We thank the reviewers for their comments and are especially grateful for Dr. Price's detailed and concrete feedback, which we respond to below.

1. Reviewer comment:
   There is no mention of any investigation of the characteristics of residuals from the two meta-regression models. Please could the authors report some details about whether the assumptions of the linear regression models have been met e.g. Normality, homoscedasticity of residuals etc.

Author response:
Thanks for drawing our attention to this point. We have done an extensive investigation of the residuals and found absolutely no evidence of any violations of the assumptions of normality. The following text was added to indicate that
linear regression assumptions were examined (Methods, Analysis section, “Meta-regression analysis” description, end of 3rd paragraph):

“Neither regression model [logit control versus ln(OR), nor re-scaled IPDAS versus ln(OR)] was found to violate the assumptions of linear regression (linearity, independence, homoscedasticity, normality) upon examination of residual plots (predicted dependent variable vs residual, independent variable vs residual, and normal probability (Q-Q) plots of residuals).”

2. Reviewer comment:
Page 10, 2nd paragraph, lines 2-3 and throughout the paper. The authors use the term “control event rate” rather than “baseline risk” to minimise confusion since risk corresponds to a favoured outcome. This is confusing as the term rate makes no sense at all in this context because the outcome is a proportion and so dimensionless. Please change to something like “baseline proportion” or “baseline score.

Author response:
We recognize that there is a strict conceptual distinction between rates and proportions. But, while imperfect, the term “control event rate” is a recognized label used currently in epidemiological literature to refer to the proportion of patients in the control group in whom an outcome event of interest is observed (see for example Cochrane Database Syst Rev 2013;1:CD009461; BMC Med Res Methodol 2012;12:111; Clin Trials 2010;7(2):109-17; J Clin Epidemiol 2004;57(12):1253-61). We have chosen this term in order to be consistent and comparable with that literature. A secondary reason for avoiding use of the word “baseline” (not mentioned in the manuscript) is that it sometimes refers to outcome measurements made before receiving either the control or experimental intervention, and we feel that “control” reduces potential for this ambiguity. Based primarily on the tradition of its use in the epidemiological literature, we respectfully submit that “control event rate” is a sufficiently recognizable and appropriate term in this context.

3. Reviewer comment:
Page 10, 2nd paragraph, lines 4-5: “Assuming the type of control intervention does not modify its effects (as confirmed in this analysis)”. The analysis does not confirm this. Please rephrase the part in brackets to something like “(and our investigations found no evidence that it does)”.

Author response:
We have revised the text in question as suggested to: “(and our investigations found no evidence that it does)”.

4. Reviewer comment:
Page 11, 2nd paragraph, lines 5-6: please change “A chi-squared test is used to detect whether heterogeneity is present” to something like “A chi-squared test is used to examine the strength of the evidence about whether heterogeneity is
present”.

Author response:
We have revised the text in question as suggested to: “A chi-squared test is used to examine the strength of evidence about whether heterogeneity is present”.

5. Reviewer comment:
Page 16, 2nd paragraph, lines 4-5. Please change “The type of control intervention does not modify” to “There was no evidence that the type of control intervention modifies”.

Author response:
We have revised the text in question as suggested to: “There was no evidence that the type of control intervention modifies”.

6. Reviewer comment:
Page 18, 3rd paragraph, lines 7-8: the phrase “they found linear correlations with ln(OR) in only 14% of meta-analysis” is unclear. Please expand on what is meant by found (p-value < 0.05 in 14% of MA’s?)

Author response:
We have added the underlined passage to clarify the sentence in question: “After correcting for dependence of the effect measure on control event rate and using a two standard error rule of significance, they found linear correlations with ln(OR) in only 14% of meta-analyses.”

7. Reviewer comment:
Discretionary revision: In my view the term statistical significance is not useful and I advise that it be removed from the document.

Author response:
We regret not defining the threshold for clinical significance in our previous draft. The following text has been added (Methods, Analysis section, end of 1st paragraph):
“The threshold for statistical significance was p < 0.05.”

To improve consistency, the word “statistically” was removed in two cases (Background, 4th para; Results, 1st para, last sentence). The words “statistically significant” were removed in one case where they were unnecessary (Discussion, 1st para, 2nd sentence). All other uses of “significance” were retained to communicate the concept of statistical significance in particular.

8. Reviewer comment:
Typographical errors.
Author response:
The errors identified have been corrected.

Stephen Gentles