

Interpretation of subgroup analyses in systematic reviews: A tutorial

Marty Richardson^{a,*}, Paul Garner^a, Sarah Donegan^b

^a Centre for Evidence Synthesis in Global Health, Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK

^b Department of Biostatistics, Block F Waterhouse Building, University of Liverpool, Liverpool, UK



ARTICLE INFO

Keywords:

Meta-analysis
Systematic reviews
Subgroup analyses

ABSTRACT

Authors of systematic reviews may perform subgroup analyses to investigate how treatment effect varies across different subgroups of patients or trials. Previous research has shown that Cochrane review authors do not sufficiently report their interpretation of subgroup analyses. Consequently, we developed a tutorial with the aim of improving the interpretation of subgroup analyses in reviews. We explain the importance of interpreting subgroup analyses, and demonstrate how to interpret subgroup analyses using theoretical examples and a real-life subgroup analysis with clinical context. Finally, we provide recommendations for the interpretation of subgroup analyses in systematic reviews.

1. Introduction

Many systematic reviews use a statistical technique called meta-analysis to combine individual study results to obtain a pooled treatment effect estimate. Review authors may also explore how this treatment effect varies across different subgroups of patients or trials, by performing subgroup analyses. In a subgroup analysis, all participant data included in the meta-analysis is split into subgroups, according to patient characteristics (such as gender) or trial characteristics (such as geographical location), and a meta-analysis is then performed on one or more of these subsets. Such analyses can be used to investigate sources of heterogeneity (differences between treatment effects from individual trials in the meta-analysis), or to provide estimates of treatment effect for clinically relevant subgroups of patients, i.e. the review authors have reason to believe that treatment effect may vary among different subgroups of patients, perhaps due to results from previously conducted studies.

A review of 52 published Cochrane reviews showed that 63% (33/52) applied subgroup analyses.¹ Similarly, a review of how heterogeneity is addressed in Cochrane reviews² found that 69% (27/39) of reviews undertook subgroup analyses. According to Donegan et al.,¹ when interpreting subgroup analyses, review authors could consider five key criteria:

- 1 whether a statistically significant subgroup difference (interaction) was detected
- 2 the covariate distribution (i.e. the number of trials and participants contributing to each subgroup)

- 3 the plausibility of the interaction or lack of interaction
- 4 the importance of the interaction or lack of interaction
- 5 the possibility of confounding

Donegan et al.¹ reported that Cochrane review authors did not sufficiently report their interpretation of subgroup analyses; only 3% (1/33) of reviews reported whether there was an interaction for each performed subgroup analysis, and 39% (13/33) reported the covariate distribution. No review authors discussed the importance or plausibility of the interaction/lack of interaction, or the possibility of confounding.

Considering these findings,¹ it seems that there is a need to improve the interpretation of subgroup analyses by review authors in systematic reviews. Firstly, we hypothesise that Cochrane review authors may not know how to interpret subgroup analyses, and we report the results of a survey that was conducted to investigate this hypothesis. Secondly, we aim to improve the interpretation of subgroup analyses in reviews by emphasising the importance of interpreting subgroup analyses, and by explaining how to interpret subgroup analyses using theoretical examples and a real-life subgroup analysis with clinical context. The introduction of a subgroup analysis with clinical context allows a demonstration of how a subgroup analysis ought to be interpreted when clinical information is known, such as the plausibility and importance of a subgroup effect, and the possibility of other factors confounding the subgroup analysis.

2. Survey of Cochrane review authors

In order to determine whether review authors know how to

* Corresponding author.

E-mail address: martha.richardson@lstm.liverpool.ac.uk (M. Richardson).

interpret subgroup analyses, we surveyed the 51 authors of the 52 Cochrane reviews identified in Donegan et al.'s study.¹ Full details of the search strategy, eligibility criteria, and review selection methods are provided in the original paper. We asked authors to interpret five subgroup analyses using an online survey site. A forest plot was presented for each of the subgroup analyses. The presented subgroup analyses had no clinical context, and consequently it was only possible to assess whether review authors knew how to interpret subgroup analyses with regards to criteria 1) and 2) of the previously listed criteria. To fulfil criteria 3) - 5), knowledge of the clinical area is required.

We received survey responses for 28/51 (55%) review authors. When asked whether there was a statistically significant subgroup effect, 17% of review authors “didn't know” whether there was a statistically significant subgroup effect for at least one of the presented subgroup analyses, while 28% of review authors answered incorrectly for at least one subgroup analysis. When asked to interpret the results of the subgroup analysis in their own words, 47.4% of review authors did not consider the covariate distribution for any of the presented subgroup analyses. These results suggest that Cochrane review authors do not know how to interpret subgroup analyses with regards to criteria 1) and 2) of the previously listed criteria.

3. Why is it important to perform and interpret subgroup analyses?

If a meta-analysis is performed across heterogeneous trials, it may be inappropriate to draw conclusions from the pooled treatment effect estimate; however, if the same trials are subgrouped and there is no heterogeneity within trials (i.e. results for individual trials within each subgroup are similar), then valid conclusions can be drawn using results from the subgroup analysis. Therefore, interpretation of the subgroup analysis can lead to informative insights into treatment effectiveness that would not be obtained from the non-subgrouped analysis.

Furthermore, review authors may choose to present results subgrouped by patient group even if there is little or no heterogeneity present, if it is clinically important to estimate treatment effectiveness for specific subgroups of patients. If subgroup analyses demonstrate that the treatment is more or less effective for certain subgroups of patients, interpretation of these subgroup analyses can provide valuable insight into how the treatment should be used in clinical practice.

4. How to interpret subgroup analyses

Here, we describe how to interpret subgroup analyses with regards to each of the previously listed criteria.

4.1. Criteria 1: report whether a statistically significant subgroup difference (interaction) was detected

A statistically significant subgroup effect means that the covariate (trial or patient characteristic) considered in the subgroup analysis statistically significantly modifies treatment effect. To determine whether a statistically significant subgroup difference was detected, the p-value from the test for subgroup differences ought to be considered. This test tests the difference between the pooled effect estimates for each subgroup. Usually, a p-value for this test of less than 0.1 indicates a statistically significant subgroup effect.

However, there are other details that it may be useful to provide when stating whether there is a statistically significant subgroup effect. It is useful to note whether the subgroup effect is qualitative (the treatment effects for each subgroup favour different treatments) or quantitative (the treatment effects for each subgroup favour the same treatment but are different sizes), and also the extent of heterogeneity (differences between treatment effects from individual trials in the meta-analysis) within each subgroup. If there is considerable heterogeneity within a subgroup, it may not be appropriate to draw

conclusions about treatment effect within this subgroup without further exploration of heterogeneity. Methods for assessing heterogeneity are provided in the Cochrane Handbook.³

If heterogeneity is identified, the review authors should consider whether it is appropriate and informative to present the analysis. If the subgroup analysis was performed to investigate sources of heterogeneity, then we would recommend a visual inspection of the forest plots to assess whether heterogeneity is lower within the subgroups than across all trials. Review authors may decide not to present the subgroup analysis in the review if the subgroup analysis has not explained heterogeneity at all. If the subgroup analysis was performed to provide estimates of treatment effects for clinically relevant subgroups of patients, then review authors may:

- i) decide that heterogeneity within each subgroup renders the results for each subgroup to be meaningless, and to not present this subgroup analysis, or
- ii) decide that it is important to show whether or not statistically significant subgroup differences exist for this covariate, and present the analysis. In this case, we recommend that review authors acknowledge the uncertainty in the evidence due to inconsistency between individual trial results.

4.2. Criteria 2: consider the covariate distribution (i.e. the number of trials and participants contributing to each subgroup)

For this criteria, the review author is required to consider the number of trials and participants contributing to each subgroup. The Cochrane Handbook³ advises that it is unlikely that an investigation of heterogeneity will produce useful findings unless there are at least 10 trials included in the meta-analysis, although even 10 trials may be too few if the covariate is unevenly distributed (i.e. if there is a limited amount of data for a particular subgroup).

4.3. Criteria 3: consider the plausibility of the interaction or lack of interaction

Considering the plausibility of an observed interaction, or lack of interaction, can help review authors decide how believable the results of subgroup analyses are.⁴ Considering the plausibility of an interaction or lack of interaction may substantiate the finding or raise the possibility that the finding may be spurious. Review authors may consider evidence that could demonstrate plausibility of an interaction or lack of interaction such as: studies of different populations (including animal studies); studies of similar interventions; studies of other, related outcomes.⁴

4.4. Criteria 4: consider the importance of the interaction or lack of interaction

Considering the extent of biological variability, treatment effect is highly likely to vary according to patient and/or trial characteristics, such as age, gender, and drug dosage. Therefore, it would be surprising not to observe subgroup effects. However, if these differences in treatment effect are not so large that they would impact clinical decisions, then it would be unnecessary to consider these subgroup effects further. As a general rule, the larger the difference between the effect in a particular subgroup and the overall effect, the more important the result.⁴ However, it is essential to consult with clinical experts in the relevant area of research to determine whether the subgroup analysis result is a clinically important finding.

4.5. Criteria 5: consider the possibility of confounding

Review authors ought to consider the possibility that some confounding factor may be influencing the results of the subgroup analysis,

leading to incorrect conclusions. Two covariates are confounded if their influences on treatment effect cannot be separated.³ For example, consider a meta-analysis that includes trials comparing Treatment A to Treatment B. Treatment A varies in intensiveness, so some trials compare “Intensive Treatment A” to Treatment B, and other trials compare “Non-intensive Treatment A” to Treatment B. Now, suppose a subgroup analysis was performed to investigate whether the intensiveness of Treatment A modifies treatment effect, and this subgroup analysis demonstrated a statistically significant treatment effect. It would be important for the review authors to consider whether there was some other factor that may be causing the subgroup difference. For example, it may be that the trials that used a more intensive version of Treatment A also recruited patients with more severe disease, and it may be that severity of disease is the true effect modifier. If this were true, it would be misleading to conclude that intensiveness of Treatment A modifies treatment effect. Review authors ought to use their expertise in the clinical area to identify patient or trial characteristics that might be confounded with one another. Review authors may also compute correlations between trial or patient characteristics to identify confounding factors.³

5. Theoretical examples of subgroup analyses

Considering the previously listed criteria, this section of the tutorial will help the reader to interpret the results of subgroup analyses with regards to criteria 1) and 2), and we do this with 5 theoretical examples. Criteria 3) - 5) can only be considered as part of a specific clinical question and accompanying meta-analysis, and are considered later.

The five theoretical examples cover the following scenarios:

- 1 Statistically significant, quantitative subgroup effect
- 2 Statistically significant, qualitative subgroup effect, with substantial unexplained heterogeneity
- 3 No subgroup effect
- 4 No subgroup effect, moderate unexplained heterogeneity
- 5 Statistically significant subgroup effect, uneven covariate distribution

The hypothetical scenario for all the subgroup analyses presented in this section of the tutorial is that the review authors want to know how treatment effect varies according to gender. The subgroup analyses were performed and the forest plots produced using Review Manager 5.3 software.⁵

5.1. Subgroup analysis 1: statistically significant, quantitative subgroup effect

The subgroup analysis presented in Fig. 1 shows the treatment effect of Intervention A versus Intervention B on a dichotomous outcome, for males and females separately. The results from individual trials are presented for males and females separately, and a pooled treatment effect estimate is provided for each of these subgroups. The test for subgroup differences is also provided at the bottom of the forest plot.

To report whether there is a statistically significant subgroup effect (criteria 1) the review author might state:

“The test for subgroup differences suggests that there is a statistically significant subgroup effect ($p = 0.04$), meaning that gender statistically significantly modifies the effect of Intervention A in comparison to Intervention B. The treatment effect favours Intervention A over Intervention B for both males and females, although the treatment effect is greater for males than females; therefore, the subgroup effect is quantitative. There is no heterogeneity between results from the trials within each subgroup that requires further exploration.”

For subgroup analysis 1, there are 10 trials included in the meta-

analysis, with 5 trials contributing data to each subgroup (> 1800 participants in each subgroup), so although clinical input is usually required to determine whether the covariate is evenly distributed, it is safe to assume in this scenario that the covariate distribution is not concerning. In addition to the above interpretation, the review author might state:

“A sufficient number of trials (5) and participants (> 1800) were included in each subgroup, so the covariate distribution is not concerning for this subgroup analysis.”

5.2. Subgroup analysis 2: statistically significant, qualitative subgroup effect, with substantial unexplained heterogeneity

For subgroup analysis 2 (Fig. 2), there is a statistically significant subgroup effect ($p = 0.04$); however, there is a substantial amount of heterogeneity within each subgroup.³ In this case, it does not appear that the subgroup analysis has explained heterogeneity (confidence intervals for the results of individual trials have poor overlap). However, it might be decided that it is important to show that there is a statistically significant subgroup effect. A possible interpretation of subgroup analysis 2 (if it was decided to present the analysis) is as follows:

“The test for subgroup differences suggests that there is a statistically significant subgroup effect ($p = 0.04$), meaning that gender significantly modifies the effect of Intervention A in comparison to Intervention B. Intervention A is favoured over Intervention B for males, while Intervention B is favoured over Intervention A for females; therefore the subgroup effect is qualitative. A sufficient number of trials (5) and participants (> 1700) were included in each subgroup, so the covariate distribution is not concerning for this subgroup analysis. However, there is substantial unexplained heterogeneity between the trials within each of these subgroups (males: $I^2 = 67\%$; females: $I^2 = 71\%$). Therefore, the validity of the treatment effect estimate for each subgroup is uncertain, as individual trial results are inconsistent.”

5.3. Subgroup analysis 3: No subgroup effect

For subgroup analysis 3 (Fig. 3), the test for subgroup differences indicates that there is no statistically significant subgroup effect ($p = 0.16$). There are more trials (and participants) contributing data to the female subgroup than to the male subgroup, and in a real-life setting, clinical input would be required to determine whether the covariate distribution is an important issue when interpreting this subgroup analysis.

Since the results from all trials included in this analysis are relatively homogenous, it is unlikely that this analysis would have been performed to investigate sources of heterogeneity. It is more likely that this analysis would have been performed to provide estimates of treatment effect for clinically relevant subgroups of patients. Review authors would need to decide whether it is important to present treatment effect estimates for these subgroups, or if it is not important to present this subgroup analysis since no significant subgroup effect was observed. Either way, the review authors ought to consider the covariate distribution before drawing any conclusions based on the results of this subgroup analysis.

A possible interpretation of subgroup analysis 3 (if it was decided to present the analysis) is as follows:

“The test for subgroup differences indicates that there is no statistically significant subgroup effect ($p = 0.16$), suggesting that gender does not modify the effect of Intervention A in comparison to Intervention B. However, a smaller number of trials and participants contributed data to the female subgroup than to the male subgroup, meaning that the analysis may not be able to detect subgroup differences. It is interesting to note that the pooled effect estimate for the males favours Intervention A but the pooled

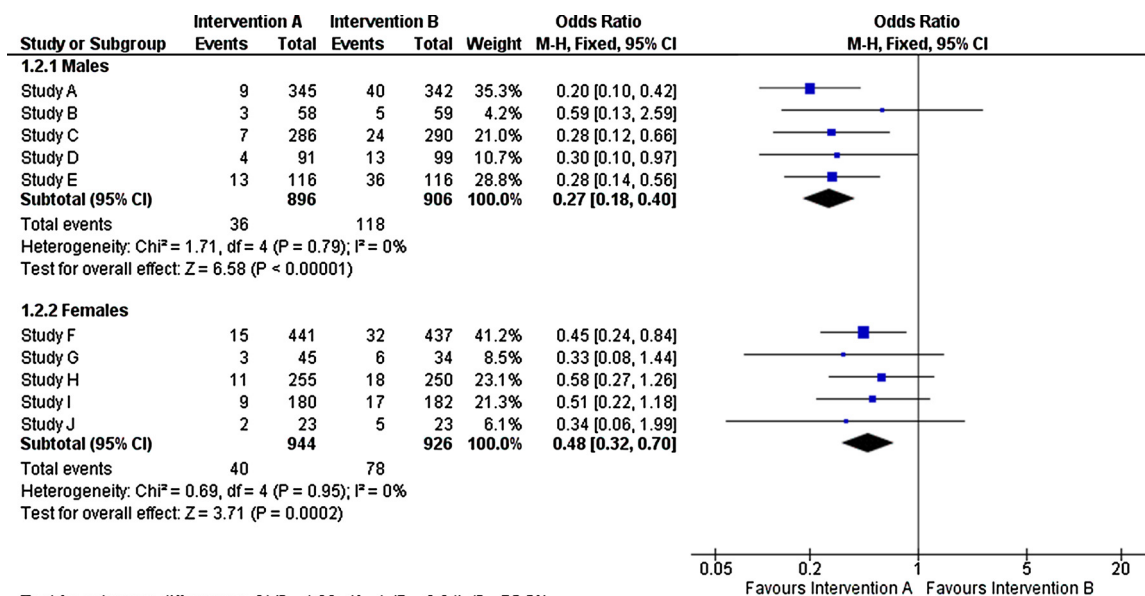


Fig. 1. Subgroup analysis 1: Statistically significant, quantitative subgroup effect.

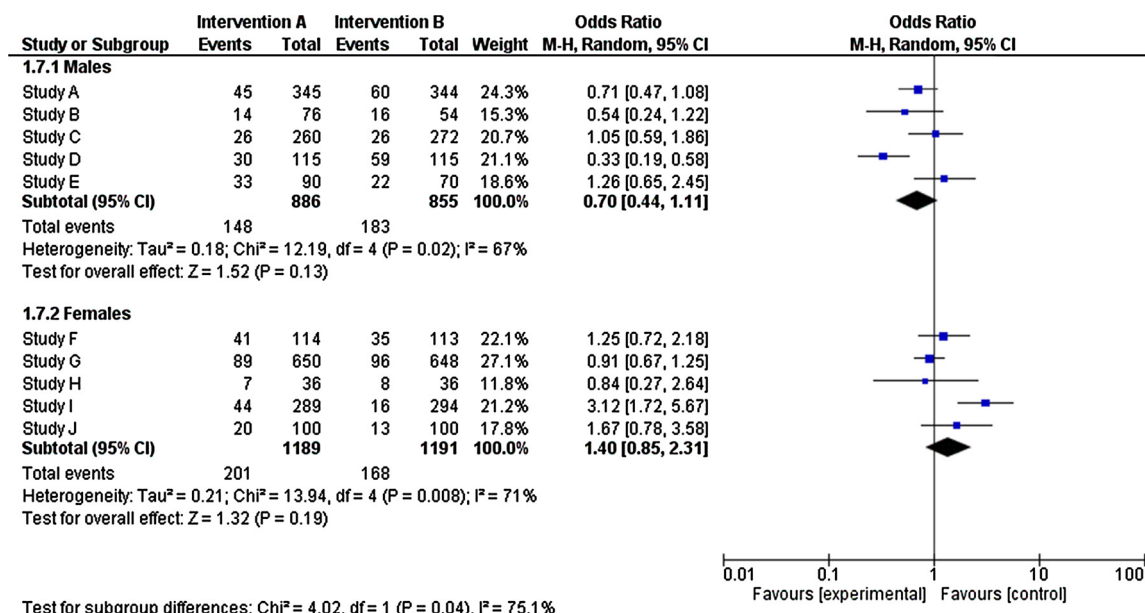


Fig. 2. Subgroup analysis 2: Statistically significant, qualitative subgroup effect, substantial unexplained heterogeneity.

effect estimate for females favours Intervention B”.

5.4. Subgroup analysis 4: No subgroup effect, moderate unexplained heterogeneity

For subgroup analysis 4 (Fig. 4), the test for subgroup differences indicates that there is no statistically significant subgroup effect. There are more trials (and participants) contributing data to the female subgroup than to the male subgroup, and clinical input may be required to determine whether the covariate distribution is an important issue when interpreting this subgroup analysis. There is also moderate heterogeneity between trials reporting data for the male subgroup.

Review authors would again need to decide whether it is informative and appropriate to present this subgroup analysis. If this subgroup analysis was performed to investigate sources of heterogeneity, then it might not be informative to present this subgroup analysis since heterogeneity does not appear to be lower within

subgroups than across all trials. Additionally, it may have been decided that the uneven covariate distribution means that the subgroup analysis would be unable to produce valid results.

If the subgroup analysis was performed to provide estimates of treatment effect for clinically relevant subgroups of patients, then the review authors may decide that heterogeneity renders the results for the male subgroup to be meaningless, or that the covariate distribution is concerning, and to not present this subgroup analysis. They may also decide that it is important to present estimates of treatment effect for these subgroups; in this case, we would recommend that review authors acknowledge the uncertainty in the evidence when interpreting results for a heterogeneous subgroup, and to consider the covariate distribution.

A possible interpretation of subgroup analysis 4 (if it was decided not to present the analysis) is as follows:

“The test for subgroup differences indicated that there is no statistically significant subgroup effect (p = 0.15, analysis not presented), suggesting

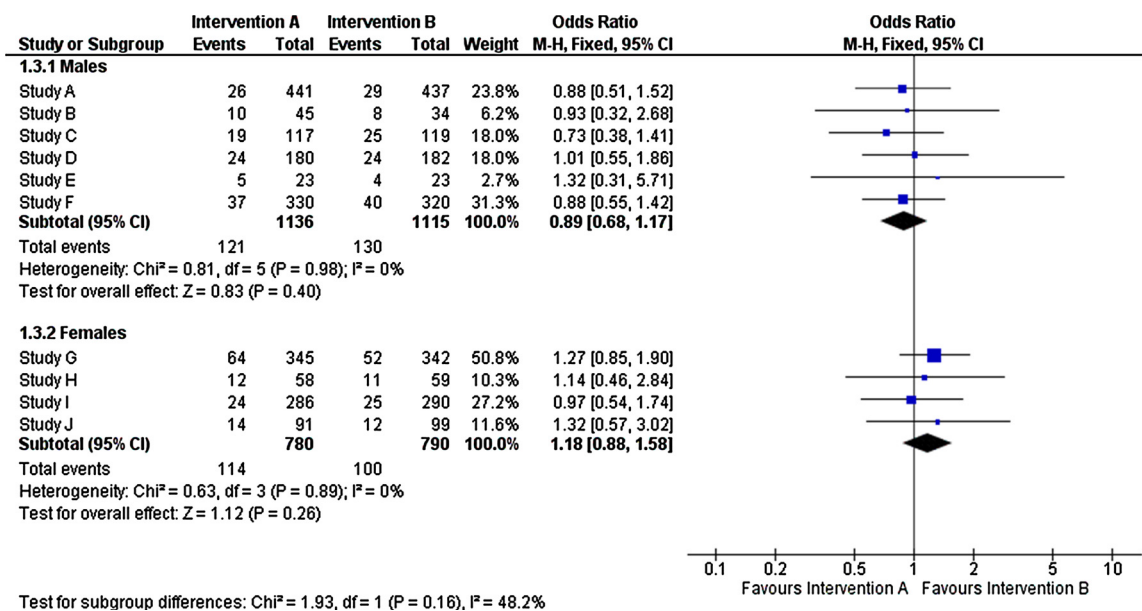


Fig. 3. Subgroup analysis 3: No subgroup effect.

that gender does not modify the effect of Intervention A in comparison to Intervention B. However, a smaller number of trials and participants contributed data to the male subgroup than to the female subgroup, meaning that the analysis may not be able to detect subgroup differences.”

“A subgroup analysis was performed to test whether gender modifies the effect of Intervention A in comparison to Intervention B (analysis not presented). However, a far smaller number of trials and participants contributed data to the female subgroup (2 trials, 186 participants) than to the male subgroup (5 trials, 2012 participants), meaning that the analysis is unlikely to produce useful findings.”

5.5. Subgroup analysis 5: statistically significant subgroup effect, uneven covariate distribution

For subgroup analysis 5 (Fig. 5), the test for subgroup differences indicates that there is a statistically significant subgroup effect. However, there are far more trials (and participants) contributing data to the male subgroup (5 trials, 2012 participants) than to the female subgroup (2 trials, 186 participants), and it seems unlikely that this subgroup analysis can be relied on to produce valid results.

A possible interpretation of subgroup analysis 5 (if it was decided not to present the analysis) is as follows:

6. Real-life subgroup analysis with clinical context

We have previously discussed how to interpret subgroup analyses (with regards to criteria 1) and 2) of the interpretation criteria) using hypothetical scenarios. Here we present an example of a subgroup analysis from a published Cochrane review, to demonstrate how to apply this knowledge of subgroup analyses in practice. We also demonstrate how to adhere to criteria 3) to 5) of the interpretation criteria, as we now have clinical context for the subgroup analyses, and so

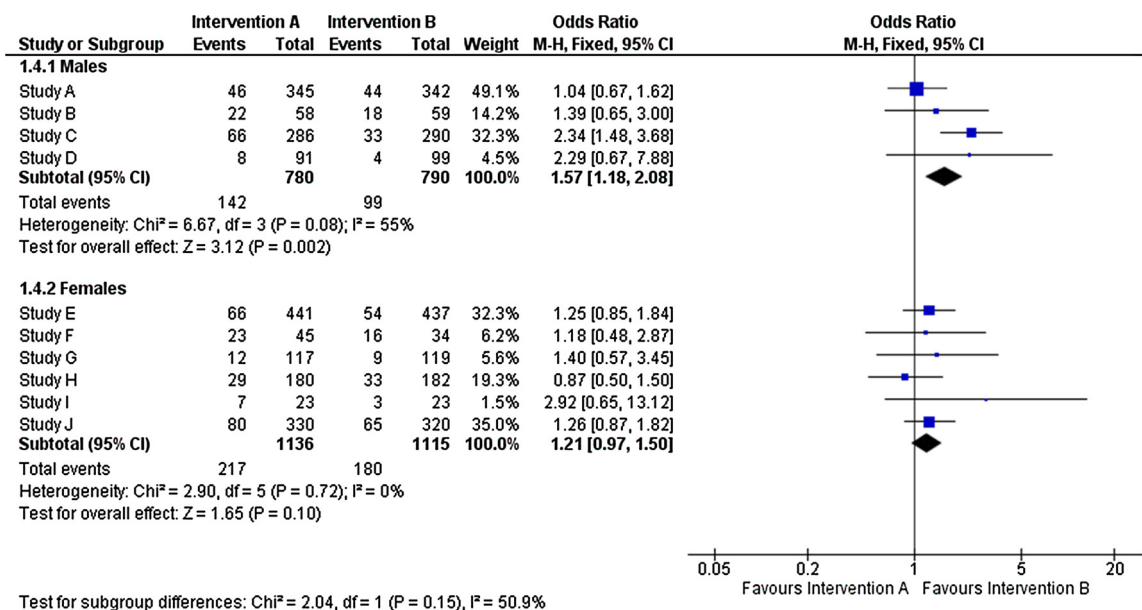


Fig. 4. Subgroup analysis 4: No subgroup effect, moderate unexplained heterogeneity.

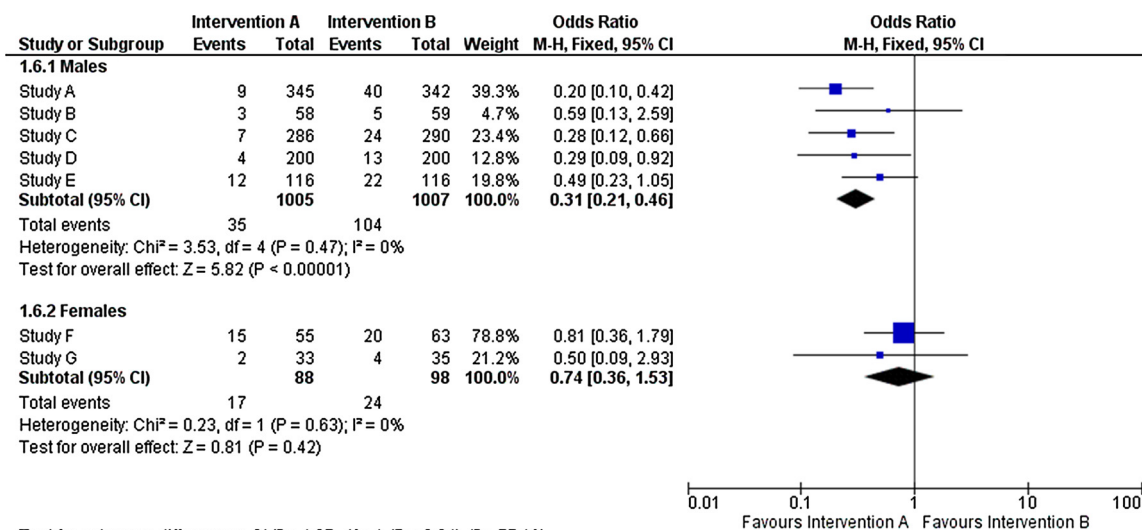


Fig. 5. Subgroup analysis 5: Statistically significant subgroup effect, uneven covariate distribution.

can consider the importance and plausibility of subgroup analyses results, and the possibility of confounding factors.

Fig. 6 presents a subgroup analysis taken from the Cochrane review “Artemisinin-based combination therapy for treating uncomplicated malaria”, published by Sinclair et al.⁶ The subgroup analysis compares the effect of two malaria treatments, artemether-lumefantrine (AL6) and amodiaquine plus sulfadoxine-pyrimethamine (AQ + SP), on the outcome of total treatment failure at day 28. The subgroup analysis was conducted to investigate whether geographical region modifies treatment effect.

The results of the subgroup analysis suggest that there is a statistically significant subgroup effect (p < 0.00001), meaning that geographic region significantly modifies the effect of AL6 in comparison to AQ + SP. AL6 is favoured over AQ + SP for East African populations, while AQ + SP is favoured over AL6 for West African populations; therefore, the subgroup effect is qualitative. There is a relatively small amount of heterogeneity between results from the trials within the West Africa subgroup (34%). However, a visual inspection of the forest plot confirms that heterogeneity is lower within the subgroups than across all trials, and so the subgroup analysis explains heterogeneity in the

overall analysis.

Considering the covariate distribution, only five trials (six cohorts of patients, since two cohorts are included from the Zongo trial⁷) are included in the analysis. Three trials contribute data to the East Africa subgroup, and three cohorts of patients contribute data to the West Africa subgroup. Since the number of trials included in the analysis is small, we do not have enough evidence to confidently conclude that there is a true subgroup effect. It is therefore useful to consider the plausibility of the demonstrated subgroup effect. It is highly plausible that the effectiveness of AL6 in comparison to AQ + SP is different in East and West Africa due to known differences in resistance to the antimalarial drugs. This plausibility of the demonstrated subgroup effect adds credibility to the results of the subgroup analysis.

Therefore, although quite possibly underpowered, this subgroup analysis suggests that AL6 has considerable advantages in East Africa, where absolute failure rates of AQ + SP are high. However, this advantage is not seen in West Africa, where cure rates with AQ + SP remain high. The importance of this subgroup analysis (i.e. criteria 4) is high. It would be beneficial for more trials to be conducted in these areas to confirm the subgroup effect.

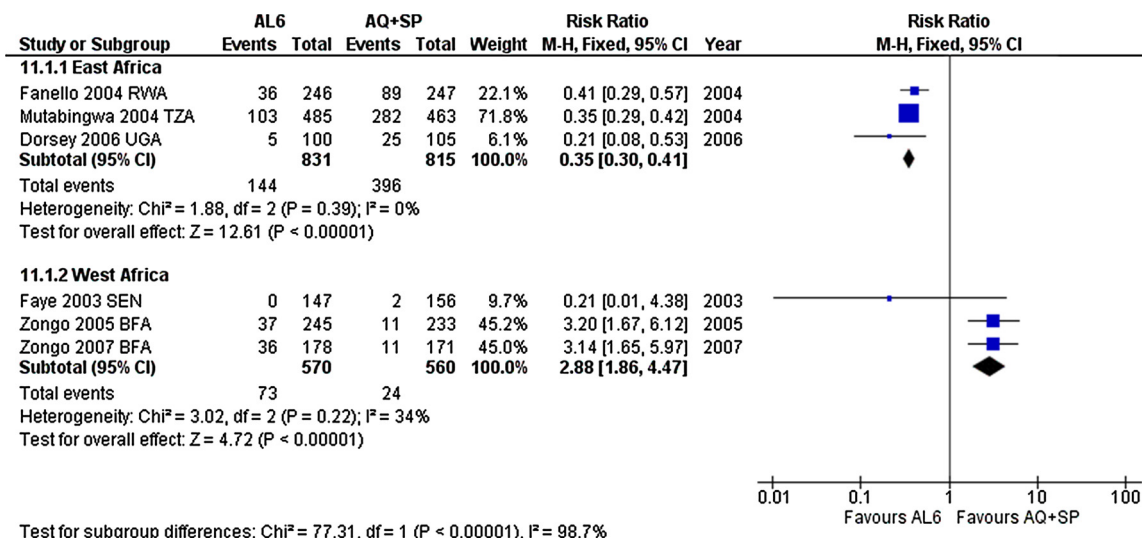


Fig. 6. AL6 versus AQ + SP; total treatment failure at day 28.

N.B.: The review did not present the test for subgroup differences in their review, but this has been presented here to aid the reader’s interpretation. Source: Artemisinin-based combination therapy for treating uncomplicated malaria, published by Sinclair et al.⁶; Analysis 11.1.1.

Finally, we consider the possibility of whether some confounding factor may be influencing the results of the subgroup analysis, leading to incorrect conclusions. Considering the subgroup analysis from the Sinclair et al. review,⁶ the authors of the Cochrane review considered key characteristics of the trials included in this subgroup analysis, and did not identify any potential confounding factors that could cause differences in treatment effect between the East African and West African trials. Therefore, confounding was not thought to be an issue of concern.

7. Summary

Previous research has shown that Cochrane review authors do not sufficiently report their interpretation of subgroup analyses in relation to five key criteria.¹ The results of a survey we conducted suggest that many Cochrane review authors do not know how to interpret subgroup analyses in relation to criteria 1) and 2); it was not possible to assess whether review authors knew how to interpret subgroup analyses with regards to criteria 3) - 5) as part of this survey.

Consequently, the aim of this tutorial is to improve the interpretation of subgroup analyses in reviews. For each subgroup analysis, we recommend that review authors:

- Report whether there is a statistically significant subgroup effect using the p-value from the test for subgroup differences
- Report whether there are sufficient, evenly distributed trials for the subgroup analysis to produce meaningful results
- Report whether the interaction, or lack of interaction, is plausible, and provide justification for this judgement
- Report whether the interaction, or lack of interaction, is a clinically important finding, and provide justification for this judgement
- Report whether any patient or trial characteristics were identified that might be confounded with the covariate of interest in the

subgroup analysis

We hope that careful interpretation of subgroup analyses in reviews will enable health policy makers to make the most of these valuable analyses when determining global health policy. We also hope that our tutorial will help readers of reviews to understand subgroup analyses, and be able to critically appraise the interpretation of subgroup analyses in systematic reviews.

Funding

MR, PG and SD are supported by the Effective Health Care Research Consortium, which is funded by UKAid from the UK Government Department for International Development (Grant number, 5242). SD is also funded by the Medical Research Council (Grant number, MR/K021435/1).

References

1. Donegan S, Williams L, Dias S, Tudur-Smith C, Welton N. Exploring treatment by covariate interactions using subgroup analysis and meta-regression in cochrane reviews: a review of recent practice. *PLoS One*. 2015;10(6):e0128804.
2. Higgins J, Thompson S, Deeks J, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy*. 2002;7(1):51–61.
3. Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. The Cochrane Collaboration; 2011. Available from: <http://handbook.cochrane.org>.
4. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med*. 1992;116(1):78–84.
5. Review Manager (RevMan) [Computer Program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.
6. Sinclair D, Zani B, Donegan S, Olliaro P, Garner P. Artemisinin-based combination therapy for treating uncomplicated malaria. *Cochrane Database Syst Rev*. 2009(3).
7. Zongo I, Dorsey G, Rouamba N, et al. Randomized comparison of amodiaquine plus sulfadoxine-pyrimethamine, artemether-lumefantrine, and dihydroartemisinin-piper-quine for the treatment of uncomplicated plasmodium falciparum malaria in Burkina Faso. *Clin Infect Dis*. 2007;45(11):1453–1461.