

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Association between alcohol consumption and impaired work performance (presenteeism): A systematic review
AUTHORS	Thørrisen, Mikkel; Bonsaksen, Tore; Hashemi, Neda; Kjekken, Ingvild; Van Mechelen, Willem; Aas, Randi

VERSION 1 – REVIEW

REVIEWER	Ryan Martin East Carolina University Greenville, North Carolina United States of America
REVIEW RETURNED	31-Jan-2019

GENERAL COMMENTS	<p>The manuscript was well-written and easy to follow. The review and description of the review, including the methods, results and discussion, was thorough. In the Quality Assessment section (page 10), it would be helpful if the authors:</p> <ol style="list-style-type: none">1. Explained the sample size thresholds utilized (low, moderate, high - were they arbitrary or based on some standard).2. Explained what they meant by adjusted/unadjusted (I understood it after reviewing Supplementary File 2, but was not sure what they were referring to in the MS).
-------------------------	--

REVIEWER	Robert Heirene University of South Wales Treforest Campus Cardiff United Kingdom
REVIEW RETURNED	05-Feb-2019

GENERAL COMMENTS	<p>Introduction</p> <p>Overall comment: The authors have integrated and discussed a range of relevant research here and build a clear rationale for their review. All of the key concepts appear to be well-defined. They also contextualise the review, noting previous review in this area and its limitations. Finally, the aim of the review is clearly stated.</p> <p>Page 4, Lines 38 to 40: Can the authors provide some indication of the levels of alcohol consumption needed to achieve these blood alcohol content levels? An important point is being made here, but it feels abstract at present. By providing say, the number of units needed to achieve these levels, it might make these points more concrete for the reader.</p>
-------------------------	---

	<p>Page 5, Line 26/27: Should this be stands* not stand?</p> <p>Page 7, Lines 52-57: Could the authors include in their aim the specific type of alcohol consumption that they are interested in investigating? Earlier in the introduction the authors distinguish between two types of alcohol consumption in relation to employment: overall consumption & work-related alcohol consumption – it seems important to clarify which is the focus early on.</p> <p>General comments: The authors use the abbreviation “i.e.” quite regularly in this section and in the abstract. Could this be replaced with say, “that is” on line 26/7 or “defined as” on line 42/43 (both page 4) to avoid repetition and appearing as if shortcuts are being taken.</p> <p>Methods</p> <p>Overall comment: The review has been registered with PROSPERO, which appears to have been kept up-to-date and is consistent with this manuscript. The eligibility criteria are mostly (see comment below) clear and transparent. There are other considerable strengths to the methods used, including the use of more than one reviewer at all stages of study selection, data extraction, and quality assessment, a thorough search strategy including hand searching of reference lists, and an appropriate and well justified strategy for analysis. However, some of the sub-sections of the methods require more detail in order to be able to fully understand the methods used by the authors in their review– see comments below:</p> <p>Page 8, Lines 46/47: Why were only studies published during or after 1990 included? The authors need to state the justification for excluding studies when the reasoning is not transparent.</p> <p>Literature search: Did the authors input each of the search blocks individually into all of the databases searched and screen the outcomes from each? This seems like it would result in huge numbers of returned studies, most of which would be irrelevant. It seems only logical to search combinations of two or more of the search blocks together: population, exposure, and outcome. Some clarification here is necessary.</p> <p>Study and data selection: How were studies screened? Were they exported to specific systematic review screening software (e.g., Covidence) or to any reference management programmes such as EndNote? If they were simply screened when studies were returned from the database searches, how did independent reviewers screen separately and how did you then ensure consistency between reviewers? I can’t find this information in your manuscript or PROSPERO registration.</p> <p>Does BMJ require the specific authors involved in each task to be identified by their initials (e.g., “... studies were independently screened by two reviewers [GD & RB]”)?</p> <p>Did the authors formally assess the consistency between reviewers involved in the study selection process by, for example, calculating a Kappa statistic? This is not essential, but can be informative if conducted.</p>
--	--

The description of how data was extracted is inadequate. What pieces of information were consistently extracted from studies? And how was this done? Did the authors use standardised data extraction form, did they develop their own, or was data simply inputted into a spreadsheet? I see nothing wrong with the latter two of these options, provided the authors provide a justification for this and no standardised extraction forms are available (though, I don't imagine any existing extraction forms would be suitable for this review). I have found more information on this on the PROSPERO registration, but it still doesn't answer all of the above questions. Even referring readers to your PROSPERO page here might be useful.

Quality assessment: The method of assessing the quality of studies based on the statistical robustness of the associations appears appropriate for this review. Are the suggested sample sizes for low, medium, and high-quality studies based on a statistical principles or existing research findings?

Results

Overall comment: The authors provide a detailed description of the included studies and the selection process. The cross-tabulation analyses relating to the associations' direction, publication year, study quality, and significance are a helpful way of understanding more about the studies reviewed, particularly in relation to the effects of study quality on the outcomes. Despite considerable heterogeneity across studies, the authors have cleverly synthesised the findings in their tables in order to make sense of the various relationships/ associations studied.

One thing that appears conspicuously absent from the results section is the inclusion of effect sizes. These have been reported in supplemental document 2, but not included or discussed anywhere in the main manuscript. The authors focused on significance and direction only, though the limitations of P values are now well understood. Integrating effect sizes into the result section would allow for a much greater understanding of the associations reviewed and the differences between the associations made between each type of consumption variable and each type of work performance measure.

There is considerable heterogeneity in the types of assessments of alcohol consumption across the studies—more than anticipated after reading the abstract and the analysis sections above. For example, the number of binge drinking episodes in the past three months, weekly consumption in grams, and DSM diagnoses of alcohol abuse or dependence are barely comparable exposure variables. However, Table 4 and the accompanying text is particularly useful for separating out and different types of alcohol measures and the association with work performance.

Nonetheless, the bulleted section following the abstract that states the studies reviewed were heterogeneous should be amended to be more informative in regards to the actual nature of the heterogeneity in alcohol measures (i.e., stating that “alcohol consumption” as it is called in the title, actually included a wide variety of variables [consumption status, frequency, volume, hangovers, dependence/ abuse diagnoses]) as this has important implications for understanding results of the review and the extent of its application. Ideally, some mention of this in the main abstract

would be better in order to give the reader a more accurate understanding of the review.

Discussion

General Comments:

The authors make some logical recommendations for future work in this area. However, in the conclusion here and in the abstract the authors suggest that longitudinal studies are required: could the authors clarify in the “Implications” subsection of the discussion how future research could use longitudinal designs in this area and why this would be useful, considering the prominence of their recommendation regarding longitudinal research?

Although the other alcohol exposure variables are clearly linked to diagnoses of alcohol abuse and dependence, these diagnoses require that individuals meet a host of criteria related to the use and the harms associated with it, and they seem conceptually distinct enough to require specific discussion dedicated to the four studies that used DSM diagnoses as the exposure measure. Were the effects observed in these studies of greater magnitude? Also, as this is a categorical variable, who were they compared against: abstinent colleagues or just those without a diagnosis of dependence or abuse but who also drink? If the latter, then this could considerably influence the findings and interpretation of these studies, given that the other studies show an association between consumption and performance in those who are not dependent. This warrants further discussion and attention.

Page 28, line 38: There is no need for the “i.e.,” here—it could be removed in the sentence would be easier to read.

Page 28, Lines 45 to 56: What is the point that is trying to be made by including this information? I presume it's to highlight the generalisability of findings, but regardless of what point is trying to be made, it needs to be more explicitly stated; otherwise this information simply seems like description that could've been in the methods section or results.

Page 30, Lines 5 to 10: That some of the studies found alcohol consumption was positively associated with work performance requires further discussion here, even if only two of the associations were statistically significant. I can see that you start to discuss positive relationship between low levels of consumption and work performance (i.e., the J shaped relationship), but the few associations that suggest improved work performance seem to be largely ignored in your discussion.

In relation to your discussion of the J shaped relationship between consumption and work performance, it may be worth noting that several large meta-analyses (see example references below) have suggested a similar relationship exists between consumption and cognitive outcomes. Again, this could be related to confounds, as you suggest could be the case for the consumption and work performance relationship at low levels. Relatedly, heavy consumption/ AUD diagnoses have been consistently associated with a variety of cognitive deficits that endure long after abstinence (e.g., Stavro et al., 2013 – below) and that could potentially impact work performance—this may be worthy of mention somewhere in your discussion.

Neafsey, E. J., & Collins, M. A. (2011). Moderate alcohol consumption and cognitive risk. *Neuropsychiatric Disease And Treatment*, 7, 465-484. Retrieved <https://www.ncbi.nlm.nih.gov/pmc/PMC3157490/>. doi:10.2147/NDT.S23159

Anstey, K. J., Mack, H. A., & Cherbuin, N. (2009). Alcohol Consumption as a Risk Factor for Dementia and Cognitive Decline: Meta-Analysis of Prospective Studies. *The American Journal of Geriatric Psychiatry*, 17(7), 542-555. Retrieved from <https://doi.org/10.1097/JGP.0b013e3181a2fd07>.

Stavro, K., Pelletier, J., & Potvin, S. (2013). Widespread and sustained cognitive deficits in alcoholism: a meta-analysis. *Addiction Biology*, 18(2), 203-213. doi:0.1111/j.1369-1600.2011.00418.x.

Page 30, Lines 28 to 31: Is the term “goofing off” used by Moore et al.? If so, I would recommend placing the term in quotation marks so it is clear that the informal language is not your own. If not, then I would suggest using a different term.

Page 31, Lines 24-26: Can the authors state how their findings could be used to inform the workplace interventions they refer to? The possible benefits of this piece of work are not really ‘sold’ in this “Implications” section at present.

Page 31, line 33: The authors state that “one cannot plausibly conclude that alcohol consumption constitutes a risk factor for impaired work performance” – given the findings of their review I would argue that this claim is certainly plausible and that the phrasing here is too harsh. Whilst I commend the authors for exercising caution in their conclusions, the statement is so strong that it appears to contradict other statements in their conclusion (e.g., lines 19-24). This paragraph would likely read better if this sentence was omitted as sufficient caution is exercised in the surrounding sentences.

Abstract

Although the authors define the association they are studying in the first sentence of the abstract, when the authors begin to talk about associations within the reviewed studies I would recommend explicitly stating what is meant by this so that the abstract is easier to understand without having to look to the result section of the paper. Simply changing “tested associations in the included studies were quality assessed, and analysed...” to “tested associations between alcohol consumption and work performance within the included studies were quality assessed, and analysed...”.

Overall comments

The supplemental files are clearly laid out and helpful additions to the manuscript.

Overall, this is a methodologically rigorous piece of work. It is well written throughout and the authors appear to make logical and necessarily cautious interpretations of the findings. However, the changes recommended above are necessary; particularly: [1] including effect sizes where possible and adjusting the discussion to reflect the knowledge gained by their inclusion; [2] clarifying the specifics of the methodology used; [3] discussing the distinction between DSM diagnoses and the other consumption measures

	<p>and how the two might differently affect the outcomes; and [4] highlighting the heterogeneity in the alcohol consumption variables studied in the abstract.</p> <p>All of these recommendations could be achieved in a moderate revision of manuscript.</p>
--	--

REVIEWER	Lode Godderis Center for Environment and Health, KU Leuven, Belgium
REVIEW RETURNED	17-Feb-2019

GENERAL COMMENTS	<p>Association between alcohol consumption and impaired work performance (presenteeism): A systematic review</p> <p>MAJOR comment regarding search string This systematic review addresses the association between alcohol consumption and impaired work performance. Considering the severe impact of alcohol use in society, and job related effects, this is an important however underestimated topic. Therefore, we also underline the need for more robust research.</p> <p>However, we think an important element in the introduction is missing, namely the consequences of alcohol use on job performance in terms of safety. Alcohol use, both by occasional and chronic drinkers, is an important factor in workplace accidents. Therefore, also safety reasons are important arguments for making and implementing an alcohol policy at work. Infra, you can find two specific aspects related to a broader definition of job performance, i.e. alcohol problems not only are health but also safety problems.</p> <p>- p.5: 'on-the-job-drinking'. The authors state that 'reduced performance are thought mainly to be due on-the-job drinking'. We think 'on-the-job' needs some additional information because one might think this only refers to drinking at the workplace. Frone (reference 5, used by the authors) defines 'on-the-job substance use' as 'the consumption of alcohol and illicit drugs at times that occur just before or during a person's workday (Frone, 2013, p.132). On page 24, Frone also gives a definition of 'workplace substance use' which refers to 'the consumption of alcohol or illicit drugs (a) within 2 hours of starting one's work shift, (b) during a lunch break, (c) during other work breaks, or (d) while performing one's job. In the discussion section, the authors mention that 'on-the-job performance outcomes may be more affected by on-the-job drinking' (p.30). Therefore, the discussion section could benefit from describing these definitions in a more exhaustive way.</p> <p>- p.6: The authors mention 'In this systematic review, we understand presenteeism as reduced on-the-job performance due to health problems. As such, presenteeism constitutes a link between on the job productivity and employee health, addressing the grey area between optimal work performance and the absence of productivity (i.e., absenteeism)'. In the discussion section (p.29), the authors mention that 'such short-term impairment-producing consumption may be more predictive of work impairments than for instance typical drinking</p>
-------------------------	--

	<p>frequency, which instead may be more predictive of long-term ill-health consequences.</p> <p>Therefore, we think that 'job safety' as an additional work performance outcome in the search strategy might have given additional information.</p> <p>Minor issues</p> <ul style="list-style-type: none"> - p.30: J-shape: a recent reference is lacking, i.e. Stockwell, T., Zhao, J., Panwar, S., Roemer, A., Naimi, T., & Chikritzhs, T. (2016). Do 'moderate' drinkers have reduced mortality risk? A systematic review and meta-analysis of alcohol consumption and all-cause mortality. <i>Journal of studies on alcohol and drugs</i>, 77(2), 185-198. - p.32: we support the suggestion that the measurement of alcohol consumption could benefit from differentiating between several types of drinking. In that sense, we think the AUDIT-C could be an alternative as the AUDIT-C score identifies at-risk drinkers (i.e., binge drinking) who are not necessarily alcohol-dependent.
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer #1:

The manuscript was well-written and easy to follow. The review and description of the review, including the methods, results and discussion, was thorough.

Authors:

We appreciate the positive feedback.

Reviewer #1:

In the Quality Assessment section (page 10), it would be helpful if the authors:

1. Explained the sample size thresholds utilized (low, moderate, high - were they arbitrary or based on some standard).
2. Explained what they meant by adjusted/unadjusted (I understood it after reviewing Supplementary File 2, but was not sure what they were referring to in the MS).

Authors:

1. The application of sample size thresholds/categories was not arbitrary, but based on an assumption of alcohol-related presenteeism being a relatively low-prevalent phenomenon, with reference to similar thresholds utilised in a recent association-based review of alcohol-related absenteeism. Accordingly, we have explained the choice of sample size categorisations in the 'Quality assessment' paragraph in the Methods section (p. 13, lines 1-6).

2. Furthermore, in the same paragraph, we have now clarified that adjusted/unadjusted refers to the extent to which associations between exposure and outcome were controlled for possible confounding variables (p. 12, lines 23-24).

PART D: RESPONSE TO REVIEWER #2

Reviewer #2:

Overall comment: The authors have integrated and discussed a range of relevant research here and build a clear rationale for their review. All of the key concepts appear to be well-defined. They also

contextualise the review, noting previous review in this area and its limitations. Finally, the aim of the review is clearly stated.

Authors:

We appreciate the positive feedback.

Reviewer #2:

Page 4, Lines 38 to 40: Can the authors provide some indication of the levels of alcohol consumption needed to achieve these blood alcohol content levels? An important point is being made here, but it feels abstract at present. By providing say, the number of units needed to achieve these levels, it might make these points more concrete for the reader.

Authors:

We agree that reference to BAC levels remain abstract if not related to amount of alcohol intake. Of course, the relationship between BAC and alcohol intake is strongly moderated by factors such as age, gender and weight. Therefore, we have now added some examples, based on a male (age 40, body weight 80 kg) and a female (age 40, body weight 60 kg) (p. 5, lines 11-14).

Reviewer #2:

Page 5, Line 26/27: Should this be stands* not stand?

Authors:

'Stands' is the correct form. We have revised accordingly (p. 6, line 9).

Reviewer #2:

Page 7, Lines 52-57: Could the authors include in their aim the specific type of alcohol consumption that they are interested in investigating? Earlier in the introduction the authors distinguish between two types of alcohol consumption in relation to employment: overall consumption & work-related alcohol consumption – it seems important to clarify which is the focus early on.

Authors:

We have now specified in the aim formulation that 'employee alcohol consumption' comprises overall as well as work-related alcohol consumption (p. 9, line 7).

Reviewer #2:

General comments: The authors use the abbreviation "i.e." quite regularly in this section and in the abstract. Could this be replaced with say, "that is" on line 26/7 or "defined as" on line 42/43 (both page 4) to avoid repetition and appearing as if shortcuts are being taken.

Authors:

We agree. Revised as suggested (p. 4, line 15; p. 5, line 15).

Reviewer #2:

Methods. Overall comment: The review has been registered with PROSPERO, which appears to have been kept up-to-date and is consistent with this manuscript. The eligibility criteria are mostly (see comment below) clear and transparent. There are other considerable strengths to the methods used, including the use of more than one reviewer at all stages of study selection, data extraction, and quality assessment, a thorough search strategy including hand searching of reference lists, and an appropriate and well justified strategy for analysis. However, some of the sub-sections of the methods require more detail in order to be able to fully understand the methods used by the authors in their review—see comments below.

Authors:

We appreciate the positive feedback.

Reviewer #2:

Page 8, Lines 46/47: Why were only studies published during or after 1990 included? The authors need to state the justification for excluding studies when the reasoning is not transparent.

Authors:

A justification for the time restriction is included in the Methods section describing the eligibility criteria (p. 10, lines 8-11). Moreover, we have included an additional paragraph in the 'Methodological considerations' subsection (p. 37, lines 3-11) discussing the probability of having missed relevant studies published prior to 1990. Here, we emphasise that this is not very likely, given that the time restrictions were not imposed at the literature search stage (rather at the study selection stage).

Reviewer #2:

Literature search: Did the authors input each of the search blocks individually into all of the databases searched and screen the outcomes from each? This seems like it would result in huge numbers of returned studies, most of which would be irrelevant. It seems only logical to search combinations of two or more of the search blocks together: population, exposure, and outcome. Some clarification here is necessary.

Authors:

We have now, in the Methods section ('Study and data selection') clarified that study selection was based on the results of combining the three main search blocks in the database search strategy (population, exposure and outcome) (p. 11, lines 13-14).

Reviewer #2:

Study and data selection: How were studies screened? Were they exported to specific systematic review screening software (e.g., Covidence) or to any reference management programmes such as EndNote? If they were simply screened when studies were returned from the database searches, how did independent reviewers screen separately and how did you then ensure consistency between reviewers? I can't find this information in your manuscript or PROSPERO registration.

Authors:

In the Methods section ('Literature search') we have now clarified that database search results were transferred to EndNote (p. 11, line 3).

Reviewer #2:

Does BMJ require the specific authors involved in each task to be identified by their initials (e.g., "... studies were independently screened by two reviewers [GD & RB]")?

Authors:

We have not received feedback from BMJ Open stating that we should include such initials in the main text. We would, however, like to emphasise that this information is already presented in the Declarations section ('Contributors') (p. 38, lines 16-22)

Reviewer #2:

Did the authors formally assess the consistency between reviewers involved in the study selection process by, for example, calculating a Kappa statistic? This is not essential, but can be informative if conducted.

Authors:

We did not conduct a statistical assessment of the consistency between reviewers.

Reviewer #2:

The description of how data was extracted is inadequate. What pieces of information were consistently extracted from studies? And how was this done? Did the authors use standardised data extraction form, did they develop their own, or was data simply inputted into a spreadsheet? I see nothing wrong with the latter two of these options, provided the authors provide a justification for this and no standardised extraction forms are available (though, I don't imagine any existing extraction forms would be suitable for this review). I have found more information on this on the PROSPERO registration, but it still doesn't answer all of the above questions. Even referring readers to your PROSPERO page here might be useful.

Authors:

We agree that the data extraction procedure was somewhat sparsely described in the original manuscript. Accordingly, we have now included data extraction as a separate subheading in the Methods section, wherein we describe both what pieces of information were extracted and how this was done (p. 11, lines 23-25; p. 12, lines 1-12).

Reviewer #2:

Quality assessment: The method of assessing the quality of studies based on the statistical robustness of the associations appears appropriate for this review. Are the suggested sample sizes for low, medium, and high-quality studies based on a statistical principles or existing research findings?

Authors:

The application of sample size thresholds/categories was not arbitrary, but based on an assumption of alcohol-related presenteeism being a relatively low-prevalent phenomenon, with reference to similar thresholds utilised in a recent association-based review of alcohol-related absenteeism. Accordingly, we have explained the choice of sample size categorisations in the 'Quality assessment' paragraph in the Methods section (p. 13, lines 1-6).

Reviewer #2:

Results. Overall comment: The authors provide a detailed description of the included studies and the selection process. The cross-tabulation analyses relating to the associations' direction, publication year, study quality, and significance are a helpful way of understanding more about the studies reviewed, particularly in relation to the effects of study quality on the outcomes. Despite considerable heterogeneity across studies, the authors have cleverly synthesised the findings in their tables in order to make sense of the various relationships/ associations studied.

Authors:

We appreciate the positive feedback.

Reviewer #2:

One thing that appears conspicuously absent from the results section is the inclusion of effect sizes. These have been reported in supplemental document 2, but not included or discussed anywhere in the main manuscript. The authors focused on significance and direction only, though the limitations of P values are now well understood. Integrating effect sizes into the result section would allow for a much greater understanding of the associations reviewed and the differences between the associations made between each type of consumption variable and each type of work performance measure.

Authors:

The following effect sizes have been included in the Abstract: (1) odds ratio, phi-coefficient and p-value for the association between direction and significance (p. 2, lines 20-21), and (2) proportion of significant positive associations with moderate/high quality based on exposure measures of hangover episodes and composite instruments (p. 2, line 23).

The following effects sizes have been included in the Results section: (1) five associations characterised by positive direction, statistically significant and high quality (p. 22, lines 16-25; p. 23, lines 1-3), (2) two associations characterised by negative direction and statistically significant (p. 23, lines 7-8), and (3) five associations characterised by not being classified as positive or negative (p. 23, lines 13, 14, 16, 19).

Reviewer #2:

There is considerable heterogeneity in the types of assessments of alcohol consumption across the studies—more than anticipated after reading the abstract and the analysis sections above. For example, the number of binge drinking episodes in the past three months, weekly consumption in grams, and DSM diagnoses of alcohol abuse or dependence are barely comparable exposure variables. However, Table 4 and the accompanying text is particularly useful for separating out and different types of alcohol measures and the association with work performance. Nonetheless, the bulleted section following the abstract that states the studies reviewed were heterogeneous should be amended to be more informative in regards to the actual nature of the heterogeneity in alcohol measures (i.e., stating that “alcohol consumption” as it is called in the title, actually included a wide variety of variables [consumption status, frequency, volume, hangovers, dependence/ abuse diagnoses]) as this has important implications for understanding results of the review and the extent of its application. Ideally, some mention of this in the main abstract would be better in order to give the reader a more accurate understanding of the review.

Authors:

We agree that the great heterogeneity in alcohol measures, in particular, should be emphasised for the reader early in the article. We have, as suggested, included a statement on this in the bulleted section following the Abstract (p. 3, lines 22-25).

Reviewer #2:

Discussion. General Comments: The authors make some logical recommendations for future work in this area. However, in the conclusion here and in the abstract the authors suggest that longitudinal studies are required: could the authors clarify in the “Implications” subsection of the discussion how future research could use longitudinal designs in this area and why this would be useful, considering the prominence of their recommendation regarding longitudinal research?

Authors:

We have restructured and extended the 'Implications' subsection in order to elaborate our recommendations for future research, including information on how longitudinal designs in this area may be conducted, and how exposure and outcome could be measured in a more serviceable manner (p. 34, lines 23-25; p. 35, lines 1-5)

Reviewer #2:

Although the other alcohol exposure variables are clearly linked to diagnoses of alcohol abuse and dependence, these diagnoses require that individuals meet a host of criteria related to the use and the harms associated with it, and they seem conceptually distinct enough to require specific discussion dedicated to the four studies that used DSM diagnoses as the exposure measure. Were the effects observed in these studies of greater magnitude? Also, as this is a categorical variable, who were they compared against: abstinent colleagues or just those without a diagnosis of dependence or

abuse but who also drink? If the latter, then this could considerably influence the findings and interpretation of these studies, given that the other studies show an association between consumption and performance in those who are not dependent. This warrants further discussion and attention.

Authors:

We agree that alcohol-related diagnoses may be conceptually quite different from more direct measures of alcohol consumption, even though the four studies in the review utilising such exposure measures do not differ considerably in terms of overall conclusion regarding the relationship between exposure and outcome. Moreover, we agree that this conceptual difference deserves attention in the discussion. We have now included a new paragraph in the Discussion relating to the studies utilising alcohol-related diagnosis as measure of exposure (p. 31, lines 20-25; p. 32, lines 1-6).

Reviewer #2:

Page 28, line 38: There is no need for the “i.e.,” here—it could be removed in the sentence would be easier to read.

Authors:

We have revised (removed) as suggested.

Reviewer #2:

Page 28, Lines 45 to 56: What is the point that is trying to be made by including this information? I presume it's to highlight the generalisability of findings, but regardless of what point is trying to be made, it needs to be more explicitly stated; otherwise this information simply seems like description that could've been in the methods section or results.

Authors:

We agree that this information may seem somewhat 'out of context' in the Discussion, as it basically refers to results from included studies. Hence, we have removed these lines, and instead described these results more appropriately in the Results section (with effect sizes) (pp. 22-23).

Reviewer #2:

Page 30, Lines 5 to 10: That some of the studies found alcohol consumption was positively associated with work performance requires further discussion here, even if only two of the associations were statistically significant. I can see that you start to discuss positive relationship between low levels of consumption and work performance (i.e., the J shaped relationship), but the few associations that suggest improved work performance seem to be largely ignored in your discussion.

Authors:

The two significant negative associations are now discussed more thoroughly in the Discussion (p. 32, lines 14-21), with an emphasis on the fact that these findings are based on drinking duration and, hence, may reflect increased tolerance/coping among more experienced drinkers (as compared to less experienced drinkers).

Reviewer #2:

In relation to your discussion of the J shaped relationship between consumption and work performance, it may be worth noting that several large meta-analyses (see example references below) have suggested a similar relationship exists between consumption and cognitive outcomes. Again, this could be related to confounds, as you suggest could be the case for the consumption and work performance relationship at low levels. Relatedly, heavy consumption/ AUD diagnoses have been consistently associated with a variety of cognitive deficits that endure long after abstinence (e.g., Stavro et al., 2013 – below) and that could potentially impact work performance—this may be worthy of mention somewhere in your discussion.

Neafsey, E. J., & Collins, M. A. (2011). Moderate alcohol consumption and cognitive risk. *Neuropsychiatric Disease And Treatment*, 7, 465-484. Retrieved <https://www.ncbi.nlm.nih.gov/pmc/PMC3157490/>. doi:10.2147/NDT.S23159

Anstey, K. J., Mack, H. A., & Cherbuin, N. (2009). Alcohol Consumption as a Risk Factor for Dementia and Cognitive Decline: Meta-Analysis of Prospective Studies. *The American Journal of Geriatric Psychiatry*, 17(7), 542-555. Retrieved from <https://doi.org/10.1097/JGP.0b013e3181a2fd07>.

Stavro, K., Pelletier, J., & Potvin, S. (2013). Widespread and sustained cognitive deficits in alcoholism: a meta-analysis. *Addiction Biology*, 18(2), 203-213. doi:0.1111/j.1369-1600.2011.00418.x.

Authors:

We appreciate the suggestions and references, and have now mentioned that similar J-shaped relationships have been found between alcohol and cognitive outcomes (p. 33, lines 3-4), and that cognitive deficits due to heavy drinking may endure long after abstinence (p. 33, lines 6-7).

Reviewer #2:

Page 30, Lines 28 to 31: Is the term “goofing off” used by Moore et al.? If so, I would recommend placing the term in quotation marks so it is clear that the informal language is not your own. If not, then I would suggest using a different term.

Authors:

The term is used by Moore et al. We have, as suggested, placed the term in quotation marks (p. 33, lines 1-2).

Reviewer #2:

Page 31, Lines 24-26: Can the authors state how their findings could be used to inform the workplace interventions they refer to? The possible benefits of this piece of work are not really ‘sold’ in this “Implications” section at present.

Authors:

We have now elaborated on the results' practical implications for workplace interventions (p. 34, lines 7-15).

Reviewer #2:

Page 31, line 33: The authors state that “one cannot plausibly conclude that alcohol consumption constitutes a risk factor for impaired work performance” – given the findings of their review I would argue that this claim is certainly plausible and that the phrasing here is too harsh. Whilst I commend the authors for exercising caution in their conclusions, the statement is so strong that it appears to contradict other statements in their conclusion (e.g., lines 19-24). This paragraph would likely read better if this sentence was omitted as sufficient caution is exercised in the surrounding sentences.

Authors:

The sentence is now omitted as suggested.

Reviewer #2:

Abstract. Although the authors define the association they are studying in the first sentence of the abstract, when the authors begin to talk about associations within the reviewed studies I would recommend explicitly stating what is meant by this so that the abstract is easier to understand without having to look to the result section of the paper. Simply changing “tested associations in the included studies were quality assessed, and analysed...” to “tested associations between alcohol consumption and work performance within the included studies were quality assessed, and analysed...”.

Authors:

We have revised as suggested (pp. 2-3).

Reviewer #2

The supplemental files are clearly laid out and helpful additions to the manuscript.

Authors:

We appreciate the positive feedback.

Reviewer #2:

Overall, this is a methodologically rigorous piece of work. It is well written throughout and the authors appear to make logical and necessarily cautious interpretations of the findings. However, the changes recommended above are necessary; particularly: [1] including effect sizes where possible and adjusting the discussion to reflect the knowledge gained by their inclusion; [2] clarifying the specifics of the methodology used; [3] discussing the distinction between DSM diagnoses and the other consumption measures and how the two might differently affect the outcomes; and [4] highlighting the heterogeneity in the alcohol consumption variables studied in the abstract.

All of these recommendations could be achieved in a moderate revision of manuscript.

Authors:

We appreciate the constructive feedback and useful suggestions. We believe we have revised the manuscript in accordance with the points raised by the reviewer.

PART E: RESPONSE TO REVIEWER #3

Reviewer #3:

This systematic review addresses the association between alcohol consumption and impaired work performance. Considering the severe impact of alcohol use in society, and job related effects, this is an important however underestimated topic. Therefore, we also underline the need for more robust research.

Authors: We agree.

Reviewer #3:

However, we think an important element in the introduction is missing, namely the consequences of alcohol use on job performance in terms of safety. Alcohol use, both by occasional and chronic drinkers, is an important factor in workplace accidents. Therefore, also safety reasons are important arguments for making and implementing an alcohol policy at work. Infra, you can find two specific aspects related to a broader definition of job performance, i.e. alcohol problems not only are health but also safety problems.

Authors:

We agree that safety issues, such as occupational accidents/injuries, constitute important aspects of alcohol-related occupational outcomes. In order to (a) clarify accidents/injuries as an important outcome, and (b) clarify that our review is primarily concerned with task performance outcomes, we have included a new paragraph in the Introduction section that emphasises, with reference to Frone (2013), that performance outcomes at work comprise task performance, contextual performance, counterproductive behaviour and accidents/injuries (p. 7, lines 19-25; p. 8, lines 1-3).

We also agree that safety concerns, in itself, represent important arguments making and implementing alcohol policies in the workplace. In the Discussion section (under the subheading 'Implications'), we have emphasised that both absenteeism and injuries/accidents (in addition to presenteeism) represent arguments for targeting alcohol consumption within workplace interventions (p. 34, lines 7-15).

Reviewer #3:

p.5: 'on-the-job-drinking'. The authors state that 'reduced performance are thought mainly to be due on-the-job drinking'. We think 'on-the-job' needs some additional information because one might think this only refers to drinking at the workplace. Frone (reference 5, used by the authors) defines 'on-the-job substance use' as 'the consumption of alcohol and illicit drugs at times that occur just before or during a person's workday (Frone, 2013, p.132). On page 24, Frone also gives a definition of 'workplace substance use' which refers to 'the consumption of alcohol or illicit drugs (a) within 2 hours of starting one's work shift, (b) during a lunch break, (c) during other work breaks, or (d) while performing one's job.

In the discussion section, the authors mention that 'on-the-job performance outcomes may be more affected by on-the-job drinking' (p.30). Therefore, the discussion section could benefit from describing these definitions in a more exhaustive way.

Authors:

We agree that the term 'on-the-job drinking' should have been operationalised with reference to Frone (2013). We have now included Frone's operationalisation, both in Introduction where his model is thematised (p. 6, lines 3-4), as well as in the Discussion (p. 33, line 15).

Reviewer #3:

p.6: The authors mention 'In this systematic review, we understand presenteeism as reduced on-the-job performance due to health problems. As such, presenteeism constitutes a link between on the job productivity and employee health, addressing the grey area between optimal work performance and the absence of productivity (i.e., absenteeism)'.
In the discussion section (p.29), the authors mention that 'such short-term impairment-producing consumption may be more predictive of work impairments than for instance typical drinking frequency, which instead may be more predictive of long-term ill-health consequences.

Therefore, we think that 'job safety' as an additional work performance outcome in the search strategy might have given additional information.

Authors:

We agree that additional information would have been retrieved if the term 'job safety' was included in the search strategy. In the initial phases of this review work, we had several discussions regarding conceptualisation, scope and relevant outcomes. We decided to focus our review on task performance, this being the outcome most directly related to our understanding of presenteeism. Other performance outcomes – e.g., accidents/injuries/safety, workplace aggression, job engagement – were of course considered for inclusion, but ultimately excluded in order to avoid ending up with a too broad scope. As should be evident, the research relating to task performance alone is very heterogeneous. That being said, naturally we appreciate job safety as an important alcohol-related occupational outcome, and we do appreciate the reviewer's emphasis on this matter. Therefore, as previously noted, we have thematised occupational injuries as an important topic, both in the Introduction (p. 7, lines 19-25; p. 8, lines 1-3) and the Discussion (p. 34, lines 7-15). In order to avoid any possible confusion, we have also clarified that we, within the scope of this review, focus on work performance as task performance (p. 8, lines 1-3).

Reviewer #3:

p.30: J-shape: a recent reference is lacking, i.e. Stockwell, T., Zhao, J., Panwar, S., Roemer, A., Naimi, T., & Chikritzhs, T. (2016). Do 'moderate' drinkers have reduced mortality risk? A systematic review and meta-analysis of alcohol consumption and all-cause mortality. Journal of studies on alcohol and drugs, 77(2), 185-198.

Authors:

We appreciate this suggestion, and have included this reference in the Discussion (p. 33, lines 9-11).

Reviewer #3:

p.32: we support the suggestion that the measurement of alcohol consumption could benefit from differentiating between several types of drinking. In that sense, we think the AUDIT-C could be an alternative as the AUDIT-C score identifies at-risk drinkers (i.e., binge drinking) who are not necessarily alcohol-dependent.

Authors:

As suggested, we have included the AUDIT-C as an alternative to the full AUDIT (p. 35, lines 20-21).

VERSION 2 – REVIEW

REVIEWER	Ryan J. Martin East Carolina University, USA
REVIEW RETURNED	20-Mar-2019

GENERAL COMMENTS	The authors have adequately addressed my concerns in this revision.
-------------------------	---

REVIEWER	Robert Heirene University of South Wales Addictions Research Group School of Psychology and Therapeutic Studies Wales, UK
REVIEW RETURNED	20-Mar-2019

GENERAL COMMENTS	The authors appear to have appropriately addressed all of the concerns raised during the review process and the paper has improved considerably as a result.
-------------------------	--