PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A meta-analysis of cognitive-based behaviour change techniques as interventions to improve medication adherence
AUTHORS	Easthall, Claire; Song, Fujian; Bhattacharya, Debi

VERSION 1 - REVIEW

REVIEWER	Peter Watson Statistician MRC Cognition and Brain Sciences Unit I have no competing interests with the research in this article.
REVIEW RETURNED	27-Feb-2013

THE STUDY	I felt the outcome measure (e.g. percentages or numbers of medications used) could be discussed more in the text e.g. an idea of types of medication, ranges of how often they were used used. The low kappa (page 8) makes we wonder how well defined these types of intervention studies are to yield such a low inter-rater agreement on deciding at an initial stage whether or not to include a study in this meta-analysis and if other reviewers would have argued that other studies might have been excluded/included in this analysis. Most of the responses (Table 2 on page 15) appear to take the form of proportions. In this case the effect size of interest should be based upon either arcsine transformed proportions or odds ratios. I am not sure if these have been used here. The reason for the need to use special effect sizes with proportions is that the variance of a proportion is itself related to the value of the proportion so mean and variance are not independent of each other. Such an approach also requires that all responses take the same form e.g. are expressed as proportions/percentages.
RESULTS & CONCLUSIONS	A main concern is that most of the studies show some bias in at least one of the five bias categories (top of page 14 and Figure 1 page 30) which cannot be adjusted for which may have led to biases in the effect size estimates. Despite reassurance in the discussion at the end of the first full paragraph on page 19, I still wonder given the resultant number of inherent differences between the studies due to the 'broad inclusion criteria' (page 8) if it is sensible to give pooled effect sizes such as those on page 14 (last two paragraphs) since the studies may all have inherently different effect sizes due to their provenance. Perhaps further explanation of the subgroup analyses and the robustness of the effect sizes indicated by their results (first full paragraph on page 18) will help allay this concern.
GENERAL COMMENTS	This meta-analysis study aims to estimate intervention effects using 23 reported randomised control trials to investigate the effectiveness of cognitive techniques in improving medication adherence. Adjustments are made for the heterogeneity in the sizes of differences between the intervention and control groups in the studies (first full paragraph on page 8)

	and for the evident publication bias in the reporting of these differences (second full paragraph on page 8) using a trim and fill adjusted funnel plot. The various biases in the studies (second paragraph page 8 and top of page 14) and the broad inclusion criteria (first full paragraph on page 8) acknowledged by the authors make a meta-analysis of this data challenging and the authors do flag up these weaknesses. Two main concerns are the inherent biases in the studies reported on page 14 and the low inter-rater agreement on choosing which studies should be used in the meta-analysis which may bring into question the accuracy and reliability of the estimates of group differences presented in this study. One might argue that if the studies considered in this paper are themselves biased then the analysis based upon these will also be biased. It is also not clear, assuming the responses are (at least mostly) in the form of proportions, if effect sizes appropriate for proportions have been computed (I suggest a website giving more details, as needed, below). The usual effect size for proportions is the pooled odds ratio, rather than Hedges's g, which is available in most meta analysis software. Page 7. Could you explain why a weighted kappa (second full paragraph) was used rather than the usual unweighted kappa? The unspoken assumption used here is that some disagreements between the reviewers may be considered more serious than others. Page 8. The method used to obtain the pooled Hedges's g effect size for the random effects model should be stated. The usual estimate for incorporating between study variances when there are heterogeneous studies is the Der Simonian-Laird so I assume this was the pooled estimate used using inverse variances of the effect sizes as the study weights which are referred to on page 28 in the right hand column which do seem related to the inverse of the effect size variances? Page 8 (last full paragraph) There seems a low agreement (kappa=0.515) between reviewers on an initial filtering out stage of papers for inclusion in the meta-analysis. Does this imply that the type of intervention is open to interpretation and is loosely defined? My worry here is that other expert reviewers in this area might have reservations about some of the papers excluded and perhaps also about some of the papers included in the meta-analysis as representing the type of intervention which this article is assessing and, thus, might have come up with a different set of papers to include in the meta-analysis if they had been involved in the choice of papers. Pages 13 (bottom paragraph), 14 (top paragraph) and 30 (Figure 1). Figure 1 seems to suggest that most of the studies considered in the meta-analysis in this article suffer from some form of bias in at least one of the five bias categories due in part to the use of self-report measures (top paragraph on page 14). This may suggest a shortcoming in the ability of the papers used in the meta-analysis to accurately measure the effects of interventions reported in this paper. Given the effects reported are low in any case (final paragraph on page 14) it may be argued that any extra unadjusted bias could influence the conclusions, perhaps even the direction of difference, concerning the effectiveness of the intervention. Pages 14 and 29. There is an acknowledged asymmetry in the funnel plot with a bias towards larger effects with smaller standard errors. I'd be interested to see the funnel plot on page 29 with the extra studies obtained using the trim and fill method (as mentioned in the second paragraph on page 8) added to the plot to see to what extent there has been an adjustment for asymmetry. I am not sure of the value of quoting the Hedges's g for the

	forest plots if the effect size adjusted for publication bias is also given. Surely the former effect size from the forest plot is misleadingly high as it has seemingly not been adjusted for publication bias? I don't have a feel for how large an effect size the quoted 0.20 is (third last line on page 14). I know there are rules of thumb from Cohen for what is a meaningful standardized difference but I wondered, given the specific clinical setting, if this could be translated into a more clinically meaningful number such as number needed to treat ie to give an indication of the effectiveness of the intervention in terms of the average number of patients that would be need to be treated using the intervention so that one more patient would take their medicine compared to the control group. Is an effect size of 0.20 a clinically meaningfully sized improvement? Page 15. I don't really get a feel from reading the text in the article what the response variable is (ie what the means represent in Table 2 on page 15). I assume, for example, the responses are mostly the percentages mentioned for most of the studies in the 'Adherence definition' columns in Table 2? In this case arcsine transforms or odds ratios should be used to compute effect sizes (see an example at http://www.statsdirect.com/help/meta_analysis/proportion_meta_analysis.htm for a meta-analysis of proportions based on what looks to me to be a similar study to this article looking at adherence with medication). This approach assumes that all the responses take the same form e.g. of proportions. Not sure if this is the case looking at Table 2 (Page 15) e.g. on the fourth row the adherence definition is the 'mean number of missed medicines' although this and others may be expressable as a proportion. Page 18. Given the heterogeneity of the study differences I'd like a little more detail about the sub-group analyses in the second paragraph which show effect size is not influenced by possible study confounders as this seems an important point particularly bearing in mind the discussion on page 19 defending the robustness of the results despite the heterogeneity of the study effect sizes. I think this analysis probably relates to the random effects meta-regression mentioned in the third paragraph on page 8 which includes a presumably significant variance component for between studies variation. I assume you are saying that none of the regression estimates for the seven predictors described on page 8 were statistically significant so the effect sizes are robust to different values of the predictors which is a 'plus' for this study? It could, however, also be acknowledged here that there are, given the 'broad inclusion criteria' (page 8), other reasons why the studies may have different effect sizes which are not considered by the random meta-regression. I do, therefore, still wonder given the acknowledged inherent heterogeneity resulting from the 'broad inclusion criteria' if the magnitude of a pooled point estimator (such as the 0.20 given in the final paragraph on page 14) is so interpretable in this article given other subgroups of studies which have not been considered may differ in the size of the effect of the intervention. Is the realistic conclusion of this article really, therefore, just to say that in most cases the intervention is helping (ie a constant direction of difference) albeit by varying degrees? Page 19. I think a simpler way of saying (first paragraph, lines 8-10) that the largest study increased the heterogeneity in study effect sizes is that the largest study (judging from the size of the Solomon effect size in the forest plot on page 28) had a comparatively small standardized group difference compared to most of the other studies. I think you could also motivate the results by saying here that all but three of the SMDs in Figure 2 are positive ie in favour of the intervention, although by differing amounts, which is far more than would be expected by chance.

	Page 28. 'Forest' rather than 'Forrest' for the plot title. I am not sure why +/- 2.12 are used as tick marks on the x axis rather than integers as is usual. These seem to be rather odd choices of values? I suspect SMD stands for standardised mean difference but this should be stated perhaps in a footnote on the plot so we know what the numbers in the column mean by just looking at the plot. Is SMD the same as Hedges's g referred to elsewhere in the article? More general points Hedges and Olkin (1985) suggest an unbiased form of g (used in this article) defined as {1 – 3/[4(n1+n2)-9]}g where n1 and n2 are the group sample sizes which the authors may wish to use or comment upon to further adjust their effect size. Slightly pedantic here but I think if one is using apostrophes it should be Hedges's g rather than Hedges' g since this statistic was first proposed by Larry Hedges in 1981. Reference Hedges, L and Olkin I (1985) Statistical methods for meta-analysis. Orlando, FL:Academic Press.

REVIEWER	Stephen Sutton Professor of Behavioural Science Institute of Public Health University of Cambridge UK Possible competing interest: The Farmer et al 2012 trial published in BMC Family Practice was not included in the review. I was a co-author of that paper.
REVIEW RETURNED	08-Mar-2013

THE STUDY	1. The definition of "cognitive-based" techniques should be re-visited and possibly revised or extended, and more examples should be given of what are -- and what are not -- cognitive-based interventions. The review compares cognitive-based interventions with educational and behavioural interventions but "providing education" is listed as a possible component of cognitive-based interventions (p.9) and the term "behavioural" is widely understood to include behaviour change techniques such as encouraging the formation of implementation intentions which the authors categorise as a cognitive-based technique. Also, Figure 1 includes a reference to "CBT" (cognitive behavioural therapy) which has both the cognitive and behavioural aspects but which the authors are presumably treating as a cognitive-based technique. ("Not MI or CBT" in Figure 1 doesn't seem to be an adequate description of the exclusion criterion based on type of intervention; also, weren't any papers excluded because they were not RCTs?). All this needs to be clarified throughout the paper so that the reader is clear what is being compared with what i.e. what is "cognitive" (but not educational or behavioural) and what is educational or behavioural but not cognitive? 2. What are "routine" healthcare providers (key messages; Conclusion section of the Abstract; Methods section; and Implications section of the Discussion)? 3. In the "key messages" it is stated that "Brief interventions are seemingly effective too". It's not clear where this conclusion comes

	from. 4. One of the “key messages” is that “Health care providers may wish to consider incorporation of these techniques into their medication adherence consultations”. I would suggest that this message is removed or modified. The interventions in most cases comprised several components (Supplementary table 1). This message (which also appears in the Implications section of the Discussion) could encourage practitioners to “pick and choose” single techniques from a multiple-component intervention, which may or may be effective on their own. 5. In the Results section of the Abstract, techniques such as aiming to increase the patient’s confidence and sense of self-efficacy are described as “more generalised” than MI (motivational interviewing). I would say that they were more specific than MI. 6. In the Introduction, it is stated that “Evidence suggests that complex, multi-faceted interventions, tailored to meet individual needs are most likely to be efficacious which is intuitive given the complex, multi-stage process that is medication taking.” In the Discussion, the authors should re-visit this statement and discuss whether their findings are consistent with it. 7. In the Methods section (p.6), it is stated that “Where multiple measures were reported, the percentage of patients achieving a specified adherence level was selected as this was common to more studies.” But in Table 2, there appear to be only 7 studies in which adherence was defined in terms of a percentage of patients (e.g. Berger et al 2005: % of patients discontinuing treatment) i.e. where the outcome was binary. Is that correct? In these cases, means are reported in Table 2 but not standard deviations; presumably these mean values are the percentages reported in the paper? How were the differences in percentages or, in one case, the odds ratio, converted to a standardised mean difference (smd)? Perhaps this information could be provided in a footnote to Table 2, so the reader knows how binary outcomes were handled. In the Data Analysis section of the Methods, it states that “Calculation of the effect size as Hedges’ g…enabled continuous adherence outcome measures of differing definition and measure to be combined, transforming this data into a common metric.”, but binary outcomes are not mentioned. 8. In other cases in Table 2 where standard deviations are not reported but the adherence measure was continuous (e.g. Van Es et al 2001), presumably the standard deviations were not reported in the original paper and the review authors have used the reported p-value to obtain the smd? Again, this could be mentioned in a footnote to Table 2. (I think it’s important for the reader to know how the study effect size was derived from the information reported in the primary source.) 9. How was the list of intervention components (Supplementary table 1) derived? What was the agreement between independent reviewers with regard to coding of intervention components? 10. The list of subgroup analyses on page 8 should map onto Table 3. 11. Ioannides has recommended that confidence intervals for I-squared are routinely reported. Especially when the number of studies included in a meta-analysis is small, the 95% confidence interval around the point estimate of I-squared may be wide, suggesting that the amount of heterogeneity is unknown. If the CI is not taken into account, meta-analysts may mistakenly conclude that heterogeneity is large when a more accurate conclusion is that we don’t know how much heterogeneity there is. There is a simple STATA procedure for calculating the CI.

	12. On page 17, it is stated that “Differences in sub-groups were not found to account for any notable degree of the observed heterogeneity.” How was this assessed? 13. In the footnote to Table 3, it is stated that the coefficient refers to the difference in effect size between the subgroups but it is not clear whether the coefficient is subgroup1 minus subgroup 2 or vice versa. Some of the coefficients are large but not statistically significant e.g. 0.828 for risk of bias. The authors should comment on the power of the meta-regression. It’s also not clear whether separate meta-regressions were conducted for each subgroup comparison. 14. Discussion. “Principle” should be “Principal”. 15. Figure 2. “Forrest” should be “Forest”. 16. In the Discussion (p.19), the authors interpret the pooled effect size to mean that “receipt of a cognitive-based technique improved adherence (% of doses taken) by 5.46%...”. But the effect size was based partly on studies that reported a dichotomous outcome (see point #7 above). So is this interpretation justified? 17. It’s not clear why the Farmer et al 2012 trial in BMC Family Practice was not included in the meta-analysis.

REVIEWER	Marie Krousel-Wood MD, MSPH Ochsner Clinic Foundation USA
REVIEW RETURNED	11-Mar-2013

THE STUDY	Details about other cognitive-based techniques should be provided.
RESULTS & CONCLUSIONS	There is concern about the notable heterogeneity of the studies included in the meta-analysis and lack of clarity about what is meant by cognitive-based intervention. These concerns raise questions about the appropriateness of combining these studies.
GENERAL COMMENTS	Overall: This is a well-written manuscript addressing a very important topic with the main objective to describe and evaluate the use of cognitive-based techniques as interventions to improve medication adherence. The authors have followed appropriate methodological steps for this meta-analysis, identified notable heterogeneity and publication bias and thus used random effects models and adjustments for publication bias. They concluded that cognitive-based interventions (including MI) were associated with a small, but statistically significant, improvement in medication adherence. While there is much interest in better understanding the effectiveness of interventions to improve medication adherence, there are concerns about the heterogeneity and publication bias that dampen enthusiasm for these results. Areas in need of clarification: 1. Page 3, results section: Consider noting that the “adjustment for publication bias” generated a more conservative estimate. 2. Page 6: The authors note that the focus of the analysis is cognitive-based interventions. However, only motivational interviewing is described. Details about other cognitive-based techniques should be provided. On page 9, it appears that nearly half of the studies did not use a ‘clearly cognitive-based technique’ and there is substantial variability in the number and type of components used in the interventions. [It may be that this lack of clarity may have accounted for the suboptimal agreement (K = 0.515) between reviewers re the abstract screening.] 3. Heterogeneity is a key concern. The authors used the appropriate random effects model which resulted in a more conservative

	estimate. Nevertheless, the notable heterogeneity raises questions about the appropriateness of combining these studies. On page 17, the authors note that differences in subgroups were not found to account for any notable degree of the observed heterogeneity. The authors have provided information in the discussion on page 19 which outlines numerous between study differences (type of intervention, definitions of adherence and assessment tools used). The information provided in the methods does not provide a convincing argument that "the core of each intervention was the use of a cognitive-based technique to improve medication adherence". Although the authors have provided information about the appropriateness of combining of the results, the justification does not appear adequate.

REVIEWER	Bond, Christine University of Aberdeen
REVIEW RETURNED	11-Mar-2013

GENERAL COMMENTS	In general this is a well written paper and an important and interesting meta-analysis. I have very few comments to make and recommend acceptance. 1. There is a typo page 7 second line- suitably instead of suitability 2. In the Results for the over view of papers I always prefer it if the citation numbers of the specific paper are listed after each statement eg 'Just over half of the studies described an intervention with a clearly defined cognitive based techniqueb,c,f,d,e', 3. Also in this overview I would like to see some commentary on geographical location, dates of publication and disease 4. Page 9 lines 8-9 I am not quite convinced how vague terms such as 'providing education' or 'increasing patient knowledge' can be described as MI? Would a sensitivity analysis with these removed change the findings? Also increasing 'social support'. I am not sure if I am missing something but this seems to indicate that any one taking an interest might help! 5. Another interesting sub group analysis might be comparing objective and subjective adherence measures

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Peter Watson
Statistician

1.1 I felt the outcome measure (e.g. percentages or numbers of medications used) could be discussed more in the text e.g. an idea of types of medication, ranges of how often they were used.

We agree that this information will add 'richness' to the data. Unfortunately, details of medication regimens such as frequency of use could not be elicited in a consistent manner from the original articles. Information regarding the types of conditions being treated was, however, available from the included articles. This information may additionally provide some indication of regimen

complexity, for example for studies considering HIV therapy, it is reasonable to assume that multiple medications and multiple daily dosing with Highly Active Anti-Retroviral Therapy were investigated. Similarly, for studies considering osteoporosis treatments, it is reasonable to assume this would involve a far simpler regimen of once weekly bisphosphonate therapy. Information regarding the conditions for which medications were prescribed has therefore been added to the second paragraph of page eight.

1.2 The low kappa (page 8) makes we wonder how well defined these types of intervention studies are to yield such a low inter-rater agreement on deciding at an initial stage whether or not to include a study in this meta-analysis and if other reviewers would have argued that other studies might have been excluded/included in this analysis.

The problem of defining interventions and determining their primary components is well documented by many other authors and primarily attributed to poor study reporting. In recognition of this problem, notable work has been undertaken to develop a taxonomy of behaviour change techniques to enable some standardisation of terms to describe interventions[1-3]. In addition, the development of the MRC framework for designing and evaluating complex behaviour change interventions has endeavoured to address the deficits in study reporting to enable greater clarity and transparency with regard to intervention definition and key components. Unfortunately, many of the studies included in the meta-analysis pre-date the recent movements to improve study reporting and therefore often yielded a paucity of clear information to judge our decisions.

Whilst the paucity of information was problematic, the present study used a structured screening tool to guide inclusion and both independent reviewers were clear regarding the type of interventions eligible for inclusion: any intervention that aimed to address a patient's thoughts, feelings or motivation towards taking their medicines.

However, the paucity of information, primarily in the abstracts often yielded differing 'interpretations' of the intervention components between the reviewers, leading to our discrepancies. Whilst we recognise that these discrepancies are far from ideal, we feel that the robust and meticulous methods employed following the abstract screening stage were sufficient to overcome this problem. For abstracts where a discrepancy between reviewers was attributed to differing interpretations of unclear information (which was the vast majority of cases), abstracts were included until full text articles could clarify any ambiguities. This conservative approach ensured that no study would have been excluded unless definitive information was available to confirm ineligibility at full text screening. As clarification, the following text had been added to paragraph 2 of "Strengths and weaknesses of our work" section of the discussion.

"The conservative approach to abstract screening prevented study exclusion if disagreement was associated with insufficient information and thus prevented exclusion in error."

Furthermore, information regarding inter-rater agreement at full text is now reported and the following text added to 'Results', paragraph 1:

"At the full text screening stage, agreement between the two independent reviewers was much higher with a kappa value of 0.91 indicative of almost perfect agreement."

1.3 Most of the responses (Table 2 on page 15) appear to take the form of proportions. In this case the effect size of interest should be based upon either arcsine transformed proportions or odds ratios. I am not sure if these have been used here. The reason for the need to use special effect sizes with proportions is that the variance of a proportion is itself related to the value of the proportion so mean and variance are not independent of each other. Such an approach also requires that all responses take the same form e.g. are expressed as proportions/percentages.

Odds ratio could be used for some of the included studies. However, the outcome measures used in many included studies were not the typical binary or categorical outcomes. For example, percentage prescribed doses taken or % of days in which medication was taken were continuous

variables. Since the different outcome measures were used in the included studies, we have to use the standardised mean difference in the meta-analysis.

1.4 A main concern is that most of the studies show some bias in at least one of the five bias categories (top of page 14 and Figure 1 page 30) which cannot be adjusted for which may have led to biases in the effect size estimates.

The primary category that accounted for risk of bias was 'blinding of outcome assessment' which is a known problem in this field where self-reported measures are most frequently used to assess adherence. It is essential that the risk of bias is acknowledged but we feel it should be recognised as an accepted limitation of this field of research. Moreover, positive adherence effects have been seen consistently across all but three of the studies, irrespective of the outcome measure used. This observation adds further weight to the argument that the risk of bias from 'blinding of outcome assessment' has not adversely influenced the integrity of the effect sizes calculated. To add further confidence to these findings, subgroup analysis to explore the effect of outcome measure has been undertaken and reported in table3 of the results. The following information has been added:

Variable	**Sub-group-A vs. subgroup-B**	**No. of studies (no. of participants) in each sub-group**	**Co-efficient (95% CI)**	**P-value**
Outcome measures	Objective vs. subjective measured outcomes	14 (3850) Vs. 12 (1366)	-0.16 (-0.44 to 0.11)	0.225

1.5 Despite reassurance in the discussion at the end of the first full paragraph on page 19, I still wonder given the resultant number of inherent differences between the studies due to the 'broad inclusion criteria' (page 8) if it is sensible to give pooled effect sizes such as those on page 14 (last two paragraphs) since the studies may all have inherently different effect sizes due to their provenance. Perhaps further explanation of the subgroup analyses and the robustness of the effect sizes indicated by their results (first full paragraph on page 18) will help allay this concern.

We have made the following amendments to the discussion text (P20) to help allay concerns:

"Interestingly, the largest study had a small standardized group difference compared to most of the other studies which contributed substantially to the heterogeneity.[42] Furthermore, results from all but three of the studies indicate positive effects of the intervention. Aside from these between study differences, the actual interventions were variable, as were the definitions of adherence and assessment tools used. The differences between subgroups were statistically non-significant in terms of disease area, intervention components, delivery methods, delivery personnel, intensity, usual care and risk of bias. However, the statistical power was limited by the small number of studies included in the subgroup analyses. The analyses may therefore have failed to detect some important subgroup differences."

1.6 The various biases in the studies (second paragraph page 8 and top of page 14) and the broad inclusion criteria (first full paragraph on page 8) acknowledged by the authors make a meta-analysis of this data challenging and the authors do flag up these weaknesses. Two main concerns are the inherent biases in the studies reported on page 14 and the low

inter-rater agreement on choosing which studies should be used in the meta-analysis which may bring into question the accuracy and reliability of the estimates of group differences presented in this study. One might argue that if the studies considered in this paper are themselves biased then the analysis based upon these will also be biased.

In response to the comment regarding bias in the reported studies, please refer to response 1.4. In response to the comment regarding low inter-rater agreement, please refer to response 1.2.

1.7 It is also not clear, assuming the responses are (at least mostly) in the form of proportions, if effect sizes appropriate for proportions have been computed (I suggest a website giving more details, as needed, below). The usual effect size for proportions is the pooled odds ratio, rather than Hedges's g, which is available in most meta analysis software.

Odds ratio could be used for some of the included studies. However, the outcome measures used in many included studies were not the typical binary or categorical outcomes. For example, percentage prescribed doses taken or percentage of days in which medication was taken were continuous variables. Since the different outcome measures were used in the included studies, we had to use standardised mean difference in the meta-analysis.

1.8 Page 7. Could you explain why a weighted kappa (second full paragraph) was used rather than the usual unweighted kappa? The unspoken assumption used here is that some disagreements between the reviewers may be considered more serious than others.

Apologies; this was an error in the manuscript which has now been corrected. Weighted Kappa was not used; as rightly pointed this would assume certain discrepancies were more serious than others and this was not so. The manuscript has been amended to remove the word 'weighted'.

1.9 Page 8. The method used to obtain the pooled Hedges's g effect size for the random effects model should be stated. The usual estimate for incorporating between study variances when there are heterogeneous studies is the Der Simonian-Laird so I assume this was the pooled estimate used using inverse variances of the effect sizes as the study weights which are referred to on page 28 in the right hand column which do seem related to the inverse of the effect size variances?

DerSimonian-Laird method was used to conduct random-effects meta-analysis. This has been clarified in the method section.

1.10 Page 8 (last full paragraph) There seems a low agreement (kappa=0.515) between reviewers on an initial filtering out stage of papers for inclusion in the meta-analysis. Does this imply that the type of intervention is open to interpretation and is loosely defined? My worry here is that other expert reviewers in this area might have reservations about some of the papers excluded and perhaps also about some of the papers included in the meta-analysis as representing the type of intervention which this article is assessing and, thus, might have come up with a different set of papers to include in the meta-analysis if they had been involved in the choice of papers.

This relates closely to item 1.2. Whilst the type of intervention that we sought to include was well defined, the information provided in abstracts was often lacking yielding the aforementioned discrepancies and subsequent low Kappa value. As mentioned in our response to 1.2, discrepancies were resolved to achieve consensus amongst reviewers, and full text reviewing yielded far higher agreement.

1.11 Pages 13 (bottom paragraph), 14 (top paragraph) and 30 (Figure 1). Figure 1 seems to suggest that most of the studies considered in the meta-analysis in this article suffer from some form of bias in at least one of the five bias categories due in part to the use of self-

report measures (top paragraph on page 14). This may suggest a shortcoming in the ability of the papers used in the meta-analysis to accurately measure the effects of interventions reported in this paper. Given the effects reported are low in any case (final paragraph on page 14) it may be argued that any extra unadjusted bias could influence the conclusions, perhaps even the direction of difference, concerning the effectiveness of the intervention.

Please see our response to item 1.4 as the issue raised is very similar. The additional subgroup analysis comparing the effect of subjective versus objective measures of adherence responds to the reviewer's concerns regarding the ability of the outcome measures to accurately measure the effects of interventions. Whilst studies adopting subjective measures of adherence have been correctly categorised as being at risk of bias, when compared with objective measures not susceptible to this bias, no significant impact on effect size was identified.

1.12 **Pages 14 and 29. There is an acknowledged asymmetry in the funnel plot with a bias towards larger effects with smaller standard errors. I'd be interested to see the funnel plot on page 29 with the extra studies obtained using the trim and fill method (as mentioned in the second paragraph on page 8) added to the plot to see to what extent there has been an adjustment for asymmetry. I am not sure of the value of quoting the Hedges's g for the forest plots if the effect size adjusted for publication bias is also given. Surely the former effect size from the forest plot is misleadingly high as it has seemingly not been adjusted for publication bias?**

As requested, the funnel plot produced using Trim and Fill techniques has been added to the manuscript as a supplementary figure demonstrating the extent of adjustment for asymmetry. Whilst we recognise that reporting both the unadjusted and adjusted hedges' g values could be misleading, we feel it is important to report both in order to offer transparency of our finding.

1.13 **I don't have a feel for how large an effect size the quoted 0.20 is (third last line on page 14). I know there are rules of thumb from Cohen for what is a meaningful standardized difference but I wondered, given the specific clinical setting, if this could be translated into a more clinically meaningful number such as number needed to treat ie to give an indication of the effectiveness of the intervention in terms of the average number of patients that would be need to be treated using the intervention so that one more patient would take their medicine compared to the control group. Is an effect size of 0.20 a clinically meaningfully sized improvement?**

We were mindful that for general medical readership, an effect size of 0.20 would not be particularly meaningful and have made attempts to address this. Paragraph 4 of the 'strengths and weaknesses of our work' section of the discussion describes that receipt of a cognitive based technique was associated with a 6% increase in the percentage of doses taken correctly and argues that for some conditions this could be clinically meaningful. This estimate is derived from the data of six studies that used the percentage of prescribed dose taken. The pooled standard deviation of this outcome was 30.7%. Then we estimated that a standardised mean difference of 0.205 (0.084 to 0.326) corresponds to a difference of 6.3% (2.6% to 10.0%) in the percentage of doses taken between the intervention and control group.

1.14 **Page 15. I don't really get a feel from reading the text in the article what the response variable is (ie what the means represent in Table 2 on page 15). I assume, for example, the responses are mostly the percentages mentioned for most of the studies in the 'Adherence definition' columns in Table 2? In this case arcsine transforms or odds ratios should be used to compute effect sizes (see an example at http://www.statsdirect.com/help/meta_analysis/proportion_meta_analysis.htm for a meta-analysis of proportions based on what looks to me to be a similar study to this article looking at adherence with medication). This approach assumes that all the responses take the same form e.g. of proportions. Not sure if this is the case looking at Table 2 (Page 15)**

e.g. on the fourth row the adherence definition is the 'mean number of missed medicines' although this and others may be expressable as a proportion.

Please refer to the response to item 1.7 regarding reporting of odds ratios. In response to the comment regarding 'what the means represent', please find below an extract of table 2:

Adherence definition (assessment measure)	**Extracted data**		
	Intervention group	**Control group**	**P-value**
% of patients scored as adherent on all 6 items of a self-report scale (based on Morisky's self-reported scale)	Mean = 91.9	Mean = 61.7	0.001
% of patients discontinuing treatment by study endpoint (patient interview)	Mean = 98.8	Mean = 91.3	0.001
% of prescribed doses taken over a month (electronic monitoring)	Mean (SD) = 93.4 (12.3)	Mean (SD) = 79.1 (28.1)	
Mean number of missed medicines in the last 30 days (self-report questionnaire)	Mean (SD) = 0.13 (0.35)	Mean (SD) = 0.98 (1.48)	

The column describing the adherence definition explains the reported mean.

1.15 Page 18. Given the heterogeneity of the study differences I'd like a little more detail about the sub-group analyses in the second paragraph which show effect size is not influenced by possible study confounders as this seems an important point particularly bearing in mind the discussion on page 19 defending the robustness of the results despite the heterogeneity of the study effect sizes. I think this analysis probably relates to the random effects meta-regression mentioned in the third paragraph on page 8 which includes a presumably significant variance component for between studies variation. I assume you are saying that none of the regression estimates for the seven predictors described on page 8 were statistically significant so the effect sizes are robust to different values of the predictors which is a 'plus' for this study? It could, however, also be acknowledged here that there are, given the 'broad inclusion criteria' (page 8), other reasons why the studies may have different effect sizes which are not considered by the random meta-regression. I do, therefore, still wonder given the acknowledged inherent heterogeneity resulting from the 'broad inclusion criteria' if the magnitude of a pooled point estimator (such as the 0.20 given in the final paragraph on page 14) is so interpretable in this article given other subgroups of studies which have not been considered may differ in the size of the effect of the intervention. Is the realistic conclusion of this article really, therefore, just to say that in most cases the intervention is helping (ie a constant direction of difference) albeit by varying degrees?

We have now added the following text to the 'Sub-group analyses via meta-regression' section of the results:

"…….to emphasise that the number of studies in the subgroup analyses was small, so that the results may have failed to detect some important differences between subgroups. "

1.16 Page 19. I think a simpler way of saying (first paragraph, lines 8-10) that the largest study increased the heterogeneity in study effect sizes is that the largest study (judging from the size of the Solomon effect size in the forest plot on page 28) had a comparatively small standardized group difference compared to most of the other studies. I think you could also motivate the results by saying here that all but three of the SMDs in Figure 2 are positive ie in favour of the intervention, although by differing amounts, which is far more than would be expected by chance.

Thank you for this suggestion; the text has been amended to read:

"Interestingly, the largest study had a small standardized group difference compared to most of the other studies which contributed substantially to the heterogeneity. Furthermore, results from all but three of the studies indicate positive effects of the intervention."

1.17 Page 28. 'Forest' rather than 'Forrest' for the plot title. I am not sure why +/-2.12 are used as tick marks on the x axis rather than integers as is usual. These seem to be rather odd choices of values? I suspect SMD stands for standardised mean difference but this should be stated perhaps in a footnote on the plot so we know what the numbers in the column mean by just looking at the plot. Is SMD the same as Hedges's g referred to elsewhere in the article?

The spelling of 'Forest plot' has been amended and we have altered the figure to include more appropriate axis labels as advised. The reference to 'SMD' has been replaced with 'Hedges' g' and the axis scale now runs from -1 to 2.

1.18 Hedges and Olkin (1985) suggest an unbiased form of g (used in this article) defined as (1 – 3/[4(n1+n2)-9])g where n1 and n2 are the group sample sizes which the authors may wish to use or comment upon to further adjust their effect size.

Thanks for suggesting some further adjustment. The calculation of Hedges' g is based on a recent text book by Borenstein et al 2009: "Introduction to Meta-analysis (Wiley)" (reference 23), and we feel this commonly used method (Hedges's g) is acceptable to this meta-analysis.

1.19 Slightly pedantic here but I think if one is using apostrophes it should be Hedges's g rather than Hedges' g since this statistic was first proposed by Larry Hedges in 1981.

Whilst different reference sources have written this as Hedge's, Hedges's and Hedges', the most consistent form is Hedges' g, as we have used in our manuscript. The apostrophe after the 's' indicates that 'g' belongs to Hedges.

VERSION 2 – REVIEW

REVIEWER	Stephen Sutton Professor of Behavioural Science University of Cambridge UK
REVIEW RETURNED	10-Jun-2013

GENERAL COMMENTS	1. I would add a sentence in the Results section of the Abstract stating that there was high heterogeneity and reporting the I-squared value. 2. The last sentence of the Conclusions section of the Abstract should be removed or re-worded. "Can be delivered" implies feasibility rather than efficacy. If a conclusion is to be drawn from the

	subgroup analyses, the relevant results should be included in the Results section of the Abstract. The statement in the Results that "No statistically significant differences were observed in a range of subgroup analyses" is misleading. 3. The Cochrane Library should be added to the list of databases on page 6. 4. Page 7.The subgroup analysis by outcome measure should be described as post hoc or exploratory (not prespecified like the other subgroup analyses). 5. Page 8. In 12 studies a clearly defined CBCT could not be identified. It is not clear how/why such interventions were classified as using cognitive-based behaviour change techniques. 6. Page 8. As I mentioned in my previous review, it is confusing to include 'providing education' and 'increasing patient knowledge' as cognitive based behaviour change techniques (CBCT), given the distinction that has been made between CBCT and education. Perhaps include a sentence of explanation. 7. Page 8. Aren't implementation intentions and if-then plans clearly defined CBCTs? Why aren't they mentioned in the paragraph that describes the intervention components? 8. In Table 1, several interventions are described as involving non-specific techniques. This needs to be explained. 9. Page 19. 1st para. Should be "this bias" not "these biases". And, further down, "implies" not "infers". 10. Page 19-20. It's fine to compare the findings for MI with those in other healthcare domains but it's perhaps also worth emphasising that the non-MI interventions appeared to be no less effective. 11. Key messages. "Brief interventions are seemingly effective too". The subgroup analysis for intervention exposure compared four sessions or fewer with five sessions or more. I'm not sure that it is appropriate to describe four sessions or fewer as "brief". 12. Consistency between Abstract, Discussion and Key messages could be improved.

VERSION 2 – AUTHOR RESPONSE

Response document for BMJ Open resubmission version 2
1. I would add a sentence in the Results section of the abstract stating that there was high heterogeneity and reporting the I-squared value.

Thank you for this suggestion, this has been added to the relevant section.

2. The last sentence of the Conclusions section of the Abstract should be removed or re-worded. "Can be delivered" implies feasibility rather than efficacy. If a conclusion is to be drawn from the subgroup analyses, the relevant results should be included in the Results section of the Abstract. The statement in the Results that "No statistically significant differences were observed in a range of subgroup analyses" is misleading.

Thank you for these comments; they are useful in improving our manuscript. The statement "no statistically significant differences were observed in a range of sub-group analyses" has been removed as we agree that this was misleading. Instead this statement has been replaced with:

"The majority of sub-group analyses produced statistically non-significant results"

We would like to add:

“for example, there were no significant differences between interventions delivered by specialists compared to interventions delivered by routine healthcare providers (co-efficient value (95% CI) = -0.10 (-0.46 to 0.26) P=0.561) and intervention exposure was not statistically associated with efficacy (co-efficient value (95% CI) =0.07 (-0.35 to 0.50) P=0.728)”.

However, the strict word limit for the abstract means this level of detail could not be added.

The conclusion has also been revised accordingly. The sentence containing the words “can be delivered” has been removed as we agree this relates to feasibility not efficacy. It has been replaced with:

“Sub-group analyses suggest that these interventions are amenable to use across different populations and in differing manners without loss of efficacy. These factors may facilitate incorporation of these techniques into routine care.”

3. The Cochrane Library should be added to the list of databases on page 6.

Thank you for spotting this omission, this has been added.

4. Page 7.The subgroup analysis by outcome measure should be described as post hoc or exploratory (not pre-specified like the other subgroup analyses).

Thank you for this suggestion, the wording of this section has been amended accordingly.

5. Page 8. In 12 studies a clearly defined CBCT could not be identified. It is not clear how/why such interventions were classified as using cognitive-based behaviour change techniques.

Thank you for highlighting to us that this section is not clear. The following sentence has been added to improve clarity:

“ All studies within this group included one or more components that aimed to alter the patients, thoughts, feelings, motivation or confidence towards adherence and that could therefore be classified as a cognitive-based behaviour change technique”

6. Page 8. As I mentioned in my previous review, it is confusing to include ‘providing education’ and ‘increasing patient knowledge’ as cognitive based behaviour change techniques (CBCT), given the distinction that has been made between CBCT and education. Perhaps include a sentence of explanation.

Thanks again for highlighting this source of ambiguity. The following sentence has been added to the end of the aforementioned paragraph:

“Many studies combined cognitive-based behaviour change techniques with more traditionally used educational (e.g. increasing patient knowledge) and behavioural (e.g. regimen simplification and provision of dosing aids) components”

7. Page 8. Aren’t implementation intentions and if-then plans clearly defined CBCTs? Why aren’t they mentioned in the paragraph that describes the intervention components?

We agree that III are clearly defined CBCTs. In this paragraph we aimed to summarise the most

commonly used techniques to provide an overview of the data. As only three studies used III it did not seem intuitive to specifically mention this. However having reconsidered this point in light of this comment, we agree that it may be useful information to our readers. The following sentence has therefore been added:

“A further three (11.5%) studies used Implementation Intention Interventions (III, also known as if-then planning) as a clearly defined CBCT”

8. In Table 1, several interventions are described as involving non-specific techniques. This needs to be explained.

We have added a sentence to the relevant part of the text to reference the table and make it clear to which studies these relate. The full sentence now reads:

“For 12 (46.2%) studies, a clearly defined CBCT such as MI could not be identified32-3545-52, these studies are identified in table 1 as ‘multiple components; non-specific techniques’.”

9. Page 19. 1st para. Should be “this bias” not “these biases”. And, further down, “implies” not “infers”.

Both of these have been amended as suggested.

10. Page 19-20. It’s fine to compare the findings for MI with those in other healthcare domains but it’s perhaps also worth emphasising that the non-MI interventions appeared to be no less effective.

Thank you for this suggestion. The following sentence has been added to the end of the paragraph:

“The overlapping confidence intervals of the effect sizes calculated for MI-based and non-MI based interventions suggests that MI based interventions are unlikely to be superior in their efficacy compared to those based on other cognitive-based behaviour change techniques”.

11. Key messages. “Brief interventions are seemingly effective too”. The subgroup analysis for intervention exposure compared four sessions or fewer with five sessions or more. I’m not sure that it is appropriate to describe four sessions or fewer as “brief”.

Thank you for highlighting this problem to us. We agree that the classification of brief interventions as four sessions or fewer is inappropriate. The majority of studies provided information regarding the number of sessions over which the interventions were delivered but this is not a reliable proxy for intervention exposure as an intervention of ten half hour sessions would be equivalent to an intervention of five one hour sessions in terms of ‘exposure time’. The total number of minutes spent delivering the intervention is therefore a more reliable measure of intervention exposure but this information was inconsistently reported in the studies. However, for 16 studies a reasonable estimate of the number of minutes spent on the intervention could be calculated. The following paragraph has been added to the first part of the results section to reflect the analysis as intervention exposure by number of minutes:

“Intervention exposure as the total number of minutes spent delivering the intervention could be estimated for 16 studies. In the remaining 10 studies this data was not available. Intervention exposure ranged from thirty minutes to eight hours and fifteen minutes. The median (IQR) intervention exposure was 175 (118 to 263) minutes”

As there is currently no widely accepted definition for what constitutes a brief intervention, determining an appropriate cut-off point for classification of interventions as brief or otherwise has been

problematic. This difficulty is augmented by the paucity and variability of data that could be extracted from the various studies. An arbitrary cut off of three hours has however been used to create two subgroups of roughly equal study number to explore this. The appropriate section explains that this is common meta-analytical practice.

“The classification of studies into sub-groups was largely intuitive. However, as a continuous rather than categorical variable, ‘total intervention exposure’ was less amenable to intuitive dichotomisation. In such instances, it is standard practice to create two sub-groups by distributing a roughly equal number of studies to each group. An arbitrary cut off point of three hours was therefore used to split the data into two sub-groups”.
We are mindful that this arbitrary cut off of three hours may not seem intuitive and so have undertaken an additional post hoc meta-regression to explore the variable ‘intervention exposure’ as a continuous variable. The following has been added to the results section:

“As the variable ‘intervention exposure’ was a continuous variable, an additional post-hoc analysis was undertaken. This allowed the variable to be analysed in it ‘natural’ continuous state rather than two sub-groups. This exploratory analysis was undertaken to ensure that the arbitrary cut off point of three hours had not adversely influenced the data. A co-efficient value (95% CI) of 0.001 (-0.001 to 0.002) suggested that there was no association between intervention exposure and effect size. A non-significant p-value of 0.540 confirmed this and demonstrates comparable results to the sub-group analysis for this variable”.
As there is no clear cut-off that constitutes a brief intervention, as advised, the message has been revised as:

“According to the results of sub-group analyses, cognitive-based behaviour change techniques can be effectively delivered by routine healthcare providers, and the effectiveness of interventions is not associated with intervention exposure.”

12. Consistency between Abstract, Discussion and Key messages could be improved.