism.* 2012; 64(10 (Supl)):S21.
22. Choudhry NK, Avorn J, Glynn RJ, Antman EM, Schneeweiss S, Toscano M, et al. Full coverage for preventive medications after myocardial infarction. *N Engl J Med.* 2011; 365(22):2088–2097. Epub 2011/11/15. [PubMed: 22080794]
23. Thorlund K, Mills EJ. Sample size and power considerations in network meta-analysis. *Systematic reviews.* 2012; 1(1):41. Epub 2012/09/21. [PubMed: 22992327]
24. Mills EJ, Bansback N, Ghement I, Thorlund K, Kelly S, Puhan MA, et al. Multiple treatment comparison meta-analyses: a step forward into complexity. *Clinical epidemiology.* 2011; 3:193–202. Epub 2011/07/14. [PubMed: 21750628]
25. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2011; 14(4):417–428. Epub 2011/06/15. [PubMed: 21669366]
26. Schmitz S, Adams R, Walsh CD, Barry M, FitzGerald O. A mixed treatment comparison of the efficacy of anti-TNF agents in rheumatoid arthritis for methotrexate non-responders demonstrates differences between treatments: a Bayesian approach. *Ann Rheum Dis.* 2012; 71(2):225–230. Epub 2011/10/01. [PubMed: 21960560]
27. Embi PJ, Barnes I, Myslinski R, Ervin W, Stevens W, Borlawsky T, et al. The Rheumatology Informatics System for Effectiveness (RISE): Enabling Data Access Across Disparate Sites for Quality Improvement and Research. *Arthritis & Rheumatism.* 2012; 64(10 (supplement)):S1101–S1102.
28. Greenberg JD, Reed G, Decktor D, Harrold L, Furst D, Gibofsky A, et al. A comparative effectiveness study of adalimumab, etanercept and infliximab in biologically naive and switched rheumatoid arthritis patients: results from the US CORRONA registry. *Annals of the rheumatic diseases.* 2012; 71(7):1134–1142. Epub 2012/02/02. [PubMed: 22294625]
29. Soliman MM, Hyrich KL, Lunt M, Watson KD, Symmons DP, Ashcroft DM, et al. Rituximab or a second anti-tumor necrosis factor therapy for rheumatoid arthritis patients who have failed their first anti-tumor necrosis factor therapy? Comparative analysis from the British Society for Rheumatology Biologics Register. *Arthritis Care Res (Hoboken).* 2012; 64(8):1108–1115. Epub 2012/03/17. [PubMed: 22422731]
30. Soliman MM, Hyrich KL, Lunt M, Watson KD, Symmons DP, Ashcroft DM, et al. Effectiveness of rituximab in patients with rheumatoid arthritis: observational study from the British Society for

- Rheumatology Biologics Register. *J Rheumatol*. 2012; 39(2):240–246. Epub 2011/12/17. [PubMed: 22174201]
31. DeWitt EM, Kimura Y, Beukelman T, Nigrovic PA, Onel K, Prahalad S, et al. Consensus treatment plans for new-onset systemic juvenile idiopathic arthritis. *Arthritis Care Res (Hoboken)*. 2012; 64(7):1001–1010. Epub 2012/02/01. [PubMed: 22290637]
 32. Wasserman RC. Electronic medical records (EMRs), epidemiology, and epistemology: reflections on EMRs and future pediatric clinical research. *Academic pediatrics*. 2011; 11(4):280–287. Epub 2011/05/31. [PubMed: 21622040]
 33. Herrinton LJ, Curtis JR, Chen L, Liu L, Delzell E, Lewis JD, et al. Study design for a comprehensive assessment of biologic safety using multiple healthcare data systems. *Pharmacoepidemiol Drug Saf*. 2011; 20(11):1199–1209. Epub 2011/09/16. [PubMed: 21919113] .
* This study demonstrated methods to pool masked, person-level individual data across different data systems to conduct large-scale comparative effectiveness analyses.
 34. Grijalva CG, Chen L, Delzell E, Baddley JW, Beukelman T, Winthrop KL, et al. Initiation of tumor necrosis factor-alpha antagonists and the risk of hospitalization for infection in patients with autoimmune diseases. *JAMA : the journal of the American Medical Association*. 2011; 306(21): 2331–2339. Epub 2011/11/08. [PubMed: 22056398]
 35. Mini-Sentinel. Data Activities http://mini-sentinel.org/data_activities/.
 36. Greenberg JD, Kremer JM, Curtis JR, Hochberg MC, Reed G, Tsao P, et al. Tumour necrosis factor antagonist use and associated risk reduction of cardiovascular events among patients with rheumatoid arthritis. *Ann Rheum Dis*. 2011; 70(4):576–582. Epub 2010/11/27. [PubMed: 21109516]
 37. Yun H, Xie F, Delzell E, Chen L, Yang S, Saag K, et al. The Comparative Effectiveness of Anti-TNF Medications Among Older and Disabled Rheumatoid Arthritis Patients in the U.S. Medicare Population. *Arthritis & Rheumatism*. 2012; 64(10 (suppl)):S213–S214.
 38. Curtis JR, Chen L, Beukelman T, Bharat A, Xie F, Saag KG, et al. Methods to Link a U.S. Arthritis Cohort with Medicare Administrative Claims Data. *Arthritis & Rheumatism*. 2012; 64(10 (Supplement)):S401.
 39. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *Journal of clinical epidemiology*. 2005; 58(4):323–337. Epub 2005/05/03. [PubMed: 15862718]
 40. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic & clinical pharmacology & toxicology*. 2006; 98(3):253–259. Epub 2006/04/14. [PubMed: 16611199]
 41. Rassen JA, Schneeweiss S. Using high-dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system. *Pharmacoepidemiol Drug Saf*. 2012; 21(Suppl 1):41–49. Epub 2012/01/25. [PubMed: 22262592]
 42. Glynn RJ, Gayne JJ, Schneeweiss S. Role of disease risk scores in comparative effectiveness research with emerging therapies. *pharmacoepidemiology and drug safety*. 2012; 21(S2) 238–147.
 43. Curtis JR, Xie F, Chen L, Muntner P, Grijalva CG, Spettell C, et al. Use of a disease risk score to compare serious infections associated with anti-tumor necrosis factor therapy among high- versus lower-risk rheumatoid arthritis patients. *Arthritis Care Res (Hoboken)*. 2012; 64(10):1480–1489. Epub 2012/07/27. [PubMed: 22833479]
 44. Crowson CS, Hoganson DD, Fitz-Gibbon PD, Matteson EL. Development and validation of a risk score for serious infection in patients with rheumatoid arthritis. *Arthritis Rheum*. 2012; 64(9): 2847–2855. Epub 2012/05/12. [PubMed: 22576809]
 45. Strangfeld A, Eveslage M, Schneider M, Bergerhausen HJ, Klopsch T, Zink A, et al. Treatment benefit or survival of the fittest: what drives the time-dependent decrease in serious infection rates under TNF inhibition and what does this imply for the individual patient? *Ann Rheum Dis*. 2011; 70(11):1914–1920. Epub 2011/07/28. [PubMed: 21791449]
 46. Fang G, Brooks JM, Chrischilles EA. Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. *American journal of epidemiology*. 2012; 175(1):60–65. Epub 2011/11/17. [PubMed: 22085626]

47. Kawatkar AA, Hay JW, Stohl W, Nichol MB. Incremental Expenditure of Biologic Disease Modifying Antirheumatic Treatment Using Instrumental Variables in Panel Data. *Health economics*. 2012 Epub 2012/06/22.

Key points

New designs of head-to-head randomized clinical trials using active comparison groups provide central evidence for comparative effectiveness research in RA.

Cluster-randomized trials are increasingly used to compare the effectiveness of interventions that operate at a group level.

Indirect and network comparisons provide sophisticated methods for comparisons of RA medications when head-to-head trials are not available.

Observational studies have broad application for comparative effectiveness in RA with real-world non-experimental data.

Table 1

Comparative effectiveness research on biologic DMARDs versus biologic DMARDs for disease activity and remission.

Study	Study Design N Duration	Study Population	Comparison (dose)	Results
Abatacept vs. Infliximab				
Schiff, et al., 2008 ATTEST study	RCT 431 1 year	Patient who have failed MTX; active RA; mean disease duration 7.3–8.4 years	ABA (~10mg/kg every 4 weeks) vs. INF (3mg/kg every 8 weeks)	Greater decrease in DAS28 for ABA vs. INF at 1 year (estimate of difference –0.62 95% CI, –0.96 to –0.29) Trend toward greater remission for ABA vs. INF at 1 year (estimate of difference 6.5%; 95% CI, –2.2 to 15.2, <i>P</i> =NS, NR)
Hetland et al., 2010 DANBIO	Prospective cohort 2,326 6 months 48 months	Patients with RA initiating therapy with biologic DMARDs	ADA (mean: 40 mg) wks) vs. ETN (mean 45 mg) vs. INF (mean 3.5 mg/kg)	Higher achievement of ACR 70 response after 6 months for ADA than INF (OR 2.05, 95% CI, 1.52 to 2.76, <i>P</i> =NR)
Kievit et al., 2008	Prospective cohort 707 1 year	Population-based, Netherlands, Anti-TNF naïve patients, failed at least 2 DMARDs; mean disease duration 6–7.7 years	ADA vs. INF (dosages NR)	Greater decrease in DAS28 for ADA vs. INF at 1 year (ADA – 1.8, INF –1.2; <i>P</i> <0.05)
Adalimumab vs. Etanercept				
Hetland et al, 2010 DANBIO	Prospective cohort 2,326 6 months 48 months	Patients with RA initiating therapy with biologic DMARDs	ADA (mean: 40 mg) wks) vs. ETN (mean 45 mg) vs. INF (mean 3.5 mg/kg)	No difference in achievement of ACR 70 response after 6 months for ADA vs. ETN (OR 1.15, 95% CI, 0.82 to 1.60, <i>P</i> =NR)
Etanercept vs. Infliximab				
Geborek et al., 2002	Nonrandomized, open-label trial 369 1 year	Population-based; active RA; had failed at least 2 DMARDs; mean disease duration 14.5 years	ETN (25 mg twice weekly) vs. INF (3 mg/kg or higher)	Higher ACR 20 responses for ETN at 3 (data NR; <i>P</i> <0.02) and 6 months (data NR; <i>P</i> =0.05); no significant differences in ACR response rates at 1 year (data NR)
Fernande z-Nebro et al., 2007	Prospective cohort 161 6 years	Tertiary care center, Spain; Anti-TNF naïve patients; mean disease duration 9.5– 9.9 years	ETN vs. INF (dosages NR)	Significantly greater decrease t in DAS28 at 6 months for ETN vs. INF (–1.7 vs/ –1.3; <i>P</i> =0.03); No difference in EULAR responses between ETN vs. INF at 6 months
Hetland et al., 2010 DANBIO	Prospective cohort 2,326 6 months 48 months	Patients with RA initiating therapy with biologic DMARDs	ADA (mean: 40 mg) wks) vs. ETN (mean 45 mg) vs. INF (mean 3.5 mg/kg)	Higher achievement of ACR 70 response after 6 months for ETN vs. INF (OR 1.78, 95% CI, 1.28 to 2.50, <i>P</i> =NR)

Study	Study Design N Duration	Study Population	Comparison (dose)	Results
Kievit et al, 2008	Prospective cohort 707 1 year	Population-based Netherlands, Anti-TNF naïve patients failed at least 2 DMARDs; mean disease duration 6–7.7 years	ETN vs. INF (dosages NR)	Greater decrease in DAS28 for ETN vs. INF at 1 year (ETN – 1.8, INF – 1.2; $P<0.05$)
Kristensen et al., 2002	Prospective cohort 949 3 years	Population-based inadequate response to at least 2 DMARDs	ETN (25 mg twice weekly) vs. INF (3 mg/kg or higher)	No difference in ACR 50 response at 3 years (data NR)
Weaver et al., 2006	Prospective cohort 1,371 1 year	Population-based; patients with active RA who required change in therapy; mean disease duration 9.3 years	ETN (25 mg twice weekly) vs. INF (3.8 mg/kg or higher)	Higher ACR 20 response rates for ETN than INF at 1 year (41% vs. 26%; $P=NR$)
Rituximab vs. Anti-tumor Necrosis Factor Therapies				
Finckh et al., 2007	Prospective cohort 116 6+ months	Population based, Switzerland; Patients who have failed at least 1 Anti- TNF treatment; mean disease duration 9–10 yrs.	RTX (2 infusions, 1000 mg) vs. Anti-TNF agent (INF, ETN or ADA, dosages NR)	Greater reduction in DAS28 at 6 months for RTX vs. Anti-TNF (–.6 vs. –0.98; $P=0.01$)

ACR = American College of Rheumatology; ADA = adalimumab; ATTEST = Abatacept or infliximab vs. placebo, a trial for tolerability, efficacy, and safety in treating rheumatoid arthritis; CI = confidence interval; DAS = disease activity score; DMARD = disease modifying antirheumatic drug; ETN = etanercept; INF = infliximab; mg/kg = milligram/kilogram; mACR = modified American College of Rheumatology; NR = not reported; NS = not significant; PCS = physical component score; RA = rheumatoid arthritis; TNF = tumor necrosis factor; US = United States; vs. = versus Source: adapted from Table 13 in Donahue KE, Jonas DE, Hansen RA, et al. Comparative effectiveness of drug therapy for rheumatoid arthritis in adults—update. AHRQ comparative effectiveness review No. 55. April 2012 (4).

Table 2

Strengths and limitations by types of study design and information source for arthritis research

	RCT	Observational Cohort/Registry	Claims Data	Indirect/Network Comparisons
Questions Optimized to Address	+++ Efficacy	± Effectiveness Safety Practice Patterns	± Effectiveness Safety Practice Patterns	Variable
Data Capture	+++ phenotype ++ completeness	+++ phenotype + completeness	± phenotype ++ completeness	++ phenotype +completeness
Internal Validity	+++	++	±±	++
Cost	+++	++	+	+
Generalizability	+	+++	+++	++