Published in final edited form as:

Nat Rev Rheumatol. 2015 July ; 11(7): 437–441. doi:10.1038/nrrheum.2015.30.

Active-comparator design and new-user design in observational studies

Kazuki Yoshida, M.D., M.P.H.^{1,2}, Daniel H. Solomon, M.D., M.P.H.^{1,3}, and Seoyoung C. Kim, M.D., Sc.D., M.S.C.E.^{1,3}

¹Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, United States

²Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States

³Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Boston, Massachusetts, United States

SUMMARY

Over the past decade, an increasing number of observational studies have examined the effectiveness or safety of rheumatoid arthritis treatments. However, unlike randomized controlled trials (RCTs), observational studies of drug effects face methodological challenges including confounding by indication. Two design principles - *active comparator design* and *new user design* can help mitigate such challenges in observational studies. To improve validity of study findings, observational studies should be designed in such a way that makes them more closely approximate RCTs. The *active comparator design* compares the drug of interest to another commonly used agent for the same indication, rather than a 'non-user' group. This principle helps select treatment groups similar in treatment indications (both measured and unmeasured characteristics). The *new user design* includes a cohort of patients from the time of treatment initiation, so that it can assess patients' pretreatment characteristics and capture all events occurring anytime during follow-up.

MeSH KEY WORDS

Rheumatoid Arthritis; Epidemiology; Pharmacoepidemiology; Cohort Studies

INTRODUCTION

Since the first introduction of tumor necrosis factor (TNF)- α inhibitors more than a decade ago, a growing number of biological disease-modifying antirheumatic drugs (bDMARDs) have become available for rheumatoid arthritis (RA). Consequently, the need for high-quality data on comparative effectiveness and safety of bDMARDs has increased. Observational studies such as cohort studies provide complementary information regarding

CORRESPONDING AUTHOR (and reprint request): Kazuki Yoshida, M.D., M.P.H., Address: Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, 75, Francis Street, Boston, Massachusetts, 02115, USA, kazuki.yoshida@mail.harvard.edu.

CONTRIBUTORSHIP: KY, DHS, and SCK planned the paper. KY conducted the literature search. All authors read and approved the final draft.

the effectiveness or safety of these drugs after their approval, as randomized controlled trials (RCTs) are often limited by lack of statistical power for non-primary endpoints, relatively short follow-up duration for assessing long-term safety¹, limited generalizability to daily practice, and infrequent head-to-head comparison^{2,3}.

Nonetheless, RCTs remain as the gold standard when available. The strength of RCTs comes from the random allocation of treatment assignment. Such randomization ensures the group-level balance of patients between the treatment arms. However, in clinical practice physicians carefully choose who should or should not be treated with the drug of interest, causing imbalance in the baseline risk for the outcome under study between treatment groups in observational studies of clinical practice data (*i.e.*, confounding by indication)^{4,5}. For example, a higher risk profile for gastrointestinal (GI) bleeding can lead to preferential prescription of cyclooxygenase (COX)-2-selective inhibitors, and also increase the risk of subsequent GI bleeding. This can result in a spurious association between COX-2 inhibitor use and higher incidence of subsequent GI bleeding if confounding by indication (higher GI bleeding risks) is not fully accounted for (Figure 1).

To address this issue, designing and analysing observational studies in ways similar to how RCTs are designed and analysed is important. Two design principles – *active comparator design* and *new user design* – can help investigators design observational studies that resemble RCTs, thus, improving the quality of comparison. These two design principles can address methodological issues in observational studies that statistical adjustment alone cannot address.

In this article, we discuss the advantages of these two design principles, using examples from the recent literature on the risk of infection and cancer associated with use of bDMARDs in patients with RA.

ACTIVE COMPARATOR DESIGN

The active comparator design refers to a study that compares the effect of ‘Drug A’, study drug of interest, to ‘Drug B’, another active drug used in clinical practice, instead of ‘no use’ (non-users). Non-users are subjects who have the disease of interest, but are not on treatment for the disease. Non-users likely include subjects with no indication for any treatment (*e.g.*, very mild disease) as well as those with contraindication for all treatment (*e.g.*, very severe coexisting conditions). These types of subjects are generally not included in RCTs.

Three main advantages that the active comparator design provides are: (1) increasing overlap of measured characteristics between the groups, (2) reducing potential for unmeasured confounding, and (3) possibly improving the research question.

(1) Increasing overlap of measured characteristics

Measured differences in the pretreatment patient characteristics can be accounted for by various statistical methods, including propensity score methods^{6,7} (Figure 2 “above surface”). However, one important assumption often hidden in the black box of multivariable analysis is that valid statistical adjustment requires sufficient overlap in patient

characteristics across treatment arms^{8,9}. Also the more overlap there is, the more efficient the statistical adjustment. Choosing an active comparator group that receives a drug with the same or a similar indication makes the treatment groups similar in terms of treatment indications, and should increase the overlap of patient characteristics. For example, in one UK study by Dixon et al¹⁰, comparisons of tuberculosis risks were made both within bDMARDs and with synthetic DMARDs. As one may expect, the measured baseline characteristics, including disease activity, disease duration, previous treatments, and comorbidities were more similar among different types of bDMARDs than synthetic DMARDs.

(2) Reducing potential for unmeasured confounding

In addition to the measured patient characteristics mentioned above, unmeasured differences between treatment groups pose threats to the validity of study results because unmeasured factors cannot be accounted for statistically, and can only be addressed by study design (Figure 2 “below surface”). These unmeasured characteristics may be measurable variables missed in the particular dataset in use or latent variables that are hard or impossible to measure. Frailty¹¹, age-related overall decline in physical function and health status, is a characteristic hard to measure, and is often suspected as the main reason of highly protective effects of preventive measures, which are received by those who are healthy enough¹². Influenza vaccination compared to non-use, for example, was found protective against all-cause mortality during the summer season when influenza infection is rare¹³. The non-user group likely includes those who are too sick to be considered for preventive interventions. Although difficult in the case of influenza vaccination studies, providing an active comparator with similar indications whenever possible should attenuate such differences in unmeasured patient baseline characteristics, and reduce unmeasured confounding.

(3) Possibly improving the research question

The question being answered by an active comparator study is “how does *this* drug compare to *another* drug that has similar indications.”¹⁴ This is different from the question answered by a non-user comparator study, and is especially relevant to drug safety studies. For example, comparing the safety of long-term bisphosphonate use to that of long-term non-use of osteoporosis medication may not be useful because patients with osteoporosis should be treated one way or another. An active comparator study gives an insight into *which drug is safer* with respect to the safety outcome in question when choosing a treatment for a patient with osteoporosis.

NEW USER DESIGN

The new user design¹⁵ (also known as *incident user design*¹⁶ or *initiator design*) identifies a cohort of patients who initiate a drug of interest and begins follow up after treatment initiation, similar to RCTs (Figure 3, Panel B). It is contrasted to the prevalent user design, which includes both current and new users of a drug of interest during the study period. Thus, the follow up starts at a different time point in each individual during the treatment course (Figure 3, Panel C).

Three main advantages that the new user design provides are: (1) assessing time-varying hazards and drug effects associated with treatment duration (2) ensuring appropriate confounding adjustment by capturing pretreatment variables; and (3) reducing potential for immortal time bias when combined with the active comparator design.

(1) Assessing time-varying hazards and drug effects associated with treatment duration

The new user design^{15,16} plays a particularly important role in drug safety studies. Rates of some adverse events change over time. This phenomenon is described as depletion of the susceptible^{17,18}, *i.e.*, early loss of patients at risk for the adverse outcomes from the cohort during the drug use, leaving those who can tolerate the drug well in the cohort at a later time. We identified several studies that examined this phenomenon in the literature^{17,19,20}. In a study by Dixon *et al.*²⁰, the risk for severe infection was highest in the first 90 days of treatment with TNF- α inhibitor versus synthetic DMARDs with an incidence rate ratio of 4.6 [95% confidence interval (CI) 1.8-11.9], while the incidence rate ratio during the entire follow-up time was 1.22 (95% CI 0.88-1.69) in the British Biologic Register²⁰. Similarly, Strangfeld *et al.*¹⁷ and Curtis *et al.*¹⁹ also noted a decline in the risk of infection associated with TNF- α inhibitors versus synthetic DMARDs over time. Thus, if the prevalent user design were used to examine the risk of infection associated with bDMARD use, this notably increased risk for infection early in the treatment course would have been missed (Figure 3, Panel C). The new user design, but not the prevalent user design, can also examine the cumulative effects or risks of drugs related to treatment duration, as seen in a Swedish cohort study by Askling *et al.*²¹ which reported no increased cancer risk over the first 6 years of treatment with TNF- α inhibitor in patients with RA. In the prevalent user design, treatment duration cannot be accurately defined, as the time of initiation is often unknown.

(2) Ensuring appropriate confounding adjustment by capturing pretreatment variables

Another problem in the prevalent user design is that their “baseline” characteristics are not always captured before initiation of the treatment. Multivariable statistical adjustment methods, including propensity score methods⁶ can control for imbalance in measured characteristics between the groups, but investigators should carefully choose what variables to adjust for. It is the *pretreatment* covariates that should be adjusted for in statistical methods. Adjusting for *posttreatment* variables may lead to overadjustment²², *i.e.*, inappropriate statistical adjustment for intermediate variables (mediators) lying between the exposure-outcome causal pathway. For example, the disease activity after bDMARDs initiation is the *result* of the treatment, which in turn may affect the subsequent infection risks (Figure 4). This represents a part of the effect of the treatment on the outcome of interest; thus, conventional statistical adjustment should be avoided. Such adjustment blocks the true causal relationship between the treatment and outcome of the interest (Figure 4). In the new user design, investigators have a clear idea regarding which variables are *pretreatment* and which are *posttreatment*, as the information preceding the treatment initiation is available (Figure 3, Panel A, thin line preceding treatment within the observation period box). This helps investigators choose appropriate variables to adjust for.

(3) Reducing potential for immortal time bias when combined with the active comparator design

Immortal time bias^{18,23,24} is another important form of bias that can invalidate findings from observational studies. *Immortal time* is defined as the period during which the outcome of interest cannot occur because of the study design^{23,24}. Immortal time bias is typically introduced when the start of follow-up time is defined differently between the treated and untreated groups, or there is a typical sequence of treatments (e.g., starting bDMARDs only after synthetic DMARDs)²⁴. For example, in a cohort study that compares the risk of death among new users of bDMARDs with non-users, there is a wait time (*i.e.*, immortal time) for the bDMARD group because patients need to survive alive or *event-free* to initiate a bDMARD, but not for the non-users. In other words, if patients have an event prior to getting a bDMARD, their person-time and the event are only attributed to the non-user group. Such differential distribution of follow-up time and events between the user and non-user groups leads to immortal time bias favouring bDMARDs compared to non-users. However, when a cohort study compares patients who switched to get a bDMARD versus switchers to other synthetic DMARDs (*i.e.*, combining the new user design with the active comparator design), the potential for immortal time bias is reduced as the start of follow-up time can be defined as the switch date for both groups.

USE OF ACTIVE COMPARATOR DESIGN AND NEW USER DESIGN IN THE RECENT LITERATURE

We systematically reviewed the literature published in the last decade to examine their use of the active comparator design and new user design. Two areas were examined: association between use of bDMARDs and 1) risk of infections or 2) risk of cancers in patients with rheumatoid arthritis. Sixty articles among 631 initial search results for infections and 15 articles among 443 articles for cancers were cohort studies, and were assessed for their use of active comparator design and new user design.

Our review of the recent literature for cohort studies that examined the risk of infection or cancer in RA patients treated with bDMARDs revealed inadequate use of these designs, as 21 out of 52 (40%) studies for infection risks and 4 of 15 (27%) studies for cancer risks fully utilized both the active comparator design and the new user design (Supplementary Tables; freely available at the publisher website). Among the studies that utilized the active comparator design, the most commonly selected comparator was the synthetic DMARDs group (42% in infection studies and 73% in cancer studies reviewed). Several papers used the general population as the comparator via standardization methods^{9,25}. Many studies included new users of bDMARDs but compared with prevalent users of synthetic DMARDs, whereas some studies defined a new-user synthetic DMARD comparison group by choosing those who switch to or use another synthetic DMARD^{26,27}.

LIMITATIONS OF DESIGN PRINCIPLES AND CONSIDERATIONS

Even with these design principles carefully applied and statistical analyses meticulously conducted, biases of observational studies are not completely solved. If an RCT answering the same clinical question is feasible, it should be given priority. While the *active*

comparator design and *new user design* have the aforementioned methodological strengths and surely improve the quality of observational studies, these designs also have potential disadvantages. If the drug of interest is the most commonly used drug for a given condition and the alternatives are used infrequently, selecting an active comparator can be challenging. There may be even no active comparator for a newly marketed drug for a given condition. Also it may be difficult to interpret the relative risk of the drug of interest compared to an active comparator if the effect or risk of the active comparator drug is unknown. Having multiple comparators including both an active comparator and a non-user group may help interpret the findings in such a case. The sample size can be limited by restricting the cohort to drug initiators only (Figure 3, Panel B). The prevalent user analysis should give a similar estimate of drug effects to those from the new user analysis when the risk is relatively constant over time. However, whether the drug effects for a specific outcome are time-varying is mostly unknown. Thus, one should consider a sensitivity analysis restricted to new users only even in a study that primarily includes prevalent users due to a feasibility issue.

CONCLUSIONS

In this article, we discussed how observational studies and RCTs should complement each other, what biases are inherently associated with observational studies, and how the *active comparator design* and the *new user design* help improving the quality of observational studies (Table 1. Key Messages).

The *active comparator design* potentially helps to reduce both measured and unmeasured confounding, and may also improve the research question being answered, while the ‘non-user’ comparator design is affected by more confounding by indication and may not answer the relevant clinical question. The *new user design* helps assess time-varying hazards and drug effects associated with treatment duration, and ensures appropriate adjustment for confounding by establishing clear temporal sequence between pretreatment variables and drug exposure, and decreases the potential for immortal time bias when combined with the active comparator design, whereas the *prevalent user design* may miss early adverse events and is more susceptible to overadjustment and immortal time bias.

As noted in our literature review, these principles are generally underused in recent literature related to treatment of rheumatoid arthritis. To address critical methodological issues in observational studies such as confounding by indication, inappropriate statistical adjustment of intermediate variables, and immortal time bias and to improve the validity of observational studies of drug effects, these two principles – *active comparator design* and *new user design* – should be considered in designing or reviewing all observational studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING INFORMATION:

KY receives tuition support jointly from Japan Student Services Organization (JASSO) and Harvard T. H. Chan School of Public Health (partially supported by training grants from Pfizer, Takeda, Bayer and PhRMA). DHS receives research grants from Eli Lilly, Amgen, and CORRONA. He also receives royalties from UpToDate, and serves in unpaid roles in studies funded by Pfizer, Novartis, and Eli Lilly. SCK receives research grants from Pfizer, Inc and Lilly. SCK is supported by the NIH grant K23AR059677.

References

1. Chan KA, Hernandez-Diaz S. Pharmacoepidemiology and rheumatic disorders. *Rheum Dis Clin North Am.* 2004; 30:835–850. vii. [PubMed: 15488696]
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. *Clin Pharmacol Ther.* 2011; 90:777–790. [PubMed: 22048230]
3. Strom, BL.; Kimmel, SE.; Hennessy, S. *Textbook of Pharmacoepidemiology.* Wiley-Blackwell; 2013.
4. Walker AM, Stampfer MJ. Observational studies of drug safety. *Lancet.* 1996; 348:489. [PubMed: 8757145]
5. Psaty BM, Siscovick DS. Minimizing bias due to confounding by indication in comparative effectiveness research: the importance of restriction. *JAMA.* 2010; 304:897–898. [PubMed: 20736474]
6. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983; 70:41–55.
7. Crump RK, Hotz VJ, Imbens GW, Mitnik OA. Dealing with limited overlap in estimation of average treatment effects. *Biometrika.* 2009; 96:187–199.
8. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res.* 2012; 21:31–54. [PubMed: 21030422]
9. Hernan, MA.; Robins, JM. *Causal Inference.* Chapman & Hall/CRC; 2015. at <<http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>>
10. Dixon WG, et al. Drug-specific risk of tuberculosis in patients with rheumatoid arthritis treated with anti-TNF therapy: results from the British Society for Rheumatology Biologics Register (BSRBR). *Ann Rheum Dis.* 2010; 69:522–528. [PubMed: 19854715]
11. Sternberg SA, Wershof Schwartz A, Karunanathan S, Bergman H, Mark Clarfield A. The identification of frailty: a systematic literature review. *J Am Geriatr Soc.* 2011; 59:2129–2138. [PubMed: 22091630]
12. Glynn RJ, Knight EL, Levin R, Avorn J. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology.* 2001; 12:682–689. [PubMed: 11679797]
13. Eurich DT, Marrie TJ, Johnstone J, Majumdar SR. Mortality reduction with influenza vaccine in patients with pneumonia outside ‘flu’ season: pleiotropic benefits or residual confounding? *Am J Respir Crit Care Med.* 2008; 178:527–533. [PubMed: 18556629]
14. Camelo Castillo W, Delaney JAC, Stürmer T. The challenges of comparing results between placebo controlled randomized trials and non-experimental new user, active comparator cohort studies: the example of olmesartan. *Pharmacoepidemiol Drug Saf.* 2014; 23:357–360. [PubMed: 24590633]
15. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol.* 2003; 158:915–920. [PubMed: 14585769]
16. Johnson ES, et al. The incident user design in comparative effectiveness research. *Pharmacoepidemiol Drug Saf.* 2013; 22:1–6. [PubMed: 23023988]
17. Strangfeld 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; 70:1914–1920. [PubMed: 21791449]
18. Choi HK, Nguyen U-S, Niu J, Danaei G, Zhang Y. Selection bias in rheumatic disease research. *Nat Rev Rheumatol.* 2014; 10:1038. nrrheum.2014.36

19. Curtis JR, et al. Risk of serious bacterial infections among rheumatoid arthritis patients exposed to tumor necrosis factor alpha antagonists. *Arthritis Rheum.* 2007; 56:1125–1133. [PubMed: 17393394]
20. Dixon WG, et al. Serious infection following anti-tumor necrosis factor alpha therapy in patients with rheumatoid arthritis: lessons from interpreting data from observational studies. *Arthritis Rheum.* 2007; 56:2896–2904. [PubMed: 17763441]
21. Askling J, et al. Cancer risk in patients with rheumatoid arthritis treated with anti-tumor necrosis factor alpha therapies: does the risk change with the time since start of treatment? *Arthritis Rheum.* 2009; 60:3180–3189. [PubMed: 19877027]
22. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology.* 2009; 20:488–495. [PubMed: 19525685]
23. Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf.* 2007; 16:241–249. [PubMed: 17252614]
24. Levesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ.* 2010; 340:b5087–b5087. [PubMed: 20228141]
25. Rothman, KJ.; Professor, TLLA.; Greenland, S. *Modern Epidemiology.* LWW; 2012.
26. Grijalva CG, et al. Initiation of tumor necrosis factor-alpha antagonists and the risk of hospitalization for infection in patients with autoimmune diseases. *JAMA.* 2011; 306:2331–2339. [PubMed: 22056398]
27. Strangfeld A, et al. Risk of incident or recurrent malignancies among patients with rheumatoid arthritis exposed to biologic therapy in the German biologics register RABBIT. *Arthritis Res Ther.* 2010; 12

FIGURE 1.

Schematic illustration of how confounding by indication can cause a spurious association. Gastrointestinal bleeding risks raise the likelihood of COX-2 inhibitor prescription (short arrow). Such gastrointestinal risk factors, by definition, predispose patients to subsequent bleeding (long arrow). Spurious statistical association (broken line) arises between COX-2 inhibitor use and gastrointestinal bleeding the gastrointestinal bleeding risks (confounders; common causes of both COX-2 prescription and subsequent bleeding). Adjustment for gastrointestinal risks is necessary.

Measured pretreatment characteristics “above the surface”: age, gender, etc. (These can be adjusted statistically.)

FIGURE 2.

Greater difference in patient characteristics between the main treatment group vs non-users rather than active comparator.

Compared to the non-users, the users of drug B prescribed for the same indication are more similar to the users of drug A in measured pretreatment characteristics, such as age and gender, and more importantly in unmeasured pretreatment characteristics, such as frailty, life style, disease activity. It is likely that non-users include subjects who have no indication for any treatment (*e.g.*, very mild disease) or contraindication for all treatment (*e.g.*, very severe coexisting conditions). Statistical analysis can only adjust for characteristics “visible” as variables (“above the surface”), and is efficient if the distributions of these characteristics are similar. Differences in unmeasured pretreatment characteristics (“below the surface”) cannot be addressed by statistical adjustment; therefore, such unmeasured differences need to be addressed by the study design.

FIGURE 3.

Comparison of how observations are utilized in new user design and prevalent user design.

Panel A: The colored line represents drug use and the gray line is for non-use. The red, yellow, and green color blocks represent different risk period for infection. The box is the study period captured in the database, and the shaded area is not.

Panel B: Only those who start the drug during the study period are included (subjects 2, 3, and 6), thus, the reduced sample size. The follow up starts at the drug initiation for all remaining patients. The initial high risk period is well represented.

Panel C All users can be included, thus, the sample size is six. However, the index date is the drug initiation date only for subjects 2, 3, and 6, and for others an arbitrary time point. Thus, the early follow up period is a mixture of the initial high risk period and later low risk period.

FIGURE 4.

Schematic illustration of difference between confounders and mediators.

Pretreatment disease activity is a potential confounder (common cause of both treatment choice and subsequent infection), thus, adjustment is necessary. Posttreatment disease activity is a result of the treatment choice (mediator). Conventional statistical adjustment is inappropriate for mediators, as such adjustment blocks the true causal relationship between the treatment and outcome of the interest, and will bias the result (overadjustment).

Table 1

Key messages

	Active comparator design	New user design
Benefits	(1) Increase overlaps in measured pretreatment characteristics (confounding), helping more efficient statistical analysis. (2) Reduce potential for unmeasured confounding, which cannot be addressed by statistical analysis. (3) Possibly improve the research question by asking more clinically relevant question, “Which treatment is more effective/ safe for a patient with RA?”	(1) Assess time-varying hazards and drug effects associated with treatment duration. (2) Ensure statistical control of confounding by capturing <i>pretreatment</i> variables in all patients. (3) Reduce potential for immortal time bias when combined with the active comparator design by enforcing similar definitions of the index date across groups.
Disadvantages	The effectiveness/safety profile of the active comparator must be well established to make the result interpretable.	Sample size issue may occur because only those who started their treatment during the study period can be included.
Consideration	Report the absolute risk in all groups in addition to relative risks. If the risk in the active comparator arm has uncertainty, including an additional non-user arm may help.	If risk is time-constant, prevalent user design should give similar results. A sensitivity analysis using the new user subset is still recommended.