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)	Video Decision Aids to Assist with Advance Care Planning: a Systematic Review and Meta-analysis
AUTHORS	Jain, Ashu; Corriveau, Sophie; Quinn, Kathleen; Gardhouse, Amanda; Brandt Vegas, Daniel; You, John

VERSION 1 - REVIEW

REVIEWER	Karen Detering Austin Health, Melbourne, Australia
REVIEW RETURNED	01-Feb-2015

GENERAL COMMENTS	This paper is well written and methodologically sound, and this topic is of great interest. I however found myself wondering about what this paper really adds to this topic, particularly given that contemporary ACP focuses on much more than whether a person would like CPR, and some further discussion about this would be of benefit. The literature is already full of research showing discussion about specific treatment alone does little to improve uptake of ACP, and improve care, incl. EOL care. I also think some further discussion regarding the limitations of the current studies is important. You do touch on this early in the article, but I think the limitations are very important - specifically the fact that the recent studies are all from 1 research group (Volandes, et al) and all are US based (and there is significant variation in ACP uptake, and availability etc.) and therefore it is unclear whether this research is relevant to other countries, , and the fact that the earlier studies are over 10 years old, and this is extremely relevant in the field of ACP which is changing over time. Also no studies really addresses the topic of ACP and the process, but rather they primarily address specific treatments (mostly CPR), or completing advance care directives. You do discuss this briefly in the first paragraph of the discussion, but I think it is very important given the title of your article, and given the fact that it is already known that these things alone do not improve care.

REVIEWER	Carmen H.M. Houben CIRO+, centre of expertise for chronic organ failure
REVIEW RETURNED	02-Feb-2015

GENERAL COMMENTS	This is an interesting review on the issue of the impact of video decision aids on patients' preferences regarding life-sustaining treatments. Overall, it is an logically written manuscript. 1. The definition of "advance directive" should be provided. Does this

	also include durable powers of attorney and signing limitations of care “code status” forms for example? 2. Two pairs of reviewers assessed risk of bias using Cochrane systematic review guidelines. A statistical measure of inter-rater agreement (i.e. Cohen’s kappa) should be provided. 3. Six of the included studies were conducted by the same group of investigators. Do these studies included the same study population? If yes, what is the effect on the results of the present review? 4. The length of the videos used varied between 2 and 15 minutes. Is there a possible association between length of the video and effect of the video decision tool on for example patients’ knowledge related to advance care planning? 5. The loss to follow-up varies from 0% to 55%. Is there any explanation for the variation in loss to follow-up between the included studies? 6. In the discussion section, strengths of the current review and limitations of the studies included in the current review were provided. Are there also any limitations of the current review? 7. The risk of bias was mainly due to the lack of blinding of outcome assessment and lack of allocation concealment. How could future studies overcome these methodological shortcomings? 8. P7, line 20: change “advanced” in “advance”. 9. P17, line 11: please provide references of the 4 studies with poolable data.

REVIEWER	Peter Watson Medical Research Council UK
REVIEW RETURNED	08-Apr-2015

GENERAL COMMENTS	Meta-analyses are performed to obtain pooled random effect estimates comparing the effectiveness of groups of people using video decision aids to a control group who do not have video aids. The analyses themselves are straightforward with pooled effect sizes and 95% confidence intervals given. I am, though, not completely convinced by the strength of the results of the video aids having better patient preferences due to the acknowledgement of a bias favouring the video aids since the same information is being presented twice in the video intervention arms and just once in the control arm (discussion on page 20), am also not clear of the impact of only using subjects with complete data (page 9) and, also, wondered if a trim and fill procedure or similar could have been used to adjust for any publication bias. Having said that all but one of the effect sizes in Figures 3 and 4 favour video aids which probably suggests some systematic difference between video and non-video aid usage although the size of this difference is not clear and may only be marginal. Page 9, line 32. What percentage of ‘outcome data’ was missing and for what reasons? Does just using complete cases introduce some bias into the results as we are only considering a subset of people? Page 9. I was slightly confused by lines 54-56 on page 9 because these mention both the Mantel-Haenszel method which is usually used for calculating the weighted pooled odds ratio under the FIXED effects model and RANDOM effects models which implies two forms of pooled odds ratios were used. DerSimonian and Laird pooled estimates are usually used for random effects models but are not mentioned here or elsewhere. In the on-line documentation of Review Manager at http://handbook.cochrane.org/chapter_9/9_4_4_3_random_effects_method.htm it suggests that a comparison is made between each study odds ratio and the

	pooled Mantel-Haenszel odds ratio although it stops short of saying if this is then used as an adjustment for any between study heterogeneity to confidence intervals for the pooled odds ratio. I also notice from the above website that an inverse-variance approach may be used as an alternative to the Mantel-Haenszel in testing for heterogeneity between study odds ratios. Did you consider using the inverse variance approach for looking at odds ratios using random effects as opposed to the Mantel-Haenszel to be consistent since an inverse variance approach is used in the random effects analysis for continuous data. Page 16. I am not sure of the logic of choosing to include a study (Yamada et al.) in a meta analysis, reporting it and then dropping it from the meta-analysis because the analysis tells us it is an outlier. The decision to include the study should be made prior to analysis and should not be changed because it causes a problem due to a lack of fit. With a small sample of 7 studies it may be that there are other studies which can be regarded as in some way similar to Yamada (a sampling bias) but just have not been able to be included in the meta-analysis. Did you consider if there was a publication bias using funnel plots, for example, to assess effect size asymmetry and especially consider adjust for such bias using Duval and Tweedie's trim-and-fill procedure which adds extra effects to adjust for bias rather than simply dropping apparent outliers as is done on page 16. I am not sure what a post-hoc sensitivity analysis (page 16, line 25) is. It appears to be simply dropping the aforementioned outlying Yamada study because of its lack of fit. Page 20. I find it worrying, as is acknowledged in lines 23-35 on page 20, that there is likely to be a bias favouring the video aids due to the same information being presented twice in the video intervention arms and just once in the control arm in as many as 4 (over half) of the 7 studies reported in Figure 3 (page 28) on patient preferences which, as it is, only shows a lukewarm effect favouring the use of video aids. If this effect is biased upwards favouring the video arm, as is implied, it may be that if the information had been given the same number of times there might have been no difference found in patient preferences between the video and non-video aids. Indeed, the lukewarm strength of the results favouring video decision aids is reinforced in the abstract (page 4) which suggests more evidence is needed to see if video decision aids improve ACP-related outcomes (presumably before one can recommend their use). Page 29 One could remove the strange reference (zero SDs) to the Epstein 2013 study since (page 17, line 6) there was insufficient detail for inclusion in the meta-analysis in Figure 4.

VERSION 1 – AUTHOR RESPONSE

Reviewer # 1 Comments (Reviewer Name: Karen Detering)

1. **This paper is well written and methodologically sound, and this topic is of great interest. I however found myself wondering about what this paper really adds to this topic, particularly given that contemporary ACP focuses on much more than whether a person would like CPR, and some further discussion about this would be of benefit. The literature is already full of research showing discussion about specific treatment alone does little to improve uptake of ACP, and improve care, incl. EOL care.**

Response: We thank the reviewer for her positive comments. We continue to believe that this paper does make a unique and valuable contribution to the literature since there is substantial interest in the adoption of these sorts of videos to assist with advance care planning but there has not been, to date, a comprehensive systematic review of the evidence to support their uptake. However, we also agree with the reviewer that ACP has evolved beyond a narrow focus on decisions about CPR and agree that further discussion about a broader focus of ACP is warranted. We have added new text to the beginning of the Discussion section in our revised manuscript to elaborate on this point (pages 19-20):

> "Although an important aspect of ACP is to clarify patients' preferences for life-sustaining treatments, including CPR, ACP involves several other important processes. In contemporary thinking, the focus of ACP is shifting away from making decisions about future treatment choices and putting more emphasis on the need to prepare future surrogate decision makers for "in-the-moment" decision-making.[2] In this way, ACP can be seen as a broader set of behaviours, including: choosing a surrogate decision maker, deciding what matters most in life (clarifying values and, in some cases, future wishes for treatments such as CPR), and communicating these values and wishes to surrogate decision makers to better prepare them to engage in future "in-the-moment" medical decision-making when the patient becomes incapable. We did not find any RCTs of video decision aids that examined ACP from this perspective. However, web-based decision aids have recently been developed to change these different behaviours related to ACP.[29-31]"

2. **I also think some further discussion regarding the limitations of the current studies is important. You do touch on this early in the article, but I think the limitations are very important - specifically the fact that the recent studies are all from 1 research group (Volandes, et al) and all are US based (and there is significant variation in ACP uptake, and availability etc.) and therefore it is unclear whether this research is relevant to other countries, and the fact that the earlier studies are over 10 years old, and this is extremely relevant in the field of ACP which is changing over time.**

Response: We agree that there are limitations to the studies included in our systematic review and have further expanded the discussion of these in our revised manuscript (page 21):

> "This narrower focus of the existing trials on the elicitation of treatment preferences or creation of advance directives, rather than the broader range of activities that are part of ACP, and the fact that all the included studies were done in the U.S. raise questions about the applicability of the available evidence in other countries and in the context of changing definitions of ACP."

3. **Also no studies really addresses the topic of ACP and the process, but rather they primarily address specific treatments (mostly CPR), or completing advance care directives. You do discuss this briefly in the first paragraph of the discussion, but I think it is very important given the title of your article, and given the fact that it is already known that these things alone do not improve care.**

Response: See responses to Reviewer #1, Comments #1 and #2 above, in which we discuss contemporary thinking about the nature and role of ACP and how we have revised our manuscript accordingly to highlight this issue.

Reviewer #2 Comments (Reviewer Name: Carmen H.M. Houben)

This is an interesting review on the issue of the impact of video decision aids on patients' preferences regarding life-sustaining treatments. Overall, it is an logically written manuscript.

Response: We thank the reviewer for her positive comments on our manuscript.

1. **The definition of "advance directive" should be provided. Does this also include durable powers of attorney and signing limitations of care "code status" forms for example?**

Response: For our review, we used the definition of "advance directive" as provided by the authors of the individual studies included in our review. In one study (Epstein 2013), advance directives were defined as "any document that instructed caregivers on details of future care". In another study (Yamada 1999), no definition was provided. Two other studies (Brown 1999 and Landry 1997) reported data on the completion of both living wills and durable powers of attorney for health care. For the latter 2 studies, we extracted data related to the completion of living wills for our outcome of "advance directive completion". This has been described this in the Methods and Results sections of our revised manuscript (pages 8 and 18, respectively):

> "Secondary outcomes of interest were: patients' knowledge related to ACP (including knowledge about life sustaining treatments), patients' confidence in any decision made about future use of life-sustaining treatments, completion of an advance directive **(as defined by the authors of the individual studies)**, actual use of life-sustaining treatments at end of life, whether the use of life-sustaining treatments at end of life was congruent with patients' prior expressed wishes, health resource use at end of life, and, for patients allocated to the video intervention arm, patients' comfort watching the video."
>
> "One study defined advance directives as "any document that instructed caregivers on details of future care",[20] another study provided no definition,[17] and two other studies reported data

on the completion of both living wills and durable powers of attorney for health care. For the latter two studies, we used the data related to the completion of living wills.[25,26]„

2. **Two pairs of reviewers assessed risk of bias using Cochrane systematic review guidelines. A statistical measure of inter-rater agreement (i.e. Cohen's kappa) should be provided.**

Response: Unfortunately, we have not kept separate copies of the individual data extraction forms for the assessment of risk of bias, only the final consensus ratings which are reported in our manuscript. (Having said that, we are unsure of the value in reporting kappa on a small number of observations, i.e. n=10 articles, since a change in the classification of a single article is likely to have a very large impact on the observed kappa value and so we would expect our point estimate of kappa to be quite unstable). We are able to report on chance-corrected, inter-rater agreement for the independent duplicate review of titles and abstracts for potential eligibility of the articles (kappa 0.29), and the assessment of eligibility of full-text articles for inclusion in our systematic review (kappa 0.48):

> "We used Cohen's kappa to assess chance-corrected inter-rater agreement of reviewers' decisions about potential eligibility of articles at title and abstract screening, and about the eligibility of articles at full-text review." (page 9)

> "Of 3,980 citations identified from primary electronic databases, 583 were duplicates, leaving 3,397 original publications. Of these, 125 were deemed potentially eligible (kappa 0.29) and underwent full-text screening. Of these 125 full-text articles, 10 RCTs were eligible for our review (kappa 0.48)." (page 10)

3. **Six of the included studies were conducted by the same group of investigators. Do these studies included the same study population? If yes, what is the effect on the results of the present review?**

Response: As described in Table 1 of our original manuscript, the study populations for the six studies that were conducted by the same group of investigators were different, and included: patients with progressive pancreatic or hepatobiliary cancer (Epstein 2013), patients with advanced cancer—from a different institution than those in the Epstein 2013 study (Volandes 2013), patients from skilled nursing facilities (Volandes 2012), patients from a rural primary care clinic (Volandes 2011), patients with malignant glioma (El-Jawahri 2010), patients from primary care clinics associated with academic teaching hospitals (Volandes 2009). For the outcome of patient preferences for life-sustaining treatment (Figure 3), the pooled estimate of effect from these 6 studies indicates that the video intervention results in a statistically significant difference in patient preferences (those allocated to video were less likely to express a preference for CPR) compared to the non-video arm (risk ratio, 0.42 [95% CI, 0.26 to 0.67]; I^2 = 0%; heterogeneity P = 0.44). This finding was reported on page 17 of our original manuscript and was the result of our *post hoc* sensitivity analysis in which we excluded

the study by Yamada et al (and thus included only these 6 studies from the same group of investigators). However, because of concerns raised by Reviewer #3 (Comment #4), we have removed this sensitivity analysis in our revised manuscript. To be consistent throughout our revised manuscript in responding to Reviewer #3's concern, we have elected not to add new *post hoc* subgroup analyses in which we explore potential differences in the effect of video-based interventions based on different study characteristics. To do so in a rigorous fashion would require the use of meta-regression techniques and a much larger sample of included studies than the total of 10 studies included in our review. If the editors feel that it is crucial that we conduct such analyses in a revised manuscript, we would be willing to re-visit this issue.

4. **The length of the videos used varied between 2 and 15 minutes. Is there a possible association between length of the video and effect of the video decision tool on for example patients' knowledge related to advance care planning?**

Response: See response to Reviewer #2, Comment #3 above. While an interesting question, for the reasons stated above we do not believe it is prudent to conduct meta-regression analyses using a small sample of 10 studies to examine a possible relationship between study characteristics and our outcomes of interest. In fact, the effective sample size for a meta-regression in our dataset would be n=7 since for any given outcome of interest we have at most 7 available studies. Qualitatively, it would appear that the effect of video interventions on knowledge related to advance care planning was very consistent across the 5 studies that reported on this outcome, despite using videos that ranged from 3 minutes to 14 minutes in length; see Figure 4.

5. **The loss to follow-up varies from 0% to 55%. Is there any explanation for the variation in loss to follow-up between the included studies?**

Response: We agree with the reviewer that the proportion of participants loss to follow-up in the included studies does vary appreciably. However, the authors of these individual studies did not report the reasons for loss to follow-up. Therefore, we can only speculate about the reasons why. We do not believe that such speculation would add much value to our manuscript. However, if the editors feel it is important to include these opinions in a revised manuscript we would be happy to re-visit this issue.

6. **In the discussion section, strengths of the current review and limitations of the studies included in the current review were provided. Are there also any limitations of the current review?**

Response: To maximize the methodological rigour of our study, we adhered to contemporary standards for the conduct and reporting of systematic reviews (ref #28, Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.

Ann Intern Med 2009;151:264-9), and we would note that Reviewer #1 assessed our study to be "methodologically sound". For instance, we reported the details of our literature search strategy, which included a wide range of relevant databases, did not exclude articles on the basis of language of publication, performed title and abstract screening and data extraction independently and in duplicate, and used a widely accepted methodology (GRADE; Guyatt GH, Oxman AD, Kunz R, et al. What is "quality of evidence" and why is it important to clinicians? BMJ 2008;336:995-998) to assess the quality of evidence. Nonetheless, our review does have some potential limitations. First, we restricted our search to articles published in 1980 or later. However, and none of the articles included in our review were published before 1996 and advance directives and advance care planning are relatively new concepts, e.g., the U.S. Patient Self-Determination Act was an important instigator of activity related to advance directives and was not passed into law until 1990, and the term "advance care planning" did not appear in the peer-reviewed literature until the mid-1990s with increased attention only in the current decade. Second, for studies with missing outcome data we did not use imputation methods but rather used a complete case analysis. This could have potentially introduced bias into our pooled estimates of effect if the reasons for "missing-ness" were related to our outcomes of interest, e.g., if those lost to follow-up were systematically more or less likely to express a preference for CPR. We have discussed these potential limitations of our review in our revised manuscript (pages 20-21):

> "Our systematic review also has limitations. First, we limited our search strategy to articles published in 1980 or later and it is possible that we missed older, relevant articles. However, none of the trials included in our review were published before 1996 and the concepts of advance directives and advance care planning did not gain widespread attention until the 1990s (e.g., after the introduction of the U.S. Patient Self-Determination Act in 1990). Second, for studies with missing outcome data, we did not use imputation methods. By including only patients with non-missing data (complete case analysis) in our meta-analyses, our resultant estimates of effect could be biased if patients lost to follow-up were systematically different in ways that were related to our outcomes of interest (e.g., if they were systematically more or less likely to prefer CPR)."

7. The risk of bias was mainly due to the lack of blinding of outcome assessment and lack of allocation concealment. How could future studies overcome these methodological shortcomings?

Response: To overcome lack of blinding of outcome assessment, e.g. for the primary outcome of patient preferences, future studies could ensure that the individual administering the outcome assessment instrument (e.g., a study questionnaire) was blinded to allocation of the patient to intervention or control. In several of the included studies, the outcome assessor was aware of the patient's allocation to intervention (video) or control which potentially could have introduced bias into the assessment of patients' preferences. To preserve concealment of allocation, which is fundamental

to providing an unbiased assessment of the effect of the intervention, we recommend that if investigators use envelopes to conceal allocation, that they use sequentially numbered, opaque, sealed envelopes, or, more preferably, that they use central, automated, telephone-based or web-based randomization services, which are now easily accessible, affordable, and much less prone to deciphering (Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet 2002;359:614-18). We have included a discussion of these points in the Discussion section of our revised manuscript (page 22) and cited the paper by Schulz et al:

> "Future trials could overcome these methodological limitations by using centralized telephone or web-based randomization to preserve allocation concealment and blinding the outcome assessors to allocation.[32]"

8. **P7, line 20: change "advanced" in "advance".**

Response: We have corrected this in the revised manuscript.

9. **P17, line 11: please provide references of the 4 studies with poolable data.**

Response: We have provided these references in our revised manuscript.

Reviewer #3 Comments (Reviewer Name: Peter Watson)

1. **Meta-analyses are performed to obtain pooled random effect estimates comparing the effectiveness of groups of people using video decision aids to a control group who do not have video aids. The analyses themselves are straightforward with pooled effect sizes and 95% confidence intervals given. I am, though, not completely convinced by the strength of the results of the video aids having better patient preferences due to the acknowledgement of a bias favouring the video aids since the same information is being presented twice in the video intervention arms and just once in the control arm (discussion on page 20), am also not clear of the impact of only using subjects with complete data (page 9) and, also, wondered if a trim and fill procedure or similar could have been used to adjust for any publication bias. Having said that all but one of the effect sizes in Figures 3 and 4 favour video aids which probably suggests some systematic difference between video and non-video aid usage although the size of this difference is not clear and may only be marginal.**

Response: We appreciate these thoughtful comments. We respond more specifically to these points below.

2. **Page 9, line 32. What percentage of 'outcome data' was missing and for what reasons? Does just using complete cases introduce some bias into the results as we are only considering a subset of people?**

Response: We did report the percentage of outcome data that was missing (i.e., percentage of patients lost to follow-up) on page 11 in the Results section of our original manuscript: "Loss to follow-up was low (0% to 11%) in 7 of the 10 included studies, whereas the studies by Siegert et al,[24] Yamada et al,[17] Volandes et al,[19] had higher rates of 14%, 23%, and 55% loss to follow-up, respectively". We have modified this sentence as follows in our revised manuscript to clarify for the reader that "loss to follow-up" carries the same meaning as "missing outcome data" (pages 11-12):

> "Loss to follow-up, **i.e., the percentage of participants with missing outcome data,** was low (0% to 11%) in 7 of the 10 included studies, whereas the studies by Siegert et al,24 Yamada et al,17 Volandes et al,19 had higher rates of 14%, 23%, and 55% loss to follow-up **(missing outcome data)**, respectively."

We have discussed the potential bias introduced by using only complete cases in our analysis within our response to Reviewer #2, Comment # 6 regarding the limitations of our systematic review.

3. **Page 9. I was slightly confused by lines 54-56 on page 9 because these mention both the Mantel-Haenszel method which is usually used for calculating the weighted pooled odds ratio under the FIXED effects model and RANDOM effects models which implies two forms of pooled odds ratios were used. DerSimonian and Laird pooled estimates are usually used for random effects models but are not mentioned here or elsewhere. In the on-line documentation of Review Manager at http://handbook.cochrane.org/chapter_9/9_4_4_3_random_effects_method.htm it suggests that a comparison is made between each study odds ratio and the pooled Mantel-Haenszel odds ratio although it stops short of saying if this is then used as an adjustment for any between study heterogeneity to confidence intervals for the pooled odds ratio.**

   I also notice from the above website that an inverse-variance approach may be used as an alternative to the Mantel-Haenszel in testing for heterogeneity between study odds ratios. Did you consider using the inverse variance approach for looking at odds ratios using random effects as opposed to the Mantel-Haenszel to be consistent since an inverse variance approach is used in the random effects analysis for continuous data.

Response: We apologize for the confusion. We carefully checked this query from the reviewer and consulted with a biostatistician who specifically has experience conducting systematic reviews and meta-analyses. When reading the Review Manager on-line documentation referenced by the reviewer, which we have included verbatim below in italics, the random-effects method is attributed to DerSimonian (as pointed out by the reviewer); however, in the 2nd paragraph of the Review Manager documentation, it also clearly states that the Mantel-Haenszel method (for setting up weights) is one of two options for the calculation of a pooled estimate for dichotomous data (with the alternative

method being the inverse variance method, as the reviewer points out). We used the Mantel-Haenszel method to set up weights in our analyses of dichotomous outcome data. To minimize confusion, in our revised manuscript we have only described the type of model used (random effects model using the method of DerSimonian and Laird) and removed the descriptions of how weights were set up (page 10):

> "For dichotomous outcome data, we used random effects models (using the method of DerSimonian and Laird) to calculate pooled risk ratios and 95% confidence intervals."

In response to the reviewer's query about the possibility of using an inverse variance method, our pooled estimates and 95% confidence intervals do not change meaningfully if we use the inverse variance method to assign weights. For the outcome of patient preferences for CPR, using an inverse variance method to assign weights, the pooled estimate is: risk ratio 0.53, 95% CI 0.31 to 0.90 (compared to what we are reporting in our manuscript using the Mantel-Haenszel method to assign weights: pooled risk ratio 0.50, 95% CI 0.27 to 0.95). For the outcome of completion of advance directives, the pooled risk ratio and 95% CI using the inverse variance method are identical to the result we present in our manuscript using the Mantel-Haenszel method.

9.4.4.3 Random-effects method

The random-effects method (DerSimonian 1986) incorporates an assumption that the different studies are estimating different, yet related, intervention effects. As described in Section 9.4.3.1, the method is based on the inverse-variance approach, making an adjustment to the study weights according to the extent of variation, or heterogeneity, among the varying intervention effects. The random-effects method and the fixed-effect method will give identical results when there is no heterogeneity among the studies. Where there is heterogeneity, confidence intervals for the average intervention effect will be wider if the random-effects method is used rather than a fixed-effect method, and corresponding claims of statistical significance will be more conservative. It is also possible that the central estimate of the intervention effect will change if there are relationships between observed intervention effects and sample sizes. See Section 9.5.4 for further discussion of these issues.

RevMan implements two random-effects methods for dichotomous data: a Mantel-Haenszel method and an inverse-variance method. The difference between the two is subtle: the former estimates the amount of between-study variation by comparing each study's result with a Mantel-Haenszel fixed-effect meta-analysis result, whereas the latter estimates the amount of variation across studies by comparing each study's result with an inverse-variance fixed-effect meta-analysis result. In practice, the difference is likely to be trivial. The inverse-variance method was added in RevMan version 5.

4. **Page 16. I am not sure of the logic of choosing to include a study (Yamada et al.) in a meta analysis, reporting it and then dropping it from the meta-analysis because the analysis tells us it is an outlier. The decision to include the study should be made prior to analysis**

and should not be changed because it causes a problem due to a lack of fit. With a small sample of 7 studies it may be that there are other studies which can be regarded as in some way similar to Yamada (a sampling bias) but just have not been able to be included in the meta-analysis. Did you consider if there was a publication bias using funnel plots, for example, to assess effect size asymmetry and especially consider adjust for such bias using Duval and Tweedie's trim-and-fill procedure which adds extra effects to adjust for bias rather than simply dropping apparent outliers as is done on page 16. I am not sure what a post-hoc sensitivity analysis (page 16, line 25) is. It appears to be simply dropping the aforementioned outlying Yamada study because of its lack of fit.

Response: First, with respect to the reviewer's comment about publication bias, we have inspected funnel plots for each outcome of interest and have not found compelling evidence of publication bias, which in our case is particularly challenging because of the small number of eligible studies (7 at most depending on the outcome of interest). Therefore, we do not see a need to proceed with adjusting for publication bias in further sensitivity analyses using a trim-and-fill procedure. With respect to the Yamada study, it was deemed eligible for inclusion in our review based upon our *a priori* eligibility criteria. Therefore, it would not seem logical to have excluded the study prior to our analysis. Our rationale for excluding this study in an exploratory, *post hoc* sensitivity analysis was that this single study appeared to explain the appreciable heterogeneity that we observed across the included studies for the outcome of patient preferences for CPR. However, we agree with the reviewer that the overall sample of studies is small (n=7) and that this *post hoc* analysis does have limitations. Therefore, we have omitted the reporting of this *post hoc* analysis in our revised manuscript. Since we believe that many readers may still be curious about the substantial heterogeneity observed for this outcome, we have still included text in our results section to qualitatively describe how and why the Yamada study may be an outlier (page 17):

> "The substantial heterogeneity across studies may have been driven by the study by Yamada et al,[17] which was published over a decade earlier than the 6 other RCTs which reported on this outcome, and had a much higher proportion of patients indicating a preference for CPR than the other studies."

5. **Page 20. I find it worrying, as is acknowledged in lines 23-35 on page 20, that there is likely to be a bias favouring the video aids due to the same information being presented twice in the video intervention arms and just once in the control arm in as many as 4 (over half) of the 7 studies reported in Figure 3 (page 28) on patient preferences which, as it is, only shows a lukewarm effect favouring the use of video aids. If this effect is biased upwards favouring the video arm, as is implied, it may be that if the information had been given the same number of times there might have been no difference found in patient preferences between the video and non-video aids. Indeed, the lukewarm strength of the results favouring video decision aids is reinforced in the abstract (page 4) which suggests more**

evidence is needed to see if video decision aids improve ACP-related outcomes (presumably before one can recommend their use).

Response: We agree with the reviewer that this "double dosing" in some of the included trials is a limitation of the existing evidence, as we discussed in lines 23-35 on page 20 of our original manuscript. We have reinforced the Conclusion statement in our revised Abstract the reviewer's point that more evidence is needed before recommending the use of video decision aids in routine clinical practice.

> "**Before recommending their use in clinical practice**, more evidence is needed to confirm these findings and to evaluate the impact of video decision aids when integrated into patient care."

6. **Page 29 One could remove the strange reference (zero SDs) to the Epstein 2013 study since (page 17, line 6) there was insufficient detail for inclusion in the meta-analysis in Figure 4.**

Response: In our revised manuscript, we have simply stated that one of the studies (Epstein 2013) did not report sufficient detail for inclusion in the meta-analysis.

VERSION 2 – REVIEW

REVIEWER	Karen Detering Respecting Patient Choices program Austin Health Melbourne, Australia
REVIEW RETURNED	24-May-2015

GENERAL COMMENTS	I am happy with the response of the authors to my concerns, and now recommend publishing this manuscript.

REVIEWER	Carmen H.M. Houben Department of Research & Education, CIRO, Centre of Expertise for Chronic Organ Failure, Horn, The Netherlands
REVIEW RETURNED	27-May-2015

GENERAL COMMENTS	The authors have improved the paper and replied to the queries appropriately.