Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHODOLOGY

Interpreting discordant indirect and multiple treatment comparison meta-analyses: an evaluation of direct acting antivirals for chronic hepatitis C infection

This article was published in the following Dove Press journal:
Clinical Epidemiology
19 June 2013
Number of times this article has been viewed

Eric Druyts[1]
Kristian Thorlund[2,3]
Samantha Humphreys[4]
Michaela Lion[4]
Curtis L Cooper[5]
Edward J Mills[1,3]

[1]Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada; [2]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; [3]Stanford Prevention Research Centre, Department of Medicine, Stanford University, Palo Alto, CA, USA; [4]Merck Sharp and Dohme Ltd, UK; [5]Division of Infectious Diseases, Department of Medicine, The Ottawa Hospital, University of Ottawa, Ottawa, Ontario, Canada

Abstract: Indirect treatment comparison (ITC) and multiple treatment comparison (MTC) meta-analyses are increasingly being used to estimate the comparative effectiveness of interventions when head-to-head data do not exist. ITC meta-analyses can be conducted using simple methodology to compare two interventions. MTC meta-analyses can be conducted using more complex methodology, often employing Bayesian approaches, to compare multiple interventions. As the number of ITC and MTC meta-analyses increase, it is common to find multiple analyses evaluating the same interventions in similar therapeutic areas. Depending on the choice of the methodological approach, the conclusions about relative treatment efficacy may differ. Such situations create uncertainty for decision makers. An illustration of this is provided by four ITC and MTC meta-analyses assessing the efficacy of boceprevir and telaprevir for chronic hepatitis C virus infection. This paper examines why these evaluations provide discordant results by examining specific methodological issues that can strengthen or weaken inferences.

Keywords: indirect treatment comparison, multiple treatment comparison, meta-analysis, hepatitis C virus

Background

Indirect treatment comparison (ITC) and multiple treatment comparison (MTC) meta-analyses are relatively new approaches to evaluate the relative treatment effect when two or more interventions have not been compared directly.[1] These approaches are being increasingly used by health technology appraisal (HTA) agencies as new and existing drugs must be placed within the context of all available evidence for technology appraisals.[2] Many national authorities, including the National Institute for Health and Clinical Excellence in the UK and the Agency for Health Research and Quality in the US, have issued guidance on the conduct and reporting of ITC and MTC meta-analyses.[3,4] In addition, the International Society for Pharmacoeconomics and Outcomes Research has provided guidance for the undertaking and reporting of ITC and MTC meta-analyses.[5–7] An extension of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for systematic reviews is now being developed to incorporate ITC and MTC meta-analyses.[8]

Both ITC and MTC meta-analyses are relatively new statistical techniques that permit the comparison of treatments that may or may not have been compared directly in randomized controlled trials (RCTs).[9] MTC meta-analyses build on the simple ITC of two different treatments that have a mutual control, first reported by Bucher et al,[1]

Correspondence: Edward J Mills
Faculty of Health Sciences, University of Ottawa, 75, Laurier Ave East, Ottawa, Ontario, Canada K1N 6N5
Email edward.mills@uottawa.ca

to allow multiple interventions to be compared. The MTC approach has an inherent appeal for decision makers because it is a statistically valid tool to compare the relative effects of multiple interventions simultaneously.

The widespread interest in ITC and MTC meta-analyses has resulted in a multitude of such analyses in the published literature and in HTA submissions. In some circumstances, there will be multiple ITC or MTC meta-analyses evaluating the same interventions; however, the results reported may be inconsistent. For example, 13 published MTCs of biologics for rheumatoid arthritis were identified by Thorlund et al, of which several reported divergent results.[10] No guidance exists to assess methodologies that lead to different results. In this article, methodological issues that should be considered when interpreting comparative ITC or MTC meta-analyses are reviewed using the example of direct acting antiviral agents (DAAs) for chronic hepatitis C infection. Our discussion uses the basic principles of a guide to interpreting discordant systematic reviews, originally developed by Jadad et al,[11] and modifies it to the scenario of ITC and MTC meta-analyses.

Direct acting antivirals

Boceprevir and telaprevir, two new DAAs, were recently approved in Europe and North America for the treatment of chronic hepatitis C genotype 1 infection. These two treatments, when added to the peginterferon alpha and ribavirin combinations, are more efficacious than the peginterferon alpha and ribavirin combination alone.[12–20]

Clinicians are faced with the choice of which DAA to prescribe to their patients. In the absence of head-to-head RCTs comparing boceprevir and telaprevir, ITC and MTC meta-analyses have been conducted to determine the relative efficacy of these two DAAs. To date, four ITC or MTC meta-analyses comparing the relative efficacy of boceprevir and telaprevir have been presented in the peer reviewed literature.[21–24] However, there are key methodological differences between each of the ITC or MTC meta-analyses that have resulted in each coming to results about the relative efficacy of boceprevir and telaprevir that are discordant. These methodological differences are not necessarily apparent when first reviewing the ITC or MTC meta-analyses, and thus, the discordant results may be confusing to some readers.

Using this example, the application of ITC or MTC meta-analyses to assess the relative efficacy of boceprevir and telaprevir is discussed. The appraisal of a set of ITC and MTC meta-analyses and the underlying sources of observed discrepancies in findings are structured into seven main categories: the clinical question, the study selection and inclusion criteria, the outcomes definition and measurement, the statistical approach, the statistical models and heterogeneity, the effect measures, and the funding source. Further, the type of discordance observed in the results is categorized into three main categories: direction of the effect, magnitude of the effect, and statistical significance. Figure 1 displays the key considerations relevant to interpreting discordant reports. Table 1 reports the characteristics of each of the four publications.

Potential sources of discordance among published indirect and multiple treatment comparisons

Are the clinical questions similar?

For ITC and MTC reports to be potentially comparable, they need to include similar populations (P), interventions (I), controls (C), and outcomes (O). The use of PICO is relevant here as even small differences in the PICO may explain why findings across reports are different.

In the DAA example, each of the four reports assessed adult patients with chronic hepatitis C genotype 1 infection. The considered interventions were boceprevir or telaprevir in combination with standard of care (peginterferon alpha plus ribavirin) versus standard of care alone. The definition of control, however, differed between reports. In reports by Cooper et al[22] and Kieran et al,[24] peginterferon alpha-2a plus ribavirin and peginterferon alpha-2b plus ribavirin were considered to have equivalent treatment effects, and therefore were not evaluated separately in the analyses. In contrast, these two interventions were considered as separate in reports by Cooper et al[21] and Cure et al.[23] Of note, although large clinical trials and meta-analyses have indicated that there is no statistical difference between peginterferon alpha-2a plus ribavirin and peginterferon alpha-2b plus ribavirin, those patients provided with peginterferon alpha-2a plus ribavirin appear to fare slightly better in terms of clinically meaningful virologic end points.[25,26] All ITC and MTC meta-analyses assessed sustained virologic response (SVR) as their primary outcome.

Are the study selection and inclusion criteria similar?

Whether a report has included the same RCTs and treatment arms as other reports may explain why conclusions differ. If an ITC or MTC meta-analysis excludes certain trials or trial arms that another report has included, perhaps for reasons of study quality or otherwise, it may be reasonable to expect that results from these ITC or MTC meta-analyses will differ.

Potential sources of discordance
Clinical question (PICO) – are the clinical questions similar? • Patient population (P) – are the defined patient populations similar? • Interventions (I) – are the interventions similar? • Controls (C) – are the control interventions similar? • Outcomes (O) – are the chosen outcomes similar?
Study selection and inclusion criteria – are study selection and inclusion criteria similar? • Study design – are the considered study designs similar? • Literature search – are databases searched similar? • Selection criteria – are the included trials and intervention/comparator arms used in the primary analysis similar?
Outcomes definition and measurement – are outcomes defined and measured similarly? • Outcomes definition – are the outcomes defined similarly? • Methods to measure outcomes – are the methods used to measure the outcomes similar?
Statistical approach – are the statistical approaches used similar?
Statistical models and heterogeneity – are the statistical models and exploration of heterogeneity similar?
Effect measures – are the measures and statistics for establishing comparative superiority or inferiority similar?
Funding source – who funded each study?
Types of discordance
Results – are the results different? • Direction of effect – are there discrepancies in the direction of effect? • Magnitude of effect – are there discrepancies in the magnitude of effect? • Statistical significance – is there discordance among claims of statistical significance?
Interpretation – are the interpretations of similar results different?

Figure 1 Potential sources of discordance and types of discordance that should be investigated when evaluating indirect and multiple treatment comparisons.
Abbreviation: PICO, population, intervention, control, outcomes.

In the DAA example, all ITC and MTC meta-analyses included only RCTs assessing boceprevir or telaprevir plus standard of care versus standard of care alone. Comprehensive literature searches were conducted in each ITC and MTC, however, a number of discrepancies occurred between the ITC and MTC meta-analyses due to varying definitions of product labels. Cure et al[23] and Kieran et al[24] state that they only included RCT arms that corresponded with the approved product labels of boceprevir and telaprevir in their primary analysis (refer to Table S1 for a list of the approved product labels in Europe and North America). However, they do not explicitly state which country's product labels they are assessing. For example, in one analysis, Cure et al[23] has compared 48-week standard duration therapy (SDT) telaprevir to response guided (RGT) boceprevir for experienced patients classified as partial responders or relapsers. However, this is not a clinically meaningful comparison because prior relapsers are to be provided telaprevir RGT, not 48-week telaprevir SDT, according to both product labels in Europe and North America. Similarly, Kieran et al[24] pooled trial arms that both correspond and do not correspond to the approved product labels in Europe and North America.

The primary analyses in both studies by Cooper et al[21,22] examined the relative efficacy of boceprevir and telaprevir by including only 48-week SDT and RGT treatment arms. Such an analysis may only be meaningful in scenarios where the product label indicates that both boceprevir and telaprevir follow comparable treatment durations. For example, in the UK, such comparisons would be relevant for treatment naïve patients (where both boceprevir and telaprevir are given as RGT), treatment experienced prior partial responders, and prior null responders (where both boceprevir and telaprevir are given as SDT), but not treatment experienced prior relapsers (where boceprevir is given as SDT and telaprevir is given as RGT). (Note, however, that in the US, treatment experienced prior relapsing patients given either boceprevir or telaprevir will follow a RGT.)

Both Cure et al[23] and Kieran et al[24] assessed the quality of their included studies according to guidelines in the Cochrane Handbook for Systematic Reviews of Interventions.[27] Cure et al[23] deemed all included studies to be of an acceptable quality for analysis. Kieran et al[24] conducted a sensitivity analysis to determine if the removal of two studies deemed to have an increased risk of bias would impact the results, and found that these results were similar to those of the primary analysis. Both studies by Cooper et al[21,22] did not provide an assessment of study quality.

Are outcomes defined and measured similarly?

It is possible that the results of ITC and MTC analyses are inconsistent because the outcomes included may differ. To illustrate, one analysis may allow a surrogate outcome as

Table 1 Potential sources of discordance among the indirect and multiple treatment comparisons assessing direct acting antivirals

	Indirect and multiple treatment comparisons			
	Cooper et al[22]	**Cooper et al[21]**	**Cure et al[23]**	**Kieran et al[24]**
Clinical question (PICO)				
Are the defined patient populations similar?	Adult patients with chronic hepatitis C genotype 1 infection	Adult patients with chronic hepatitis C genotype 1 infection	Adult patients with chronic hepatitis C genotype 1 infection	Adult patients with chronic hepatitis C genotype 1 infection
Are the interventions similar? Are the control interventions similar?	• Boceprevir + peginterferon alpha + ribavirin versus peginterferon alpha + ribavirin • Telaprevir + peginterferon alpha + ribavirin versus peginterferon alpha + ribavirin	• Boceprevir + peginterferon alpha + ribavirin versus peginterferon alpha + ribavirin • Telaprevir + peginterferon alpha + ribavirin versus peginterferon alpha + ribavirin • Peginterferon alpha-2a + ribavirin versus peginterferon alpha-2b + ribavirin	• Boceprevir + peginterferon alpha + ribavirin versus peginterferon alpha + ribavirin • Telaprevir + peginterferon alpha + ribavirin versus peginterferon alpha + ribavirin • Peginterferon alpha-2a + ribavirin versus peginterferon alpha-2b + ribavirin	• Boceprevir + peginterferon alpha + ribavirin versus peginterferon alpha + ribavirin • Telaprevir + peginterferon alpha + ribavirin versus peginterferon alpha + ribavirin
Are the chosen outcomes similar?	Primary outcome: Sustained virologic response	Primary outcome: Sustained virologic response	Primary outcome: Sustained virologic response	Primary outcome: Sustained virologic response
Study selection and inclusion criteria				
Are the considered study designs similar?	Randomized controlled trials	Randomized controlled trials	Randomized controlled trials	Randomized controlled trials
Are databases searched similar?	MEDLINE, Embase, Cochrane Library, AMED, CINAHL, TOXNET, Development and Reproductive Toxicology, Hazardous Substances Data Bank, PsycINFO, Web of Science	MEDLINE, Embase, Cochrane Library, AMED, CINAHL, TOXNET, Development and Reproductive Toxicology, Hazardous Substances Data Bank, PsycINFO, Web of Science	MEDLINE, Embase, Cochrane Library, Center for Review and Dissemination	MEDLINE, Embase, Cochrane Library, Science Citation Index
Are the intervention/comparator arms used in the primary analysis similar?	Trials assessing only standard duration therapy and response guided therapy of boceprevir and telaprevir	Trials assessing only standard duration therapy and response guided therapy of boceprevir and telaprevir	Trials only assessing approved treatment durations of boceprevir and telaprevir	Trials only assessing approved treatment durations of boceprevir and telaprevir
Are the included trials of comparable quality?	Not assessed	Not assessed	Assessed using the criteria outlined in the Cochrane Handbook for Systematic Reviews of Interventions[27] The authors deemed all studies to be of acceptable quality and sufficiently homogenous	Assessed using the criteria outlined in the Cochrane Handbook for Systematic Reviews of Interventions.[27] In a sensitivity analysis, the authors excluded those deemed to have an increased risk of bias. The results of this analysis were not different from those of the primary analysis
Outcomes definition and measurement				
Are the outcomes defined similarly?	Primary outcome: • Sustained virologic response (undetectable HCV-RNA at the end of the 24-week posttherapy follow-up period)	Primary outcome: • Sustained virologic response (undetectable HCV-RNA at the end of the 24-week posttherapy follow-up period)	Primary outcome: • Sustained virologic response (undetectable HCV-RNA at the end of the 24-week posttherapy follow-up period)	Primary outcome: • Sustained virologic response (undetectable HCV-RNA at the end of the 24-week posttherapy follow-up period)

	Secondary outcomes: • Relapse (the reoccurrence of HCV-RNA within the 24-week posttherapy follow-up period) • Treatment discontinuation (the discontinuation of all assigned drugs during the set treatment period) • Anemia (hemoglobin $<$100 g/L) • Neutropenia (an absolute neutrophil count $<0.75 \times 10^9$/L) • Thrombocytopenia (a platelet count $<$150,000/mL) • Rash (any, as reported by site investigators) • Pruritus (any, as reported by site investigators)	Secondary outcomes: • Relapse (the reoccurrence of HCV-RNA within the 24-week posttherapy follow-up period) • Anemia (defined as hemoglobin $<$100 g/L) • Neutropenia (an absolute neutrophil count $<0.75 \times 10^9$/L) • Thrombocytopenia (a platelet count $<$150,000/mL) • Rash (any, as reported by site investigators) • Pruritus (any, as reported by site investigators)		
Are the methods used to measure the outcomes similar?	Primary outcome: • COBAS TaqMan HCV assay (Roche)	Primary outcome: • COBAS TaqMan HCV assay (Roche)	Primary outcome: • COBAS TaqMan HCV assay (Roche)	Primary outcome: • COBAS TaqMan HCV assay (Roche)
Statistical approach				
Are the statistical approaches used similar?	Adjusted indirect comparison	Bayesian multiple treatment comparison (odds ratio and credible intervals)	Bayesian multiple treatment comparison (odds ratio and credible intervals)	Bayesian multiple treatment comparison (odds ratio and credible intervals)
Statistical model and heterogeneity				
Are the statistical models and exploration of heterogeneity similar?	Random effects model	Random effects model when more than 1 trial arm was available and fixed effects model when only 1 trial arm was available	Fixed effects model (a secondary analysis for the random effects model to the base-case analysis)	Fixed effects model
Effect measures				
Are the measures and statistics for establishing comparative superiority or inferiority similar?	Relative risk	Odds ratio	Odds ratio	Odds ratio
Funding source				
Who funded each study?	Merck	Merck & Co	Janssen	Not stated

Abbreviations: HCV-RNA, hepatitis C virus RNA; PICO, population intervention control outcomes.

evidence of an event, whereas another report may require a clinical event or more stringent criteria of an event. For example, in an MTC evaluating smoking cessation therapies, Eisenberg et al only considered cessation events if they were continuous abstinence at specific time points, the most rigorous end point that can be used in smoking cessation.[28] Other ITC and MTC meta-analyses on this topic have included both continuous abstinence and point prevalence (an older but less reliable end point) and conducted sensitivity analyses to determine if choice of end point mattered.[29,30]

In the DAA example, all ITC and MTC meta-analyses used SVR as the primary outcome. SVR was consistently defined as an undetectable level of hepatitis C virus ribonucleic acid (HCV-RNA) at the end of the 24-week posttherapy follow-up period. HCV-RNA was measured using the COBAS TaqMan HCV-RNA assay in all the RCTs assessing boceprevir and telaprevir that were included in the ITC and MTC analyses.

Are the statistical approaches used similar?

The specific choice of ITC or MTC is both a choice of the authors and directed by the data. If there are no head-to-head trials available, an ITC should be conducted. If the analysis includes direct and indirect evidence, then an MTC is usually preferred. In some circumstances, an ITC will be conducted within a Bayesian framework, but the results should be nearly identical to a frequentist approach.

To assess the relative efficacy of the two DAAs, Cooper et al[22] and Kieran et al[24] modeled their analyses after the ITC approach displayed in Figure 2 (panel A). Their analyses focused on an ITC of boceprevir and telaprevir assuming that peginterferon alpha-2a and peginterferon alpha-2b provide comparable efficacy. The Cooper et al[21] and Cure et al[23] reports modeled their analyses of treatment naïve patients after the MTC approach displayed in panel B of Figure 2. There was insufficient information on the comparisons of peginterferon alpha-2a and peginterferon alpha-2b for the network of treatment experienced patients, and thus, these investigators modeled the analysis of treatment experienced patients on the ITC approach displayed in panel A of Figure 2.

Are the statistical models and heterogeneity exploration similar?

A recurring theme in meta-analytic studies, including ITC and MTC meta-analyses, is the choice between the fixed effect and random effects model when deciding whether to account for unexplained heterogeneity in the employed statistical model. The fixed effect model only accounts for one source of variation between results, within trial sampling error, whereas the random effects model accounts for variation between trials that is due to underlying differences of the trials rather than sampling error, ie, the random effects model includes an extra variance term in the model, and thus, has more variation associated with the estimated treatment effects. For this reason, a fixed effect model will always provide tighter confidence intervals (or credible intervals) around the treatment effect estimates when compared with the random effects model. The choice of one model over the other can often explain discordance in the statistical significance of observed treatment effects.

Of the DAA evaluations, Cooper et al[22] used a random effects model for all analyses. Cooper et al[21] also utilized a random effects model, but only for analyses that included more than one trial arm (since between trials variation

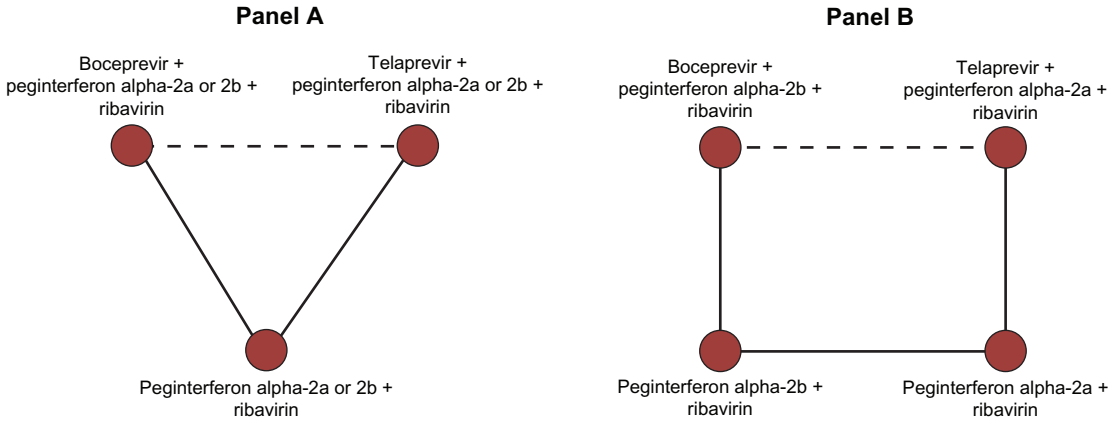

Figure 2 Indirect treatment comparison (**Panel A**) and multiple treatment comparison (**Panel B**) meta-analyses of boceprevir and telaprevir.
Notes: The solid lines between interventions represent direct evidence. The dashed lines between interventions represent indirect evidence.

cannot be estimated when only one trial is available). In Cooper et al,[21] a fixed effects model was therefore used when only one trial was available. The Cure et al[23] and Kieran et al[24] papers both used a fixed effect model for all analyses. In a secondary analysis, however, Cure et al[23] used the random effects model for the base case scenario, and did not note any differences in the results when compared to the fixed effects model. It should be noted that it is possible that by using only a fixed effect model, the results of the Cure et al[23] and Kieran et al[24] studies do not account for heterogeneity in patient populations across trials, and as a result, the 95% credible intervals could spuriously narrow if such heterogeneity exists.

Are the measures and statistics for establishing comparative superiority or inferiority similar?

The choice of effect measures can result in different interpretations of the same data. Using binary outcomes, the most commonly used effect measures in clinical medicine are odds ratios or relative risks. Odds ratios are statistically advantageous over relative risks, but less easy to interpret for clinicians.[31] Typically, odds ratios will be larger than relative risks and should not be interpreted similarly, therefore, when comparing across reviews the reader should be aware that a relative risk will be more conservative than odds ratios.

In addition to measures of relative effects, analysis using a Bayesian framework can report estimates of treatment rank probabilities. This type of measure is derived from the probability distributions around the associated odds ratios, and provides the probability of observing the largest odds ratio estimate with each treatment when sampling odds ratios from their probability distributions. While such probabilities can be valuable supplements to reported relative effects, it should be noted that they are subject to misinterpretation (eg, 75% probability of being best is not necessarily convincing) especially when conventional interpretation of significance is not presented. Further, treatment rank probabilities are highly sensitive to the data included and statistical models employed.

In the DAA example, Cooper et al[22] reported relative risk estimates to illustrate the comparative efficacy and safety for all considered outcomes. Cooper et al,[21] Kieran et al,[24] and Cure et al[23] all reported odds ratios. Only the study by Cure et al[23] reported treatment probabilities, and these treatment probabilities were a modification of the probabilities conventionally used in MTC meta-analyses. In particular, Cure et al[23] reported that the probability that the odds ratio between telaprevir and boceprevir was larger than 1, in contrast to conventional MTC meta-analyses that report probabilities of each intervention yielding larger effect estimates (eg, odds ratio) than all other considered treatments. The study found no significant difference in the primary analysis (odds ratio 1.42, 95% credible interval (CrI), 0.89–2.25), but because the probability of being the best favored telaprevir (93%) over boceprevir (7%), the study authors concluded in their abstract "an indirect comparison based on Bayesian network meta-analysis suggests better efficacy for telaprevir than boceprevir in both treatment-naïve and treatment-experienced patients."[23] This is an example of the problem of overinterpretation of treatment probabilities.

Are the results and interpretation of the results different?

The results of ITC and MTC meta-analyses can differ in two important domains: the actual calculated results of comparative efficacy and safety, and the accompanying interpretation of these results. While determining whether studies differ in terms of results is a matter of methodological approaches, the interpretation of study findings may be motivated by other factors. Therefore, readers should be aware of the funder of such reports and interpret study findings according to the data analysis findings rather than narrative conclusion.

Table 2 presents the results of each of the ITC and MTC meta-analyses, and the types of discordance that should be considered for the example of DAAs. For the analyses of treatment naïve patients, no statistical differences were observed between boceprevir and telaprevir in any of the ITC and MTC meta-analyses. Although the effect was not significant, the interpretation of nonsignificance is variable across reports.

For the analyses of treatment experienced patients, the reports by Cooper et al[21,22] found no statistical difference between boceprevir and telaprevir for both SDT and RGT. The results reported by Cure et al[23] favored telaprevir in treatment experienced patients overall and in the subgroup of patients who were prior relapsers. This result was only statistically significant when SDT telaprevir was compared to RGT boceprevir. However, SDT telaprevir is not the licensed treatment for prior relapsers in Europe and North America, nor is RGT boceprevir the licensed treatment for prior relapsers in Europe; therefore, this is not necessarily an appropriate or relevant comparison. Kieran et al[24] also reported results in favor of telaprevir for treatment experienced prior relapsers. This study combined both the RGT and SDT treatments for

Table 2 Discordance for the primary outcome (sustained virologic response) among the indirect and multiple treatment comparisons assessing direct acting antivirals

	Indirect and multiple treatment comparisons			
	Cooper et al[22]	**Cooper et al[21]**	**Cure et al[23]**	**Kieran et al[24]**
Analysis of treatment naïve patients				
Are there discrepancies in the direction of effect?	• Favors boceprevir (48-week SDT) • Boceprevir and telaprevir equivalent (RGT)	• Favors telaprevir (48-week SDT) • Favors telaprevir (RGT)	• Favors telaprevir (RGT) • Favors telaprevir (RGT telaprevir vs 48-week SDT boceprevir)	• Favors telaprevir (RGT telaprevir vs combined regimens boceprevir)
Are there discrepancies in the magnitude of effect?	• Small effect (48-week SDT) • No effect (RGT)	• Small effect (48-week SDT) • Small effect (RGT)	• Small effect (RGT) • Small effect (RGT telaprevir vs 48-week SDT boceprevir)	• Small effect (RGT telaprevir vs combined regimens boceprevir)
Is there discordance among claims of statistical significance?	• Not significant (48-week SDT) • Not significant (RGT)	• Not significant (48-week SDT) • Not significant (RGT)	• Not significant (RGT) • Not significant (RGT telaprevir vs 48-week SDT boceprevir)	• Not significant (RGT telaprevir vs combined regimens boceprevir)
Analysis of treatment experienced patients				
Are there discrepancies in the direction of effect?	• Favors telaprevir (48-week SDT)	• Favors telaprevir (48-week SDT)	• Favors telaprevir (48-week SDT) • Favors telaprevir (48-week SDT telaprevir vs RGT boceprevir)	• Favors telaprevir (48-week SDT telaprevir vs combined regimens boceprevir)
Are there discrepancies in the magnitude of effect?	• Small effect (48-week SDT)	• Small effect (48-week SDT)	• Small effect (48-week SDT) • Moderate effect (48-week SDT telaprevir vs RGT boceprevir)	• Small effect (48-week SDT telaprevir vs combined regimens boceprevir)
Is there discordance among claims of statistical significance?	• Not significant (48-week SDT)	• Not significant (48-week SDT)	• Not significant (48-week SDT) • Significant (48-week SDT telaprevir vs RGT boceprevir)	• Not significant (48-week SDT telaprevir vs combined regimens boceprevir)
Subgroup analysis of treatment experienced prior relapsers				
Are there discrepancies in the direction of effect?	• Favors telaprevir (48-week SDT)	–	• Favors telaprevir (48-week SDT) • Favors telaprevir (48-week SDT telaprevir vs RGT boceprevir)	• Favors telaprevir (48-week SDT telaprevir vs combined regimens boceprevir)
Are there discrepancies in the magnitude of effect?	• Small effect (48-week SDT)	–	• Moderate effect (48-week SDT) • Moderate effect (48-week SDT telaprevir vs RGT boceprevir)	• Moderate effect (48-week SDT telaprevir vs combined regimens boceprevir)
Is there discordance among claims of statistical significance?	• Not significant (48-week SDT)	–	• Not significant (48-week SDT) • Significant (48-week SDT telaprevir vs RGT boceprevir)	• Significant (48-week SDT telaprevir vs combined regimens boceprevir)
Subgroup analysis of treatment experienced prior partial responders				
Are there discrepancies in the direction of effect?	–	–	• Favors boceprevir (48-week SDT) • Favors boceprevir (48-week SDT telaprevir vs RGT boceprevir)	–
Are there discrepancies in the magnitude of effect?	–	–	• Small effect (48-week SDT) • Small effect (48-week SDT telaprevir vs RGT boceprevir)	–

Is there discordance among claims of statistical significance?	–	–	–	• Not significant (48-week SDT) • Not significant (48-week SDT telaprevir vs RGT boceprevir)

Notes: We arbitrarily defined the magnitude of effect sizes as follows: Small: Odds Ratio (OR) 1.00–1.99; RR 1.00–1.49. Moderate: OR 2.00–4.99; RR 1.50–2.99. Large: OR ≥ 5.00; RR ≥3.00.
Abbreviations: RGT, response guided therapy; SDT, standard duration therapy; vs, versus.

boceprevir, which is not a clinically meaningful comparison when considering the licensed treatment regimens.

Conclusion

Interpreting discordant ITC and MTC meta-analyses requires careful consideration of a variety of methodological factors, including, but not limited to, the clinical question, study selection and inclusion, data extraction, data analysis, and presentation of results. Each ITC and MTC meta-analysis assessing the relative efficacy of two DAAs included comparable patient populations, interventions, and primary outcome measures. Each ITC and MTC meta-analysis included only RCTs, and conducted rigorous database searches to identify eligible studies. However, the ITC and MTC meta-analyses diverged with regards to the selection criteria used to identify the trial arms to be included in the analyses, the statistical methods employed in the data analysis, and the interpretation of the study findings. This paper represents a step forward in interpreting divergent ITC and MTC meta-analyses and results.

Acknowledgments

The academic researchers received funding from Merck Sharp and Dohme to support this study. No restrictions have been placed on the analysis or the interpretation on the results.

Author contributions

ED, KT, CLC, and EJM conceived the design of the study. ED drafted the first manuscript. ED, SH, ML extracted the necessary data. All authors contributed equally to the interpretation and final write up of the manuscript.

Disclosure

KT and EJM have consulted to Merck Sharp and Dohme, Inc, Pfizer Ltd, Novartis, or Takeda on MTC issues. KT and EJM have received grant funding from the Canadian Institutes of Health Research (CIHR) Drug Safety and Effectiveness Network (DSEN) to develop methods and educational materials on MTCs. DSEN had no role in the design and conduct of the study. EJM receives salary support from the CIHR through a Canada Research Chair. KT receives salary support from the CIHR DSEN Netman project. SH and ML are employees of Merck Sharp and Dohme. ED and CLC report no conflicts of interest related to this work.

References

1. Bucher H, Guyatt G, Griffith L, Walter S. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50(6):683–691.

2. Sutton A, Ades A, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics*. 2008;26(9):753–767.
3. Wells G, Sultan S, Chen L, Khan M, Coyle D. *Indirect Evidence: Indirect Treatment Comparisons In Meta-Analysis*. Ottawa, Ontario: Canadian Agency for Drugs and Technologies in Health; 2009.
4. Fu R, Gartlehner G, Grant M, et al. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol*. 2011;64(11):1187–1197.
5. Hoaglin D, Hawkins N, Jansen J, 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.
6. Jansen J, 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.
7. Dias S, Welton N, Sutton A, Ades A. *NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials*. 2011. Available from: http://www.nicedsu.org.uk/TSD2%20General%20meta%20analysis%20corrected%20Mar2013.pdf. Accessed February 12, 2013.
8. Liberati A, Altman D, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med*. 2009;151(4):W65–W94.
9. Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105–3124.
10. Thorlund K, Druyts E, Avina-Zubieta J, Wu P, Mills E. Why the findings of published multiple treatment comparison meta-analyses of biologic treatments for rheumatoid arthritis are different: an overview of recurrent methodological shortcomings. *Ann Rheum Dis*. Epub October 20, 2012.
11. Jadad A, Cook D, Browman G. A guide to interpreting discordant systematic reviews. *Can Med Assoc J*. 1997;156(10):1411–1416.
12. McHutchison J, Everson G, Gordon S, et al; PROVE1 Study Team. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N Engl J Med*. 2009;360:1827–1838.
13. Hezode C, Forestier N, Dusheiko G, et al; PROVE2 Study Team. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N Engl J Med*. 2009;360:1839–1850.
14. Jacobson I, McHutchison J, Dusheiko G, et al; ADVANCE Study Team. Telaprevir for previously untreated chronic hepatitis C virus infection. *N Engl J Med*. 2011;364:2405–2416.
15. McHutchison J, Manns M, Muir A, et al; PROVE3 Study Team. Telaprevir for previously treated chronic HCV infection. *N Engl J Med*. 2010;362:1292–1303.
16. Zeuzem S, Andreone P, Pol S, et al; REALIZE Study Team. Telaprevir for retreatment of HCV infection. *N Engl J Med*. 2011;364: 2417–2428.
17. Kwo P, Lawitz E, McCone J, et al; SPRINT-1 Investigators. Efficacy of boceprevir, an NS3 protease inhibitor, in combination with peginterferon alfa-2b and ribavirin in treatment-naive patients with genotype 1 hepatitis C infection (SPRINT-1): an open-label, randomised, multicentre phase 2 trial. *Lancet*. 2010;376:705–716.
18. Poordad F, McCone J, Bacon B, et al; SPRINT-2 Investigators. Boceprevir for untreated chronic HCV genotype 1 infection. *N Engl J Med*. 2011;364:1195–1206.
19. Bacon B, Gordon S, Lawitz E, et al; HCV RESPOND-2 Investigators. Boceprevir for previously treated chronic HCV genotype 1 infection. *N Engl J Med*. 2011;364:1207–1217.
20. Flamm S, Lawitz E, Jacobson I, et al. Boceprevir with peginterferon alfa-2a-ribavirin is effective for previously treated chronic hepatitis C genotype 1 infection. *Clin Gastroenterol Hepatol*. 2013;11(1):81–87.
21. Cooper C, Lester R, Thorlund K, et al. Direct-acting antiviral therapies for hepatitis C genotype 1 infection: a multiple treatment comparison meta-analysis. *QJM*. 2013;106(2):153–163.
22. Cooper C, Druyts E, Thorlund K, et al. Boceprevir and telaprevir for the treatment of chronic hepatitis C genotype 1 infection: an indirect comparison meta-analysis. *Ther Clin Risk Manag*. 2012;8: 105–130.
23. Cure S, Diels J, Gavart S, Bianic F, Jones E. Efficacy of telaprevir and boceprevir in treatment-naive and treatment-experienced genotype 1 chronic hepatitis C patients: an indirect comparison using Bayesian network meta-analysis. *Curr Med Res Opin*. 2012;28(11): 1841–1856.
24. Kieran J, Schmitz S, O'Leary A, et al. The relative efficacy of boceprevir and telaprevir in the treatment of hepatitis C virus genotype 1. *Clininical Infectious Diseases*. Jan 2013;56(2):228–235.
25. McHutchison J, Lawitz E, Shiffman M, et al; IDEAL Study team. Peginterferon alfa-2b or alfa-2a with ribavirin for treatment of hepatitis C infection. *N Engl J Med*. 2009;361:580–593.
26. Druyts E, Mills E, Nachega J, O'Regan C, Cooper C. Differences in clinical outcomes among hepatitis C genotype 1-infected patients treated with peginterferon alpha-2a or peginterferon alpha-2b plus ribavirin: a meta-analysis. *Clin Exp Gastroenterol*. 2012;5:11–21.
27. Higgins JPT, Green S, editors. *The Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011. Available from: http://www.cochranehandbook.org. Accessed May 6, 2013.
28. Eisenberg M, Filion K, Yavin D, et al. Pharmacotherapies for smoking cessation: a meta-analysis of randomized controlled trials. *CMAJ*. 2008;179(2):135–144.
29. Mills E, Wu P, Lockhart I, Thorlund K, Puhan M, Ebbert J. Comparisons of high dose and combination nicotine replacement therapy, varenicline and bupropion for smoking cessation: a systematic review and multiple treatment meta-analysis. *Ann Med*. 2012;44(4):588–597.
30. Mills E, Wu P, Spurden D, Ebbert J, Wilson K. Efficacy of pharmacotherapies for short-term smoking abstinance: a systematic review and meta-analysis. *Harm Reduct J*. 2009;6:25.
31. Walter S. Choice of effect measure for epidemiological data. *J Clinical Epidemiol*. 2000;53(9):931–939.

Supplementary table

Table S1 Approved treatment labels for boceprevir and telaprevir in Europe and North America

	Europe		North America	
	Boceprevir	**Telaprevir**	**Boceprevir**	**Telaprevir**
Treatment naïve	PR 1–4 → BPR 5–28 or PR 1–4 → BPR 5–36 → PR 37–48*	TPR 1–12 → PR 13–24 or TPR 1–12 → PR 13–48*	PR 1–4 → BPR 5–28 or PR 1–4 → BPR 5–36 → PR 37–48*	TPR 1–12 → PR 13–24 or TPR 1–12 → PR 13–48*
Treatment experienced – prior relapsers	PR 1–4 → BPR 5–36 → PR 37–48	TPR 1–12 → PR 13–24 or TPR 1–12 → PR 13–48*	PR 1–4 → BPR 5–36 or PR 1–4 → BPR 5–36 → PR 37–48*	TPR 1–12 → PR 13–24 or TPR 1–12 → PR 13–48*
Treatment experienced – prior partial responders	PR 1–4 → BPR 5–36 → PR 37–48	TPR 1–12 → PR 13–48	PR 1–4 → BPR 5–36 or PR 1–4 → BPR 5–36 → PR 37–48*	TPR 1–12 → PR 13–48
Treatment experienced – prior null responders	PR 1–4 → BPR 5–48	TPR 1–12 → PR 13–48	PR 1–4 → BPR 5–48	TPR 1–12 → PR 13–48
All cirrhotic patients	PR 1–4 → BPR 5–48	TPR 1–12 → PR 13–48	PR 1–4 → BPR 5–48	TPR 1–12 → PR 13–48

Note: * denotes response guided therapy.
Abbreviations: PR, Peginterferon–ribavirin; BRP, Boceprevir + peginterferon-ribavirin; TPR, Telaprevir + peginterferon–ribavirin.