Practical Issues When Conducting Network Meta-Analyses With a Limited Number of Studies

Dawn Odom, Costel Chirila, Beth Sherrill, Jianmin Wang, Bintu Sherif

RTI Health Solutions, Research Triangle Park, NC, United States

OBJECTIVES

- Meta-analysis, or the synthesis of results across multiple trials, is being conducted extensively in part due to requirements from health care decision making agencies. Often trials are designed to compare the experimental drug against placebo or best supportive care. In these cases, health care agencies may require as part of their review process a comparison against an active comparator, which can be accomplished using indirect comparison meta-analysis methods.
- Indirect comparison meta-analysis techniques continue to develop, and software now exists to model networks of randomized clinical trials (RCTs) using Bayesian or frequentist approaches with trial effects treated as fixed or random. The anchored indirect [treatment] comparison (AIC) method, which is not model based, is also suitable for making these treatment comparisons.
- Many meta-analyses are performed with a very limited number of studies. In a review of thousands of meta-analyses, Davey et al. (2011)[1] reported that over one-third included the minimum requirement of two studies only, and just under three-quarters contained five or fewer studies. However, practical issues emerge, particularly when the network comprises a limited number of studies.[2-4]
- Our goal was to investigate the performance and interpretation of different indirect comparison meta-analysis methods when few studies are available. Of special interest is the situation in which a star network contains only one trial for each given treatment comparison.

METHODS

- A star network based on two trials anchored by placebo was created for a binary endpoint (Figure 1).
- Using this network, the odds ratio from Trial #1 (Drug A vs. placebo) can be compared with the odds ratio from Trial #2 (Drug B vs. placebo) to create an indirect comparison of Drug A versus Drug B.
- Sample data were created for nine different scenarios in the network by varying sample size and placebo response (see Table 1).

Figure 1. A Star Network Comprising Two Trials

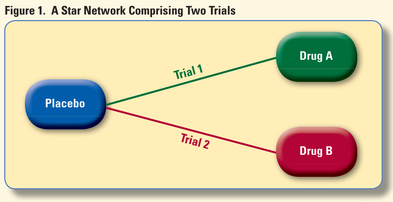


Table 1. Number of Events per Trial and per Drug, With Corresponding Odds Ratio, for Each Scenario

Scenario	Trial 1			Trial 2		
	Drug A	Placebo	OR	Drug B	Placebo	OR
1. Equal placebo response, same sample size	85/100 (85%)	45/100 (45%)	6.93	70/100 (70%)	45/100 (45%)	2.85
2. Small difference in placebo response, same sample size	85/100 (85%)	45/100 (45%)	6.93	70/100 (70%)	37/100 (37%)	3.97
3. Large difference in placebo response, same sample size	85/100 (85%)	45/100 (45%)	6.93	70/100 (70%)	32/100 (32%)	4.96
4. Equal placebo response, small difference in sample size, trial #2	85/100 (85%)	45/100 (45%)	6.93	42/60 (70%)	27/60 (45%)	2.85
5. Small difference in placebo response, small difference in size, trial #2	85/100 (85%)	45/100 (45%)	6.93	42/60 (70%)	22/60 (37%)	4.03
6. Large difference in placebo response, small difference in sample size, trial #2	85/100 (85%)	45/100 (45%)	6.93	42/60 (70%)	19/60 (32%)	5.04
7. Equal placebo response, large difference in sample size, trial #2	85/100 (85%)	45/100 (45%)	6.93	140/200 (70%)	90/200 (45%)	2.85
8. Small difference in placebo response, large difference in sample size, trial #2	85/100 (85%)	45/100 (45%)	6.93	140/200 (70%)	74/200 (37%)	3.97
9. Large difference in placebo response, large difference in sample size trial #2	85/100 (85%)	45/100 (45%)	6.93	140/200 (70%)	64/200 (32%)	4.96

OR = odds ratio.

- For each scenario, indirect-comparison meta-analyses examining the odds ratio of drug A versus drug B were produced using the following three methods:
  - Non-model based method
    1. AIC method[5,6]: The AIC method is equivalent to a generalized linear model with a logit link in which treatment and trial are fixed effects.
  - Model-based methods: For a binary endpoint that follows a binomial distribution (logit link) with a fixed effect for treatment and a random effect for trial. Note that the Bayesian approach and the frequentist approach both use the same fundamental statistical model, which is a generalized linear mixed model (GLMM). The primary difference is in the approach to parameter estimation.
    2. Bayesian approach using WinBUGS software (Bayesian inference using Gibbs sampling): Noninformative prior distributions used with 10,000 burn-in simulations followed by 10,000 simulations for estimation; model was thinned by a factor of 5.
    3. Frequentist approach using PROC GLIMMIX in SAS software (version 9.3, SAS Institute, Cary, North Carolina, USA): A no-intercept model with logit link, estimation method=RSPL, and degrees of freedom method=NONE. Note that if a random effect for trial was not estimable in a model, then it was treated as a fixed effect instead.
- The odds ratio estimates for the indirect comparisons are generally not expected to be equal in the AIC and model-based methods, since the AIC method assumes a fixed effect for trial and the model-based methods specify a random effect for trial.

RESULTS

- Estimated odds ratios from the indirect comparison (A vs. B) were examined to identify patterns of performance of the three methods across the nine scenarios (see Table 2 and Figure 2a-2c).

General Observations

Scenarios 1, 4, 7: Placebo Response Rates Were the Same across the Two Studies

- The GLIMMIX approach could not estimate a random effect for trial, and it was necessary to run a model with a fixed effect for trial. As expected, GLIMMIX and AIC results were identical since the AIC method is equivalent to the GLIMMIX approach.
- The WinBUGS model (with random effect for trial) produced odds ratios very similar to those produced by the GLIMMIX and AIC methods.
- The confidence intervals do not indicate a significant difference across the methods.

Scenarios 2, 3, 5, 6, 8, 9: Placebo Response Rates Were Not the Same Across the Two Studies

- The GLIMMIX approach produced consistently higher odds ratio estimates than the WinBUGS approach, which in turn produced consistently higher odds ratio estimates than the AIC method. Again, the confidence intervals overlapped between all three methods, indicating that they are not statistically different.
- Before conducting the analyses, we expected that the GLIMMIX and WinBUGS approaches would produce similar odds ratio estimates because they were based on the same fundamental model. However, slight differences in results were noted between the two approaches:
  - When the difference between the placebo response rates was "small" (45% vs. 37%), the odds ratio from the GLIMMIX approach demonstrated the greatest difference from the odds ratio produced by the WinBUGS approach (e.g., scenario 2: WinBUGS OR=1.87, GLIMMIX OR=2.24), compared with when the difference between the placebo response rates was "large" (45% vs 32%) (e.g., scenario 3: WinBUGS OR=1.52, GLIMMIX OR=1.63).

Scenarios 2, 5, 8: Placebo Response Rates Were Not the Same across the Two Studies, and the Sample Sizes Were Different

- When placebo response rates and sample sizes are different between studies, the GLIMMIX approach is more affected than the WinBUGS approach by fluctuations in sample sizes (e.g., scenario 5: WinBUGS OR=1.87, GLIMMIX OR=2.37; vs. scenario 8: WinBUGS OR=1.91 GLIMMIX OR=2.10).

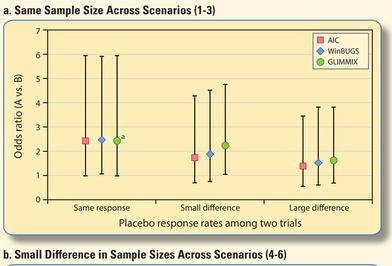
Table 2. Estimated Indirect Comparison Odds Ratio (95% Confidence Interval) of Drug A Versus Drug B for Each Scenario

Scenario	Analysis Method		
	AIC	WinBUGS (Bayesian GLMM)	GLIMMIX (Frequentist GLMM)
1. Equal placebo response, same sample size	2.429 (0.996-5.922)	2.471 (1.075-5.892)	2.429[a] (0.996-5.922)
2. Small difference in placebo response, same sample size	1.743 (0.711-4.274)	1.872 (0.763-4.505)	2.240 (1.059-4.740)
3. Large difference in placebo response, same sample size	1.397 (0.566-3.447)	1.524 (0.621-3.813)	1.632 (0.698-3.813)
4. Equal placebo response, small difference in sample size, trial #2	2.429 (0.885-6.668)	2.441 (0.928-6.538)	2.429[a] (0.885-6.668)
5. Small difference in placebo response, small difference in size, trial #2	1.719 (0.621-4.758)	1.873 (0.668-4.872)	2.374 (1.071-5.264)
6. Large difference in placebo response, small difference in sample size, trial #2	1.376 (0.492-3.846)	1.510 (0.533-4.325)	1.688 (0.658-4.333)
7. Equal placebo response, large difference in sample size, trial #2	2.429 (1.101-5.356)	2.463 (1.188-5.386)	2.429[a] (1.101-5.356)
8. Small difference in placebo response, large difference in sample size, trial #2	1.743 (0.788-3.856)	1.913 (0.861-4.156)	2.101 (1.038-4.252)
9. Large difference in placebo response, large difference in sample size trial #2	1.397 (0.629-3.101)	1.511 (0.682-3.417)	1.566 (0.729-3.368)

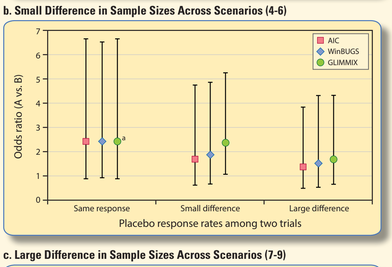
[a] A random effect for trial was not estimable and instead, a fixed effect for trial was used.

Figure 2. Indirect Comparison Odds Ratios of Drug A Versus Drug B

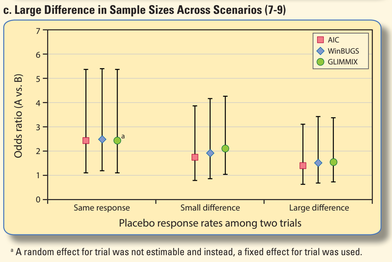
a. Same Sample Size Across Scenarios (1-3)



b. Small Difference in Sample Sizes Across Scenarios (4-6)



c. Large Difference in Sample Sizes Across Scenarios (7-9)



[a] A random effect for trial was not estimable and instead, a fixed effect for trial was used.

Other Observations

- For the frequentist GLIMMIX approach, the impact on the results based on the estimation method and degrees of freedom calculation method was also assessed:
  - Although Quadrature and Laplace estimation methods are recommended for binomial data,[7] the models using these methods often did not converge, and when they did, produced odds ratio estimates that were consistently higher than the odds ratio estimates obtained when RSPL estimation or Bayesian methods were used. The odds ratio estimates were closer as the sample size of individual trials increased (data not shown).
  - Kenward-Roger or residual methods for calculating denominator degrees of freedom produced consistently wider confidence intervals than method=None. The confidence intervals narrowed as the individual trial sample size increased (data not shown).

CONCLUSIONS

- The odds ratio point estimates for indirect comparisons in meta-analyses with small studies should be viewed with caution as differences may occur depending on methods used, assumptions, and underlying data patterns. The magnitude of difference depends not only on the various estimation approaches but also on (1) the actual differences between the rates of active treatments and placebo and (2) the relative sample sizes between the studies.
- When the placebo rate is the same across the individual studies, it is necessary to treat the trial effect as fixed in the GLMMIX approach; thus it is equivalent to the AIC method.
- Treating the trial effect as fixed will produce odds ratio estimates closest to the simple ratio of results from the two trials; the rationale and justification for using a random-effects method may not be supported when only two studies are available.
- With a limited network, a random-effects model using GLIMMIX or WinBUGS can produce different odds ratios estimates. As sample size increases within the individual studies, these estimates should become closer.
- The joint ISPOR-AMCP-NPC task force recently published a questionnaire to assess the relevance and credibility of indirect treatment comparisons.[8] The authors recommend using a random-effects model whenever possible but concede that it may not be feasible in certain situations such as our example network. In this case, a fixed-effect model is reasonable, but the authors prefer a random-effects model where one assumes a value for the study-level heterogeneity (i.e., random study effect).
- The limitations of meta-analyses with a small number of studies are well-documented. Researchers participating in indirect-comparison meta-analyses under these conditions should understand the differences in results when using the various methods and how interpretation of results is affected.

REFERENCES

1. Davey J, Turner RM, Clarke MJ, Higgins JP. Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: a cross-sectional, descriptive analysis. BMC Med Res Methodol. 2011 Nov 24;11:160.
2. Jackson D, Riley RD. A refined method for multivariate meta-analysis and meta-regression. Stat Med. 2014 Feb 20;33(4):541-54.
3. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol. 2014;14:25.
4. Follmann DA, Proschan MA. Valid inference in random effects meta-analysis. Biometrics. 1999;55(3):732-7.
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):683-91.
6. 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;326:472.
7. Jones B, Roger J, Lane PW, Lawton A, Fletcher C, Cappelleri JC; on behalf of PSI Health Technology Special Interest Group, Evidence Synthesis sub-team. Statistical approaches for conducting network meta-analysis in drug development. Pharm Stat. 2011;10:523-31.
8. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health. 2014 Mar;17(2):157-73.

ACKNOWLEDGMENTS

The authors wish to thank Adele Monroe and Candy Webster for editorial and design assistance in the preparation of this poster and would like to thank Christina Radder for her programming assistance.

CONTACT INFORMATION

Dawn Odom
Director, Biostatistics

RTI Health Solutions
200 Park Offices Drive
Research Triangle Park, NC 27709

Telephone: +919.316.3529
Fax: +919.541.7222
E-mail: dodom@rti.org

Presented at: ISPOR 19th Annual International Meeting
May 31-June 4, 2014
Montreal, QC, Canada