Author’s response to reviews

Title: Antiviral treatment for treatment-naïve chronic hepatitis B: Systematic review and network meta-analysis of randomized controlled trials

Authors:

William WL Wong (wwlwong@uwaterloo.ca)
Petros Pechivanoglou (petros.pechlivanoglou@sickkids.ca)
Josephine Wong (josephine.wong@theta.utoronto.ca)
Joanna Bielecki (joanna.bielecki@theta.utoronto.ca)
Alex Haines (alex.haines@theta.utoronto.ca)
Aysegul Erman (aysegul.erman@mail.utoronto.ca)
Yasmin Saeed (yasmin.saeed@mail.utoronto.ca)
Arcturus Phoon (rcrtrphn056@gmail.com)
Mina Tadrous (TadrousM@smh.ca)
Mona Younis (mona.younis@westmountpharmacy.ca)
Noha Rayad (drnohazaki@gmail.com)
Valeria Rac (valeria.rac@theta.utoronto.ca)
Harry Janssen (harry.janssen@uhn.ca)
Murray Krahn (murray.krahn@theta.utoronto.ca)

Version: 1 Date: 13 May 2019

Author’s response to reviews:

First, we would like to thank both editor and reviewers for taking the time to read our submission and providing constructive feedback. Based on the comments provided, we have made major revision to improve its clarity. Our responses are detailed below. We hope that this response letter addresses any issues with our submission.
Reviewer #1

Comment 1:

This systematic review is not registered. Lack of transparency. Need registration as per the journal policy.

Response:

We agree with the reviewer that systematic review should be registered and be transparent.

The protocol for this systematic review was an extension of our earlier systematic review carried out under the sponsorship of the Ontario Drug Policy Research Network (ODPRN). Our original protocol is publicly available in the ODPRN web site since 2015.


In the present systematic review, we updated the search to 2017 and included some new interventions.

Location of edits: Page 5, Line no 110-112

Comment #2:

Who are Two reviewers, need to be more specific with initials? "Two reviewers independently screened the titles and abstracts of the identified studies to determine if they met the inclusion criteria"

"Disagreements between the two reviewers were resolved by discussion"

Response:

We have inserted the initials of the two reviewers and revised the text as follows: “Two reviewers (AE and YS) independently screened the titles and abstracts of the identified studies to determine if they met the inclusion criteria"

"Disagreements between the two reviewers (AE and YS) were resolved by discussion"
Comment #3

It will be better to show kappa for the selection and data extraction. Please show the data of kappa of agreement during the systematic searches. How disagreements were solved during the systematic search among two independent reviewers?

Response:

We thank the reviewer for this helpful suggestion. We have edited the text as follows: “A total of 6319 studies were identified from the original literature (Appendix A). After screening the titles and abstracts, 1040 potentially relevant publications were retrieved for full text review (kappa=0.733; good agreement). After full text review, 42 publications were selected for inclusion (kappa=0.94; very good agreement).”

Comment #4:

How authors dealt with missing data. Did you receive all answers from authors of the studies or make some imputation?

Response:

We did not contact the authors of the included studies for missing data. Not all the studies reported all the outcomes. For each outcome, we only included those studies which reported the outcome in the analysis.
Comment #5:

It will be better to describe the excluded article after assessing for eligibility using the full text, and reference it in the manuscript.

Response:

We thank the reviewer for this suggestion. We have added a description of the excluded articles in the manuscript as follows: “As for the 998 articles excluded, the reasons for exclusion were: lack of comparators or control (36%), no outcomes of our interest reported (28%), duplicate papers, secondary analysis, modeling studies or abstracts (20%), and studies on special subsets of patients such as co-infections, decompensated cirrhosis, hepatocellular carcinoma, hemodialysis, and transplant (5%).”

Location of edits: Page 12, Line no 255-259

Comment #6:

Authors should discuss the reason of heterogeneity.

Response:

We thank the reviewer for this suggestion. We have briefly mentioned heterogeneity under the section on Endpoint definitions (p.6). We have edited the Discussion (p.20) to elaborate on heterogeneity as follows: “Secondly, in order to include as many interventions as possible to inform the network, we did not specify participants’ baseline characteristics in our eligibility criteria. This could be an important source of heterogeneity. However, we tried to reduce baseline variability by excluding studies with decompensated cirrhosis, hepatocellular carcinoma, and liver transplant cohorts. …… This variation in baseline clinical characteristics is a recognized source of heterogeneity in NMA.”

Location of edits: Page 20, Line no 447-461
Reviewer #2

Comment #1:

The description of the data analysis methods needs to be more comprehensive so that the appropriateness of the methods can be judged:

a) It is stated that WinBUGS code was used, but a reference for this code is not given, nor is the code provided for review. Many different codes are available or could be written by the authors some of which may be appropriate and some may not. The authors should either, cite the source of the code used, or if they used custom written code, this should be made available for review. Ideally, all code should also be provided as Supplementary material.

Response:

We thank the reviewer for this suggestion. We have rerun all the analyses and revise the section to address the concern accordingly:

In the originally submitted manuscript, we relied on peer-reviewed publically available code to conduct the NMA. However, for this revised version of the manuscript, we rely on the package “gemtc” and JAGS to conduct the network meta-analysis. We have revised the paragraph as suggested:

“Network meta-analysis (NMA) methods were used to synthesize the evidence from the RCTs on relative effectiveness across the treatment options. Binomial likelihood functions were assumed for the binary events of interest, and a random effects models was used for all the outcomes. The NMA was conducted within a Bayesian framework using JAGS v.4.3.0 (18) and the R package “gemtc”.(19) Under this framework the distribution of each parameter of interest (posterior distribution) was estimated through a Markov Chain Monte Carlo simulation method”

Location of file: Page 9, Line no 182-190

Comment #2:

The authors state that they considered both fixed and random effects "based on the data availability in the included studies" - this statement does not make it clear how fixed or random effects models were chosen and in fact, it is unclear throughout the whole manuscript which models were fitted and if there was any heterogeneity. The authors should detail the model selection criteria used and provide an assessment of model fit for their chosen model.
Response:

We relied on random effect models as they were conceptually more appropriate for the synthesis of possibly heterogeneous data. For each outcome, a common heterogeneity parameter was assumed across treatment comparisons. We compared the Deviance information criterion (DIC) between a fixed and a random effect model for each outcome in a sensitivity analysis. An appendix F has been created to illustrate the fixed effect model.

Location of file: Appendix F

Comment #3:

When using Markov chain Monte Carlo methods, details on how convergence of the algorithms was assessed and the number of burn-in and subsequent sample iterations must be provided.

Response:

We assumed a burn-in of 20,000 and we ran three Markov Chains per outcome. We sampled 50,000 values from each chain after convergence has been achieved. We tested convergence using the Gelman and Rubin convergence diagnostic. We have revised the paragraph as suggested:

“Under this framework the distribution of each parameter of interest (posterior distribution) was estimated through a Markov Chain Monte Carlo simulation method. We assumed three chains and performed 70,000 simulations for each chain for each outcome, and we excluded the first 20,000 simulations to ensure that we selected only converged values. We further assessed convergence for all models through the Gelman and Rubin diagnostic.(20)”

Location of file: Page 9, Line 188-193

Comment #4:

Full details of the model and prior distributions used must be given (perhaps in a technical appendix with references, where appropriate)
Response: We have revised the paragraph to include this important information:

“Vague priors were assumed throughout the model. Normal priors were assumed for treatment
effectiveness parameters with a mean zero and a variance of 10,000, while a uniform prior
U(0,2) was assumed for the standard deviation of the heterogeneity parameter.”

Location of file: Page 9, Line 193-196

Comment #5:

The authors chose to assess inconsistency by looking at all loops and comparing results using
"direct" and "indirect" information within those loops. However, it is not made clear how the
"direct" estimates were obtained (or the estimates used to calculate the indirect estimates). In
addition, the selected method has the following important drawbacks: it does not consider
"indirect" evidence provided by the whole network, but only the evidence on a specific loop,
which typically will not encompass all important indirect evidence for all comparisons

Response:

We thank the reviewers for the comments. We have shifted towards using the gemtc R package
to better address inconsistency. Now, we are using a "node-splitting" method to assess
inconsistency. The node splitting method provided in the gemtc package, to our understanding,
does not consider inconsistency that is occurring due to multi-arm trials. Once multi-arm trials
are excluded, some outcomes in our analysis do not have an estimate of inconsistency attached to
them. We present all the new inconsistency results in Appendix E.

Location of file: Appendix E

Comment #6:

if a random effects model is chosen for the network meta-analysis, different estimates of
heterogeneity may be calculated for each direct and indirect piece of evidence used, with some of
these analysis possibly using fixed effect models, which means the comparisons made will not
reflect the true variability in the network.
The authors should consider using a different consistency checking method such as the node-split method (which is implemented in the gemtc package in R - see eg Dias et al https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704208/, for a description of other methods). In any case, more details need to be provided on the method used.

Response:

We thank the reviewers for the suggestions. We have used the node-splitting method in the gemtc package. We have followed the recommendations and all inconsistency estimation was now done through the R package. We have revised the inconsistency paragraph:

“An important assumption underlying NMA is that of consistency between the direct and indirect evidence of relative efficacy. Inconsistency evaluates the validity of a network meta-analysis by assessing the compatibility of direct and indirect evidence. If the results from direct evidence conflict against the results from indirect evidence then this highlights a problem of inconsistency in the network. In this case one could argue that the direct and indirect evidence may not be compatible. Any presence of inconsistency in closed loops was assessed through the use of the node-split method.(21) The difference in effect size between the direct and indirect evidence was estimated and tested to see if the difference was significantly different from zero. Any p-values below 0.05 would indicate that the difference in effect size between direct and indirect evidence was statistically significant from zero and therefore inconsistency existed in the model. In the case where inconsistency was identified, the studies involved in the inconsistent loops of the network were reviewed to detect any clinical heterogeneity that might have caused the inconsistency.”

Location of file: Page 10, Line 203-217

Comment #7:

It should be stated which treatment was taken as the reference for each of the networks analysed.

Response:

We thank the reviewers for the suggestions. The reference for each network was set by the gemtc package. In addition, as suggested below, we no longer present the predicted posterior probabilities of each outcome, thus, treatment reference was not mentioned.
Comment #8:

On page 11 "predicted posterior probabilities of each outcome" are mentioned. I am not sure what these are and no details are provided on how they are calculated. For example, it is not clear whether a higher or lower value of the probabilities is better, or whether they are expected to add to one over all interventions (which they don't).

In addition, the use of "predicted" and "posterior" is confusing: do you use the predictive distribution or the posterior distribution, to calculate these probabilities? On line 29 it is stated "…calculated from the posterior predictive distribution…” which suggests the predictive distribution is used. Usually, rank probabilities are calculated using the posterior distribution as this summarises what is known based on current evidence, although a case can be made for using the predictive distribution instead (but I am not clear if the authors calculate are these rank probabilities). Details of what the probabilities are and how the distributions are obtained need to be given, perhaps in a technical appendix with a brief description in the main text.

The authors may want to look at plots that can be used to represent the probability that each intervention is ranked best, second best etc up to worst (rankograms, see eg Ades et al https://www.ncbi.nlm.nih.gov/pubmed/20825617), as well as summarising the mean/median rank and 95%CI for each intervention in tables. Rankograms in particular are useful to display the different ranks of treatments on multiple outcomes and might also facilitate the discussion, which at the moment is not very clear. Related to point 4 above, the use of predictive distributions, only makes sense when a random effects model is used. However, it is not clear if this is the case in the manuscript. If random effects models were used to obtain results, then the estimated between-study heterogeneity and its 95% CI should be presented for each network.

Response:

We thank the reviewers for the comments. We have agreed with the reviewer and have addressed them by presenting ranks probabilities instead. In particular, we have presented rankograms for all outcomes and all treatments. We have also considered estimating SUCRAs, however, after coming across some criticism in the literature (e.g. L. Mbuagbaw 2017) and we have felt that one number was not indicative of the “story” behind the ranking; we have decided to only show the rankograms.

We have included the following sentence in the paragraph:

“In addition, the probability of each treatment to be the best, second best, third best, etc. was presented for all treatment options using rankograms.”
Comment #9:

In the Abstract it is stated that "NMA was performed…” but it is not clear that this was done for both the positive and negative population. The relative effects between interventions should also be reported in the Results section, as the probabilities on their own are not sufficient to judge the efficacy of treatments.

Response:

We thank the reviewer for this suggestion. We have edited the Abstract as follows: “Network meta-analysis was performed to synthesize evidence on the efficacy of treatment for both the positive and negative populations separately.”

In the Results section, we have added the odds ratios where appropriate as follows: “In pairwise comparisons across all drugs (Table 3a), TDF had statistically significantly higher odds of achieving viral suppression than the rest except for ETVTDF and TAF, e.g., ETV vs TDF, OR = 0.46, 95% CrI: 0.25-0.86; TAF vs TAF vs TDF, OR = 0.88, 95CrI: 0.38-1.99. …………”

Comment #10:

In the Methods section, it should be made clear whether the protocol for the systematic review was previously published, and if so, it should be cited.

Response:

We thank the reviewer for this suggestion. We have added in the Methods section the following: “The protocol for this systematic review was an extension of our earlier systematic review carried out under the sponsorship of the Ontario Drug Policy Research Network (ODPRN) and reported in 2015)”
Comment #11:

On page 11, lines 38-44 and page 13, line 49, a "subgroup analysis" is mentioned to refer to an analysis of a subset of the interventions. I would not call this a "subgroup" analysis as this terminology usually refers to subsets of patients divided by some patient or study level characteristic.

Response:

We thank the reviewer for pointing this out. We have removed such analysis.

Comment #12:

On page 12, lines 15-17, it is stated that some studies assessed both positive and negative populations, but it is not clear which network these studies were included in - did they report separate outcomes for each sub-population? Were they included in both networks?

Response: We thank the reviewer for pointing this out. We have edited the text as follows: “Six publications assessed both populations (58-63) and were included in both networks as they reported the outcomes separately for each population.”

Comment #13:

The treatment networks for the main analyses (currently in appendix G) should be in the main paper, if possible.
Response:

We thank the reviewer for this helpful suggestion. We have inserted the figures of the treatment networks in the main paper, and edited the text as follows:

“The evidence networks are presented in Figures 1a and 1b for virologic response, 2a and 2b for ALT normalization, Figures 3a and 3b for HBeAg loss and seroconversion respectively and Figure 4 for HBsAg loss. In these evidence networks, the size of the circles corresponds to the number of patients exposed to a treatment and the thickness of the connecting lines corresponds to the number of studies comparing the treatments directly. The relative efficacy in terms of odds ratios and 95% credible intervals of pairwise comparisons for each outcome are presented in Tables 3a and 3b. The probability of each treatment being the 1st, 2nd, 3rd, etc. are provided in the form of rankograms in Figures 5a and 5b for the e-positive and e-negative populations respectively.”

Location of file: Page 13, Para 2, Line no 281-289

Comment #14:

The tables of relative effects (Table 4) are not a very effective display - they take up a lot of space and most of the cells are blank! These tables can be very useful to display two outcomes together (eg one on the upper diagonal and one on the lower diagonal) or to display results from two different analyses. Currently they are not used for this purpose and therefore they waste space. I would suggest presenting the comparisons in standard tables (ie with the different comparisons on different rows) and perhaps putting different outcomes on different columns (with some blank rows if the network for that outcome did not include a particular comparison). In addition, the table caption stating "pairwise comparisons" is misleading as it suggests the results are from separate pairwise meta-analyses, whereas I believe these are from the network meta-analyses. I suggest changing this to read something like "relative effects of all pairs of interventions calculated from the network meta-analyses" or similar. As noted above, if a random effects model was used, the between-study heterogeneity could also be displayed in these tables, for example as a footnote or additional row.

Response:

We thank the reviewer for this helpful suggestion.
We have reformatted Table 4 (now Table 3) as you recommended, and changed the table caption to “Relative effects of all pairs of interventions in Odds Ratios (95% credible intervals) calculated from the network meta-analyses (Random effects model)” for Table 3a and 3b.

Location of file: Page 41-44