OPINION

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

Kazuki Yoshida, Daniel H. Solomon and Seoyoung C. Kim

Abstract | Over the past decade, an increasing number of observational studies have examined the effectiveness or safety of treatments for rheumatoid arthritis. Unlike randomized controlled trials (RCTs), however, observational studies of drug effects have methodological limitations such as confounding by indication. Active-comparator designs and new-user designs can help mitigate such biases in observational studies and improve the validity of their findings by making them more closely approximate RCTs. In an active-comparator study, the drug of interest is compared with another agent commonly used for the same indication, rather than with no treatment (a 'non-user' group). This principle helps to ensure that treatment groups have similar treatment indications, attenuating both measured and unmeasured differences in patient characteristics. The new-user study includes a cohort of patients from the time of treatment initiation, enabling assessment of patients' pretreatment characteristics and capture of all events occurring during follow-up. These two principles should be considered when designing or reviewing observational studies of drug effects.

Yoshida, K. *et al. Nat. Rev. Rheumatol.* **11**, 437–441 (2015); published online 24 March 2015; doi:10.1038/nrrheum.2015.30

Introduction

Since the introduction of TNF inhibitors in the 1990s, a growing number of biologic DMARDs have become available for the treatment of rheumatoid arthritis (RA). Consequently, the need for high-quality data on the comparative effectiveness and safety of biologic DMARDs has increased. Randomized controlled trials (RCTs) are often limited by insufficient statistical power to detect differences in non-primary endpoints, relatively short follow-up durations (which limit assessment of long-term safety),[1] lack of generalizability to a wide range of patients in daily practice, and infrequent inclusion of head-to-head comparisons.[2,3] Observational studies, such as cohort studies, provide complementary information regarding the effectiveness or safety of these drugs after their approval. Nonetheless, RCTs remain the 'gold standard' for clinical research, when feasible.

The strength of RCTs comes from the random allocation of treatment assignment. Such randomization ensures the group-level balance of patients in the treatment groups. In clinical practice, however, physicians carefully choose who should or should not be treated with the drug of interest, causing imbalance between treatment groups in the baseline level of risk for the outcome of interest in observational studies—that is, confounding by indication.[4,5] For example, the presence of risk factors for gastrointestinal bleeding can lead to preferential prescription of cyclooxygenase 2 (COX-2)-selective inhibitors, and also, by definition, increase the risk of subsequent gastrointestinal bleeding. A spurious association between COX-2 inhibitor use and increased incidence of subsequent gastrointestinal bleeding can result, if confounding by indication (that is, increased risk of gastrointestinal bleeding) is not fully accounted for (Figure 1).

To address the limitations of observational studies, including confounding by indication, it is important that observational studies be designed and analysed in ways similar to RCTs. Two principles—active-comparator design and new-user design—can help investigators construct observational studies that more closely approximate RCTs, thus improving the quality of the comparison. These two design principles can address methodological issues in observational studies that cannot be addressed by statistical adjustment alone. In this article, we discuss the advantages of these two design principles, and their application in recent studies on the risk of infection and cancer associated with use of biologic DMARDs in patients with RA.

Active-comparator design

An active-comparator study compares the effect of 'drug A', the study drug of interest, with 'drug B', another active drug used in clinical practice, instead of with a 'non-active comparator'. Patients in the non-active comparator group (non-users), who have the disease of interest but who are not receiving active treatment, could in theory include those with no indication for any treatment (for example, if they have very mild disease) as well as those for whom all treatment is contraindicated (for example, if they have very severe coexisting conditions). These particular groups of non-users are generally not included in RCTs, and should also be avoided in observational studies.

The active-comparator design has three main advantages: increased similarity in measured patient characteristics between treatment groups; reduced potential for unmeasured confounding; and possibly improving the clinical relevance of the research question.

Increasing between-group overlap

Differences in measured pretreatment patient characteristics ('above the surface'; Figure 2) can be accounted for by use of various statistical methods, including those based on propensity scores.[6,7] 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 groups.[8] Also, the more extensive the overlap, the more efficient the

Competing interests
K.Y. has received tuition support jointly from Japan Student Services Organization (JASSO) and Harvard T. H. Chan School of Public Health (supported in part by training grants from Bayer, PhRMA, Pfizer and Takeda). D.H.S. has received research or funding support from Amgen, Lilly and CORRONA, received royalties from UpToDate and served in unpaid roles in studies funded by Lilly, Novartis and Pfizer. S.C.K. has received research support from Lilly and Pfizer.

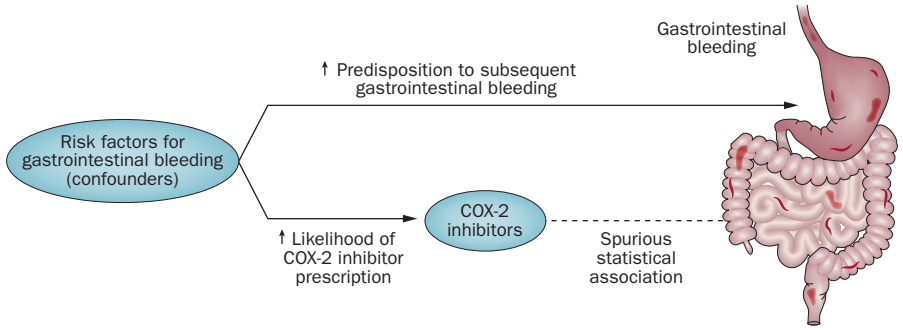

Figure 1 | Schematic illustration of how confounding by indication can cause a spurious statistical association. Risk factors for gastrointestinal bleeding increase the likelihood of COX-2 inhibitor prescription. Such gastrointestinal risk factors, by definition, also predispose patients to subsequent bleeding. A spurious statistical association (dashed line) arises between COX-2 inhibitor use and gastrointestinal bleeding. The risk factors for gastrointestinal bleeding are confounders (common causes of both COX-2 prescription and subsequent bleeding), and so appropriate statistical adjustment for these risk factors is necessary. Abbreviation: COX-2, cyclooxygenase 2.

statistical adjustment. Choosing an active-comparator group that receives a drug with the same or similar indication as the study drug means the treatment groups will be similar in terms of treatment indications, and should increase the overlap in patient characteristics between the groups. For example, a UK study by Dixon *et al.*[9] compared the risk of tuberculosis in patients receiving various biologic DMARDs and in those receiving biologic versus synthetic DMARDs. As one might expect, the measured baseline characteristics, including disease activity, disease duration, previous treatments and comorbidities, were more similar among patients receiving the different types of biologic DMARDs than in those being treated with biologic versus synthetic DMARDs.

Reducing unmeasured confounding

In addition to the measured patient characteristics mentioned above, unmeasured differences between treatment groups ('below the surface'; Figure 2) can undermine the validity of study results because these factors cannot be accounted for statistically. However, they can be addressed by study design. Unmeasured characteristics could be measurable variables not included in the particular dataset in use, or might be latent variables that are challenging or impossible to measure. Frailty,[10] which is defined as age-related overall decline in physical function and health status, is an example of a characteristic that is difficult to measure but that can influence treatment choices. Frailty is often suspected to account for the highly protective effects of preventive measures—which are more likely to be used by those who are healthy enough already than by frail patients, for whom the preventive treatment would be futile (an example of 'healthy user' bias).[11] Influenza vaccination compared with non-vaccination, for example, was associated with reduced all-cause mortality during the summer season, when influenza infection is rare.[12] The non-user group probably included individuals who were too ill 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 differences in unmeasured patient baseline characteristics, and reduce unmeasured confounding.

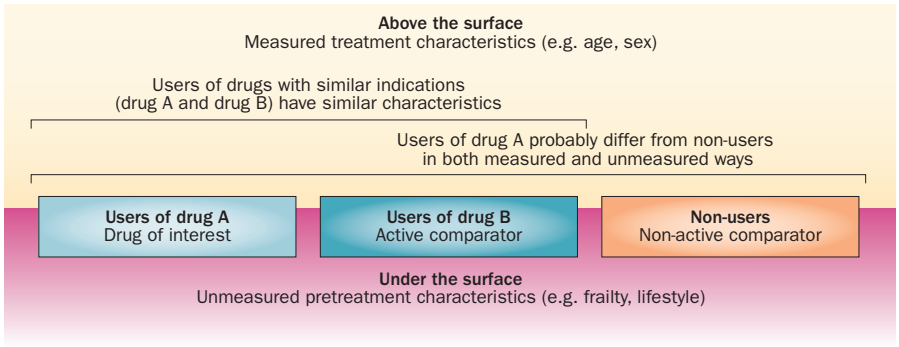

Figure 2 | Differences in patient characteristics are greater between users of the study drug and non-users, than between users of the study drug and users of an active comparator. Compared with non-users, users of drug B are more similar to users of drug A (prescribed for the same indication) in measured pretreatment characteristics, such as age and sex, and, more importantly, in unmeasured pretreatment characteristics, such as frailty and lifestyle. The group of non-users is likely to include individuals who have no indication for any treatment (e.g. very mild disease) or in whom all treatment is contraindicated (e.g. very severe coexisting conditions). Statistical analysis can only adjust for characteristics 'visible' as variables ('above the surface'), and is more efficient if the distributions of these characteristics are similar across treatment groups. 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.

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?"[13] This question differs from that being answered by a non-user comparator study, and is especially relevant to drug safety studies. For example, the safety of long-term bisphosphonate use would not be compared to that of long-term non-use of osteoporosis medication because patients with osteoporosis should be treated with one drug or another. When choosing a treatment for a patient with osteoporosis, an active-comparator study can provide insight into which drug is superior with respect to the safety outcome in question.

New-user design

A study using the new-user design,[14] also known as incident-user design[15] or initiator design, includes a cohort of patients initiating treatment with a drug of interest who are followed up from treatment initiation, similar to RCTs (Figure 3b). By contrast, the prevalent-user design includes both current and new users of a drug of interest within the study period, and follow-up thus starts at a different time point in the course of each individual's treatment (Figure 3c). The new-user design has three main advantages: time-varying hazards and drug effects associated with treatment duration can be assessed; appropriate adjustment for confounding

is ensured by capturing pretreatment variables; and potential for immortal time bias is reduced when this design is combined with the active-comparator design.

Assessing time-dependent effects

The new-user design[14,15] has a particularly important role in drug safety studies. Rates of some adverse events change over time, a phenomenon described as 'depletion of the susceptible'.[16,17] That is, the patients who develop drug-related adverse outcomes are lost from the cohort early on, leaving only those who tolerate the drug well in the cohort at a later time. We identified several studies that examined this phenomenon.[16,18,19] In a study of data from the British Society for Rheumatology Biologics Register by Dixon *et al.*,[19] the risk of severe infection was highest in the first 90 days of treatment with a TNF inhibitor versus synthetic DMARDs, with an incidence rate ratio of 4.6 (95% CI 1.8–11.9), whereas the incidence rate ratio during the entire follow-up period was 1.22 (95% CI 0.88–1.69).[19] Similarly, Strangfeld *et al.*[16] and Curtis *et al.*[18] also noted a decline in the risk of infection associated with TNF inhibitors versus synthetic DMARDs over time. Thus, if the prevalent-user design were used to examine the risk of infection associated with biologic DMARD use, this notably increased risk of infection early in the treatment course would have been missed (Figure 3).

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.*,[20] who reported no increased cancer risk over the first 6 years of anti-TNF treatment in patients with RA. In studies with a prevalent-user design, treatment duration cannot be accurately defined, as the time of initiation is often unknown.

Appropriate adjustment for confounding

Another problem inherenent in the prevalent-user design is that 'baseline' patient characteristics are not always captured before initiation of the treatment. Multivariable statistical adjustment methods, including propensity scores,[6] can control for imbalance in measured characteristics between treatment groups, but investigators should carefully choose what variables to adjust for. Pretreatment covariates, rather than post-treatment covariates, should be subject to statistical adjustment. Adjustment for post-treatment variables could lead to overadjustment;[21] that is, inappropriate statistical adjustment for intermediate variables (mediators) in the exposure–outcome causal pathway. For example, disease activity measured after the initiation of biologic DMARDs is a result of the treatment that can, in turn, affect the patient's subsequent risk of infection (Figure 4). This causal pathway through post-treatment disease activity represents a part of the effect of the treatment on the outcome of interest; thus, conventional statistical adjustment should be avoided. Such adjustment would obscure the true causal relationship between the treatment and the outcome of the interest (Figure 4). The new-user design gives investigators a clear idea regarding which variables are pretreatment factors (potential confounders) and which are post-treatment factors (mediators), as relevant information is available from the period preceding treatment initiation (Figure 3). This knowledge helps investigators choose appropriate variables for statistical adjustment.

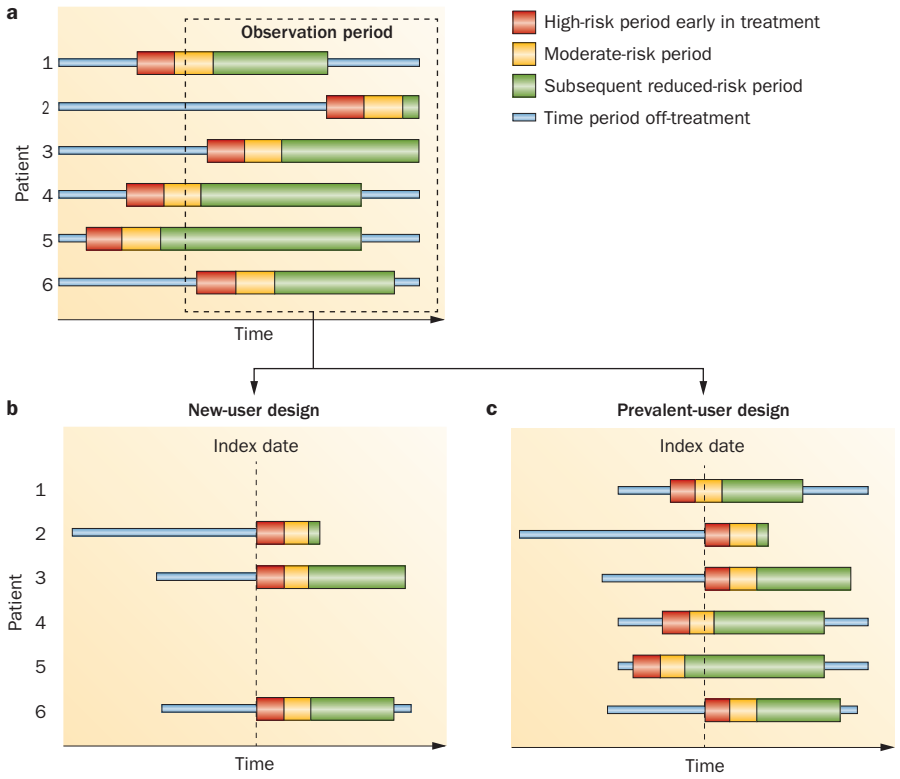

Figure 3 | Comparison of how observations are utilized in new-user design and prevalent user design. **a** | The area within the box represents the study period captured in the database; events that occur in the shaded area are not captured. **b** | In the new-user design, the sample is limited to only those patients who start the drug within the study period are included (patients 2, 3 and 6). For these patients, the index date marking the start of the follow-up period is the time of drug initiation. The initial high-risk period is well-represented. **c** | In the prevalent use design, all patients using the drug can be included in the study. However, the index date is the drug initiation date only for patients 2, 3 and 6; for the other patients, an arbitrary time point is selected. Thus, the 'early' follow-up period is a mixture of the initial high-risk period and the later low risk period.

Reducing immortal time bias

Immortal time bias[17,22,23] is another important form of bias that can invalidate findings from observational studies. 'Immortal time' is defined as a period during which the outcome of interest cannot occur because of the study design.[22,23] Immortal time bias is typically introduced when the start of follow-up is defined differently in the study drug and comparator groups, or when treatments are administered in a typical sequence (for example, starting biologic DMARDs only after synthetic DMARDs).[23] In a cohort study comparing mortality among new users of biologic DMARDs and non-users, for instance, the biologic DMARD group, but not the non-users, experience a wait time (immortal time) because these patients must necessarily have remained alive (or event-free) until the time they initiated a biologic DMARD. In other words, if patients have an event prior to beginning biologic DMARD treatment, their person-time and the event

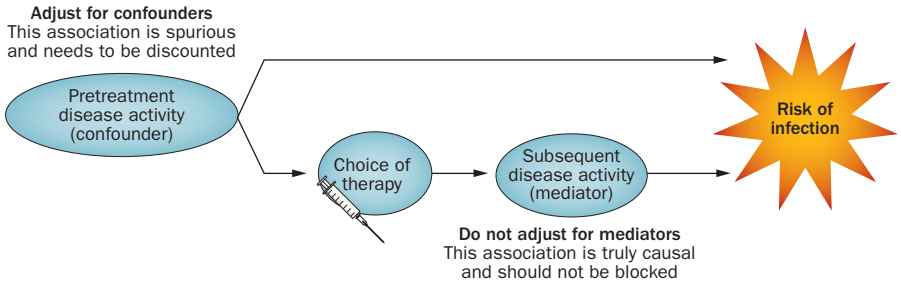

Figure 4 | Schematic illustration of the difference between confounders and mediators. Pretreatment disease activity is a potential confounder as it influences both treatment choice and risk of subsequent infection; thus, adjustment for pretreatment disease activity is necessary. Post-treatment disease activity is a result of the treatment choice and is a mediator of the outcome. Conventional statistical adjustment is inappropriate for mediators, as such adjustment obscures the true causal relationship between the treatment and the outcome of the interest, and will bias the result (overadjustment).

Table 1 | Features of active-comparator and new-user design principles

Features	Active-comparator design	New-user design
Advantages	Increases overlap in measured pretreatment characteristics between groups, enabling more-efficient statistical analysis	Assesses time-varying hazards and drug effects associated with treatment duration
	Reduces potential for unmeasured confounding, which cannot be addressed by statistical analysis	Ensures statistical control of confounding by capturing pretreatment variables in all patients
	Possibly improves the clinical relevance of the research question, by asking "Which treatment is more effective or safe for a patient with rheumatoid arthritis?"	When combined with the active-comparator design, reduces potential for immortal time bias by enforcing consistent definitions of the index date across groups
Disadvantages	The effectiveness and/or safety profile of the active comparator must be well-established to make the result interpretable	Sample size issues might occur because only those who initiated treatment during the study period can be included
Considerations	Report absolute risk in all groups in addition to relative risks; if the risk in the active-comparator arm is uncertain, including an additional non-user arm could be beneficial	If risk is time-constant, new-user design and prevalent-user designs should give similar results; a sensitivity analysis using the new-user subset is still recommended

are attributed only to the non-user group. Such differential distribution of follow-up time and events between user and non-user groups leads to immortal time bias favouring biologic DMARDs over non-use. However, when a cohort study compares patients who switched to a biologic DMARD versus those who switched to another synthetic DMARD (thereby 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 in published cohort studies

We systematically reviewed the literature published in PubMed since 2005 to examine the use of the active-comparator design and the new-user design in cohort studies examining the association between use of biologic DMARDs and either risk of infections or risk of cancers in patients with RA. Among the 1,074 research reports identified in the initial search, 52 cohort studies assessed infection risk and 15 cohort studies assessed cancer risk. These cohort studies were assessed for their use of active-comparator and new-user designs (Supplementary Tables 1 and 2).

Our analysis revealed inadequate use of these designs, as only 21 (40%) of the 52 studies on infection risks and four (27%) of the 15 studies on cancer risks fully implemented both the active-comparator and the new-user principles (Supplementary Tables 1 and 2). The most commonly selected comparator for biologic DMARDs was synthetic DMARDs (42% of the infection studies and 73% of the cancer studies reviewed). Several studies used the general population as the comparator group via standardization methods.[24] Many studies included new users of biologic DMARDs but compared them with prevalent users of synthetic DMARDs, whereas some studies defined a synthetic DMARD new-user comparison group by selecting those who switched to or added another synthetic DMARD.[25,26]

Limitations and considerations

Even with careful application of these design principles and meticulous statistical analyses, the biases inherent in observational studies are not completely resolved. If an RCT to answer the same clinical question is feasible, that type of study should be given priority over an observational study. Although the aforementioned methodological strengths of the active-comparator and new-user designs 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 infrequently used, selecting an active comparator can be challenging. With regard to a newly marketed drug, an active comparator for the same indication might not even exist. Also, interpreting the relative risk of the drug of interest compared with an active comparator can be difficult if the effect or risk of the comparator is unknown. Having multiple comparators, including both an active comparator and a non-user group, could help interpretation of the findings in such a case. With the new-user design, the sample size can be limited by restricting the cohort to drug initiators only (Figure 3b). Analysis of the outcomes of prevalent users should give a similar estimate of a drug's effects to that obtained by a new-user analysis when the effect is relatively constant over time; however, whether the drug's effects on a specific outcome vary with time is mostly unknown. Thus, investigators should consider including a sensitivity analysis restricted to new users only, even in a study that primarily includes prevalent users due to feasibility issues.

Conclusions

Observational studies and RCTs should complement each other. Although some biases are inherently associated with observational studies, using the active-comparator and new-user designs can help improve the quality of observational studies (Table 1).

The active-comparator design could help to reduce both measured and unmeasured

confounding, and might also improve the interpretability of the data presented, whereas the 'non-user' comparator design carries an increased likelihood of confounding by indication and might not even answer the relevant clinical question. The new-user design helps with the assessment of time-varying hazards and drug effects associated with treatment duration, and ensures appropriate adjustment for confounding by establishing a clear temporal sequence between pretreatment variables and drug exposure. When combined with the active-comparator design, the new-user design results in a decreased potential for immortal time bias. By contrast, the prevalent-user design could exclude some early adverse events and is susceptible to over-adjustment and immortal time bias.

As is evident from our review of the relevant literature, these principles are generally underused in cohort studies related to treatment of RA. To improve the validity of studies of drug effects and address critical methodological issues in observational studies—such as confounding by indication, inappropriate statistical adjustment of intermediate variables and immortal time bias—the active-comparator and new-user principles should be considered in designing or reviewing all observational studies.

*Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA (**K.Y., D.H.S., S.C.K.**).*
Correspondence to: K.Y.
kazukiyoshida@mail.harvard.edu

1. Chan, K. A. & Hernandez-Diaz, S. Pharmacoepidemiology and rheumatic disorders. *Rheum. Dis. Clin. North Am.* **30**, 835–850 (2004).
2. Schneeweiss, S., Gagne, J. J., Glynn, R. J., Ruhl, M. & Rassen, J. A. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. *Clin. Pharmacol. Ther.* **90**, 777–790 (2011).
3. Strom, B. L., Kimmel, S. E. & Hennessy, S. (Eds) *Textbook of Pharmacoepidemiology* (Wiley-Blackwell, 2013).
4. Walker, A. M. & Stampfer, M. J. Observational studies of drug safety. *Lancet* **348**, 489 (1996).
5. Psaty, B. M. & Siscovick, D. S. Minimizing bias due to confounding by indication in comparative effectiveness research: the importance of restriction. *JAMA* **304**, 897–898 (2010).
6. Rosenbaum, P. R. & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55 (1983).
7. Crump, R. K., Hotz, V. J., Imbens, G. W. & Mitnik, O. A. Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96**, 187–199 (2009).
8. Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y. & van der Laan, M. J. Diagnosing and responding to violations in the positivity assumption. *Stat. Methods Med. Res.* **21**, 31–54 (2012).
9. Dixon, W. G. *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.* **69**, 522–528 (2010).
10. Sternberg, S. A., Wershof Schwartz, A., Karunananthan, S., Bergman, H. & Clarfield, A. M. The identification of frailty: a systematic literature review. *J. Am. Geriatr. Soc.* **59**, 2129–2138 (2011).
11. Glynn, R. J., Knight, E. L., Levin, R. & Avorn, J. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology* **12**, 682–689 (2001).
12. Eurich, D. T., Marrie, T. J., Johnstone, J. & Majumdar, S. R. Mortality reduction with influenza vaccine in patients with pneumonia outside 'flu' season: pleiotropic benefits or residual confounding? *Am. J. Respir. Crit. Care Med.* **178**, 527–533 (2008).
13. Camelo Castillo, W., Delaney, J. A. C. & 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.* **23**, 357–360 (2014).
14. Ray, W. A. Evaluating medication effects outside of clinical trials: new-user designs. *Am. J. Epidemiol.* **158**, 915–920 (2003).
15. Johnson, E. S. *et al.* The incident user design in comparative effectiveness research. *Pharmacoepidemiol. Drug Saf.* **22**, 1–6 (2013).
16. 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.* **70**, 1914–1920 (2011).
17. Choi, H. K., Nguyen, U.-S., Niu, J., Danaei, G. & Zhang, Y. Selection bias in rheumatic disease research. *Nat. Rev. Rheumatol.* **10**, 403–412 (2014).
18. Curtis, J. R. *et al.* Risk of serious bacterial infections among rheumatoid arthritis patients exposed to tumor necrosis factor α antagonists. *Arthritis Rheum.* **56**, 1125–1133 (2007).
19. Dixon, W. G. *et al.* Serious infection following anti-tumor necrosis factor α therapy in patients with rheumatoid arthritis: lessons from interpreting data from observational studies. *Arthritis Rheum.* **56**, 2896–2904 (2007).
20. Askling, J. *et al.* Cancer risk in patients with rheumatoid arthritis treated with anti-tumor necrosis factor α therapies: does the risk change with the time since start of treatment? *Arthritis Rheum.* **60**, 3180–3189 (2009).
21. Schisterman, E. F., Cole, S. R. & Platt, R. W. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* **20**, 488–495 (2009).
22. Suissa, S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol. Drug Saf.* **16**, 241–249 (2007).
23. Levesque, L. E., Hanley, J. A., Kezouh, A. & Suissa, S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ* **340**, b5087 (2010).
24. Rothman, K. J., Greenland, S. & Lash, T. L. *Modern Epidemiology* (Lippincott Williams & Wilkins, 2012).
25. Grijalva, C. G. *et al.* Initiation of tumor necrosis factor-α antagonists and the risk of hospitalization for infection in patients with autoimmune diseases. *JAMA* **306**, 2331–2339 (2011).
26. 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.* **12**, R5 (2010).

Acknowledgements
S.C.K.'s work is supported by a grant from the NIH (K23AR059677).

Author contributions
K.Y. researched data for the article. All authors provided substantial contributions to discussion of the content, writing and reviewing or editing of the manuscript before submission.

Supplementary information is linked to the online version of the paper at www.nature.com/nrrheum.