PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Safety of Corticosteroids in Young Children with Acute Respiratory Conditions: A Systematic Review and Meta-Analysis
AUTHORS	Fernandes, Ricardo; Wingert, Aireen; Vandermeer, Ben; Featherstone, Robin; Ali, Samina; Plint, AMy; Stang, Antonia; Rowe, Brian; Johnson, David; Allain, Dominic; Klassen, Terry; Hartling, Lisa

VERSION 1 - REVIEW

REVIEWER	Yoon K Loke Norwich Medical School UK
REVIEW RETURNED	22-Jan-2019

GENERAL COMMENTS	Thank you for giving me the opportunity to comment on this submission. This review illustrates the problems with broad AE reviews which aim to synthesize data based on what is reported in the study manuscript, rather than pre-specifying outcomes of interest in the review. Basically, the reviewers end up in the unfortunate position of being hamstrung by heterogeneity and selective non-reporting bias from the primary studies. There's not much that reviewers can do to overcome this major problem, but I have a few suggestions. 1. Please list the AE that were specified a priori in the six studies, and what the findings were. I would consider this set as being able to yield more reliable data. 2. It would help to have some discussion of the numerous studies that say 'no significant AE'. These studies presumably measured and analysed the AE but selectively chose not to report it because of their opinion regarding statistical or clinical effect. This is unfortunate because there may have been an effect that could be pooled in meta-analysis, even though the study itself was under-powered to detect statistically significant findings. 3. Biological plausibility of short course of corticosteroids causing reduction in height should be considered. There are dangers of relying on single, possibly selectively reported outcomes , in the face of possibly several unknown or unclear studies where they did not report the height because they found no difference. This needs to be contrasted with meta-analyses of long-term corticosteroid use in children with asthma.

REVIEWER	Giorgio Piacentini University of Verona – Italy GP has served in advisory boards for Chiesi, MSD and GSK.
REVIEW RETURNED	22-Jan-2019

GENERAL COMMENTS	The article is well written. The authors' aim is to perform a systematic review of the literature regarding adverse events linked to a short-course of oral or high-dose inhalatory corticosteroid use for respiratory infections in young children less than 6 years old. The conclusion of the authors is referred to the absence of adverse effects attributable to a short course of oral or high-dose inhalatory corticosteroid. In the intentions of the authors, this systematic review add some evidence: an increase in the safety of the use of a short course of oral corticosteroid and the absence of significant correlation between a short course of high-dose inhaled corticosteroid and significant adverse events. Although this manuscript is characterized by a number of limiting factors, these two evidences have not been demonstrated by previous studies. The articles analyzed in this systematic review came from randomized controlled trial or observational studies, although the methodology is not in line with GRADE criteria. The sample considered (11000 children) is a strength of this systematic review; the methodology seems to be a weakness. In conclusion, this article, because of the big number of the sample considered, can represent a significant contribution to the scientific knowledge. Major comments: The major criticism regards the missing use of the GRADE scale for the selection of papers; another criticism resides into the “conclusions” section, where authors claim that there is no correlation between short-course of oral or high-dose inhalatory corticosteroid and significant adverse events. Moreover, I have some major concerns that need to be addressed to the authors: • Page 5 "Search strategies combined index terms and keywords for respiratory illnesses, children and drug classes identified in the Global Initiative for Asthma (GINA) guidelines" • Page 5 "Original database searches were conducted September 2014 in Ovid Medline [...] Update searches were executed in Medline and CENTRAL in February 2016, and then again in July 2017" It would be better to clearly specify ”search” terms; they seem too broad in the way described by the authors,. Moreover, why was the literature search carried out on different platforms in 2014, 2016 and 2017? Finally, references 14 and 112 refer to paper published in 2018; the authors write that the last literature update was in 2017. • Page 6 "Studies that did not report or mention AEs were excluded". It is necessary to specify how the authors made search for adverse events in non-considered articles. • Page 14. "A common concern when using corticosteroids in young children is effect on growth. Results from a single, small trial (n = 129) of recurrent high-dose inhaled fluticasone propionate in

	wheezing preschoolers were heterogeneous across outcome measures, but suggested a small significant risk of growth suppression. " Please explain adequately what the author means by the term "recurring". • Page 17. "While the McHarm scale is to be used in conjunction with other quality assessment tools to evaluate the broader elements of the study quality of the study. [...] • Page 18. Two to the variation in corticosteroids and an extended range of reported AEs among varied study designs of overall poor quality, we did not attempt to rate the quality of the body of evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach The McHarm scale has been validated by Cochrane (Santaguida P, Keshavarz H, MacQueen G, Levine M, Beyene J, Raina P. Development of the McHarm: A tool evaluating validity of the collection and reporting of harms. In:" Abstracts of the 19th Cochrane Colloquium"; 2011 19-22 Oct; Madrid, Spain. John Wiley & Sons; 2011). The GRADE scale would be the most appropriate for systematic reviews. The authors point out that they did not follow the GRADE scale because of the intrinsic difficulties of their literature research. This aspect represents a strong weakness of this article. • Page 18. CONCLUSION The consideration that it is difficult to obtain incontrovertible evidence regarding the correlation between short-course of oral or high-dose inhaled corticosteroid and adverse events appears to be justified. The conclusion that their use is not related to a significant increase in significant adverse events appears to be foolhardy; specifically, the term "significant" should be specified. Moreover, the authors underline an increase in the incidence of vomiting episodes after a short-course of oral corticosteroid in the text (page 9). Minor Issues • References 2, 11, 18, 27, 32, 50, 53, 58, 74, 79 are not in line with the editorial guidelines. • Reference 14 comes from GINA guidelines 2018, while the authors have specified in the text that the last update was in 2017. • Same consideration is referred to reference 112, whose publication year is 2018. • Page 14. "Importantly, these results can help future research in the collection and reporting of AEs, particularly concerning the effects of growth and behavioral outcomes; "this in turn is needed to help inform decision making between clinicians and parents / caregivers of young children." This sentence could be inserted in the "conclusions" section; it is not remarkable in the "discussion" section.

REVIEWER	David C. Hoaglin Adjunct Professor Department of Population and Quantitative Health Sciences University of Massachusetts Medical School Worcester, Massachusetts, USA
REVIEW RETURNED	30-Mar-2019

GENERAL COMMENTS	As requested, I focused mainly on the statistical methods and analyses. The authors undertook a daunting task. As they explain under Strengths and limitations (on page 17), their extensive systematic review "was limited by the quality of the primary literature, particularly regarding the definition, assessment and reporting of AEs. This underscores the challenges researchers encounter when attempting to synthesize safety data due to sparse and poor reporting, and highlights the urgent need to enhance detection and reporting of AEs." Worse, they are in a situation where meta-analysis provides little help. From Supplement 3c, I have the impression of substantial heterogeneity among the studies in characteristics such as design, doses, and conditions. Thus, even if the studies available for a particular meta-analysis contributing to Figure 2, Figure 3, or Figure 4 show little or no statistical heterogeneity, the justification for combining their effects may be weak. The statistical methods have serious shortcomings. The Peto method produced all the odds ratios reported in Figures 2, 3, and 4; but Greenland and Salvan (1990) studied its behavior and concluded, "The one-step (Peto) method for obtaining pooled effect estimates can yield extremely biased results when applied to unbalanced data. Even for balanced studies, the one-step method may incorporate an unacceptable degree of bias." I was not able to examine the degree of imbalance (if any) in the individual meta-analyses, because Figures 2, 3, and 4 and Supplement 5 give only the total numbers of children and AEs in a meta-analysis, and not the numbers in the individual studies. The authors say (page 7, lines 33-35) that they pooled risk difference by using "a Mantel-Haenszel random effects model." The statement is problematic, because no such model exists, despite the impression created by the Review Manager software. According to the memorandum "Statistical algorithms in Review Manager 5," the weights and pooled estimate from the fixed-effect Mantel-Haenszel method are used in an alternative version of the heterogeneity statistic Q, which is then used in estimating the between-study variance for use in the inverse-variance weights of a random-effects pooled estimate. That is the extent of the difference between the "Mantel-Haenszel random effects method" and the usual inverse-variance random-effects method, introduced by DerSimonian and Laird, and the resulting estimates generally only slightly. As far as I am aware, the "Mantel-Haenszel random effects method" exists only in Review Manager 5. No detailed specification of it appears in the meta-analysis literature, and it is not supported by any theoretical or empirical analysis of its properties. Thus, users of Review Manager who choose the "M-H random" option should not assume that, by doing so, they can avoid the well-documented shortcomings of the DerSimonian-Laird method (see, for example, IntHout et al. 2014).

	The authors say (page 7, lines 40-42), “One AE (growth) was reported as a continuous outcome and data were pooled using a mean difference.” They do not document the method that they used for this (the Mantel-Haenszel method is applicable only to odds ratio, risk ratio, and risk difference). Unfortunately, there are more difficulties. The usual test for statistical heterogeneity, based on Q, uses an incorrect null distribution (Hoaglin 2016). For that reason and because the correct null distribution differs substantially among measures of effect, I2 unfortunately has no useful interpretation (Hoaglin 2017). It is discouraging that the authors (and many others) have been so ill-served by the Cochrane Collaboration. Users should be able to count on up-to-date software, documentation, and guidance. Despite these criticisms of the statistical methods, the Results section contains valuable summaries of the available evidence. It may be possible to preserve that contribution while de-emphasizing the meta-analyses (e.g., include appropriate caveats on the Peto method and avoid the “Mantel-Haenszel random effects method”). The discussion (page 16, lines 8 to 10) repeats the result that “evidence favored oral dexamethasone over oral prednisone for vomiting.” In view of the sizable number of comparisons (Figures 2, 3, and 4 contain total of 33 confidence intervals), it seems likely that the authors are capitalizing on chance. In the interest of transparency and reproducibility, it would be a good idea to include (among the supplements) the study-level data for each of the various meta-analyses. Interested readers should not have to request those data from the corresponding author (page 20, line 31). I noticed some rough edges in the manuscript. Pages 32 and 33: “Table 5” should be “Table 1”. Are the very small numbers of participants (2 for Respiratory distress and 1 for each of Psychosis, Positive wheal, Hematology, and Tumor cell proliferation) correct? Also, “Table 1” shows a total of 1635 participants (in 5 studies) for Systemic infections, but Figure 2 and Supplement 6b (which should be named 5b) show 1095 + 1083 = 2178 (in 4 studies). Figures 2, 3, and 4 should say that all the odds ratios are pORs from meta-analyses (except when the number of studies is 1). Why does Supplement 3b account for only 83 studies? The parts of Supplement 5 are numbered incorrectly, as Supplement 6a, etc. References Greenland S, Salvan A (1990). Bias in the one-step method for pooling study results. Statistics in Medicine 9:247-252. Hoaglin DC (2016). Misunderstandings about Q and ‘Cochran’s Q

	Test' in meta-analysis. Statistics in Medicine 35:485-495. Hoaglin DC (2017). Practical challenges of I2 as a measure of heterogeneity. Research Synthesis Methods 8:254. IntHout J, Ioannidis JPA, Borm GF (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Medical Research Methodology 14:25.

VERSION 1 – AUTHOR RESPONSE

Authors' Responses to Reviewers' Comments

Reviewers' Comment	Authors' Response	Reference
Reviewer #1		
R1 General Comment Thank you for giving me the opportunity to comment on this submission. This review illustrates the problems with broad AE reviews which aim to synthesize data based on what is reported in the study manuscript, rather than pre-specifying outcomes of interest in the review. Basically, the reviewers end up in the unfortunate position of being hamstrung by heterogeneity and selective non-reporting bias from the primary studies. There's not much that reviewers can do to overcome this major problem, but I have a few suggestions.	Thank you for your comment and suggestions.	Not applicable
R1.1 Please list the AE that were specified a priori in the six studies, and what the findings were. I would consider this set as being able to yield more reliable data.	We are uncertain which six studies are being referred to, and have made the assumption this is based on the findings of the studies comparing dexamethasone with prednisone that reported on vomiting: Aljebab 2017, Altamimi 2006, Cronin 2016, Fifoot 2007, Garbutt 2013, and Paniagua 2017. We have conducted a subgroup analysis (see p. 19 below), identifying/separating studies that pre-specified vomiting as an outcome of interest (versus studies that did not pre-specify this	Post-hoc subgroup analysis (p. 19, last page of current document);

	outcome). The pooled estimate for the studies that pre-specified vomiting (4 studies of 1164 children; RD -0.07, 95% CI -0.12, -0.02) is not substantially different from the combined effect estimate when studies that do not pre-specify vomiting are also included (6 studies of 1373 children; RD -0.06, 95% CI -0.09, -0.02). We commented on reporting differences among studies depending on their intentions, in the Discussion: Further, safety reporting was not a primary focus of the studies, AEs were rarely defined a priori, and methods for ascertaining AEs were usually absent.	Discussion (p. 19)
R1.2 It would help to have some discussion of the numerous studies that say 'no significant AE'. These studies presumably measured and analysed the AE but selectively chose not to report it because of their opinion regarding statistical or clinical effect. This is unfortunate because there may have been an effect that could be pooled in meta-analysis, even though the study itself was under-powered to detect statistically significant findings.	Thank you for raising this point. We agree that it is important to bring attention to studies that report “no significant AE”, with respect to “significance” at the statistical level (i.e., sample sizes in studies being under-powered to detect adverse events) and in terms of importance/severity of adverse event(s). We have added the following sentence in the Discussion: For example, it is worthwhile noting that 26 studies reported ‘no AEs’ or ‘no significant AEs’ which could not be included in pooled estimates; this may be a reflection of these studies being under-powered to detect statistically significant findings (especially for rare AEs) and/or AEs that may or may not be considered of special interest and/or clinically important. Such blanket statements are problematic for interpretation, highlighting the need for study authors to clearly report AEs of interest pre- and post-study conduct.	Discussion (p. 18-19)
R1.3 Biological plausibility of short course of corticosteroids causing	We agree with the reviewer that use of short course corticosteroids have	

reduction in height should be considered. There are dangers of relying on single, possibly selectively reported outcomes, in the face of possibly several unknown or unclear studies where they did not report the height because they found no difference. This needs to be contrasted with meta-analyses of long-term corticosteroid use in children with asthma.	the potential to reduce height. We attempted to provide a balanced view of the available evidence. We have drawn attention to some published literature on this in the Discussion: Observational data have also suggested that multiple corticosteroid bursts can increase the risk of growth suppression, fractures, bone mineral accretion and osteopenia in children with underlying respiratory disease.[5, 6, 109] This calls for caution and monitoring of linear growth, particularly when use of high-dose inhaled or systemic corticosteroid is recurrent. We also contrasted the evidence on short course corticosteroids with meta-analyses of long-term corticosteroid use in children with asthma: Although the present study suggests that single doses of systemic or inhaled corticosteroids may result in few AEs, recurrent courses may lead to long-term risks, as cumulative dosing has been shown to be a determinant of safety.[109] We also raise the issue of reporting in the Discussion: However, given the low quality of included studies, the heterogeneous and poor reporting of AEs, and the lack of precision of results, considerable uncertainties remain regarding the safety of high-dose inhaled or systemic corticosteroids for these indications in this age range. This review was limited by the quality of the primary literature, particularly regarding the definition, assessment and reporting of AEs. This underscores the challenges	Discussion (p. 16); Discussion (p. 19); Discussion (p. 15); Discussion (p. 18);

	researchers encounter when attempting to synthesize safety data due to sparse and poor reporting,[117] and highlights the urgent need to enhance detection and reporting of AEs. Additionally, we address the issue of reporting in our response above (R1.2).	See Response to R1.2
Reviewer #2		
R2 General Comment The article is well written. The authors' aim is to perform a systematic review of the literature regarding adverse events linked to a short-course of oral or high-dose inhalatory corticosteroid use for respiratory infections in young children less than 6 years old. The conclusion of the authors is referred to the absence of adverse effects attributable to a short course of oral or high-dose inhalatory corticosteroid. In the intentions of the authors, this systematic review add some evidence: an increase in the safety of the use of a short course of oral corticosteroid and the absence of significant correlation between a short course of high-dose inhaled corticosteroid and significant adverse events. Although this manuscript is characterized by a number of limiting factors, these two evidences have not been demonstrated by previous studies. The articles analyzed in this systematic review came from randomized controlled trial or observational studies, although the methodology is not in line with GRADE criteria. The sample considered (11000 children) is a strength of this systematic review; the methodology seems to be a weakness. In conclusion, this article, because of the big number of the sample considered, can represent a significant contribution to the	Thank you for your comment.	Not applicable

scientific knowledge.		
R2.1 The major criticism regards the missing use of the GRADE scale for the selection of papers; another criticism resides into the “conclusions” section, where authors claim that there is no correlation between short-course of oral or high-dose inhalatory corticosteroid and significant adverse events.	The GRADE approach is not intended to be used in the selection of papers. We followed standard methods for systematic reviews, where selection of papers is based on eligibility according to pre-defined criteria for study design, in addition to population, interventions, comparators, timing and setting (PICOTS). We are unclear about the reviewer’s comment regarding the concluding remark on overall findings. If this is in reference to the term “significant” in this sentence, we have revised this as per response in R2.6: While the existing evidence suggests that short-term high-dose inhaled or systemic corticosteroids is not associated with an increase in AEs across organ systems, uncertainties remain due to low quality of studies, poor reporting and lack of precision of results.	Conclusion (p. 20); See Response to R2.6
R2.2 Page 5 "Search strategies combined index terms and keywords for respiratory illnesses, children and drug classes identified in the Global Initiative for Asthma (GINA) guidelines" Page 5 "Original database searches were conducted September 2014 in Ovid Medline [...] Update searches were executed in Medline and CENTRAL in February 2016, and then again in July 2017" It would be better to clearly specify ”search” terms; they seem too broad in the way described by the authors,. Moreover, why was the literature search carried out on different platforms in 2014, 2016 and 2017? Finally, references 14 and 112 refer to paper published in 2018; the authors write that the last literature update was in 2017.	We reported the literature search according to established standards, aiming to keep this as concise as possible within the body of the main manuscript. We also reported the inclusion of the detailed search strategy in Supplement 1, which specifies terms and dates of searches for each database. The literature search was carried out on different platforms for 2014 (versus 2016 and 2017), as 2014 was the comprehensive, original search strategy. Searches were subsequently updated February 2016 and July 2017, in databases from which the included studies (2014 search) originated. References 14 and 112 are not studies included in the body of the evidence. Reference 14 (GINA) is a website resource on asthma, and contains updated and archived	Methods, Literature search (p. 6); Supplement 1. Search strategy; References (p. 24-34)

	reports (1995 to 2019); the date referenced (January 12, 2018) is when we last accessed the website. Reference 112 (Rieder 2018) is a citation included in the Discussion (p. 16).	
R2.3 Page 6 "Studies that did not report or mention AEs were excluded". It is necessary to specify how the authors made search for adverse events in non-considered articles.	The statement “Studies that did not report or mention AEs are excluded.” refers to the selection of studies during full-text screening of primary studies. Screening occurred after the literature search, and was conducted against pre-defined selection criteria. Therefore, studies that did not report or mention AEs (including adverse drug reactions, adverse drug events, medication errors, side effects or potential adverse drug events) are not eligible for inclusion in the systematic review. For English and non-English records, 20 and 163 were excluded, respectively.	Methods, Eligibility criteria (p. 7); Figure 1. PRISMA study flow selection
R2.4 Page 14. "A common concern when using corticosteroids in young children is effect on growth. Results from a single, small trial (n = 129) of recurrent high-dose inhaled fluticasone propionate in wheezing preschoolers were heterogeneous across outcome measures, but suggested a small significant risk of growth suppression. " Please explain adequately what the author means by the term "recurring".	Our eligibility criteria included single or recurrent doses of systemic corticosteroids. That is, we included more than one dose of corticosteroid treatment, as well as more than one course of treatment as long as each course was ≤14 days in duration (and also having met the criteria of treatment for an acute respiratory condition). There was no criterion for a minimum time interval between courses. Most of the included studies administered a single course of corticosteroids (single or multi-dose, up to a total of 14 days), but some studies administered more than one course over a period of a year or more for multiple respiratory episodes/exacerbations. In the sentence referenced, the study by Ducharme et al (2009) administered 750 mcg of fluticasone propionate (or placebo) twice daily, starting at the onset of an upper respiratory tract infection and continuing for 10 days, over a period of 6 to 12 months.	Supplement 2. Eligibility criteria for study inclusion; Discussion (p. 16)
R2.5 Page 17. "While the McHarm scale is to be used in conjunction	The GRADE approach is a system for rating the quality of a body of	Discussion (p. 19)

with other quality assessment tools to evaluate the broader elements of the study quality of the study. [...] • Page 18. Two to the variation in corticosteroids and an extended range of reported AEs among varied study designs of overall poor quality, we did not attempt to rate the quality of the body of evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach The McHarm scale has been validated by Cochrane (Santaguida P, Keshavarz H, MacQueen G, Levine M, Beyene J, Raina P. Development of the McHarm: A tool evaluating validity of the collection and reporting of harms. In:" Abstracts of the 19th Cochrane Colloquium"; 2011 19-22 Oct; Madrid, Spain. John Wiley & Sons; 2011). The GRADE scale would be the most appropriate for systematic reviews. The authors point out that they did not follow the GRADE scale because of the intrinsic difficulties of their literature research. This aspect represents a strong weakness of this article.	evidence in systematic review and other evidence syntheses, such as health technology assessments, and guidelines and grading recommendations in health care. It is used to rate the body of evidence at the outcome level rather than at the study level. Given the resources needed to use GRADE, we considered that this might be more valuable in case effectiveness/efficacy of corticosteroid interventions was also being addressed, and that our approach to evaluating the methodological quality of included studies highlighted the key questions in this body of evidence. If conducted, we anticipate that the certainty of evidence in the overall body of evidence would be low or very low when considering lack of or poor reporting of AEs, lack of precision among effect estimates, and heterogeneous respiratory conditions, interventions, and comparators.	
R2.6 Page 18. CONCLUSION The consideration that it is difficult to obtain incontrovertible evidence regarding the correlation between short-course of oral or high-dose inhaled corticosteroid and adverse events appears to be justified. The conclusion that their use is not related to a significant increase in significant adverse events appears to be foolhardy; specifically, the term "significant" should be specified. Moreover, the authors underline an increase in the incidence of vomiting episodes after a short-course of oral corticosteroid in the text (page 9).	Thank you for your comment. To avoid confusion and align the statement with the evidence, we have removed "significant" in the Conclusion: While the existing evidence suggests that short-term high-dose inhaled or systemic corticosteroids is not associated with an increase in AEs across organ systems, uncertainties remain due to low quality of studies, poor reporting and lack of precision of results. We report in the Results fewer cases of vomiting for children who received a short course of oral dexamethasone (compared with those who received short course of	Conclusion (p. 20);

	oral prednisone): Meta-analysis of six studies (1,373 children)[25, 27, 41, 49, 52, 80] found fewer cases of vomiting in patients who received dexamethasone compared with another corticosteroid, although the number of events was small (12/663 versus 51/710 cases; pOR 0.29, 95% CI 0.17, 0.48; I^2=0%).	Results (p. 11)
R2.7 References 2, 11, 18, 27, 32, 50, 53, 58, 74, 79 are not in line with the editorial guidelines.	Thank you for pointing these out. We could not detect how/where the errors were for these references. We have reviewed all the references and attempted revisions to ensure alignment with editorial guidelines as much as possible.	References (p. 24-34)
R2.8 Reference 14 comes from GINA guidelines 2018, while the authors have specified in the text that the last update was in 2017. Same consideration is referred to reference 112, whose publication year is 2018.	References 14 and 112 are not studies included in the body of the evidence. We responded to a similar comment above (R2.2): Reference 14 (GINA) is a website resource on asthma, and contains updated and archived reports (1995 to 2019); the date referenced (January 12, 2018) is when we last accessed the website. Reference 112 (Rieder 2018) is a citation included in the Discussion (p. 16).	Methods, Literature search (p. 6); References (p. 24-34); See Response to R2.2
R2.9 Page 14. "Importantly, these results can help future research in the collection and reporting of AEs, particularly concerning the effects of growth and behavioral outcomes; "this in turn is needed to help inform decision making between clinicians and parents / caregivers of young children." This sentence could be inserted in the “conclusions” section; it is not remarkable in the "discussion" section.	We have moved this sentence from the Discussion to the Conclusion: Importantly, these results can help guide future research in the collection and reporting of AEs, particularly concerning measures of growth and behavioral outcomes; this in turn is needed to help inform shared decision-making between clinicians and parents/caregivers of young children.	Conclusion (p. 20)
Reviewer #3		
R3 General Comment As requested, I focused mainly on the statistical methods and analyses. The authors undertook a daunting task. As they explain under Strengths and limitations (on page	Thank you for this comment. We agree with these sentiments.	Not applicable

17), their extensive systematic review "was limited by the quality of the primary literature, particularly regarding the definition, assessment and reporting of AEs. This underscores the challenges researchers encounter when attempting to synthesize safety data due to sparse and poor reporting, and highlights the urgent need to enhance detection and reporting of AEs."		
Worse, they are in a situation where meta-analysis provides little help. From Supplement 3c, I have the impression of substantial heterogeneity among the studies in characteristics such as design, doses, and conditions. Thus, even if the studies available for a particular meta-analysis contributing to Figure 2, Figure 3, or Figure 4 show little or no statistical heterogeneity, the justification for combining their effects may be weak.	While there was substantial clinical heterogeneity among the multitude of included studies, we only performed meta-analysis on smaller subsets that the review team deemed sufficiently homogeneous in terms of population characteristics, corticosteroid type, formulation, equivalent dose and duration. Also, since we were specifically looking at safety outcomes in pediatric populations, we believe the pooling of studies in the cases we did is justifiable.	Supplement 3c. Characteristics of included studies; Figure 2. Forest plot of adverse events – systemic vs. placebo; Figure 3. Forest plot of adverse events – inhaled vs. placebo; Figure 4. Forest plot of adverse events – dexamethasone vs. other
R3.1 The statistical methods have serious shortcomings. The Peto method produced all the odds ratios reported in Figures 2, 3, and 4; but Greenland and Salvan (1990) studied its behavior and concluded, "The one-step (Peto) method for obtaining pooled effect estimates can yield extremely biased results when applied to unbalanced data. Even for balanced studies, the one-step method may incorporate an unacceptable degree of bias." I was not able to examine the degree of imbalance (if any) in the individual meta-analyses, because Figures 2, 3, and 4 and Supplement 5 give only the total numbers of children and AEs in a meta-analysis, and not the numbers in the individual studies.	We are aware of potential issues with the Peto method of pooling binary data. The primary conditions that can make this method problematic are 1) unbalanced trial arms, 2) common outcomes, and 3) large effect sizes. With few exceptions, our meta-analyses had events that were quite rare, small effect sizes, and balanced trials—situations where the Peto method performs quite well. In addition, there were many trial arms with zero events; a situation where the Peto method does not require the arbitrary "add 0.5 to each cell" approach taken by the Mantel-Haenszel Method. Bradburn et al (2007) demonstrated that the Peto method often performs best in these situations. The forest plots of all meta-analyses have been included in Supplement 6. Bradburn MJ, Deeks JJ, Berlin JA, Localio AR. Much ado about	Methods, Data synthesis (p. 8); Supplement 6. Forest plots of adverse events

	nothing: a comparison of performance of meta-analytical methods with rare events. Statistic in Medicine, 2007; 26:35-77.	
R3.2 The authors say (page 7, lines 33-35) that they pooled risk difference by using "a Mantel-Haenszel random effects model." The statement is problematic, because no such model exists, despite the impression created by the Review Manager software. According to the memorandum "Statistical algofithms in Review Manager 5," the weights and pooled estimate from the fixed-effect Mantel-Haenszel method are used in an alternative version of the heterogeneity statistic Q, which is then used in estimating the between-study variance for use in the inverse-variance weights of a random-effects pooled estimate. That is the extent of the difference between the "Mantel-Haenszel random effects method" and the usual inverse-variance random-effects method, introduced by DerSimonian and Laird, and the resulting estimates generally only slightly. As far as I'm aware, the "Mantel-Haenszel random effects method" exists only in Review Manager 5. No detailed specification of it appears in the meta-analysis literature, and it is not supported by any theoretical or empirical analysis of its properties. Thus, users of Review Manager who choose the "M-H random" option should not assume that, by doing so, they can avoid the well-documented short-comings of the DerSimonian-Laird method (see, for example, IntHout et al. 2014).	We have amended the methods section with the following: Risk differences were pooled using the DerSimonian Laird inverse variance random effects method utilizing the Mantel-Haenszel Q statistic. We are aware of the potential problems of the DerSimonian Laird method and did not assume using the Mantel-Haenszel option would eliminate them. We presented the risk difference estimates more as an alternative to the Peto odds ratios numbers, since the latter could not incorporate the trials (sometimes substantial amounts) that had zero outcomes in both arms—using risk difference allowed us to include these trials in the analysis.	Methods, Data synthesis (p. 8)
R3.3 The authors say (page 7, lines 40-42), "One AE (growth) was reported as a continuous outcome and data were pooled using a mean difference." They do not document the method that they used for this (the Mantel-Haenszel method is applicable only to odds ratio, risk	We have added to our methods section that growth was analyzed using a DerSimonian Laird inverse variance random effects method: One AE (growth) was reported as a continuous outcome and data were pooled using a DerSimonian Laird	Methods, Data synthesis (p. 9)

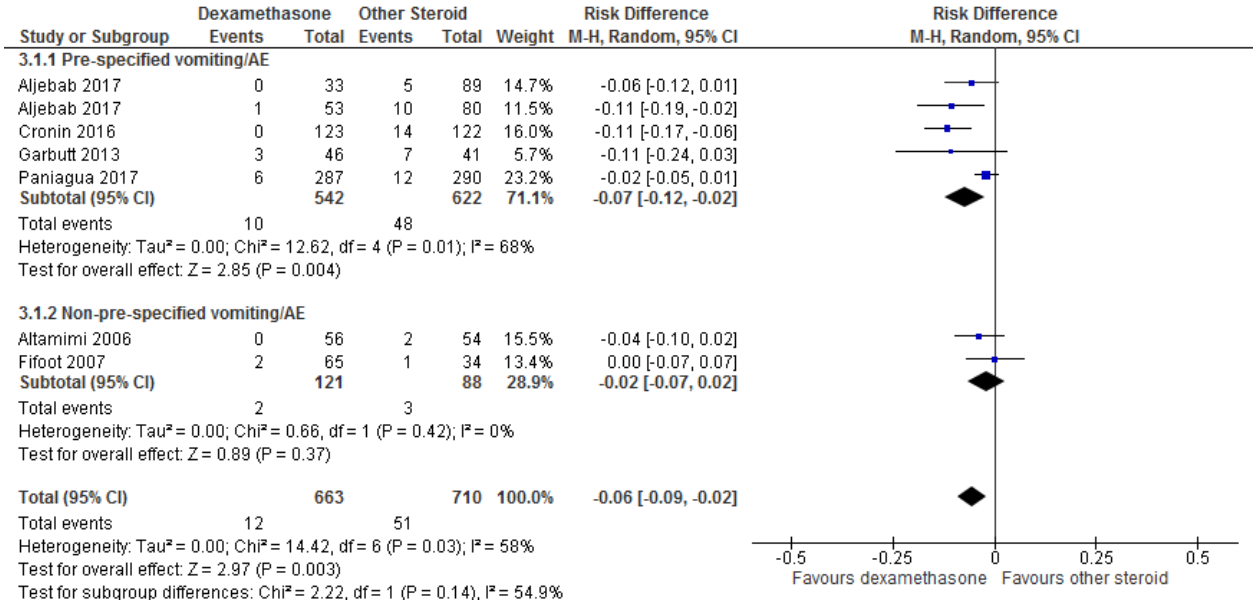
ratio, and risk difference).	inverse variance random effects method as a mean difference (MD; in cm).	
R3.4 Unfortunately, there are more difficulties. The usual test for statistical heterogeneity, based on Q, uses an incorrect null distribution (Hoaglin 2016). For that reason and because the correct null distribution differs substantially among measures of effect, I^2 unfortunately has no useful interpretation (Hoaglin 2017).	Because we presented many meta-analyses summarized in tables, we felt a need to give readers a quick summary of approximately how much heterogeneity was in each analysis, without having to consult the forest plots. While we agree that the I^2 measure has interpretation issues that continue to be revealed, it remains the best and most succinct quantification of the heterogeneity present in each analysis and is still useful to present (Hedges 2016). We did not present any confidence intervals around I^2, nor did we try to interpret them as percentage of heterogeneity due to between studies variance. We have added to our methods section a caveat about the potential danger of misinterpretation of this statistic, but that we still present them for informational purposes: The I^2 statistic was presented to quantify the magnitude of statistical heterogeneity between studies; while the I^2 has the potential to be misinterpreted, we chose to present this statistic for informational purposes.[19] Hedges LV. Comment on 'Misunderstandings about Q and "Cochran's Q Test" in meta-analysis'. Statistics in Medicine 2016;35(4);496-497.	Methods, Data synthesis (p. 9)
R3.5 It is discouraging that the authors (and many others) have been so ill-served by the Cochrane Collaboration. Users should be able to count on up-to-date software, documentation, and guidance. Despite these criticisms of the statistical methods, the Results section contains valuable summaries of the available evidence. It may be possible to preserve that contribution	We have added the appropriate caveats mentioned here about the meta-analyses in Responses R3.2-3.4.	See Responses R3.2-3.4

while de-emphasizing the meta-analyses (e.g., include appropriate caveats on the Peto method and avoid the “Mantel-Haenszel random effects method”).		
R3.6 The discussion (page 16, lines 8 to 10) repeats the result that “evidence favored oral dexamethasone over oral prednisone for vomiting.” In view of the sizable number of comparisons (Figures 2, 3, and 4 contain total of 33 confidence intervals), it seems likely that the authors are capitalizing on chance.	While we did not present p-values, this particular estimate has a very small p-value (<0.00001) and thus maintains its statistical significance even in the face of multiple testing.	Discussion (p. 17)
R3.7 In the interest of transparency and reproducibility, it would be a good idea to include (among the supplements) the study-level data for each of the various meta-analyses. Interested readers should not have to request those data from the corresponding author (page 20, line 31).	Thank you for this suggestion. We recognize that there is a push for open access to published data in the interest of greater transparency and reproducibility. We have included study level data for each meta-analysis (forest plots from RevMan) in Supplement 6.	Supplement 6. Forest plots of adverse events
R3.8 I noticed some rough edges in the manuscript. Pages 32 and 33: “Table 5” should be “Table 1”. Are the very small numbers of participants (2 for Respiratory distress and 1 for each of Psychosis, Positive wheal, Hematology, and Tumor cell proliferation) correct?	Thank you for pointing these out. We have revised the ‘Number of studies and participants reporting adverse events’ as Table 2, in order of tables reported in the main manuscript. The very small numbers (e.g., 1, 2) of participants for respiratory distress (Nahum 2009), psychosis (Lee 2001), positive wheal (Lehmann 2008), hematology (Sadowitz 2012) and tumor cell proliferation (Panigada 2014) are reported in case reports or case series.	Table 2. Number of studies and participants reporting adverse events
R3.9 Also, “Table 1” shows a total of 1635 participants (in 5 studies) for Systemic infections, but Figure 2 and Supplement 6b (which should be named 5b) show 1095 + 1083 = 2178 (in 4 studies).	The number of studies and participants in ‘Table 2. Number of studies and participants reporting adverse events’ captures adverse events reported by all of the included studies, but does not delineate the various interventions/comparators within each study (systemic vs. placebo, inhaled vs. placebo, dexamethasone vs. other corticosteroid).	Table 2. Number of studies and participants reporting adverse events; Supplement 5a. Effect estimates for all adverse events with subgroups – Infection & respiratory system

	Data on systemic infections were pooled from four studies that examined a systemic corticosteroid arm with a placebo arm involving 2178 children (Bjornson 2004, Corneli 2007, Daugbjerg 1993 and Plint 2009); this was captured under 'Systemic infections' in Figure 2 and Supplement 5a ('Systemic infection, overall>Systemic vs. placebo'). Data on systemic infections were also pooled from two studies that examined an inhaled corticosteroid arm with a placebo arm involving 129 children (Daugbjerg 1993 and Ducharme 2009) and this was captured under 'Systemic infections, overall>Inhaled vs. placebo>Multi-dose, wheeze' in Supplement 5a. Infection & respiratory system. Therefore, the multiple comparisons in Daugbjerg 1993 were captured in Supplement 5a under 'Systemic vs. placebo' and under 'Inhaled vs. placebo'. However, Table 2 only captures this study (and its participants) once, to avoid double counting its multiple contributions to pooled estimates.	
R3.10 Figures 2, 3, and 4 should say that all the odds ratios are pORs from meta-analyses (except when the number of studies is 1).	Thank you for pointing this out. We have revised Figures 2-4 to indicate Peto odds ratios.	Figure 2. Forest plot of adverse events – systemic vs. placebo; Figure 3. Forest plot of adverse events – inhaled vs. placebo; Figure 4. Forest plot of adverse events – dexamethasone vs. other
R3.11 Why does Supplement 3b account for only 83 studies?	Thank you for pointing this out. There was an omission error of two studies/comparisons in Supplement 3b. This has been corrected.	Supplement 3b. Summary characteristics of included studies – comparisons
R3.12 The parts of Supplement 5 are numbered incorrectly, as Supplement 6a, etc.	Thank you for pointing this out. Each table in Supplement 5 has been corrected to 5a, 5b, etc.	Supplement 5. Effect estimates for all adverse events with subgroups

Response to Reviewer (R1.1)

Dexamethasone vs. Other – subgroups for pre- vs. non-pre-specified vomiting/AEs



Pre-specified AEs (if done) among studies:

Aljebab 2017 – vomiting, nausea, abdominal pain

Altamimi 2006 – no AEs a priori; efficacy study

Cronin 2016 – vomiting

Fifoot 2007 - no AEs a priori; efficacy study

Garbutt 2013 – open ended for AEs, including sleep problems, mood changes, headache, dizziness, nausea, stomach pain, secondary infections, vomiting and tremor

Paniagua 2017 - vomits

Dexamethasone vs. Other – no subgroups