Dear Editors,

We are grateful for your consideration of this manuscript, and we also very much appreciate your suggestions, which have been very helpful in improving the manuscript. We also thank the reviewers for their careful reading of our text.

All the comments we received on this study have been taken into account in improving the quality of the article, and we present our reply to each of them separately. With regard to some of the suggestions, we would note the following:

- The reviewers do not always agree as to their suggestions, and in some cases they are even inconsistent.
- After updating the review, one new article was selected. Therefore, all results have been changed slightly. These changes have not been highlighted in the revised version “with changes” (to be readable).
- We are especially grateful for the suggestions of reviewers regarding certain confounders in the consideration of mortality.
- Responding to all the suggestions of the reviewers has made it necessary to increase the number of bibliographic citations and extend the length of the manuscript. We have gone from 66 to 70 citations, and from 3764 to 4622 words.
- We appreciate the suggestion with regard to the incorporation of the following reference, which permitted us to include aspects that, despite their relevance, were not taken into account in the original version:
- Moreover, additional references have been included with respect to other comments:
  - New selected article:
  - Quality instrument:
  - Discussion:
We believe that the editor and reviewers’ suggestions have been very helpful in improving the manuscript, both in the introduction and methodology (a section in which a substantial amount of information has been added) as well as the results (with a more precise and extensive description of the characteristics of the sample) and discussion (this section has also been lengthened considerably). Two versions of the manuscript are enclosed, one where all the changes have been underlined, and another version without any marks.

We hope that these changes to the manuscript will facilitate the decision to publish this study in your journal. We have made a considerable effort to take into account the interesting suggestions proposed by the reviewers. In any case, we are open to consideration of any further comment on our answers.

Sincerely,

The authors
COMMENTS FROM BMJ:

1. Members of the committee were: José Merino (Chair), Richard Riley (Statistics advisor), Elizabeth Loder, Rubin Minhas, Amy Price, Tiago Villanueva, Wim Weber.

Decision: Put points

Detailed comments from the meeting:

We thought your paper addresses an interesting and important research question.

We had the following concerns:

A. The abstract is very difficult to understand.

ANSWER: Following this suggestion, abstract has been extensively rewritten. There are reviewed conclusions. (See answer “I”)

B. The search is old.

ANSWER: The search has been updated to September 2016. One new article has been added to this meta-analysis. Consequently, there have been slight changes in Figure 1 (Flow diagram of selection process) and in the numbers in the figure results. Nevertheless, this update has not changed the meaning or interpretation of the results.

C. We wondered about possible confounders: do we know enough about the characteristics of patients and doctors in these studies. You evaluate some patient characteristics, but not others that surely have an impact on outcomes, especially psychiatric and other comorbidities. Not all doctors and health care workers are equally skilled in supervising this treatment. In the US, doctors must have special training in order to be able to prescribe suboxone. That is not true with methadone. Perhaps some outcomes reflect physician or treatment program characteristics.

ANSWER: We clearly agree with the committee, but these confounders were not included in our tables or results due to heterogeneity of studies in including this information. In fact, most of the selected articles lack this information.

Nevertheless, this editor’s suggestion about possible confounders and other similar reviewers’ comments have been taken into consideration in this version. Consequently:

1. We have reviewed all selected articles in order to include all available information on possible confounders. Therefore, Table 1 has been expanded with new information about:
   a. Patients: psychiatric and somatic comorbidities, and non-opioid drug abuse.
   b. Health care workers supervising: OST providers
   c. Inpatient induction
2. We have compared all the values of unadjusted effect indicators in articles with the adjusted ones in those articles where this information was available. In all cases, values of adjusted indicators were similar or even better in terms of our rate ratio. Only 7 articles included this kind of information and confounding factors were heterogeneous, thus they were impossible to meta-analyze. Nevertheless, we have included data on crude and adjusted rate ratio for such studies in a new table (See Table 3), where it can be appreciated that the magnitude of the effect is similar or even greater after adjusting for some specified confounders. These new results have been discussed in discussion subsection “Strengths and limitations” (page 19, line 15):

BEFORE: First, temporal-geographic differences in patients, OST and contextual characteristics could have differentially affected mortality during and after OST.

AFTER: First, characteristics of patients and OST were poorly reported or adjusted in most included studies. However, using data from seven studies, pooled effect size measures (ie rate ratio) adjusted for several potential confounders was calculated, the results showing that the effect of MMT in reducing mortality was even greater (Table 3). As mentioned above, the possible existence of differences in patient or treatment characteristics may also partly explain differences in mortality between MMT and BMT, and must be the subject of future research.

D. PRISMA is a reporting guideline not a guide to conducting systematic reviews and meta-analyses.

Indeed, the committee is right. We followed the PRISMA statement for reporting systematic reviews and metaanalyses. So the last sentence of the first paragraph has been changed (page 6, line 11).

BEFORE: We followed the PRISMA statement for conducting and reporting systematic reviews and metaanalyses.

AFTER: We followed the PRISMA statement for reporting systematic reviews and metaanalyses.

E. You present exclusion, but not inclusion criteria, unless we missed it.

In order to clarify the inclusion criteria, we have rewritten the first sentence of the second paragraph in the methods section (Page 6, line 7):

BEFORE: We sought to identify all prospective or retrospective cohort studies that provided mortality follow-up data among opioid-dependent people during and after opioid substitution treatment with methadone or buprenorphine.

AFTER: We included cohort studies comparing the mortality among opioid-dependent people. To be eligible, studies had to include follow-up data during and after opioid substitution treatment with methadone or buprenorphine.

F. We were not familiar with the quality assessment you have used.
We have created an ad hoc quality assessment based on validated instruments. This instrument has been created based on a well-known and widely used SIGN50 and the specific drug-related instrument created in Australia by the National Drug and Alcohol Research Centre. The reason for creating this instrument instead of using SIGN50 was based on the lack of specific items relating to the quality of observational drug-related articles, which we strongly believe are necessary to evaluate each selected article more accurately. The majority of existing tools are aimed at assessing the quality of reporting (see the equator collection of checklist at http://www.equator-network.org/?post_type=eq_guidelines&eq_guidelines_study_design=observational-studies&eq_guidelines_clinical_specialty=0&eq_guidelines_report_section=0&s=+&eq_guidelines_study_design_sub_cat=0), whereas we wanted to focus on the quality of studies. The use of the SIGN tool is also supported by the Cochrane drugs and alcohol group (see for example: https://www.ncbi.nlm.nih.gov/pubmed/23740538). In order to make this clearer, we have rewritten the explanation (Page 7, line 24):

BEFORE: The quality of selected studies was assessed by using criteria adapted from standardized and extensively used instruments, namely, the methodology checklist for cohort studies developed by the Scottish Intercollegiate Guidelines Network \(^{14}\) and the checklist for drug-related studies by the National Drug and Alcohol Research Centre, Australia.\(^{15}\) These criteria comprised a general appraisal of potential biases and confounding relevant to cohort studies, together with an ad hoc assessment of reporting for studies on mortality during and after opioid substitution treatment (appendix 2).

AFTER: A standardized quality assessment form for observational studies was specifically designed based on well-known and extensively used instruments: SIGN50 Scottish Intercollegiate Network 2004\(^{14}\) and the drug-related checklist developed by the National Drug and Alcohol Research Centre, Australia\(^{15}\). The design process, based on a thorough review of the above-mentioned sources, included the development of different proposals, discussion of their appropriateness and final agreement among the authors. The final version comprised separate sections according to the study design, and was based on a “star system” score approach\(^{16}\), including a general appraisal of external and internal validity and of the biases relevant to observational studies, plus an ad hoc assessment of reporting for studies on mortality during and after opioid substitution treatment (appendix 2).

G. In Table 1 we do not see any information on the proportion of patients for whom outcome data were available; in other words, how did the denominator change over time?

Following this suggestion, we have checked all included articles in order to provide this information. The results can see in table 1: “Loss to follow-up”

H. It does seem clear that people do better in treatment than out of treatment, but we are not sure we need a meta-analysis to know that.

We agree with the first sentence, but, first, most articles compare people in treatment (in general) with people without treatment (frequently: never in treatment). This article only includes studies in which patients themselves are compared between periods in and out of treatment, which reduces the likelihood that the results are affected by measured or unmeasured confounders. We really think it is quite relevant as we explain in the introduction. Second, we also think it is important to quantify the magnitude of this “better situation” as is done in our paper. And third, as the second reviewer says, the more novel aspects of the
paper pertain to differences in mortality outcomes between the two studied medications, particularly during the first weeks following OST initiation.

I. There are only 3 studies for buprenorphine, which means the CIs are very wide. We are not sure we can draw such strong conclusions in favor of buprenorphine based on this level of evidence.

Following this comment and other reviewers’ considerations, the conclusions regarding buprenorphine have been moderated both in the Abstract and discussion:

Abstract:

BEFORE: There is consistent evidence that retention in OST reduces all-cause and overdose mortality. The induction phase onto MMT and immediate periods after leaving OST are associated with elevated risk. There may be clinical benefits to initiating OST using buprenorphine; thereafter there is little evidence of any difference in mortality risk by medication type. Retention in treatment will reduce patients’ potentially repeated exposure to periods of elevated risk, both after leaving treatment and during induction onto MMT.

AFTER: Findings suggest that retention in MMT reduces all-cause and overdose mortality, avoiding about two thirds of deaths expected without MMT. BMT probably is also effective in reducing mortality. The induction phase onto MMT and immediate periods after leaving MMT or BMT are associated with high mortality risk, which should be appropriately addressed by public health and drug treatment responsible. Possible advantages of BMT compared to MMT require further research, because BMT studies are scarce and comparisons are affected by potential confounding.

Discussion:

First, to make clearer about BMT, a new subsection has been added (Page 17, line 22):

Mortality risk during specific periods in and out of buprenorphine treatment
Data from three cohorts (two from Australia) suggest that BMT could reduce all-cause mortality, although the disparity in mortality rates comparing periods in and out of treatment (4.3 vs. 9.5/1,000 py) did not reach statistical significance. Findings also suggest that BMT patients had a significantly increased mortality in the first four weeks post-treatment cessation compared to the remaining time out of treatment (31.4 vs. 10.3/1,000 py), while during the treatment period there is no difference between the first four weeks and the remaining time. It is difficult to draw firm conclusions due to limited evidence from similar settings: 96% and 98% of included deaths and py, respectively, came from Australian studies (87% and 86% from the study of Kimber et al).

Second, subsection “Comparing effectiveness of buprenorphine and methadone substitution treatment” has been reformulated (Page 18, line 8):

BEFORE: In the cohorts that directly compared BMT and MMT, results suggested that mortality is much lower in BMT than MMT patients during (4.3 vs. 11.6 deaths/1000 py) and after treatment (9.5 vs. 38.1 deaths/py). There is a higher toxicological involvement of methadone than buprenorphine in drug-related deaths among people in BMT and MMT.\textsuperscript{37} 40 41 Such findings suggest that BMT could be more effective than MMT in reducing mortality especially
from overdose, perhaps because of the ceiling of respiratory depressant effects of buprenorphine, and there is a lower risk of cardiac arrhythmias. The lower risk of BMT compared to MMT patients after treatment cessation may be partially explained by confounding by indication whereby patients with fewer comorbid problem and perhaps less severe opioid dependence may be more likely to be initiated onto BMT. Given that two out of three studies of BMT patients were conducted in Australia (where mortality in MMT patients is also lower than elsewhere); that BMT patients were somewhat younger; and there is poorer retention in BMT, patient variables might explain the differences observed between BMT and MMT. We found few data that permitted direct examination of this possibility, although in one cohort the lower mortality in BMT patients remained after adjustment for some confounders.

AFTER: Other sources of data have suggested that BMT could be more effective than MMT in reducing mortality especially from overdose. Unlike methadone, there is a ceiling for respiratory depressant effects of buprenorphine as dose increases, and the probability of triggering arrhythmias is lower for buprenorphine than methadone. When the mortality between periods in and out of treatment was compared, first within MMT and then within BMT patients, a greater reduction in mortality (measured as the rate difference or rate ratio between periods) in MMT than BMT was observed. Whereas, when mortality between BMT and MMT was compared, first within the treatment period and then within off-treatment period, a significantly lower mortality in BMT than MMT was found in both periods. Our conclusions must remain tentative, however, until further studies in varied treatment settings and contexts are undertaken to examine this issue. BMT-MMT mortality differences may reflect confounding through differences in characteristics of patients (i.e. age, opioid dependence severity, injecting drug use, other drug use, comorbidities, prison history, overdose history, patient’s preference); characteristics of treatment (i.e. previous treatment, specialization of the doctor who controls the treatment, dose, provision characteristics, co-interventions, retention or drop-outs) or sociopolitical context in which studies have been conducted. For example, the initial prognosis may be better in BMT than MMT patients (i.e. fewer comorbid problems, less severe opioid dependence), though this was not clearly found in a recent US study.

Few details on patient or treatment characteristics were reported in articles included in this meta-analysis to permit a detailed examination of this potential issue (Table 1), so it was not possible to assess the possibility of confounding. However, a sensitivity simulation analysis in the study by Kimber et al suggested that the lower mortality in BMT than MMT patients during first four week of treatment was unlikely to be caused by unmeasured confounding.

J. Should not the conclusion be that we need a randomized trial comparing these two things? Retention in treatment is lower with buprenorphine, but methadone is trickier to use.

We agree we need further research. Taking into account that Buprenorphine is widely used in several countries, there are very few observational studies of these results. We have included these considerations in the conclusions of the abstract (see point I) “Possible advantages of BMT compared to MMT would require further research, because BMT studies are scarce and affected by potential confounding.”, and in the last paragraph of the discussion (Page 21, line 13):

Finally, given the inconclusive results of the comparative effectiveness of MMT and BMT in reducing mortality, clinicians should choose BMT or MMT depending on the availability and cost of these treatments, as well as the characteristics and patient preferences. New studies are needed to assess comparative effectiveness. It is unrealistic to expect that new RCTs would be conducted: they would be prohibitively expensive and difficult to implement. An alternative
pragmatic solution is to implement more cohort studies in multiple countries, including both BMT and MMT patients, with heterogeneous characteristics, and sufficient information on potential confounders. In addition these studies could focus on the relation between time in treatment and post-interruption risks in order to identify a minimum requirement for treatment duration.

First, please revise your paper to respond to all of the comments by the reviewers. Their reports are available at the end of this letter, below.

In your response please provide, point by point, your replies to the comments made by the reviewers and the editors, explaining how you have dealt with them in the paper.

Comments from Reviewers

Reviewer: 1

Recommendation:

Comments:

A. This is an important review and meta analysis of a key public health topic. It pulls the data together in a way that is likely to be highly impactful and provides a state of the art overview of the current state of knowledge on OST mortality, before, during and after treatment. It is very well written and I support its publication in its current form. It carries a key public health message and is likely to be highly cited in the future. It extends work that has previously been done on individual cohorts into an overview of all the key cohorts available.

Thank you very much for your kind words about our paper.

Additional Questions:

Please enter your name: Michael Farrell

Job Title: Director

Institution: NDARC University of New South Wales, Australia

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: Yes

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No
Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests'target='_new'>please see BMJ policy</A> please declare them here: I note at this stage that one of the Authors Professor Degenhardt is a member of staff in the Research Centre I direct and I have coauthored publications and hold grants with her.

Reviewer: 2

Recommendation:

Comments:

1. The paper “Mortality risk during and after opioid substitution treatment: systematic review and meta-analysis of cohort studies” is a large and up-to-date review of available observational longitudinal studies in the field of opioid substitution treatment (OST). The main stated aims were to compute and compare all-cause and overdose mortality in persons during and after OST and compare risk estimates between the two medications methadone and buprenorphine, specifically focusing on time trends immediately following treatment initiation and cessation. The paper has through systematic and well-described literature searches identified 18 eligible cohort studies published during 1974-2015; including 17 cohorts of Methadone treatment comprising more than 115,000 person and 3 cohorts involving buprenorphine comprising over 15,000 participants. Generally this is a well-written paper and a well performed project with strong methodology. The main findings, that the provision of OST to persons with opioid dependence reduces mortality and that longer term retention in treatment is required to achieve risk reduction is not new. Similarly that the main difference observed between in-and-out of treatment mortality rates is mainly explained by differences in opioid related overdose deaths and reductions in these deaths during treatment are established knowledge.

Thanks a lot

2. The more novel aspects of the paper pertain to differences in mortality outcomes particularly during the first weeks following OST initiation between the two studied medications. However these comparisons rest on only one or two cohorts including buprenorphine patients including very limited numbers of “cases” for the two outcomes all-cause mortality and overdose mortality (dominated by the Australian “Kimberly study”). Therefore the comparison presented in the analysis is in reality a comparison between pooled analysis and a more real meta-analysis for methadone versus a single study dominated by the Australian setting for buprenorphine. This is a limitation to the project in terms of ground breaking news value. The additional value of this systematic review to the results presented in the original buprenorphine studies is therefore somewhat limited.

Thank you for this suggestion. This particular aspect has been pointed out. We discuss this aspect extensively in response to the editor’s point “I”.
3. In terms of interpretation of the presented results my view is that the authors give differences in biological activity of the medications too high emphasis compared with the likely differences in characteristics between the medication groups and cohorts; likely resulting in selection/confounding by indication and region, as the authors also state in the limitations section.

Thank you for this suggestion. As we discussed in our response to previous point, in this new version we have extracted new information about different studies (see new Table 1) and we have compared results (when that was possible) between adjusted and unadjusted mortality rates. Besides, we have included a new paragraph in discussion (see answers to editors, point C).

4. Figure 5 (a In methadone treatment part) as an example derived as I understand it by “a forced” trend function to smooth the curve also seems to be heavily dominated by a single study (Degenhardt et al 2009) which is not in consistence with the other 3 studies included, regarding very high risk of mortality the first week following MMT induction.

Figure 5 has been updated since, after extending the search to September 2016, we identified a new study reporting mortality data by time interval in and out of methadone substitution treatment. The updated figure still shows a marked increase in risk during the first four weeks of methadone treatment. We agree with the reviewer that study-specific trends are quite heterogeneous. However, we accounted for this between-study heterogeneity by performing a random-effects meta-regression, which resulted in an average risk trend across studies (with study-specific trends below and above the pooled trend). Indeed, the pooled trend was essentially the same under alternative model specifications and it was consistent with the more robust categorical results (≤4 and >4 weeks) presented in Figure 4. Based on these new results, a new line has been added in line 3, page 14:

*In continuous time trend analysis (fig 5), there were significant departures from linearity in all-cause mortality risk trends over time in and out of methadone treatment (P=0.04).*

5. Overall I miss a more detailed discussion of differences in drug user characteristics between regions and between the included cohorts, as well as over time (i.e. from 1960s until 2015). The clinical practices in OST provision varies greatly across settings. It is likely that the clinical practice during induction of methadone is of paramount importance to outcomes, and that these differ largely between included cohorts. For example is OST initiation initiated following an inpatient and completed detoxification or introduced in outpatient settings directly from ongoing polydrug use within a population with unstable living condition, or is the treatment provided by addiction medicine specialists or General practitioners? Similarly will drug use characteristics, such as injection practice, dominating drug use combinations etc, hugely influence the risk for overdose, and hence also the potential for risk reduction experienced by the provision of OST.

The reviewer is quite right to be concerned about these aspects, but despite having made a great effort we will not be able to fully satisfy him. A new exhaustive search for information on the aspects mentioned by the reviewer (patient and OST characteristics) has been carried out for all the articles included in this meta-analysis. As a result, the information available on aspects such as psychiatric comorbidity, somatic comorbidity, non-opioid drug use, treatment providers, existence of inpatient induction, follow-up losses, first/subsequent treatment episode, and complete/incomplete treatment has been incorporated in tables in the body of the article (See table 1 ...) or web appendices (3 and
4), and some changes have been made in the methods and discussion sections. Virtually all studies reported information on treatment providers, percentage of patients with inpatient induction and percent of patients lost to follow-up. In addition, most studies reported mean doses of MMT or BMT (13 studies out of 21, all MMT). In contrast, relatively few studies reported information on the proportion of participants who were injectors (8 studies), had psychiatric comorbidity (7 studies, all MMT), had somatic comorbidity (7 studies) or used non-opioid drugs (7 studies, all MMT). Similarly, there were also very few studies reporting information on other treatment characteristics, such as first/subsequent treatment episode (4 studies, all MMT) or complete/incomplete treatment (3 studies, all MMT). On the other hand, there could be substantial heterogeneity in the definitions and level of quality of reported information. In the 4 MMT studies that included information on first and subsequent treatment episodes, a clear disparity was not found in all-cause and overdose mortality rate between both types of episodes (Appendix 3). However, in the 3 MMT studies that included information on incomplete or complete treatment, a greater all-cause and overdose mortality rate was found in participants who did not complete the treatment than in those who completed it (Appendix 4). These new appendixs (with information about these exhaustive search) have been pointed out in Page 7, line 21:

Finally, information on first and subsequent treatment episodes (Appendix 3) and on completeness of treatment (Appendix 4) was registered.

6. Another aspect which is not clear is what has been reported and coded in the different cohorts as first day/week in OST; is it starting with the first day of medication or first day of detox prior to OST medication, or starting at stable or adequate plasma levels/stabilization. I assume it represents first day of OMT medication, but this is not specified. Nevertheless I also assume two different medications with differing biological profiles, may yield different outcomes, but I do think there are many factors involved and that the observed large differences in outcomes as presented are not necessarily as much caused by the different medications as advocated for in this paper.

As suggested by the reviewer, in each selected cohort, follow-up started with the first registered day on maintenance treatment, excluding any previous detoxification episode. This information is explicitly stated in the revised manuscript (Page 7 lines 16-18):

BEFORE: follow-up length from the start of treatment loss to follow-up, and mortality outcomes.

AFTER: follow-up length from the start of maintenance treatment (excluding any previous detoxification period), loss to follow-up, and mortality outcomes.

Regarding other involved factors, we have discussed this aspect extensively in response to the editor’s point “I”.

7. Finally if a person is in OST, and then exits treatment, that person will have a high tolerance for opioids including for heroin the first days following treatment, due to gradual decrease of the plasma levels of the OST medication over the first days (as half-life of these medications are in the range 24-36 hours). Therefore it is not likely that high overdose rates immediately following termination are a result of low opioid tolerance but maybe rather as a result of a chaotic polydrug crisis which includes the involvement of OST medications and other licit and illicit substances around the time of OST termination.
The concern of the reviewer is legitimate. It is likely that the long half-life of methadone will contribute to cushioning the rise in mortality in the first 3-4 days after cessation of treatment, but not in the remainder of the first 4 weeks after cessation. Consequently, the last paragraph under the subtitle “Mortality risk during specific periods in and out of methadone treatment” has been rewritten to reflect these considerations, as well as to emphasize that the increased mortality may be due to the increased consumption of multiple licit and illicit substances during this period and to other factors (Page 17, line 12).

BEFORE: The mortality peak immediately after MMT cessation may be explained by tolerance loss to toxic effects of heroin, assuming that many patients return to use.\(^8\)\(^9\)\(^26\)\(^44\)\(^51\)\(^52\).

AFTER: Although the long half-life of methadone could contribute to cushion the increase in mortality in the first 3-4 days after MMT cessation, a large increase in mortality was found in first four weeks after treatment compared to treatment period. Such increased mortality may be explained by tolerance loss to toxic effects of heroin, assuming that many patients return to use,\(^6\)\(^8\)\(^32\)\(^47\)\(^55\)\(^56\) and probably also by the use of multiple licit and illicit substances during that period. Similar drug-related mortality peaks occur in other situations where opioid tolerance is probably diminished (e.g. after prison release or hospital discharge\(^57\)\(^58\)). Lifestyle factors, and comorbid issues such as mental health problems, may also increase mortality from other external causes, since suicide and injury deaths are also elevated during that period.\(^53\)

8. Overall I do think the authors have done a thorough job with the review, but I find the added knowledge to be limited by small numbers of buprenorphine cohorts included and by the fact that the cohorts included are rather diverse. These contextual differences between cohorts cannot be appropriately handled without more detailed inclusion of setting specific descriptive variables, and without these the presented outcomes may well lead readers to assume results are caused by differences in medications provided rather than from other reasons. The conclusions; that BMT is superior to MMT and that “BMT could be more effective” or that the results “confirm BMTs safety” etc, are not sufficiently grounded in evidence as of yet.

Again, thank you for this suggestion. In the new version of the paper, the conclusions in relationship with MMT and BMT, have been changed. This particular aspect has been pointed out. We discuss this aspect extensively in response to the editor’s point “I”.

9. As the authors state in their introduction; “a systematic review of cohort studies can provide valuable evidence...” that is correct, but that assumption rests on the fact that included cohort studies in a meta-analysis are sufficiently representative and homogenous as well as adequately plentiful, I am not convinced that these two criteria are fulfilled in the current review. Particularly the BMT estimates seem to be hampered by low statistical power. I agree with all the limitations presented by the authors and think they are important limitations. Without measures to resolve any of them this renders the review somewhat vulnerable and therefore as strong conclusions as presented cannot be inferred in my view.

In fact, representativeness and homogeneity are important limitations that had been pointed out in the sent version (“First, temporal-geographic differences in patients, OST
and contextual characteristics could have differentially affected mortality during and after OST. Studies were conducted in different high-income countries, with the follow-up often spread over many calendar-years (MMT 1965-2010, BMT 1990-2010) and an average length (years) highly variable (MMT range 1.3-15.8, BMT range 1.1-4.5)). After this review, we have included in Table 1 and analyzed (Table 3 and appendixes) new possible confounding factors in order to show differences among included cohorts. There are further details in answer “C” to editors. Regarding BMT estimates, following this and other comments, we have changed the previous “strong” conclusions. This point has been discussed in answer “I” to editors.

10. The first “what this study adds” point presented; is already established knowledge and the second point including “4 weeks as minimum duration for MMT” seems out of context and not established current practice. Much longer methadone treatments than 4 weeks are advised and required for opioid dependent persons for them to experience benefits, and most patients likely would require lifelong OST.

Reviewer is right. Following this comment, both sentences have been replaced (Page 22):

BEFORE:
Retention in opioid substitution treatment with methadone and/or with buprenorphine reduces all-cause and overdose mortality.

In patients provided with methadone the all-causes mortality reduces sharply after 4 weeks in treatment, suggesting this as a minimum period of treatment. Differently, in the patients treated with buprenorphine, the all-cause mortality remained stable during the induction and the remaining time on buprenorphine treatment.

AFTER:
Methadone maintenance treatment could prevent an average of 26 deaths/1000 person years in treatment. Mortality risk among opioid users while in such treatment seems to be less than a third of that expected in the absence of opioid substitution treatment.

Buprenorphine maintenance treatment probably it is also effective in reducing mortality in opioid users but quantification of averted deaths requires further studies.

The mortality risk in the first four weeks (induction) in methadone maintenance treatment is high but seems to decreases substantially during this period, with a further stabilization around 6 deaths /1,000 person-years in the remaining time in treatment. Such increased mortality during induction is not observed during buprenorphine induction.

The mortality risk in the four weeks immediately following the cessation of methadone or buprenorphine treatment is very high, and may exceed 30 deaths/1,000 person-years.

Additional Questions:
Please enter your name: Thomas Clausen

Job Title: Professor

Institution: Norwegian Centre for Addiction Research, University of Oslo, Norway

Reimbursement for attending a symposium?: No
A fee for speaking?: No
A fee for organising education?: No
Funds for research?: No
Funds for a member of staff?: No
Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <a href='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests' target='_new'>please see BMJ policy</a> please declare them here: No conflicts of interest to report.

Reviewer: 3

Recommendation:

Comments:
The article 'Mortality risk during and after opioid substitution treatment: systematic review and meta-analysis of cohort studies' by Sordo et al. addresses an important topic in the treatment of opioid dependence, i.e. the mortality of patients in and out of treatment and across treatment episodes. This is of great interest for both clinicians and policy makers since results are suited to have an impact on both individual treatment decisions and strategic development of OST services from a public health perspective. To my knowledge the presented work is the first meta-analysis on this topic.

The methods are sound and, mostly, well described.

I do have some queries, however, regarding the presentation of the results, which I will detail below.

1. The general readability should be improved, especially in the discussion section, where the writing seems to reflect more a statistician's than a clinical or public health perspective. For example, sentence on p. 17: “The greater MMT relative effect size (as measured by rate ratio) for overdose than non-overdose deaths, is consistent with previous findings.” In my opinion, this kind of presentation of one of the main results is not very illustrative to the general reader.

Following this and other comments, we have improved the discussion’s readability, especially the first subsection. The sentence that this reviewer points out as confuse, has been erased (changes are shown in the next answer)
2. Moreover, the number of deaths per person years might not be the best measure to illustrate the impact of OST (MMT and BMT) on mortality for the general reader in the first sentence of the discussion section. Maybe the rate ratios as reported in the results section would be better suited here.

The sentence including quantification of the effect in absolute terms (rate difference), has been maintained, but other referring to such quantification in relative terms (rate ratio) has been added (“The in-treatment rate, was fewer than a third of the expected rate out of MMT”), all translated into a language accessible to clinicians. Discussion, Main findings:

BEFORE: Our review suggests that MMT can prevent an average of 26 deaths/1000 person years (py) in treatment (95% CI 15-38), with largest impacts in reduction of overdose deaths. The all-cause mortality risk during MMT is much higher the first four weeks (with a steep decrease across this period from 45 deaths/1000 py at first week to 8/1000 py at fourth week) than in the remainder of treatment. The greater MMT relative effect size (as measured by rate ratio) for overdose than non-overdose deaths, is consistent with previous findings.6 9 24 36 This disparity suggests the reduction in mortality during MMT is largely mediated by reductions in heroin and other illicit opioid use.24 26 Mortality risk is also higher in the first four weeks after MMT cessation (range 25-37 deaths/1000 py) than afterward. Based on three cohorts (two from Australia), findings also suggest that Buprenorphine substitution treatment (BMT) reduces mortality, with a similar risk during all BMT (=4 deaths/1000 py) confirming its safety,37-42 and a post-BMT risk higher in the first four weeks than afterward (31 vs. 10 deaths/1000 py) as with MMT.

AFTER: Our review suggests that MMT prevents an average of 26 deaths/1000 person years (py) (95% CI 15-38). The in-treatment rate, was fewer than a third of the expected rate out of MMT, with the greatest impact upon overdose deaths. The all-cause mortality risk during MMT is much higher in the first four weeks (with a steep decrease from 45 deaths/1000 py at first week to 8/1000 py at fourth week) than in the remainder of treatment. Mortality risk is also higher in the first four weeks after MMT cessation (range 25-37 deaths/1000 py) than in the remainder time out of MMT. Based on three cohorts (two from Australia), findings also suggest that BMT could reduce mortality, with a similar risk during all BMT period (=4 deaths/1000 py) confirming its safety,39-44 and an out-of BMT risk higher in the first four weeks than in the remainder time (31 vs. 10 deaths/1000 py).

3. The implications of the meta-analytic findings for public health service development and also for treatment goals (like abstinence or unrestricted access to substitution therapies) for people with opioid dependence should be discussed in more detail.

Implications of our findings for public health have been improved following this and other comments. In this sense, we have added a new sentence in the subsection “implications” (Page 20, line 15):

...policymakers, clinicians and those responsible for drug treatment systems should work to ensure the availability of OST, remove access barriers and promoting OST engagement. This means the strategic development of OST services from a public health perspective, which may be essential to reduce the social harm associated with opioid use.
4. With regards to the comparison of MMT and BMT I am concerned about the small number of BMT studies included in the 'meta-analysis' (only 3 BMT studies included, sometimes only one BMT study was used for sub-analysis). This limitation should be made more explicit in the manuscript. In addition, with regards to this and other limitations as outlined below I recommend to be cautious with respect to the conclusions and the clinical recommendations related to the comparison of MMT/BMT as, for example, prominently stated in the abstract (e.g. the recommendation of buprenorphine for induction phase) in the context of this meta-analysis, even if they might make sense from a clinical perspective and are suggested elsewhere. Additionally, the authors should address in one or two sentences potential difficulties when changing from BMT to MMT (and also vice versa). Moreover, BMT/MMT ratios seem to show great heterogeneity across countries (probably dependent on treatment systems). Transferability of BMT results mainly from Australia to other parts of the world should be discussed (2 of 3 BMT studies stem from Australia). In addition two other factors (comorbidity of study populations different for MMT/BMT, MMT studies are older with less stringent study protocols; as both stated by the authors) seem to hamper a sound comparison of BMT and MMT.

These particular aspects have been already pointed out. We widely discuss this aspect in the response to editors, section “I”

5. In general, the authors describe a very large heterogeneity across studies (p. 12). How problematic is this in the context of the present meta-analysis? Implications for the findings should be discussed.

In any meta-analysis, a large between-study heterogeneity results in wider confidence intervals for the pooled effect, which should be interpreted as an average measure of the underlying effects that vary randomly across studies, rather than a fixed effect common to all studies. Nevertheless, in the present meta-analysis, the strong heterogeneity refers to mortality rates, but rate ratios comparing periods in and out of treatment are fairly more homogeneous across studies.

6. The authors report no evidence of publication bias (p. 12). How was this evaluated?

This bias was evaluated as can be seen in the Methods section, subsection “Statistical analysis”:

BEFORE: “Publication and other related biases were assessed by using the extended Egger test allowing for heterogeneity.”

AFTER: Publication bias and genuine small-study effects were assessed by using the extended Egger test allowing for heterogeneity.

7. The authors’ attempt to evaluate the differential impact of distinct treatment phases of OST on mortality (e.g. induction phase) in this meta-analysis is of great value for the field. In addition, however, it would be interesting to know more about the impact of the situation before entering OST on mortality, especially in the induction phase. As many patients cycle in and out of treatment: Is there e.g. a higher mortality risk for people who are entering OST for the first time (compared to those with previous OST episodes, see also p. 16, first paragraph). If this information is available it would be important to integrate this aspect in the analysis.
Everything related to patient characteristics and treatment (including first / previous treatment) is mentioned in the discussion section as aspects that can influence the differences in mortality between MMT and BMT. In addition, in the same section it is noted that these characteristics are infrequently reported in the articles included in the meta-analysis and could have differentially affected mortality during and after OST. Specifically, as mentioned above, the possible existence of differences in these characteristics may partly explain differences in mortality between MMT and BMT. Furthermore, we have checked all articles in order to compare OST for the first time and subsequent episodes. Only four articles gave this information, and we have decided not to include it in the text. Nevertheless, the information will be available in Appendix 3.

Minor issues:

8. p. 8: The authors assume a Poisson distribution of their data. Why they do not take a negative binomial distribution into account that might be better suited than Poisson, which might represent an oversimplification of data structure (a topic often termed as overdispersion)?

For each selected cohort, we used exact methods (rather than large-sample approximations) to compute 95% confidence intervals for the mortality rates. These methods are based on an exact Poisson distribution for the observed number of deaths, which is an appropriate model for non-contagious events occurring in a stationary population over a fixed follow-up period and can be used for any arbitrarily low number of events.

9. p. 10: The authors report a rating of study quality on a 16-point scale, with a median of 7.5. What does this mean/imply? A qualification in words would be helpful for general readers.

Indeed, the majority of the studies had a moderate quality. Thus, of the 18 MMT studies, 2 had a low quality (0-5 points in Quality Score), 11 moderate quality (5-11 points), and 5 high quality (12-16 points), while of the three BMT studies, one had low quality, other medium quality and other high quality. The results on studies quality have been rewritten in a more understandable form for the readers. The inclusion of a new study has changed the median (Page 11, line 21).

BEFORE: The quality of reviewed studies ranged from 3 to 13 points on a 16-point scale, with a median of 7.5 points

AFTER: The majority of the studies had a moderate quality. They ranged from 3 to 13 points on a 16-point scale, with a median of 8 points.

Also in a footnote to Appendix 2: Quality Score, an explanation of the classification of studies based on their total score has been included.

10. p. 15: The authors report that MMT can prevent an average of 26 deaths per 1000 person years as outcome measure (which is simply the difference between number in and out of treatment). This measure should also be listed e.g. in Figure 2.
We partially agree with this comment. Unfortunately, we have 3 tables (one of them in two parts), 6 figures and 4 additional appendices, so we had to avoid certain repetition among tables and results.

11. p. 16, line 7: When the authors refer to studies investigating effects of cycling in and out of treatment they should also include a recent paper on this topic by Nordt et al. (2015).

Thanks. This article has been included in the new version (Page 17, line 1): “This is important, because some patients cycle in and out of OST,”

12. p. 16, line 39: This statement lacks a reference.

There is no line 39. We have assumed the reviewer meant this statement (where a reference has been added):

Lifestyle factors, and comorbid issues such as mental health problems, may also increase mortality from other external causes, since suicide and injury deaths are also elevated during this period.


13. Methods: In the flowchart illustrating study selection (Fig. 1) there seems to be a inconsistency for the number of total articles excluded (291) and the subsequent list for reasons of exclusions. This should be clarified.

Indeed, in the previous version there was a typo in figure 1. As a consequence of updating the search, figure 1 has changed (the review is now updated to September 2016). We have carefully checked possible inconsistencies.

14. In the section “What this paper adds” the advance from “Opioid substitution treatment has been shown to be safe and effective in suppressing illicit opioid use and reducing all-cause and overdose mortality” (already known) and “Retention in opioid substitution treatment with methadone and/or with buprenorphine reduces all-cause and overdose mortality” (new) is not clear to me. The authors should be more precise here.

The key messages have been extensively rewritten. We hope that the most important messages are now much clearer (See answer 10 to reviewer 2).

Additional Questions:
Please enter your name: Marcus Herdener

Job Title: MD

Institution: Center for Addictive Disorders, Psychiatric Hospital, University of Zurich

Reimbursement for attending a symposium?: No
A fee for speaking?: Yes

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: Yes

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <a href='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests' target='_new'>please see BMJ policy</a> please declare them here: I was on Advisory Boards and received a speaker fee from Lundbeck.

**Reviewer: 4**

**Recommendation:**

**Comments:**

This is a timely meta-analysis that tries to answer a highly relevant question in the field of opioid use disorder. OUD patients who are maintained on medication-assisted treatment with opioid substitution agents (opioid agonists) have better outcomes including mortality. The two widely available opioid substitution agents are the full opioid agonist methadone and the partial agonist buprenorphine. Overall clinical trial data suggests that methadone maintenance is associated with higher levels of treatment adherence (that is possibly associated with better survival and outcomes) than buprenorphine. However, use of opioid substitution therapy is inherently associated with some mortality risks by itself that varies by time of treatment (early vs. later period), treatment retention and the choice of OST agent (buprenorphine versus methadone). It has been argued that methadone maintenance treatment despite it higher drug toxicity could result in in lower mortality risks due to far higher retention compared to buprenorphine maintenance treatment that is associated with lower drug toxicity, but lower treatment retention. This meta-analysis tries to quantify the mortality risks associated with OST, and then better define the variation of mortality risks based on time of treatment, treatment retention and nature of the OST agent used.

The meta-analysis is appropriately exhaustive, strong in methodology, and patient centered in its overall approach. I commend the author in this excellent piece of work. The findings of this article have significant practice and policy implications.

Their main finding were the following

1. There is a palpable level of mortality associated with OST, and this varied substantially by the time period, treatment retention and the type of OST.
2. All-cause and overdose mortality were substantially lower among patients who were in treatment compared to those who were not.
3. All-cause and overdose mortality were substantially higher during the first 4 weeks of treatment compared to later periods.
4. Buprenorphine maintenance treatment appears to be associated with a substantially lower level of mortality compared methadone maintenance treatment.
5. Patients on Buprenorphine maintenance treatment did not seem to have the excess mortality during the initial 4 weeks as in methadone maintenance treatment. These findings have to be tempered with all the methodological limitations well described by the authors. I find these results compelling and will substantially advance the knowledge and inform the practice in this area.

Thanks a lot. Some of these comments have been quite useful in the writing of this new version

Comments:
I have no major comments, but some points to enhance the manuscript.

Introduction: Page 4, para 2:

1. “However, growing evidence suggests that mortality experience during and after opioid substitution treatment is time-varying and differs by medication type. Methadone, a full opioid agonist, might pose an excess risk of overdose death during treatment induction if initial doses are too high or coexist with illicit opioid use, whereas buprenorphine, a partial agonist, is less effective in retaining patients in treatment and might lead to an increased risk of death after after premature treatment cessation.” This statement is scientifically appropriate, but may be impenetrable to treaters and patients who access this important literature. Please try to make it more of a clinical language (an imperfect form in is provided in my introductory comments).

It is true, may be the sentence is not clear. We have changed the sentence (Page 4, line 13):

BEFORE: Methadone, a full opioid agonist, might pose an excess risk of overdose death during treatment induction if initial doses are too high or coexist with illicit opioid use, whereas buprenorphine, a partial agonist, is less effective in retaining patients in treatment and might lead to an increased risk of death after premature treatment cessation.9,11

AFTER: Methadone, a full opioid agonist, might pose an excess risk of overdose death during treatment induction if initial doses are too high or coexist with illicit opioid use, whereas buprenorphine, a partial agonist, is less effective in retaining patients in treatment and might lead to an increased risk of death after premature treatment cessation.9,11

The introduction states the problem and the purpose of analysis clearly.

2. Many parts of the world use the term “opioid agonist treatment” instead of “opioid substitution treatment”. An acknowledgement of this in the introduction would be helpful.

The term “opioid agonist treatment” is used only twice in the text and, in both instances, to explain the effects of Methadone or Buprenorphine. It is not used instead of OST. In fact, We think it can be problematic to use the term "opioid agonist treatment" because buprenorphine
is a partial agonist, which may have both agonist and antagonist effects. This is why we have chosen to always use "opiod substitution treatment".

The methodology is excellent and described in detail. The results are well laid out and supported by figures.

Discussion:

3. Page 18, para 1: Main findings. This section has some problems. As a clinician, I see the results as following:

1. There is a palpable level of mortality associated with OST, and this varied substantially by the time period, treatment retention and the type of OST.

2. All-cause and overdose mortality were substantially lower among patients who were in treatment compared to those who were not.

3. All-cause and overdose mortality were substantially higher during the first 4 weeks of treatment compared to later periods.

4. Buprenorphine maintenance treatment appears to be associated with a substantially lower level of mortality compared methadone maintenance treatment.

5. Patients on Buprenorphine maintenance treatment did not seem to have the excess mortality during the initial 4 weeks as in methadone maintenance treatment.

The authors appear to be hesitant in stating that buprenorphine maintenance is associated with lower levels of mortality than methadone maintenance, a finding that is coming across clearly to readers. If authors are reluctant to state this or are skeptical about these findings, they should state the reason. This is a key information for users of these findings. They discuss this in a limited way in a later paragraph down the discussion (page 16 last para).

The main findings section has been partially rewritten and we believe that most of the concerns of the reviewer have been considered (See answer 2 to Reviewer 3). However, the conclusion referred to a lower mortality in BMT than MMT patients not been selected as one of the important contributions of this meta-analysis, because the reasoned objections raised by other reviewers and the reasons explained in discussion section.

4. Page 15, para 1:

"The greater MMT relative effect size (as measured by rate ratio) for overdose than non-overdose deaths, is consistent with previous findings. This disparity suggests the reduction in mortality during MMT is largely mediated by reductions in heroin and other illicit opioid use. This is bit of a stretch and maybe to speculative. Anyways, this do not belong in the main findings, may be as a part of the later discussion. It is well known that the characterization of deaths as overdose deaths has its own methodological problems, and mortality is mortality in this young population no matter the mode is.

The reviewer's remarks have been considered. This sentence has been removed in the new version, and the main findings section has been partially rewritten (See answer 2 to Reviewer 3).

5. Page 17 para 2: Authors speculate that confounding by indication and patient variables may explain the lower mortality in BMT. Please note that a recent large observational study from US veterans have shown that the psychiatric and medical comorbidity, psychotropic medication use and service use characteristics does not differ much
between those being managed with BMT and MMT, bit there were significant demographic differences.1 It may be useful to include this in the discussion.

Thanks for this comment. It is true that this article can improve a little bit our discussion. Therefore, a comment and the reference (Manhapra et al 2016) have been included in discussion section (Page 18, line 25):

“For example, the initial prognosis may be better in BMT than MMT patients (i.e. fewer comorbid problems, less severe opioid dependence), 6 8 10 61 though this was not clearly found in a recent US study 62.

6. Implications for policy and practice: (page 18) This section needs to be more robust as this study carries significant policy and practice implications. I see the following implications not being discussed adequately.

a. Maintenance of treatment engagement appears to be effective in reducing mortality. So more efforts to increase treatment engagement appears to be needed.

b. Excess mortality soon after treatment induction and treatment discontinuation appears to be striking. So patient education, supportive care, and program elements should focus more on these vulnerable times of OST.

c. BMT appears superior to MMT is mortality benefits. Authors have discussed the need for research, but what about action?

Thanks a lot for these considerations. Reviewer’s suggestions have been incorporated into the section “Implications for policy and practice” (Page XX)

BEFORE: The number of deaths attributable to opioid use remains substantial in many countries, 62 63, so there is a need for preventive interventions. Given OST’s effectiveness in reducing mortality among opioid dependent people, such interventions should be expanded where coverage remains insufficient. Existing evidence suggests that coverage is low in many countries worldwide. 64 Some precautions during and after OST implementation should be considered to increase OST effectiveness. First, careful clinical assessment of opioid tolerance before OST onset, in order to fix the initial dose appear warranted. Second, monitoring during the induction period, especially for MMT, adjusting opioid doses, addressing mental and somatic problems and provision of information to patients about risk of overdose, especially when combined with other respiratory depressants 65 is warranted. Also, buprenorphine induction followed by transition to methadone might be considered. 11 66 Retention in OST reduces exposure to post- cessation mortality risk, and also to re-exposure to mortality risk during induction onto MMT, so efforts to improve retention are important as a strategy to reduce mortality.

Finally, given suggestive findings of a lower mortality risk in BMT patients, there is a need for new research, which considers important patient confounding variables to examine differences in MMT versus BMT effectiveness in reducing mortality risk.

AFTER: The number of deaths attributable to opioid use remains substantial in many countries, 65 66, while these are preventable causes of death warranting wide implementation of preventive interventions. Given effectiveness of MMT and probably BMT in reducing mortality among opioid dependent people, and that OST coverage is low in many countries worldwide, 67 policymakers, clinicians and those responsible for drug treatment systems should work to ensure the availability of OST, remove access barriers and promoting OST engagement. This means the strategic development of OST services from a public health perspective, which may be essential to reduce the social harm associated with opioid use.
Some precautions during and after OST implementation should be considered to increase OST safety. First, careful clinical assessment of opioid tolerance before OST onset to establish a safe induction dose appears warranted. Second, monitoring during the induction period, especially for MMT, adjusting opioid doses, monitoring mental and somatic problems and preventing opioid medication diversion. In addition, education of patients about overdose risk including naloxone take-home especially when OST could be combined with other respiratory depressants, is warranted. Also, buprenorphine induction followed by transition to methadone might be considered. Retention in OST reduces exposure to post-cessation mortality risk, and also to re-exposure to mortality risk during induction onto MMT, so efforts to improve retention are important as a strategy to reduce mortality. In any case, the large excess mortality during the first weeks immediately following the cessation of MMT and BMT treatment should be addressed decisively by public health authorities through drug treatment systems and harm reduction measures. More generally, establishing mechanisms for information and coordination between healthcare, social and legal services, and patient counselling while in treatment in addition to more specific overdose prevention programmes such as naloxone distribution should be considered.

Finally, given the inconclusive results of the comparative effectiveness of MMT and BMT in reducing mortality, clinicians should choose BMT or MMT depending on the availability and cost of these treatments, as well as the characteristics and patient preferences. New studies are needed to assess comparative effectiveness. It is unrealistic to expect that new RCTs would be conducted: they would be prohibitively expensive and difficult to implement. An alternative pragmatic solution is to implement more cohort studies in multiple countries, including both BMT and MMT patients, with heterogeneous characteristics, and sufficient information on potential confounders. In addition, these studies could focus on the relation between time in treatment and post-interruption risks in order to identify a minimum requirement for treatment duration.

7. The what is already known and what this study add sections should reflect the lower risk of mortality with BMT and MMT in a cleaner language. The current one is a bit confusing. The abstract should also use clear language around the lower mortality with BMT compared to MMT, both in results and conclusions.

As stated above, the conclusion referred to a lower mortality in BMT than MMT patients has not been selected as one of the important contributions of this meta-analysis, because the reasoned objections raised by other reviewers and the reasons explained in discussion section.

This point has been extensively discussed in answer to editors “I”


Additional Questions:
Please enter your name: Ajay Manhapra

Job Title: Lecturer

Institution: Department of Psychiatry, Yale School of medicine, New Haven, CT, USA

Reimbursement for attending a symposium?: No
A fee for speaking?: No
A fee for organising education?: No
Funds for research?: No
Funds for a member of staff?: No
Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests please see BMJ policy please declare them here:

Reviewer: 5

Recommendation:

Comments:
This is an interested project, and thank you for the opportunity to review it. I must admit to finding it quite hard to follow at times, because of the different comparisons and periods being made, but clearly there is a wealth of information being combined and summarised here on an important topic. I have reviewed this from a statistical perspective, and have some important comments and queries for the authors to deal with.

1) My main concern is about potential confounding. I am worried that there is no adjustment for confounders within each study, and thus none in the meta-analysis. In particular:
(i) in versus out comparison: Can we conclude anything useful if there are no confounder adjustments here with each study? Are the in and out comparisons on exactly the same people? Is there not the potential for confounding by indication, such that the ‘out’ period and ‘in’ period are not directly comparable (influenced by clinical decision making about apparent risk / health)?

Regarding comparisons between in and out of treatment or between subperiods in and out of treatment, patients are compared to themselves, which reduces the likelihood that the results are affected by measured or unmeasured confounders. However, there is little information on losses to follow and this could have affected comparisons.

(ii) MMT versus BMT: Can we compare these treatments fairly? Are there any within-cohort comparisons, or are they comparing results from different studies? If the latter, there is the large potential for study-level confounding. Statements such as “In contrast, all-cause mortality remained stable during induction and remaining time on BMT” infer a comparison
between BMT and MMT, which I do not think is fair if there is no within-cohort comparison and adjustment for confounders (and confounding by indication of which to receive).

The sentences starting with "In contrast ..." and “Differently...” have been rewritten or removed.

Further, there are some strong statements such as: “All-cause mortality was significantly lower in the three buprenorphine cohorts compared with all methadone cohorts (P=0.03) or the five methadone cohorts with more recent follow-up (P=0.045; table 2).” And “Differently, in the patients treated with buprenorphine, the all-cause mortality remained stable during the induction and the remaining time on buprenorphine treatment.”

I am very worried that such statements are too strong, given the potential for confounding by indication and that they are only making across, rather than within, cohorts (i.e. indirect comparisons). I do not see the lack of direct evidence raised anywhere as a limitation.

The findings referred to the BMT-MMT comparison, and are not selected in the new version as relevant findings of the study. The limitations of the BMT-MMT comparison are widely considered in the discussion section. Even among BMT patients, the limitations of the finding of lower mortality during the treatment period compared to the period after treatment are discussed, since the differences did not reach statistical significance, and the findings are based only on 3 cohorts with very little heterogeneity.

I see a limitation is noted in the discussion about potential confounding, but there is nothing mentioned in the abstract! Further, given the lack of confounding adjustments (as far as I can tell), I am not convinced about the clinical impact this paper would have, and thus whether it should be a high priority for the BMJ. But I leave that to the BMJ clinical editors and perhaps the authors to make a case for.

The limitation of BMT-MMT comparison has been included in Abstract

(iii) language is causal: such as “Our review suggests that MMT can prevent an average of 26 deaths/1000 person years (py) in treatment (95% CI 15-38), with largest impacts in reduction of overdose deaths.”

Given the concerns of confounding, can we really make statements such as these? Are the rate differences comparing the same people, in the same circumstances, but just in or out of treatment? Or could there be many reasons why the ‘out’ period is higher, not just due to the lack of MMT treatment?

The causal language has been avoided. Now the conclusions beginning with "Findings suggest ..."

2) Given the above, should the main message be that this overview is motivation for further research (e.g. in a trial to compare MMT and BMT) to identify which is preferred using direct evidence?

We think that several relevant messages emerge from this meta-analysis, and these are included in the Key Points and Abstract. The message mentioned by the reviewer is one of several. Perhaps in the first version we did not explain these messages clearly enough.

3) Please be consistent in using the abbreviations MMT and BMT, as sometimes the full words are still used.
Thanks, it has been done

4) It is great to see multivariate meta-analysis methods being used to synthesise the rates in and out of treatment, jointly, to account for their correlation. However, more details are needed. Such models must account for both within-cohort and between-cohort correlation. The within-cohort correlation is due to the same people (I assume?) providing data toward the in and the out periods. Whereas the between correlation is due to the correlation across studies in the true in and out rates (as one goes up, the other goes up). I cannot see how the within-study correlation can be obtained without individual patient data. Can the authors clarify?[1, 2]

For each selected cohort, the outcomes included (all-cause or overdose) mortality rates over different time intervals (periods in and out of opioid substitution treatment, or distinct subintervals in and out of treatment). Thus, the estimated mortality rates over these non-overlapping time intervals were assumed to be independent (they included distinct deaths in their numerators) and the within-cohort correlations were set to zero. This information was missing in the original manuscript and is now included in the revised version (changes in next answer).

5) Also, in my experience the between-study correlation is often poorly estimated at -1 or +1. Can the authors also give us details of the between-study correlation (as well as variance) estimates? [3]

Unstructured between-cohort covariance matrices were used in all multivariate random-effects meta-analyses. Although the estimated between-cohort correlations in the true rates in and out of treatment were high (0.94 for all-cause mortality and 0.85 for overdose mortality), there were no estimation problems in any of the fitted multivariate meta-analytic models regarding estimated between-cohort correlations at the boundary of the parameter space of -1 or 1. We have added this information to the Methods section of the revised manuscript (page 10, lines 8−21):

BEFORE: All multivariate random-effects meta-analytic models, were fitted through maximum likelihood methods using the R package mvmeta (R Foundation for Statistical Computing).

AFTER: In all the above multivariate random-effects meta-analytic models, the within-cohort correlations in the estimated mortality rates were assumed to be zero, since they were obtained over different follow-up periods in and out of treatment. All models were fitted through maximum likelihood methods with unstructured between-cohort covariance matrix using the R package mvmeta (R Foundation for Statistical Computing). The estimated between-cohort correlations in the underlying rates in and out of treatment were 0.94 for all-cause mortality and 0.85 for overdose mortality, with no estimation problems in any model regarding between-cohort correlations at the boundary of the parameter space of 1. The choice of maximum likelihood methods resulted in slightly underestimated standard errors of pooled rates (5.2 to 6.2% for all-cause mortality and 4.7 to 9.7% for overdose mortality) compared with restricted maximum likelihood, but allowed the comparison of nested models through likelihood ratio tests, which have been shown to perform better than Wald tests in multivariate meta-analysis. [4]

6) Please be extra cautious when interpreting the meta-regression results, as these are prone to low power and ecological bias / study-level confounding when examining mean patient-level covariates. [4-6]
We agree with the reviewer that there is a large potential for study-level confounding in meta-regression models assessing aggregate measures (means or percentages) of patient-level characteristics. Following reviewers’ suggestions, in addition to the original meta-predictors (study location, opioid injection, percentage of men, mean age, average methadone dose, and calendar follow-up period), we have also evaluated between-study heterogeneity in all-cause and overdose mortality rates by treatment induction method, treatment provider, and percentage loss to follow-up. Of all these covariates and outcomes, we only found significantly higher all-cause mortality rates in cohorts exclusively enrolling opioid injectors and those conducted in Europe/North America, as well as higher overdose mortality rates in cohort of patients in specialist services (see updated table 2). These three covariates are global (not aggregate) measures shared by all cohort participants. Nevertheless, we recognise that these associations are still prone to bias, since studies are hardly homogeneous or comparable with respect to other study-level or patient-level confounders. Thus, although meta-regression results are mentioned in the Results section to highlight the strong heterogeneity in cohort characteristics, they are merely exploratory and not addressed in the Discussion section.

7) Should the multivariate extension to I-squared be referenced as Jackson et al. 2012, and not Higgins? [7]

The reference by Higgins et al, which refers to the original univariate I-squared statistic, has been changed by Jackson et al 2012, which deals with the multivariate extension of the I-squared statistic. We have also added a new reference by Jackson et al 2011 with a general review of multivariate meta-analysis (line 21, page 8).

8) The authors say they use ML estimation methods for the random effects meta-analysis models. Please be clear that REML was used (I hope), as ML is downwardly biased.

As properly indicated by the reviewer, maximum likelihood (ML) methods tend to produce downwardly biased estimates of the between-cohort (co)variance components, which may result in artificially low standard errors of pooled estimates. However, even with the relatively limited number of studies included in this meta-analyses (19 for all-cause mortality and 12 for overdose mortality), the standard errors of pooled estimates based on ML were just slightly lower (5.2–6.2% for all-cause mortality and 4.7–9.7% for overdose mortality) than those based on restricted maximum likelihood (REML), with little impact on point and interval estimates for pooled mortality rates, rate differences, and rate ratios.

On the other hand, likelihood ratio tests (used in this meta-analysis to compare pooled mortality rates between methadone and buprenorphine cohorts, to explore heterogeneity of pooled rates by methadone cohort characteristics, to test for non-linear pooled risk trends over time in and out of methadone treatment, and to evaluate publication bias and small-study effects) can only be performed in ML random-effects meta-analytic models, but not in REML models, since the latter require comparing models with identical fixed-effects structures. Thus, in REML models, one is forced to use Wald tests, which are more sensitive to model misspecifications and outliers than likelihood ratio tests and have been shown to perform poorly in multivariate meta-analytic models (see reference 18 of the manuscript).

In summary, we opted to use ML estimation methods due to the slight underestimation in standard errors of pooled estimates and the ability to perform robust likelihood ratio tests. These arguments are explicitly stated in the revised manuscript (see answer 5).
9) There is growing calls in meta-analysis to account for the uncertainty in heterogeneity estimates when deriving 95% CIs for the pooled effects. Did the authors do this and, if not, would it have an impact? See for example: [8-11]

We did not account for the uncertainty in the estimated between-study covariance matrix when making inferences about the pooled effects, as this requires computationally intensive procedures that have not yet been evaluated/implemented in the multivariate setting. Nevertheless, approximate methods (which apply a scaling factor to the standard errors of pooled effects) have recently been proposed for multivariate meta-analysis. Based on simulation results (Figure 3 in Jackson et al 2014), little improvement can be expected from this approximate method in our bivariate meta-analysis based on 19/12 studies with complete outcome data.

10) Page 11: “All-cause mortality varied widely …” change to “All-cause mortality RATES varied …”

The word “rates” has been included in that sentence (line 1, page 12).

11) It is worth noting that small study effects may be due to genuine heterogeneity, rather than say publication bias.

As stated in the original manuscript, we found some evidence of small-study effects on all-cause mortality, with higher rates in small cohorts enrolling high-risk HIV-positive opioid injectors. This is likely due to genuine heterogeneity, rather than publication bias. In the Methods section of the revised manuscript, we have reworded the sentence on the extended Egger test to highlight that it allows detecting either publication bias or genuine small-study effects (line 8, page 10):

BEFORE: Publication and other related biases were assessed by using the extended Egger test allowing for heterogeneity. 20, 21

AFTER: Publication bias and genuine small-study effects were assessed by using the extended Egger test allowing for heterogeneity. 22, 23

So, in summary there is a lot to clarify and address at this stage, especially in regard confounding and thus relevance/reliability for the BMJ reader. Nevertheless, the authors have clearly worked hard on a big, complex project, and so I do hope my comments are useful to the them, and help them to refine their work going forward.

Thanks a lot.

With best wishes, Richard Riley

Reference List

Additional Questions:
Please enter your name: Richard Riley
Job Title: Professor of Biostatistics
Institution: Keele University
Reimbursement for attending a symposium?: No
A fee for speaking?: No
A fee for organising education?: No
Funds for research?: No
Funds for a member of staff?: No
Fees for consulting?: No
Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No
Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests'target='_new'>please declare them here:</A>