RESPONSE TO REVIEWER #1

"The authors have done a good job revising the document. I only have a few additional (minor) comments to make:"

Response: We thank the reviewer for their time and effort, and for the positive critique.

Comment 1. I could not find the estimate for tau in the NMA in the paper or the supplement. Please give the estimate for heterogeneity in NMA along with its 95% CrI on the main paper, and commend on this value.

• Response: We appreciate the reviewer's comment. Tau-squared is a measure of heterogeneity for pairwise meta-analysis. We have provided that estimate with 95% confidence limits for all comparisons (eTable 3). For the NMA estimates, we are unaware of a method that uses tau-squared as a measure of heterogeneity for the mixed estimate. There are methods to evaluate inconsistency (or incoherence) between direct and indirect estimate. Some use a global measure (the use of which is usually discouraged), but a more preferred approach is to compare for each comparison the direct and indirect estimates (eTables 13A-C, in revised manuscript) (Puhan MA, Schünemann HJ, Murad MH, et al; GRADE Working Group. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014 Sep 24;349:g6330).

We feel that our manuscript addresses heterogeneity sufficiently. First, in terms of traditional heterogeneity (i.e., variation between effect sizes of the different studies), we presented 3 measures: I2, tau squared and p-value for Cochrane test. Second, for NMA inconsistency, we used node-splitting approach and showed each estimate separately (direct and indirect) and demonstrated how statistically they contrast.

• Changes in manuscript: None

Comment 2. Regarding inconsistency: authors say in the main text that no inconsistency was found, but do not refer to the actual results presented in the Supplement. I suggest adding a line in the main text, referring to tables 12 A-C for the actual results, so as to help the reader.

• Response: We have added a statement saying using the node-splitting approach, no inconsistency was observed in the network meta-analysis estimates, and eTables 13A-C (in revised manuscript) have been referenced.

• Changes in manuscript: Results, Page 21, line 462

Comment 3. Page 18: The authors give the DIC (deviance information criterion) for the random effects model. DIC is a model selection criterion, i.e. it is used to compare models. Mentioning the value of DIC for a single model is meaningless, if it is not compared to something else. I suggest either mentioning the DIC for fixed effects, or deleting it.

• Response: We agree with the reviewer, and have deleted the deviance information criteria as suggested.

• Changes in manuscript: Deleted

Comment 4. The total residual deviance should in general approximate the number of unconstrained points (as the authors mention), but there are no rules on what “approximate” exactly means. So I would suggest to rephrase slightly the wording, instead of writing “Model fit was appropriate with total residual deviance approximating the number of unconstrained data points” the authors could write something less strong, e.g. that “The total residual deviance was found to be [...], implying a good model fit.”

• Response: We have revised the statement as suggested.

• Changes in manuscript: Results, Page 21, line 465

Comment 5. In the online supplement, statistical analyses, it writes: "For the prior distribution of between-trial standard deviation, we used a uniform distribution with a mean of 0 and large variance (5). This doesn't make sense for many reasons (a uniform distribution is usually described by the lower and upper limits, not the mean and variance. The standard deviation is a positive number, the distribution the authors describe also gets negative values). The authors probably mean "a uniform distribution U(0,5)" , which means a uniform distribution from 0 to 5, not one with mean 0 and variance 5

• Response: We agree with the reviewer, that uniform distribution is commonly defined by its minimum and maximum values. We have changed the statement accordingly.

• Changes in manuscript: Online supplement, eMethods, page 4, line e63

Comment 6. You use the abbreviation mITT without first defining it. Please define the term in page 10 (where you say about modified ITT analysis).

• Response: We have added the abbreviation at point of first use.

• Changes in manuscript: Methods, page 9, line 197

RESPONSE TO REVIEWER #2

Comment 1. The authors took into account my comments and they proceeded with the necessary revisions. However, I would like to re-direct the attention of authors to aspects that refer to heterogeneity and sensitivity analysis on priors. Specifically, results from the common within network heterogeneity are not provided. A textual reference in the Results section as well as a comparison with the respective predictive distribution as provided by Turner et al would suffice. Regarding the pairwise meta-analysis, it is not stated which confidence interval strategy for the heterogeneity variance the authors used.

• Response: We appreciate the reviewer's comments. We generated the confidence interval for tau squared using a frequentist approach, using the command metaan in Stata. The approach follows a profile likelihood random-effects model. We have added a sentence in the revised manuscript contrasting our heterogeneity to the empirical distribution by Turner et al. Please also note that we did not compute a global inconsistency measure for the whole network but rather compared direct and indirect estimates for each pairwise comparison (see response to reviewer #1, comment 1).

• Changes in manuscript: Online supplement, page 3, line e44

Comment 2. In terms of prior sensitivity analysis, instead of being too general on the sensitivity analysis set-up, I would
recommend to be succinct by creating a table where you can specify for each parameter which alternative priors you used (i.e. distribution and hyper-parameters). A simple reference of the results of this sensitivity analysis in the Result section will suffice. Particularly, the sensitivity of the heterogeneity variance on the prior specifications is very crucial.

Response: We appreciate the reviewer's comment. As suggested, we have created a supplementary table (eTable 12) which presents the SUCRA ranking for all agents for the primary outcome of advanced metachronous neoplasia, using 3 alternative priors (vague, tau squared 0.12, tau squared 0.20 and tau squared 0.30). Overall results were similar to the primary analysis.

Changes in manuscript: Results, page 21, line 455; Online supplement, eTable 12.

RESPONSE TO REVIEWER #3

Comment 1. The paper is very well written and the analyses well described. The authors have done a very good revision from their earlier version, and in particular responded well to the truly superb review by Efthimiou. I have some further suggestions for improvement, though these are relatively minor except for comments about ORs (See below) which need detailed thought/justification. The analysis methods appear entirely appropriate and in keeping with the current literature. It is an immense piece of work and very well reported, with excellent appendices.

Response: We thank the reviewer for their time and effort in reviewing our manuscript, and the positive critique. We have responded to all of the reviewer's comments systematically.

Changes in manuscript: Throughout the manuscript

Comment 2. ORs are provided, and I think used to translate back to the absolute risk scale (assuming a particular baseline risk), The authors are therefore assuming that ORs are close approximation to RRs. But the event %s appear quite high. I therefore worry that this is not appropriate. Please can the authors justify this, or revise accordingly.

Response: We appreciate the reviewer's comment. To calculate absolute risk reduction, odds ratios were converted to relative risk using the Zhang equation (JAMA. 1998;280(19):1690-1691) using NMA summary estimates (odds ratios) and baseline risks obtained from the National Cancer Institute pooling project. The equation is provided here:

Changes in manuscript: Methods, page 13, line 269

Comment 3. Why are ORs relevant here, when the studies have different follow up lengths? Surely hazard ratios would be preferred. Indeed, the authors often refer to rates, but simply give %s, which are not rates but risks. I worry they are confusing the two here. They say "The primary efficacy outcome was prevention of advanced metachronous neoplasia, within 3-5 years of the index colonoscopy" -- therefore, it could be that ORs for 3 years are very different to ORs for 5 years? If follow-up length was different for different agents, then this might averse bias the comparisons? Can the authors address this please, either as a limitation, or to justify why ORs were used and what time-point they exactly relate to. We certainly need to know the time in each study when the OR was measured.

Response: The reviewer brings up a good point. Unfortunately, we do not have individual participant data to produce hazard ratios, or to produce estimates at various time points. We agree that relative association measures need a time frame. In the revised manuscript, we have added the length of follow up as a range and median for each outcome and mentioned this as a limitation in the manuscript. The median initial surveillance period to assess for metachronous neoplasia was 36m (range, 24-60m).

Changes in manuscript: Discussion, page 26, line 587; Results, page 16, line 359

Comment 4. A related point, when the absolute risks are presented: what time-point is this the risk of an outcome by?

Response: As mentioned above, the median initial surveillance period to assess for metachronous neoplasia was 36m (range, 24-60m).

Changes in manuscript: Results, page 16, line 359

Comment 5. ORs require all patients to be followed until the time-point of analysis. Therefore, how were patients handled who were censored before the time-point? I don't know whether censoring was an issue or not, but with 3-5 year follow up it is possible.

Response: We appreciate the reviewer's comments. As mentioned, the median anticipated initial surveillance period to assess for metachronous neoplasia was 36m (range, 24-60m). There was insufficent data to perform time-to-event analysis. No censoring was performed in individual studies, which had a well-defined time for outcome assessment.

Changes in manuscript: None

Comment 6. SUCRA is an unusual term for the BMJ reader, and the abstract would benefit from either explaining its meaning in more detail, or removing it.

Response: We have added the definition of SUCRA in the Abstract of the revised manuscript.

Changes in manuscript:

Comment 7. Introduction says: "Network meta-analyses combine direct and indirect evidence to establish comparative efficacy and safety across a network of RCTs of all agents used in a particular condition" – the authors should note the assumptions at this point also. We gain more by assuming more.

Response: We appreciate the reviewer's comments, and have added the key assumption in NMA's, that trials are conceptually similar with regard to trial design, participants, intervention (dosing, duration), co-interventions, and outcome assessment.

Changes in manuscript: Abstract, page 4, line 88

Comment 8. The authors put most of their an analysis description in the supplementary appendix. I strongly suggest this is moved to the main paper. That is, integrate the 'methods' stats analysis section to the methods of the main paper. There are many assumptions and explanations therein that are important, and deserve to be in the paper. The BMJ has no word limit.

Response: As suggested, we have integrated large sections of the statistical methods from the online supplement to main
Comment 9. Figure 1 – why are the number of studies with direct evidence not simply added to the lines?

Response: We appreciate the reviewer's suggestion, and attempted to add the number of studies to the lines. However, given the complexity of the network, it becomes very hard to read the comparisons and corresponding number of studies. The number of studies with direct comparisons is reported in the corresponding forest plots.

Changes in manuscript: None