METHODOLOGY

Incorporating alternative design clinical trials in network meta-analyses

Kristian Thorlund[1–3]
Eric Druyts[1,4]
Kabirraaj Toor[1,5]
Jeroen P Jansen[1,6]
Edward J Mills[1,3]

[1]Redwood Outcomes, Vancouver, BC, [2]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada; [3]Stanford Prevention Research Center, Stanford University, Stanford, CA, USA; [4]Department of Medicine, Faculty of Medicine, [5]School of Population and Public Health, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada; [6]Department of Public Health and Community Medicine, Tufts University, Boston, MA, USA

Introduction: Network meta-analysis (NMA) is an extension of conventional pairwise meta-analysis that allows for simultaneous comparison of multiple interventions. Well-established drug class efficacies have become commonplace in many disease areas. Thus, for reasons of ethics and equipoise, it is not practical to randomize patients to placebo or older drug classes. Unique randomized clinical trial designs are an attempt to navigate these obstacles. These alternative designs, however, pose challenges when attempting to incorporate data into NMAs. Using ulcerative colitis as an example, we illustrate an example of a method where data provided by these trials are used to populate treatment networks.

Methods: We present the methods used to convert data from the PURSUIT trial into a typical parallel design for inclusion in our NMA. Data were required for three arms: golimumab 100 mg; golimumab 50 mg; and placebo. Golimumab 100 mg induction data were available; however, data regarding those individuals who were nonresponders at induction and those who were responders at maintenance were not reported, and as such, had to be imputed using data from the rerandomization phase. Golimumab 50 mg data regarding responses at week 6 were not available. Existing relationships between the available components were used to impute the expected proportions in this missing subpopulation. Data for placebo maintenance response were incomplete, as all induction nonresponders were assigned to golimumab 100 mg. Data from the PURSUIT trial were combined with ACT-1 and ULTRA-2 trial data to impute missing information.

Discussion: We have demonstrated methods for converting results from alternative study designs to more conventional parallel randomized clinical trials. These conversions allow for indirect treatment comparisons that are informed by a wider array of evidence, adding to the precision of estimates.

Keywords: adaptive, network meta-analysis, indirect treatment comparison, ulcerative colitis, golimumab

Introduction

Network meta-analysis (NMA) is an extension of conventional pairwise meta-analysis that allows for simultaneous comparison of multiple interventions.[1–4] When two interventions have not been compared head-to-head in randomized clinical trials (RCTs), but both have been compared to the same control interventions, NMAs can utilize such indirect evidence to establish an estimate of comparative efficacy between the two interventions. When NMAs are based solely on indirect evidence, they are known as indirect treatment comparisons (ITCs).[4–8] The validity of NMAs and ITCs hinges on the similarity of RCTs included in the analysis. In particular, it is crucial that the

Correspondence: Kristian Thorlund
Redwood Outcomes Health Consulting Inc., 302-1505 2nd Ave West, Vancouver, BC, V6H 3Y4 Canada
Email kthorlund@redwoodoutcomes.com

study designs and patient populations exhibit sufficient homogeneity to avoid confounding.[4,6]

In most common disease areas where pharmacotherapies are available (or where they are being developed), it is common that the efficacy of a class of drugs will become so well established that randomizing to placebo will be both unethical and pragmatically challenging (ie, no patient wants to be randomized to placebo when equipoise no longer exists).[9] However, head-to-head comparisons still remain rare due to the high cost of running RCTs with two expensive active agents, as well as the unfavorable risk–reward profiles of such RCTs for manufacturers of one of the active interventions. For these reasons, more efficient RCT designs, which allow patients to switch to the active interventions or higher doses, are now frequently being used. In the context of NMAs and indirect comparisons, this presents a methodological challenge, as data from such trial designs are not apparently combinable with data from conventional parallel design RCTs. Yet, simple mathematical conversions and extensions to the NMA statistical model can readily allow for a valid NMA incorporating data from RCTs that do not follow the conventional parallel design. We here describe such conversions and NMA model extensions in the case of antitumor

necrosis factor drugs (anti-TNFs) for moderately to severely active ulcerative colitis.

Illustrative example of trials assessing antitumor necrosis factor drugs (anti-TNFs) for ulcerative colitis

Three seminal trials have provided long-term (ie, 1-year) data on the efficacy of anti-TNFs when inducing remission or response in patients with moderately to severely active ulcerative colitis.[10–13] The ACT-1 trial and the ULTRA-2 trial, two parallel design trials, provided evidence for infliximab and adalimumab, respectively.[10,13] Both trials were included in a recent ITC.[14] More recently, the PURSUIT trial has provided evidence for golimumab (Merck & Co, Inc., Whitehouse Station, NJ, USA).[11,12] The PURSUIT trial, however, employed a sophisticated rerandomization scheme contingent on the induction response observed in patients who were initially randomized to either placebo or one of three loading doses of golimumab.[11,12] The trial design scheme of PURSUIT is displayed in Figure 1. The data format ensuing from this trial design does not lend itself directly to inclusion in NMAs. However, as shown in the following sections, enough data are available to approximate what the
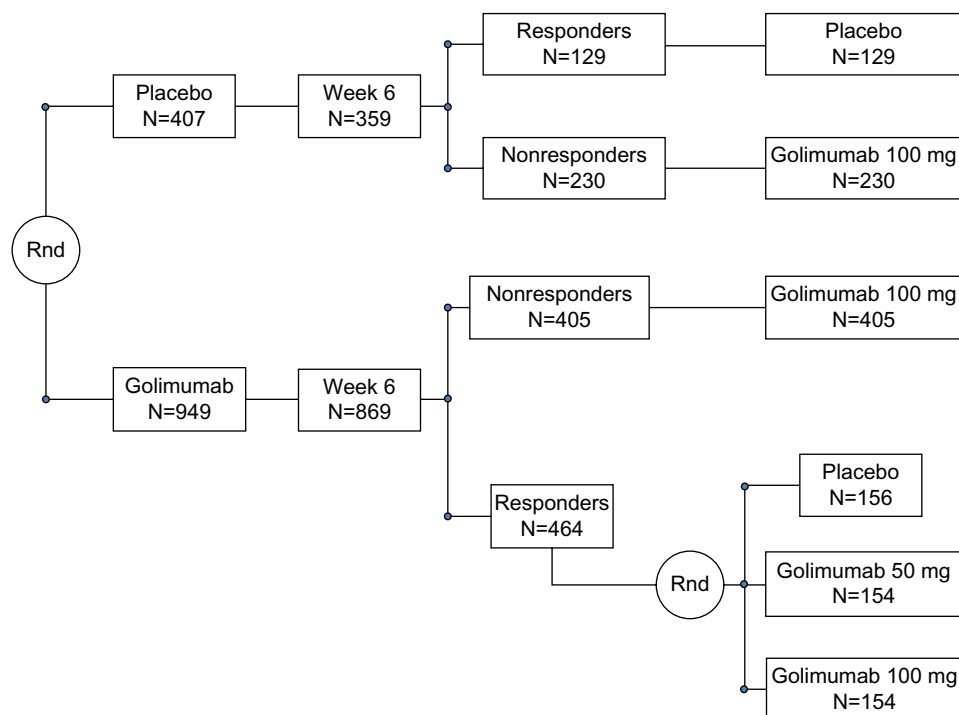
Figure 1 Schematic of the PURSUIT study design.
Notes: Circles containing "Rnd" indicates where Rnd of the patients occurred. The solid lines where no circle is present indicates where treatment continuation or switching with no Rnd occurred. The total number of patients that were randomized to continue or switch to a treatment is indicated with "N".
Abbreviations: N, number; Rnd, randomization.

results from PURSUIT would look like had the trial instead followed a conventional parallel arm study design.

Methods

While the data available from PURSUIT does not lend itself to direct valid incorporation into an indirect comparison meta-analysis, a number of simple mathematical conversions can be employed in order to convert PURSUIT into a parallel design RCT. To perform these conversions, some simple mathematical relationships are assumed. Further, incorporating evidence from external sources, in this particular case the ACT-1 and ULTRA-2 trials[10,13] can aid strongly in facilitating a valid ITC meta-analysis. Here we present a motivating example of how such methods can be employed in a Bayesian framework as well as the methodology used for each imputation.
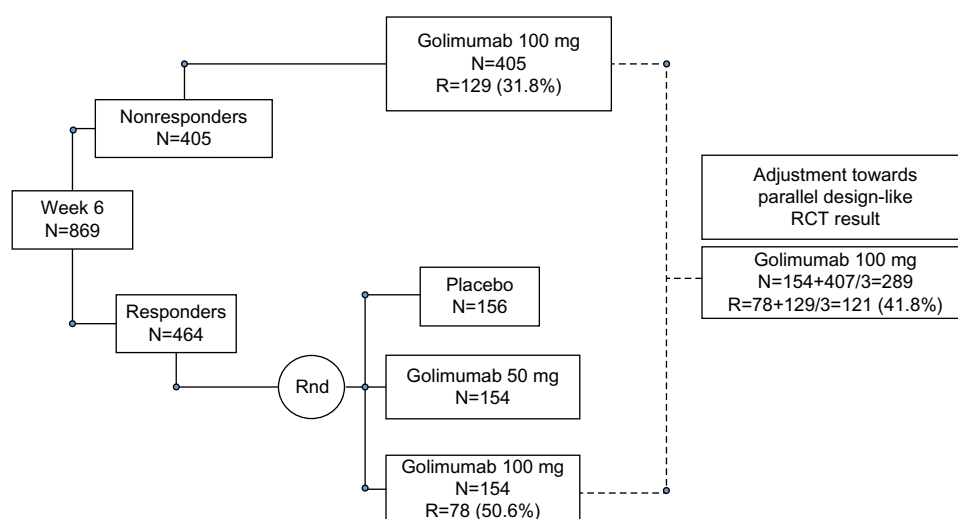
Golimumab 100 mg

In the PURSUIT trial,[12,13] patients were first randomized to receive either golimumab induction therapy or placebo induction therapy. Thus, we have data regarding the induction proportion for golimumab. Those who were nonresponders (NRs) after induction were immediately allocated to receive golimumab 100 mg (number [N] =407). Those who were considered responders (N=464) were rerandomized to receive golimumab 100 mg (N=154), golimumab 50 mg (N=154), or placebo (N=156). The reallocation and rerandomization of patients randomized to golimumab induction therapy is depicted in Figure 1. The information we require is overall maintenance proportions. In practice, maintenance response is a combination of the patients who responded at induction and the patients who did not respond at induction.

To approximate the proportion of patients who would have responded to golimumab 100 mg at maintenance if no reallocation or rerandomization had occurred after induction, the proportion of responders and NRs needs to be as similar at maintenance as after induction. With the rerandomization occurring among responders, the number of responders receiving golimumab 100 mg is diluted by a factor of three, relative to the total number of induction responders (ie, 154 patients versus 464 patients). Therefore, we are also required to dilute the number of NRs patients by a factor of three. That is, we assume that the number of NRs receiving golimumab 100 mg after induction is $407/3 \cong 135$, and dilute the observed number of events accordingly (ie, 129/3=43).[15] The observed number of maintenance responses among induction responders is 78 (50.6%), and so, when combining these numbers, we get a total proportion of 41.8% or 121/289 ([78 + 43]/[154 + 135]). All calculations are depicted in Figure 2, where "N" is the number of patients in that arm and "R" is the number of responders.

Golimumab 50 mg

With regards to golimumab 50 mg, no data are available for patients not responding at week 6 to golimumab induction therapy. Therefore, the existing relationships between the

Figure 2 Illustration of the conversion approach to golimumab 100 mg maintenance data.
Notes: Circles containing "Rnd" indicate where the Rnd of patients occurred. The solid lines where no circle is present indicate where treatment continuation or switching with no Rnd occurred. The dashed lines indicate from where and how golimumab 100 mg maintenance response is being estimated. The total number of patients who were randomized to either continue or switch to a treatment is indicated with "N", and the total number of responders is indicated with "R".
Abbreviations: N, number; R, responders; RCT, randomized clinical trial; Rnd, randomization.

available data components were used to impute the expected proportions in this missing subpopulation.

The premise of this imputation is that in patients receiving a 54-week course of golimumab, the relative difference in efficacy between induction responders and induction NRs (on a multiplicative scale) is assumed to be independent of the dose of golimumab given for 54 weeks. However, other assumptions can be applied, as discussed in the paragraphs to follow.

The PURSUIT trial[12,13] provides golimumab 50 mg maintenance data only for golimumab induction therapy responders. As seen earlier, the trial also provided data for golimumab 100 mg for both golimumab induction responders and golimumab induction NRs. To calculate the hypothetical proportion of responders among golimumab induction therapy NRs, one simply multiplies the relative efficacy between the two golimumab 100 mg groups (golimumab induction therapy responders and NRs) to the maintenance response of golimumab 50 mg, taking into account the downgrade of precision by assuming only one-third of induction therapy NRs were hypothetically randomized to golimumab 50 mg. In particular, as described earlier, we assumed that 405/3=135 induction NRs. The proportion of induction responders achieving maintenance response with golimumab 50 mg is 46.7%, but this proportion needs to be down-adjusted according to the relationship between induction NRs and responders. In induction NRs and responders receiving golimumab 100 mg, the maintenance proportions were 31.8% and 50.6%, respectively, thus yielding a relationship of 0.318/0.506=0.628. When this number is multiplied to the 46.7% maintenance response among patients rerandomized to golimumab 50 mg, we get an estimated maintenance response of $0.628 \times 0.467 = 0.293 = 29.3\%$. These calculations correspond to imputing 135 patients not responding to golimumab induction therapy, but of which $0.293 \ast 135 \cong 40$ patients responded to subsequent golimumab 50 mg maintenance therapy. A summary of these calculations is depicted in Figure 3.

As a sensitivity analysis, one can assume that the relationship among induction responders and induction NRs is not constant across golimumab doses. We can, for example, arbitrarily assume that the difference is exacerbated by 25%, yielding a relationship of 0.628/1.25=0.50, and corresponding to induction NRs doing relatively worse with a lower dose. We can also arbitrarily assume that the difference is attenuated by 25%, yielding a relationship of 0.628*1.25=0.785. Assuming a 25% exacerbation, the imputed number of events would be 135*0.50*0.467=32, yielding a (72 + 49)/289=41.8% response proportion. Assuming 25% attenuation, the imputed number of events would be 135*0.785*0.467=49, yielding a (72 + 32)/289=35.9% response proportion.

Placebo

A total of N=407 patients were randomized to placebo induction therapy, and a total of N=359 completed placebo induction therapy. Maintenance data for this treatment arm is incomplete, as no subpopulation response result was reported for patients not achieving response after 6 weeks of
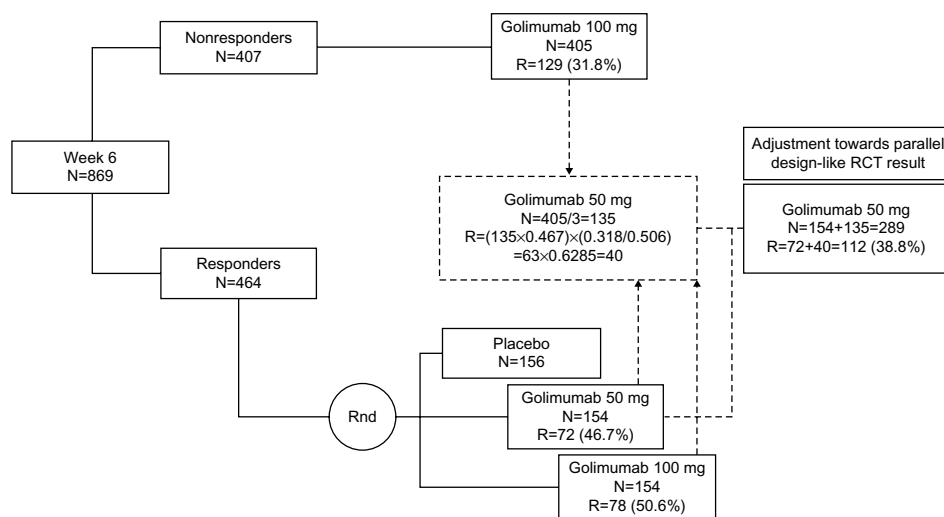
Figure 3 Illustration of the conversion approach to golimumab 50 mg maintenance data.
Notes: Circles containing "Rnd" indicate where the Rnd of patients occurred. The solid lines where no circle is present indicate where treatment continuation or switching with no Rnd occurred. The dashed lines indicate from where and how golimumab 50 mg maintenance response is being estimated. The total number of patients that were randomized to either continue or switch to a treatment is indicated with "N", and the total number of responders is indicated with "R".
Abbreviations: N, number; R, responders; RCT, randomized clinical trial; Rnd, randomization.

placebo induction therapy. This is because all such placebo induction NRs after 6 weeks (N=230) were assigned to golimumab 100 mg for the maintenance phase. However, the proportion of responders to the placebo induction therapy (N=129) that also responded at the end of maintenance was reported.[15] The proportion of responders at the end of maintenance therefore corresponds to an intention-to-treat (ITT) analysis, if we consider the NRs receiving golimumab 100 mg for maintenance therapy as dropouts. In other words, the total number of patients on placebo and responding at the end of maintenance, divided by the number originally randomized to placebo, would constitute an "ITT-like" placebo maintenance response proportion. Of course, the assumption that all these patients switching to golimumab 100 mg would have dropped out had they not switched is likely too strong. This assumption will likely cause the placebo maintenance response to be underestimated, since some of the 230 patients that switched to golimumab 100 mg could have later achieved response during the maintenance phase had they continued on placebo. To curb this downward bias, we propose combining the "ITT-like" observed proportion of maintenance responders with evidence from external sources – in particular, the ACT-1 and ULTRA-2 trials.[10,13]

In the Bayesian framework where the NMA is conducted, it is possible to incorporate external data in the form of prior distributions. In particular, the mean proportion of placebo responders from ACT-1 and ULTRA-2, as well as the uncertainty around this estimate, can be incorporated as a "prior" distribution of the placebo response in PURSUIT. NMA of binary outcomes (eg, response) is typically set up as a logistic regression. As such, when estimating the pooled logit proportion and associated standard error, a normal prior distribution can be assigned to the logit placebo response for PURSUIT. To do so, a fixed-effect meta-analysis of proportions (including ACT-1 and ULTRA-2) will provide a pooled proportion with 95% confidence intervals. These results can subsequently be transformed to the logit scale. The pooled estimate on the logit scale is the mean for the prior distribution, and the standard error on the logit scale can be approximated by dividing the width of the 95% confidence interval by 3.92. The corresponding variance (precision estimate) can then be used in the Bayesian model.

The specific calculations for the PURSUIT example are as follows. Out of the N=129 placebo induction therapy responders, a total of R=46 also maintained their response during placebo maintenance therapy. Because the total number of patients randomized to placebo was N=407, this yields an "ITT-like" placebo proportion of 46/407=11.3%. As expected, this proportion is notably lower than the placebo maintenance responses observed in the ACT-1 trial (24/121=19.8%) and in the ULTRA-2 trial (35/145=24.1%).[10,13] In the Bayesian framework, these sources of information can be combined by turning the ACT-1 and ULTRA-2 proportions into prior distributions and mixing them with the ITT-like placebo proportion from PURSUIT. The pooled proportion of the ACT-1 and ULTRA-2 proportions is 22.2% (95% credible interval [CrI]: 17.3%–27.7%), which roughly translates into a normal distribution for the corresponding logit proportion with a mean of −1.25 and a standard error of 0.153. Of course, one may want to assign less weight to this distribution as a prior in order to reflect the limited confidence that the ACT-1 and ULTRA-1 placebo maintenance responses are representative of the same population and trial design as that of the PURSUIT trial. To do so, the precision of the prior distribution can be deflated – or, in other words, the standard error can be inflated. We could, for example, inflate the standard error by a factor of two (corresponding to a variance inflation of $2^2=4$), in which case, the standard error of the prior would be 0.306. Now, we note that by the same logit transformations, the likelihood (data) distribution of the PURSUIT logit proportion corresponds to a normal distribution with a mean of −2.06 and a standard error of 0.163. In a Bayesian framework, the "mixing" of these two produces a posterior distribution for the logit proportion with an approximate mean of −1.88, with a standard error of 0.13, and where the corresponding estimated placebo proportion is 13.2% (95% CrI: 10.2%–16.8%). Figure 4 depicts how the combination of the prior distribution and the likelihood (data) distribution form the posterior distribution.

Comparison of results

Currently, one published NMA has compared the efficacy of golimumab 50 mg and 100 mg with infliximab and adalimumab using data only from the rerandomized golimumab induction therapy responders in the PURSUIT trial.[16] As we have argued, this data subset of the PURSUIT trial does represent a similar trial population as the populations from the ACT-1 and ULTRA-2 trials.[10,13] One clear indication that this is true is the 31.2% placebo maintenance response probability, which is considerably higher than the placebo maintenance response in ACT-1 and ULTRA-2. To gauge how an unadapted, or naïve use of the PURSUIT data versus the methods presented in this report can impact results, we calculated the odds ratio estimates and their associated 95% uncertainty intervals. The results are presented in Table 1.
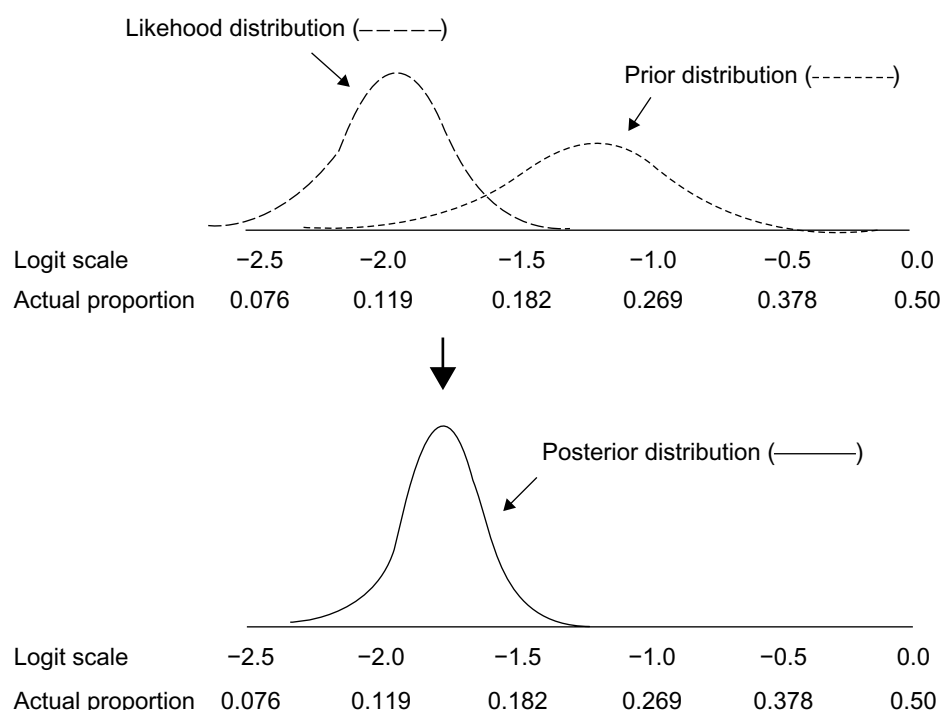
Figure 4 Illustration of the prior distribution for the placebo maintenance response shaped from the ACT-1 and ULTRA-2 data, mixed with the data from PURSUIT (the likelihood distribution).
Note: The data ultimately yield the final posterior distribution for the placebo maintenance response.

Discussion

Large, well-designed RCTs provide the basis for the investment of health care dollars and the shaping of clinical practice guidelines. However, if these data cannot be synthesized with typical trial designs, their use is substantially diminished. As alternative trial designs become more common in RCTs, systematic review investigators, health technology assessment specialists, and policy decision makers will be presented with unique challenges for incorporating these trials in NMAs and ITCs. Where newer generations of pharmacotherapies, such as biologics, demonstrate improved efficacy over the conventional standard of care, it becomes unethical to randomize patients to placebo, standard of care, or less efficient

doses of the treatment for an extended period. As such, an increasing number of trials will not include parallel design type long-term data. The PURSUIT trial provides one such example. The need for new methods to incorporate results from nonparallel design trials into ITC and NMA is therefore eminent.

We have illustrated methods for converting results from nonparallel rerandomization design RCTs to results corresponding to parallel design RCTs. While the method applied to the PURSUIT data is by no means generalizable to all nonparallel design trials that are, or will be, potentially eligible for inclusion in NMA, our wish is that the provided illustration can stimulate awareness and spawn developments

Table 1 Comparison of maintenance response proportions and odds ratio estimates for golimumab versus placebo using the proposed mathematic conversions (non-naïve approach) versus using only data from rerandomized golimumab induction responders (naïve approach)

	Proportions		Odds ratios	
	Naïve	**Non-naïve**	**Naïve**	**Non-naïve**
Placebo	31.2%	15.5%	–	–
Golimumab 50 mg				
Primary analysis	47.0%	38.8%	1.96 (1.23–3.12)	4.15 (2.88–5.98)
Sensitivity (0.75 × ratio)	–	35.9%	–	3.69 (2.55–5.32)
Sensitivity (1.25 × ratio)	–	41.8%	–	4.73 (3.28–6.80)
Golimumab 100 mg	49.7%	41.8%	2.18 (1.36–3.48)	4.73 (3.28–6.80)

Note: Data in the odds ratios, naïve and non-naïve columns are shown as odds ratio (95% credible interval).

of other solutions that are unique to other nonparallel design trials for other NMAs.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making*. 2013;33(5):607–617.
2. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value Health*. 2011;14(4):429–437.
3. Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. *BMJ*. 2013;346:f2914.
4. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ*. 2003;326(7387):472.
5. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50(6):683–691.
6. Wells GA, Sultan SA, Chen L, Khan M, Coyle D. *Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis*. Ottawa, ON: Canadian Agency for Drugs and Technologies in Health; 2009.
7. Jansen JP, Fleurence R, Devine B, 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 Health*. 2011;14(4):417–428.
8. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105–3124.
9. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med. 1987;317(3):141–145.
10. Rutgeerts P, Sandborn WJ, Feagan BG, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med*. 2005; 353(23):2462–2476.
11. Sandborn WJ, Feagan BG, Marano C, et al; PURSUIT-SC Study Group. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology*. 2014;146(1):85–95; quiz e14–e15.
12. Sandborn WJ, Feagan BG, Marano C, et al; PURSUIT-Maintenance Study Group. Subcutaneous golimumab maintains clinical response in patients with moderate-to-severe ulcerative colitis. *Gastroenterology*. 2014;146(1):96–109. e1.
13. Sandborn WJ, van Assche G, Reinisch W, et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology*. 2012;142(2):257–265. e1–e3.
14. Thorlund K, Druyts E, Mills EJ, Fedorak RN, Marshall JK. Adalimumab versus infliximab for the treatment of moderate to severe ulcerative colitis in adult patients naïve to anti-TNF therapy: an indirect treatment comparison meta-analysis. *J Crohns Colitis*. 2014;8(7):571–581.
15. Thorlund K. Incorporating adaptive clinical trials in network meta-analysis. Poster presented at: ISPOR 19th Annual Meeting; May 31-June 4, 2014; Montreal, Canada.
16. Stidham RW, Lee TC, Higgins PD, et al. Systematic review with network meta-analysis: the efficacy of anti-tumour necrosis factor-alpha agents for the treatment of ulcerative colitis. *Aliment Pharmacol Ther*. 2014;39(7):660–671.