Indirect Comparisons of Therapeutic Interventions

Introduction

Clinicians, P&T committees, payers and others, such as content developers, are interested in the comparative efficacy of therapeutic interventions. When head-to-heat trials comparing interventions are not available, medical decision-makers frequently compare results from studies of active interventions compared to placebo. For example, they may compare results of intervention A and intervention B where each intervention has been compared to placebo in separate studies. Doing so, however, can be misleading.

Study Validity

Before results are considered, each study should be evaluated for validity (closeness to truth) by evaluating the study for bias and the likelihood of chance effects. Bias is defined as anything that systematically leads away from truth—meaning anything leading away from truth other than random chance (e.g., study groups with differing important baseline characteristics). Many researchers have studied the effects of bias on study results and have observed that bias frequently tends to favor the intervention under study, making the results appear bigger than they actually are, potentially making an intervention appear to be effective when it is not. [1-4] Major biases (e.g., failure of the randomization process to result in balanced groups for study, neglecting to conceal group assignment, lack of adequate blinding, etc.) can distort results by up to, and in some cases, more than 50%.

Many payers, P & T committees, clinicians and others do not perform critical appraisals to evaluate the risk of bias, and many are not aware of the need to do so. We and many others have found that many studies—even those published in journal with the best reputations—have significant threats to study validity, rendering the results untrustable, or are reported in such a way that validity is uncertain. A major study has shown that studies of uncertain validity report results that are similarly inflated to those that are judged as at high risk of bias, suggesting that studies of uncertain validity may be just as unreliable as biased studies with better reporting. [5]

Comparisons of Study Elements in Valid Studies

Once relevant studies that are likely to be valid have been identified, the use of what are referred to as PICPOT-SD elements can help distinguish between key differences in various studies which might render the simple comparison of study comparisons inappropriate. (Note: you may be more familiar with original acronym, PICPOTS which we have updated—see reference 6 for more information.)

Table 1. Delfini PICPOT-SD Table

Element	Meaning	Examples of Study Elements to Review
P	Patient population	Inclusions, exclusions, table of baseline characteristics noting such things as proportion of screened individuals enrolled, demographics, attrition before randomization (+ reasons), severity of condition, co-morbidities, etc.
I	Intervention	Dosing, frequency, monitoring, noting deviations from current practice, duration. Note likelihood of exposure.
C	Comparison	See Intervention, plus dose equivalency.
P	Performance outcomes of study	Training of staff and quality control. Presence or

		avoidance of key biases of those studies still passing a validity screening such as likely balance in study groups, success of blinding including blinded assessment, balance in co-interventions, adherence, protocol deviations, missing information, etc.
O	Outcomes	Chosen endpoints (definitions, surrogates, composites), individual items from composite outcomes, placebo event rates, etc.
T	Time issues	Concurrence in studied groups, treatment duration, follow-up duration, seasonal issues, changes over time such as in infectious disease issues, registry issues, etc.
S	Setting	Multicenter, single center, primary, secondary versus tertiary care centers, university setting, etc., noting differences for settings of interest.
D	Design of study	Experiment or observation, randomization, run-in periods to assess likelihood of nonadherence, application of intervention, care experiences, measurement methods, analysis methods, etc.

PICPOT-SD reminds us where differences in study elements are likely to be found. Therefore, PICPOT-SD can be used as a checklist to help guide comparisons of the efficacy results of studies of different interventions where each active intervention has been compared to placebo (indirect comparisons). Go to reference 6 for more background material.

Differences in the PICPOT-SD elements (heterogeneity) can explain differences in reported results of studies and, therefore, comparison of the PICPOT-SD elements is an important step when direct comparative efficacy evidence from valid randomized controlled trials (RCTs) is lacking. Judgment is required to assess how the differences in PICPOT-SD elements in studies being compared may affect the study results and, thus, the comparative efficacy of the interventions.

Steps for Indirect Comparisons of Studies of Interventions

The following are steps to accomplish indirect comparisons of studies of interventions. Only after doing these steps should one compare the risk reductions or numbers-needed-to-treat or other results reported in the trials.

Table 2. Steps for Indirect Comparisons of Studies of Interventions

Step 1: Select the studies for comparison.
Step 2: Critically appraise the relevant studies for reliability (internal validity) before seriously considering reported results.
Step 3: Assess how PICPOT-SD differences in studies being compared may affect comparative efficacy of the interventions.

There is no generally accepted method for accomplishing step 3. The role of the critical appraiser is to clarify how various PICPOT-SD factors may limit the applicability of each study's reported results to

target populations. Of interest to many working in the health care industry is the area of network meta-analyses because they are used with increasing frequency to compare interventions when head-to-head trials are not available.

Example of Application of PICPOT-SD: Comparing Studies With Populations Having Different Baseline Risks

For example, if a study of intervention A includes sicker patients with higher baseline risks for experiencing study endpoints, this should be noted along with an assessment of how this factor might affect the comparison of results of that study of intervention A with the results of another study of intervention B which includes healthier patients.

To illustrate this, assume that intervention A and intervention B are equally effective. In study A of intervention A, the placebo fracture rate is 20%. In study B of intervention B, assume the placebo fracture rate is 5%. Clearly one or more of the PICPOT-SD elements must differ in these two studies or the placebo event rates would be the same in both studies. The net effect is that the results will be more dramatic, the higher the incidence rate.

When populations are at higher risk, more people can benefit from effective interventions, thus resulting in higher absolute risk reductions as illustrated by Table 3. below. Using a scenario of equally effective interventions, but very different risk rates, the resulting absolute risk reduction rates are more impressive for the higher risk population—again, with equally effective interventions.

Table 3. Illustration of Variation in Absolute Risk Reduction With Equally Effective Interventions, But Populations With Varying Risk

Two valid studies with equally effective interventions ➜	Study of Intervention A with 10% effect but population with 20% Placebo Event Rate		Study of Intervention B with 10% effect but population with 5% Placebo Event Rate	
Categories	Intervention A	Placebo	Intervention B	Placebo
N	1000	1000	1000	1000
Resulting Fractures	180	200	45	50
No Fractures	820	800	955	950
Absolute Risk Reduction (ARR)	2%		0.5%	

In the above table, you can see that the ARR differs in the two studies, but intervention A and intervention B are equally effective. Therefore, some other element or elements must differ in the two trials. Biases or PICPOT-SD elements are where differences are likely to be found.

Summary

Comparative effectiveness reviews rely heavily on systematic reviews of technologies, pharmaceuticals and other health care interventions. When systematic reviews and head-to-head trials are not available, reviewers frequently compare results of separate studies where an active intervention is compared to placebo. These are referred to as indirect comparisons.

In such situations, assessing validity as well as differences in PICPOT-SD elements is critically important to avoid the error of accepting differences in benefits reported in the studies as due to differences in

efficacy of the interventions. If study results from different studies with different study populations, methodologies and other elements are compared without including these steps, readers are likely to draw incorrect conclusions regarding the comparative efficacy of the interventions being compared and make errors in estimating the effects of the interventions in their own population.

References

1. Juni P, Altman DG, Egger M (2001) Systematic reviews in health care: assessing the quality of controlled clinical trials. BMJ 2001;323:42-6.

2. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998 Aug 22;352(9128):609-13.

3. Savovic J, Jones HE, Altman DG, et al. Influence of Reported Study Design Characteristics on Intervention Effect Estimates From Randomized, Controlled Trials. Ann Intern Med. 2012 Sep 4. doi: 10.7326/0003-4819-157-6-201209180-00537. [Epub ahead of print]

4. Schulz KF, Chalmers I, Hayes RJ, Altman D. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408–12.

5. Hartling L, Ospina M, Liang Y et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. BMJ. 2009 Oct 19;339:b4012. doi: 10.1136/bmj.b4012.

6. 5 "A"s of Evidence-based Medicine & PICOTS: Using "Population, Intervention, Comparison, Outcomes, Timing, Setting" (PICOTS) In Evidence-Based Quality Improvement Work http://delfini.org/blog/?p=416