We thank Professors Evans and Wilkins for their interest in our systematic review.[1]

We have reached the same conclusion as previous systematic reviews published in 2008[2] and 2014[3] and a review prepared for the New Zealand Ministry for Health in 2009[4]. Even the 'alternative systematic review' prepared by Professors Evans and Allen about which we have significant misgivings concludes that 'larger and rigorous randomised controlled trials of interventions for visual stress are required'.[5]

Professor Wilkins makes a sweeping statement that the Cochrane tools for assessing bias are 'not up to the job'. However, there is ample evidence from meta-epidemiological studies that the domains of bias featured in the Cochrane tools *do* influence trial results[6,7] and for this reason they have gained international acceptance for the purpose of assessing randomised controlled trials (RCTs). In fact, inconsistent scoring of bias reflects usually results from ambiguous descriptions of trial methodology not the shortcomings of the Cochrane tools, and if trials were reported according to the CONSORT statement[8], it would be easier to rate them in a consistent manner.

Systematic reviews, in common with all forms of research, have their limitations. However, when the results of systematic reviews with meta-analysis are followed up with large scale confirmatory RCTs, the positive effects identified by reviews comprising multiple small or underpowered studies tend not to be replicated.[9,10]

The fundamental component of a systematic review is the risk of bias (RoB) table and the results of this table should inform the discussion.[11] Professor Wilkins and Evans have left the RoB table used in our review essentially unchallenged. For this reason the accusation that we have adopted a light touch to studies with negative results compared with those that display positive results is unfounded. We would, of course, be happy to review any judgments of bias in the RoB table that are felt to be incorrect.

Professor Evans advises that our review should be read alongside their 'alternative systematic review'[5] in order to obtain 'a balanced insight'. Their review used the Critical Appraisal Skills Program checklist (CASP)[12], another recognised tool for evaluating bias. However, the domains of bias of the CASP checklist are incompletely represented in their RoB table. For this reason we do not believe readers should attach equal weight to the two reviews.[1,5] We have raised our concerns about their application of the CASP tool, as well as other concerns about their review, in a letter to the editor of the journal where their review was published.[13]

Professor Wilkins raised the issue of lack of detail on the agreement between the individuals making the RoB assessments. There was good agreement between us in the scoring of the domains of bias. The one area of difference (five studies) related to the scoring of allocation concealment. However, this occurred in

studies using a crossover methodology, which were scored at low risk of bias for allocation concealment and random sequence generation irrespective of the ambiguities frequently found in the methods sections of the papers we reviewed. This is because such studies are usually at low risk of confounding at baseline. It could be argued that by adopting this approach we have relatively disadvantaged studies such as Ritchie et al.[14] which used a rigorous central allocation technique for random sequence generation and allocation concealment, but which also received a low risk of bias rating in these domains. This is another reason why we do not accept the allegation that we have adopted a light touch to studies with 'negative' results.

The principle source of bias in studies of Intuitive Overlays and Precision Tinted Lenses, in which both respondents acknowledge a financial conflict of interest that could potentially influence their views, was a lack of masking, because coloured overlays were compared with a no overlay or clear overlay condition. It is neither 'subjective nor untraceable' to argue that these studies are at higher risk of bias in the domain 'blinding of personnel and participants' compared to studies that compare a chosen colour with a placebo colour. Similarly, a study that follows up only 52.7% of participants[15] can only be deemed to be at high risk of attrition bias. We would challenge any reviewer to score such studies differently and most systematic reviews would not even include such studies in the final synthesis. Also, as stated above, neither Professors Evans nor Wilkins indicated that different judgments of bias would have been more appropriate for any of the studies we reviewed.

One potential source of bias in RCTs is the source of funding. Evidence from two major systematic reviews has shown that this can influence the result of trial in a way that is favourable to the sponsor. [16.17] For this reason, some systematic reviews consider the source of funding under the heading 'other sources of bias. We note that in many of the studies of Intuitive overlays and Precision Tinted Lenses, the authors make financial disclosures but the source of funding for the study is often not made clear. Because we did not consider this in our review, it is possible we may even have underestimated the risk of bias of many of the studies we included.

Of course, there is no such thing as a perfect trial and for many of the studies that we reviewed limited resources were available. However, numerous participants have been enrolled onto time-costly studies comparing coloured overlays with the no overlay or clear overlay condition. The same resources, properly marshalled, could have financed studies at lower risk of bias.

It is suggested that we have confused the steps required to diagnose visual stress and place precedence on an immediate improvement with an overlay, as opposed to sustained voluntary use. However, the authors do not consistently define how long an individual has to use an overlay to have clinically meaningful visual stress. In some studies it could be as little as three weeks[15] and in others 12 weeks or more. [18] Any test that requires a lengthy assessment of all children and issuing of overlays to as many as 60% to identify a subgroup who may or may not benefit in terms of reading is not an economically viable or ethically

appropriate process. Professor Evans has himself published a study that used the criterion of immediate benefit from overlays[19] and has cited such a study favourably as evidence that overlays can improve reading.[20]. It appears that this 'confusion' is not limited to our review. A pre-requisite for more rigorous trials is an agreed definition of clinically significant visual stress. Professor Evans acknowledges as much when he states that 'improvements in the diagnosis of this condition [visual stress] are also a priority'.[4]

It is argued that the negative results of placebo controlled studies were a result of using naturalistic text which is too large and too widely spaced to demonstrate the beneficial effect of coloured filters, this being the motivation for developing the Wilkins Rate of Reading Test (WRRT). However, since none of the studies utilising the WRRT had a well-designed placebo control condition, the most likely explanation for the discrepancy between studies using naturalistic text and the WRRT has to be placebo effects. Importantly, we noted that studies utilising naturalistic text *did* show improvements in reading using the chosen filter but in contrast to poorly controlled studies this improvement was matched by the placebo filter. Furthermore, unless it can be shown that overlays or lenses produce benefits for reading the sort of text that is encountered in everyday life, the clinical or educational relevance of improvements observed with the WRRT remains questionable. In the case of children who are still undergoing reading development, one should also expect to see significantly larger gains in test scores on standardised tests of reading following longer-term use of overlays than compared to a control group. Our literature search identified no RCTs with sufficiently prolonged follow-up to answer this question.

We appreciate the desire to isolate visual factors that could be interfering with reading; however, it is not yet clear if the WRRT achieves this aim. The extent to which performance on this test is influenced by higher order language skills and/or reading experience, remains to be examined. Furthermore, unless it can be shown that increases in reading speed (on the WRRT) are accompanied by preserved (or better) comprehension then improvements in reading may be due to the reader becoming more confident and less risk averse. We acknowledge that reading accuracy is also measured in the WRRT but it is not clear if this is a sufficient measure of response-criterion. For example, a recent review found that speed reading apps improve reading speed at the expense of comprehension.[21]

We are criticised for scoring Henderson et al. [22] and Ritchie et al.[14] as being at uncertain risk of external bias even though they utilised the WRRT. However, both of these studies also used naturalistic text.

According to Professor Wilkins we make sweeping generalisations which the reader is obliged to take on trust. For example, that 'improvements have been reported with prescribed overlays and lenses but similar improvements are found with placebo colours'. In fact, the references to support this assertion are provided in the general discussion (page 537 paragraph 4).

With regard to our ignoring individual participants who showed substantial improvements in reading, it is important to remember that *post-hoc* subgroup

analyses are hypothesis generating only and have been compared with a marksman firing shots at a barn and painting a target around the bullet hole after the event.[23] The target only shows how accurate the shot was if it was in place before the shooting. For this reason, *post-hoc* subgroup analysis would not normally play any part in a systematic review and studies that contain such analyses would be scored at high risk of reporting bias. We agree, however, that the authors of those studies could usefully have looked at the subgroups who appeared to benefit to see how they differed from their peers who obtained no benefit. Such an analysis might have generated useful hypotheses that could have been tested in confirmatory trials, which might in turn have been included in a systematic review. We did attempt to further analyse the RCT published in 1994 by Wilkins et al.[15] We requested further data concerning the subgroup of 36 participants (out of 68) who completed symptom diaries to determine whether equipoise had been lost with regard to subjects using their chosen lens or placebo lens first. Unfortunately these data were not available.

Professor Evans adopts the logic that because the majority of studies show a positive effect of coloured filters this is evidence for their effectiveness. However, small studies at high risk of bias are cheaper and easier to perform for which reason they generally outnumber larger more rigorous studies. Consequently, simple counting of studies without reference to RoB will overestimate treatment effects. This so called 'vote-counting' approach might have some value if the positive studies were at high risk of bias in different bias domains. However because the studies showing positive results are all at high risk of bias in the same domain - lack of masking – the value of the approach of simply counting the number of studies with positive results is extremely limited. The 'vote-counting' approach to systematic reviews has been heavily criticised[11] and is not advocated by the Cochrane Collaboration.[24]

Scepticism is a condition of doubt or uncertainty about putative knowledge that is usually seen as a virtue; particularly so in scientific circles. If high quality trials support the use of coloured overlays and/or lenses then of course we would be happy to see them having positive effects. Equally, Professors Evans and Wilkins should acknowledge the possibility that the larger and more rigorous RCTs, that we all agree are necessary, may not support their beliefs. It is unfortunate, for all, that overlays and lenses have been marketed prematurely, before robust evidence has been obtained.

There is near universal agreement that properly designed RCTs are required and research ethics committees should carefully consider whether or not to give approval to further studies at high risk of bias due to the lack of a proper placebo control condition, including adequate masking of study personnel. As we outline in our article, future confirmatory trials should pre-register their methodology, sample size and proposed statistical tests and 'real world' outcome measures should be incorporated. The results should be reported in accordance with current standards for reporting trials.[8]

1: Griffiths PG, Taylor RH, Henderson LM & Barrett BT. The effect of coloured overlays and lenses on reading: a systematic review of the literature. Ophthalmic Physiol Opt 2016; 36: 519–544.

2: Albon E, Adi Y & Hyde C. Coloured filters for reading disability. Available from: http://www.birmingham.ac.uk/Documents/college-mds/haps/projects/WMHTAC/REPreports/2008/ColouredfiltersforreadingdisabilityFINALVERSION.pdf, accessed 5/10/2016.

3: Galuschka K, Ise E, Krick K & Schulte-Körne G. Effectiveness of treatment approaches for children and adolescents with reading disabilities: a meta-analysis of randomized controlled trials. Plos One 2014; : 9: e89900. doi: 10.1371/journal.pone.0089900.

4. Malins C. The use of coloured filters and lenses in the management of children with reading difficulties: a literature review commissioned for the Irlen Screening Research Project, Ministry of Health. [Wellington, N.Z.]: Mind Matters Psychology; 2009. Available from: http://www.moh.govt.nz/moh.nsf/pagesmh/10456/$File/use-of-coloured-filters-and-lenses-children-learning-difficulties-nov2010.pdf, accessed 15/2/2016.

5: Evans BJW & Allen PM. A systematic review of controlled trials on visual stress using Intuitive Overlays or the Intuitive Colorimeter. J Optom. 2016; 9:205-218. doi: 10.1016/j.optom.

6: Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008; 336: 601–605.

7: Balk EM. Correlation of Quality Measures With Estimates of Treatment Effect in Meta-analyses of Randomized Controlled Trials. JAMA 2002; 287: 2973–2982.

8: Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 134:663–694.

9: Egger M, Smith GD. Misleading meta-analysis. BMJ1995; 310: 752–754.

10: Kjaergard LL. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001; 135:982-989.

11: Katikireddi SV, Egan M & Petticrew M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. J Epidemiol Community Health 2015; 69:189–195.

12: CASP Tools & Checklists.Available from: http://www.casp-uk.net/#!casp-tools-checklists/c18f8, accessed 6/8/2016

13: Griffiths PG, Taylor RH, Henderson LM & Barrett BT. Letter to the Editor concerning "A systematic review of controlled trials on visual stress using intuitive overlays or colorimeter." J Optom (under review).

14: Ritchie SJ, Della Sala S &McIntosh RD. Irlen colored overlays do not alleviate reading difficulties. Pediatrics 2011;128:e932–e938.

15: Wilkins AJ, Evans BJ, Brown JA, Busby AE, Wingfield AE, Jeanes RJ, et al. Double-masked placebo-controlled trial of precision spectral filters in children who use coloured overlays. Ophthalmic Physiol Opt 1994; 14: 365–370.

16: Bekelman JE, Li Y & Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. JAMA289; : 454–465.

17: Lexchin J, Bero LA, Djulbegovic B & Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003; 326: 1167–1170.

18: Jeanes R, Busby A, Martin J, Lewis E, Stevenson N, Pointon D, et al. Prolonged use of coloured overlays for classroom reading. Br J Psychol 1997; 88: 541–548.

19: Bouldoukian J, Wilkins AJ & Evans BJW. Randomised controlled trial of the effect of coloured overlays on the rate of reading of people with specific learning difficulties. Ophthalmic Physiol Opt 2002; 22: 55–60.

20: Kriss I & Evans BJW. The relationship between dyslexia and Meares-Irlen Syndrome. J Res Read. 2005; 28: 350–364.

21: Rayner K, Schotter ER, Masson MEJ, Potter MC & Treiman R. So much to read, so little time: how do we read, and can speed reading help? Psychol Sci Public Interest 2016; 17:4–34.

22: Henderson LM, Tsogka N & Snowling MJ. Questioning the benefits that coloured overlays can have for reading in students with and without dyslexia. J Res Spec Educ Needs 2013; 13: 57–65.

23: Fletcher J. Subgroup analyses: how to avoid being misled. BMJ 2007; 335: 96–97.

24: Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011] Editors: Julian PT Higgins and Sally Green. Section 9.4.11 Use of vote counting for meta-analysis [Internet]. [cited 2016 Aug 15]. Available from: http://handbook.cochrane.org/chapter_9/9_4_11_use_of_vote_counting_for_meta_analysis.htm, accessed 10/9/2016.