PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Corticosteroids and risk of gastrointestinal bleeding: a systematic review and meta-analysis
AUTHORS	Narum, Sigrid; Westergren, Tone; Klemp, Marianne

VERSION 1 - REVIEW

REVIEWER	Mark Hannon Academic Department of Endocrinology Beaumont Hospital / RCSI Medical School Dublin Ireland
REVIEW RETURNED	18-Feb-2014

GENERAL COMMENTS	This is a good paper and worthy of publication. However, I do not agree with the conclusions drawn by the authors. Specific criticisms: 1. Page 5 line 33: poor English 2. Page 8: The table does not have any p values comparing the studies with bleeding and those without bleeding. There is an error in the numbers as line 8 says that 1 study involved ambulant patients and 12 did not, whereas the table (line 34) shows 2 studies involving ambulant patients. 3. Page 8: Given that the studies analysed were mostly performed relatively recently, why are so few patients on PPIs?? Surely there must have been a gross under-reporting of the number of patients on PPIs, this needs to be commented upon. 4. Page 9 line 48: these data need to be shown. 5. page 9 line 41 - 45: the data are analysed with/without documentation of concomitant NSAID use which is commendable. However, can the same analysis be performed for PPIs? It would be extremely useful if so. Page 10 line 23: Given my earlier points, this conclusion cannot be drawn as we are missing some fundamental data about PPI usage. Maybe most ambulatory patients were on PPIs? If these data are known they need to be shown. Page 11 line 18: Again, comment needs to be made here on PPI usage. Page 12 last paragraph / Page 13: As patients with dyspepsia were not analysed, there is no way of determining whether patients with

	dyspepsia simply went to their GP and were put on PPIs. It is not possible to say with certainty from these data that ambulatory patients on steroids do not have an increased bleeding risk given the heterogeneous nature of the studies analysed and the lack of data concerning PPI usage. This is a useful paper and appropriate for publication. However, the data presentation needs to be tightened up with more data on PPI usage if that conclusions drawn are to be supported.

REVIEWER	Katie Saunders Cambridge Centre for Health Services Research
REVIEW RETURNED	06-Mar-2014

GENERAL COMMENTS	This is a methods/statistics review: This paper represents a substantial piece of work by the study authors. I was impressed by the clarity and detail of their supplementary table of study characteristics. Extracting data from 159 studies is an impressive task. This table should form part of the supplementary material available from this publication The results and conclusions are plausible. However I have strong methodological concerns about this work: **Search strategy** The search strategy is thorough and clearly described. The authors should clearly justify why they did not search databases of clinical trials, such as clinicaltrials.gov or the European clinical trials register and could consider doing so to improve this review. I note that German and Scandinavian languages were included. Could these search terms be included in a supplementary appendix? I was also interested that the search strategy included specific conditions or diseases. Why was this done? Surely the interest of this review is for ANY use of steroids, rather than just disease specific. For example “lung diseases” is a pretty specific text string to search for. I was surprised that only one trial out of all 159 identified was for COPD, and wonder whether it is related to specificity of the search string. Please expand this justification. **Selection of clinical trials over observational data** The authors argue that clinical trials are more appropriate than observational data for their research hypothesis. In general this is a

	good suggestion for main effects. However this is a study of rare adverse events for what is essentially a common treatment. Observational data may have several strong advantages to trial adverse event data, for example single studies may have several hundred events, while many of the included studies in this meta-analysis have no events at all, and this was not the main outcome of the included studies. The authors need to strengthen their argument here. **Methodological quality assessment of included trials** The authors state in the methods that methodological assessment of eligible trials was done by including only randomised double blind studies. Although I agree with the authors that this is a good way to ensure the quality of the included trials when performing a meta-analysis of the main study outcome it is NOT the same as a rigorous methodological assessment for bias for the reporting of adverse events. There is a section in the Cochrane handbook on this: **14.6 Assessing risk of bias for adverse effects** I am copying the following section from this page: *Examples of potentially useful questions to consider in assessing the quality of evidence on adverse effects are:* *On conduct:* • *Are definitions of reported adverse effects given?* • *Were the methods used for monitoring adverse effects reported? Use of prospective or routine monitoring; spontaneous reporting; patient checklist, questionnaire or diary; systematic survey of patients?* *On reporting:* • *Were any patients excluded from the adverse effects analysis?* • *Does the report provide numerical data by intervention group?* • *Which categories of adverse effects were reported by the investigators?* The review authors should preferably consider incorporating this into their bias assessment. **Statistical analysis**

	The statistical analysis methods for meta-analysis of rare adverse events and for meta-analysis of studies with zero events are both relevant to this paper. The authors mention in the methods and argue again in the discussion that a strength of their analysis is that they applied a correction factor of 1 to both groups to overcome the issue of zero events, and that this allowed them to include a large number of studies into their meta-analysis with zero events, which would otherwise have been excluded. The section of the Cochrane Handbook which deals with zero events (reference 23 in the paper) 16.9.5 Validity of methods of meta-analysis for rare events comments that *In other circumstances (i.e. event risks above 1%,[[such as in this study]] very large effects at event risks around 1%, and meta-analyses where many studies were substantially imbalanced) the best performing methods were the Mantel-Haenszel OR without zero-cell corrections, logistic regression and an exact method. None of these methods is available in RevMan.* *Methods that should be avoided with rare events are the inverse-variance methods (including the DerSimonian and Laird random-effects method). These directly incorporate the study's variance in the estimation of its contribution to the meta-analysis, but these are usually based on a large-sample variance approximation, which was not intended for use with rare events. The DerSimonian and Laird method is the only random-effects method commonly available in meta-analytic software. We would suggest that incorporation of heterogeneity into an estimate of a treatment effect should be a secondary consideration when attempting to produce estimates of effects from sparse data – the primary concern is to discern whether there is any signal of an effect in the data.* The authors of this review need extremely strong justification for why they have not followed the recommendations of the Cochrane Handbook for analysis of this type of data, and I would not consider it a particular strength of the analysis. On a practical note did the authors used a continuity correction of 1 rather than 0.5 as this allowed them to enter the data directly into RevMan with this continuity correction already added. I would not consider this correct, and use of a specific continuity correction should be methodologically sound and justified. Sub-group and sensitivity analyses appear appropriate and well justified It should be stated whether the heterogeneity statistics presented in the forest plot have been estimated with the continuity correction already added. This is a very heterogeneous set of studies in terms of patient and treatment characteristics. I think that this needs commenting on, even if the fact that no individual study was adequately powered to detect the bleeding outcome means lack of heteroegeneity statistically between individual studies.

VERSION 1 – AUTHOR RESPONSE

Comments 2 – 11 from reviewer Mark Hannon:

2. Page 5 line 33: poor English

Our reply: We assume Mr. Hannon points to the sentence starting with “Trials, where other adverse effects…” This sentence has now been rewritten: ”In some trials, other adverse effects were reported in the results section but no gastrointestinal bleeding was listed. These studies were included only if adverse event monitoring was described in the methods section or if it was judged reasonable to expect from the adverse event monitoring system that any gastrointestinal adverse effects would have been recorded.”

3. Page 8: The table does not have any p values comparing the studies with bleeding and those without bleeding.

Our reply: P-values have been added in table 2.

4. There is an error in the numbers as line 8 says that 1 study involved ambulant patients and 12 did not, whereas the table (line 34) shows 2 studies involving ambulant patients.

Our reply: In line 8, use of gastroprotective drugs was described for hospitalized vs ambulant patients. In table 2; use of gastroprotective drugs was described for studies with bleeding vs non-bleeding. The numbers in the brackets do not give any essential information and may seem confusing. We have thus deleted the brackets with its content. In addition we have rewritten the sentences in this paragraph. When we re-read all the primary articles, we found use of gastric protection in 14 studies.

5. Page 8: Given that the studies analysed were mostly performed relatively recently, why are so few patients on PPIs?? Surely there must have been a gross under-reporting of the number of patients on PPIs, this needs to be commented upon.

Our reply: We agree with the reviewer that there might be some under-reporting or undisclosed use of gastroprotective drugs among the study participants/patients in the primary studies. Use of PPI`s was not described or part of most clinical trial protocols. In some cases, other gastroprotective agents were part of the trial medications, or were allowed in the study. We do not however think there is any gross under-reporting.

Firstly, in the period from 1983 to 1989 H2 blockers and antacids were the only gastric protection in use. Omeprazole, the first PPI, was first marketed in the United States in 1989, and became an over-the-counter drug some years later (in US and UK around 2003/ 2004). This means that the occurrence of under-reporting and undisclosed use of PPI`s might have changed over time during the study period. At the same time, the conformity of study reporting has improved. We have also recorded use of other medications such as drugs used “at the doctors discretion”, or “Additional medical therapy was provided according to standard clinical practice” etc, which might indicate a possibility of use of gastroprotective drugs. We found this described in 12 studies in addition to the 14 studies where use of gastroprotective drugs were described.

Secondly, the duration of treatment was relatively short (median 8.5 days). Therefore, we believe that many studies have been completed before the patients treated ambulatory would have developed gastric discomfort, realised it and have come to see a doctor or pharmacy to get a PPI, H2 blocker or an antacid. Many of the hospitalized patients were sedated and mechanically ventilated, and were thus not able to communicate gastric discomfort. Because we included only randomized, double-blind studies, we believe risk of bias due to selective use and reporting of gastroprotective drugs to the steroid group is low.

We have now added a few sentences regarding the possibility of PPI under-reporting to the Discussion section under “limitations of the review”.

6. Page 9 line 48: these data need to be shown.

Our reply: See table 3. The data for the subgroup analysis without newborns in prevention of bronchopulmonary dysplasia has been added. The data for the newborns in prevention of bronchopulmonary dysplasia is shown in figure 2. In addition, we have added two columns to table 3, showing number of events for each subgroup analysis and events per 1000 patients.

7. Page 9 line 41 - 45: the data are analysed with/without documentation of concomitant NSAID use which is commendable. However, can the same analysis be performed for PPIs? It would be extremely useful if so.
Our reply: Se table 3. Two rows have been added; ie. subgroup analyses with and without use of gastroprotective drugs. In addition, we have described the analysis in the text ;" When studies with peptic ulcer as exclusion criterion and studies with concomitant use of gastroprotective drugs were subsequently excluded from the analyses, there were little change in the risk of bleeding or perforation in the remaining studies (table 3)." In addition, use of concomitant drugs "according to standard clinical practice etc.", which may potentially include use of gastroprotective drugs, was described in 12 studies, and have been included in the discussion part under "limitations of the review".
8. Page 10 line 23: Given my earlier points, this conclusion cannot be drawn as we are missing some fundamental data about PPI usage. Maybe most ambulatory patients were on PPIs? If these data are known they need to be shown.
Our reply: The main focus of this review was not to study the effect of PPI-use, so we have now adjusted our conclusion. See abstract (Conclusion) and discussion (first paragraph and last paragraph).
9. Page 11 line 18: Again, comment needs to be made here on PPI usage.
Our reply: We have made changes to the last paragraph in the "Strengths and limitations section" where we now discuss possible under-reporting and undisclosed use of gastroprotective drugs.
10. Page 12 last paragraph / Page 13: As patients with dyspepsia were not analysed, there is no way of determining whether patients with dyspepsia simply went to their GP and were put on PPIs. It is not possible to say with certainty from these data that ambulatory patients on steroids do not have an increased bleeding risk given the heterogeneous nature of the studies analysed and the lack of data concerning PPI usage.
Our reply: We agree with the reviewer that patients treated ambulatory may possibly have undisclosed PPI use. However, since all the included studies are RCTs, the patients are instructed not to take any undisclosed drugs during the study period. We have now discussed this in the "Strengths and limitations section" and in the "Clinical implications section" and we have adjusted the last paragraph in the paper.
11. This is a useful paper and appropriate for publication. However, the data presentation needs to be tightened up with more data on PPI usage if that conclusions drawn are to be supported.
Our reply: We thank the Reviewer for the helpful comments. See our detailed replies above.
-
Comments 12 – 17 from reviewer Katie Saunders
This is a methods/statistics review:
This paper represents a substantial piece of work by the study authors. I was impressed by the clarity and detail of their supplementary table of study characteristics. Extracting data from 159 studies is an impressive task. This table should form part of the supplementary material available from this publication. The results and conclusions are plausible. However I have strong methodological concerns about this work:
12. Search strategy.
The authors should clearly justify why they did not search databases of clinical trials, such as clinicaltrials.gov or the European clinical trials register and could consider doing so to improve this review.
Our reply: We acknowledge that clinical trial databases are essential for literature search in fields with sparse documentation, new entities or few published results. However, for treatment with corticosteroids we found plenty of published studies. Therefore, we limited the search to results published in journals. To test if we missed many studies by not doing a search in the clinical trial databases, we did a search in Clinicaltrials.gov for 1980-2014. We searched for the same steroids AND placebo as in Medline/Embase, and got 87 records. Compared to the 159 studies included from the 3483 records identified in Medline/Embase, we do not think a search in the clinical trials

databases would change our results.
13. I note that German and Scandinavian languages were included. Could these search terms be included in a supplementary appendix?
Our reply: The Medline/Embase search terms were only in English, and were as stated in supplementary file 1: Search strategy. The title or abstract indicated if the article was in another language than English. According to the protocol, studies in these languages could be included in addition to studies in English.
14. I was also interested that the search strategy included specific conditions or diseases. Why was this done? Surely the interest of this review is for ANY use of steroids, rather than just disease specific. For example “lung diseases” is a pretty specific text string to search for. I was surprised that only one trial out of all 159 identified was for COPD, and wonder whether it is related to specificity of the search string. Please expand this justification.
Our reply: The main literature search for this systematic review was performed as stated in supplementary file 1: Search strategy. The search for specific conditions or diseases was an additional hand search we performed when we got aware of the existing Cochrane meta-analyses of corticosteroids for traumatic brain injury, meningitis and in preterm infants. For this specific search, only the Cochrane Database was searched for the 7 specific conditions other than “Miscellaneous” shown in table 1. The aim was to identify missing publications from those meta-analyses and detect similar meta-analyses for the additional four conditions.
In the 3483 records identified, there were more trials for COPD. The reason why we did not include more trials for COPD in the present systematic review, was that these trials did not fit our inclusion criteria (i.e. corticosteroid not the active ingredient, steroids in both arms, and local vs systemic treatment).
We have now highlighted in the manuscript that this was an additional search.
15. Selection of clinical trials over observational data
The authors argue that clinical trials are more appropriate than observational data for their research hypothesis. In general this is a good suggestion for main effects. However this is a study of rare adverse events for what is essentially a common treatment. Observational data may have several strong advantages to trial adverse event data, for example single studies may have several hundred events, while many of the included studies in this meta-analysis have no events at all, and this was not the main outcome of the included studies. The authors need to strengthen their argument here.
Our reply: We agree with the reviewer in that observational studies may be more appropriate than clinical trials in detecting rare adverse effects. Regarding corticosteroid use and risk of gastrointestinal bleeding, there still exists uncertainty whether this is a true risk, or if the sickest and most fragile patients are treated differently than the more healthy ones. In addition, observational studies of cohorts over time would be more exposed to confounding factors, such
as use of NSAIDs and gastroprotective agents. Both these groups of drugs are freely available as over-the-counter medications. Steroid dosages and exposure may vary to a greater extent than in an RCT, due to fluctuations in disease activity. Therefore, we found it appropriate only to include randomized controlled trials in this review. We have now added a sentence about possible advantages with observational studies when assessing rare adverse events in the Introduction section and the rationale for choosing only RCTs.
16. Methodological quality assessment of included trials
The authors state in the methods that methodological assessment of eligible trials was done by including only randomised double blind studies. Although I agree with the authors that this is a good way to ensure the quality of the included trials when performing a meta-analysis of the main study outcome it is NOT the same as a rigorous methodological assessment for bias for the reporting of adverse events.
There is a section in the Cochrane handbook on this:14.6 Assessing risk of bias for adverse effects
I am copying the following section from this page:
Examples of potentially useful questions to consider in assessing the quality of evidence on adverse effects are:

On conduct:
• Are definitions of reported adverse effects given?
• Were the methods used for monitoring adverse effects reported? Use of prospective or routine monitoring; spontaneous reporting; patient checklist, questionnaire or diary; systematic survey of patients?
On reporting:
• Were any patients excluded from the adverse effects analysis?
• Does the report provide numerical data by intervention group?
• Which categories of adverse effects were reported by the investigators?
The review authors should preferably consider incorporating this into their bias assessment.
Our reply: We have now recorded the definitions of the adverse effects and the methods used for monitoring these. Both are included as separate columns in the supplementary table of study characteristics (column Q and R). We found some diversity in the reported definitions of gastrointestinal bleeding and differences in the methods used for monitoring these. To describe and discuss this point, we have added some sentences to the Methods section, the Results section, and the Discussion section.
The numerical data by intervention group are presented in the forest plot in supplementary file 2.
17. Statistical analysis
The statistical analysis methods for meta-analysis of rare adverse events and for meta-analysis of studies with zero events are both relevant to this paper.
The authors mention in the methods and argue again in the discussion that a strength of their analysis is that they applied a correction factor of 1 to both groups to overcome the issue of zero events, and that this allowed them to include a large number of studies into their meta-analysis with zero events, which would otherwise have been excluded.
The section of the Cochrane Handbook which deals with zero events (reference 23 in the paper) 16.9.5 Validity of methods of meta-analysis for rare events comments that In other circumstances (i.e. event risks above 1%,[[such as in this study]] very large effects at event risks around 1%, and meta-analyses where many studies were substantially imbalanced) the best performing methods were the Mantel-Haenszel OR without zero-cell corrections, logistic regression and an exact method. None of these methods is available in RevMan.
Methods that should be avoided with rare events are the inverse-variance methods (including the DerSimonian and Laird random-effects method). These directly incorporate the study's variance in the estimation of its contribution to the meta-analysis, but these are usually based on a large-sample variance approximation, which was not intended for use with rare events. The DerSimonian and Laird method is the only random-effects method commonly available in meta-analytic software. We would suggest that incorporation of heterogeneity into an estimate of a treatment effect should be a secondary consideration when attempting to produce estimates of effects from sparse data – the primary concern is to discern whether there is any signal of an effect in the data. The authors of this review need extremely strong justification for why they have not followed the recommendations of the Cochrane Handbook for analysis of this type of data, and I would not consider it a particular strength of the analysis.
On a practical note did the authors used a continuity correction of 1 rather than 0.5 as this allowed them to enter the data directly into RevMan with this continuity correction already added. I would not consider this correct, and use of a specific continuity correction should be methodologically sound and justified. Sub-group and sensitivity analyses appear appropriate and well justified. It should be stated whether the heterogeneity statistics presented in the forest plot have been estimated with the continuity correction already added .This is a very heterogeneous set of studies in terms of patient and treatment characteristics. I think that this needs commenting on, even if the fact that no individual study was adequately powered to detect the bleeding outcome means lack of heterogeneity statistically between individual studies.
Our reply: We agree with the reviewer and have thus re-analysed the data without applying the correction factor to the groups with zero events. Without the correction factors, we find an even higher

risk of bleeding or perforation than in the first analysis. This means the odds ratios resulting from the meta-analysis and the sensitivity analyses have now been adjusted in the revised manuscript. Additionally, we have discussed the heterogeneity issue in some added sentences in the “limitations” section of the manuscript.