To the Editor:

We appreciate the opportunity to address the Committee's and Reviewers' comments and revise our manuscript. Below, please find item-by-item responses to the Reviewers' comments, which are included verbatim. All page and paragraph numbers refer to locations in the revised manuscript. In the interim, a new article was published on this topic and we have incorporated this into the manuscript. This new study also did not find an association between benzodiazepine use and AD.

Committee Review

1. **The finding that low benzodiazepine use was associated with higher risk of dementia isn't easy to interpret.**

   Response: We agree with the committee that this finding is hard to interpret. We do not think that the increased risk noted with the lower levels of use is causal, but more likely represents use for prodromal symptoms. This is supported by the fact that when we increase the lag time for exposure measurement to 2 years, the association is attenuated toward the null. We have included the following text in the discussion section of the manuscript to indicate that this association may be spurious (page 12, last para):

   "*This pattern does not support a causal association between cumulative benzodiazepine use and dementia risk, and the small increased risk observed with low use may represent treatment of prodromal symptoms, as supported by our sensitivity analyses*."

2. **The reviewers raise important concerns over the measurement of exposure, in particular the use of seemingly arbitrary cut-offs. Whilst the authors state that the categories of low, medium and high benzodiazepine use were based on the distribution of exposure and clinically meaningful cut-points, there is no further information on the distribution and no references to support. To avoid the problems with arbitrary categorisation of exposure, we suggest modelling with TSDD as the continuum that it is, using non-linear models as appropriate to capture the changes.**

   Response: As suggested, we have performed additional analyses modelling our exposure as a continuous variable using natural cubic splines (Figure 4). These results are consistent with our primary analysis using categorical exposure variables. We have added this to the results section (page 11, last para):

   "Figure 4 shows results from additional analyses that modeled benzodiazepine exposure as a continuous variable using natural cubic splines. These curves show the estimated HRs (and 95% CIs) for each level of exposure relative to a referent group with cumulative exposure of 0 TSDDs. For dementia, a modestly elevated HR between 1.1 and 1.4 was observed with 90 or fewer TSDD of benzodiazepine use but risk declines toward 1.0 with higher TSDD. Benzodiazepine use was not significantly associated with AD at any TSDD."

3. **Furthermore there are doubts over the appropriate assessment period to reduce the capacity for susceptibility bias in this observational study.**

Response: This comment relates to comment # 11 from Reviewer 3**.** We are unable to exclude the possibility of susceptibility bias, and in fact as explained below, we think it is likely that all studies of this topic share this limitation, given the high prevalence and typical patterns of benzodiazepine use in the general population. Susceptibility bias relates to the idea that within a pool of potentially eligible people, benzodiazepine exposures before the beginning of our study period impacted a subset of extremely susceptible people and led them to have dementia at a younger age, making them ineligible for enrollment into ACT, which would thus deplete the enrolled cohort of people most at risk for these outcomes. People who were found to have dementia at study baseline did not sign consent forms so we are not able to comment on whether benzodiazepine use was different among those with prevalent dementia than those who enrolled in the study. The pharmacy data available for this cohort is far more detailed and extensive than the data available to others who have tried to address this concern with a "new user" design (Billioti de Gage 2012), so we are able to explore this issue in some unique ways. We find that patterns of sporadic benzodiazepine prescriptions stretch back for decades, making it unlikely that other studies would be able to thoroughly and accurately identify true "new users" with more limited data such as from periodic patient interviews. Our biological model was that cumulative exposure over years was the most plausible causal mechanism by which benzodiazepine use could impact dementia risk, and designed our analyses to test that model. If this mechanism is true, then a "new user" design that excluded only people with benzodiazepine use in the recent past (e.g. 1 or 2 years) would likely be inadequate to address susceptibility bias, and the "new users" of greatest interest would be those who truly had no prior use over decades. To our knowledge no study has been able to identify such a cohort. We have added comments about susceptibility bias to the limitations section (page 16, 1st para) and a section earlier in the discussion to address the related issue of Reviewer's 3 thought about a new user design:

Page 16, 1st para: "*We are unable to exclude the possibility that within the source population, the most susceptible users of benzodiazepines may have developed dementia at a younger age and therefore been ineligible for enrollment into ACT, perhaps limiting our ability to examine this association"*

Page 14 (last para): "*Some of the challenges inherent to using an observational study design to examine benzodiazepine use and dementia risk have been discussed, such as accounting for use of these medications for prodromal symptoms and use of administrative data for outcome ascertainment. Another methodological challenge is balancing the desire for a new user design with the reality of how people use benzodiazepines (ie. potentially episodic use over several years) and sample size considerations. A new user design is one strategy proposed to address the bias caused by depletion of susceptible users; however, it is not clear that this design is well suited to examine the principal hypothesis that heavier exposure (e.g., as might be accumulated over many years), is important for increasing dementia risk. Such a design would be challenging to implement in association with prospective follow up of a cohort with a rigorous process for ascertainment of dementia outcomes such as that used in our study, thus forcing the "new user" study to rely on a less optimal outcome definition such as diagnosis codes from administrative or automated data, which have been shown to have limited sensitivity. Additionally, benzodiazepines may be used sporadically over long periods of time on an as-needed basis which could make it challenging to identify true "new users", particularly in an elderly cohort."*

"We are unable to exclude the possibility that within the source population, the most susceptible users of benzodiazepines may have developed dementia at a younger age and therefore been ineligible for enrollment into ACT, perhaps limiting our ability to examine this association. Utilizing a "new user" design is one strategy proposed to address the bias caused by depletion of susceptible users; however, it is difficult to employ this strategy in our setting to explore our particular research question. Most importantly, a new user design would not be well suited to examine our principal hypothesis: that heavier exposure (e.g., as might be accumulated over many years), would be associated with dementia. Such a design would be challenging to implement in association with prospective follow up of a cohort with a rigorous process for ascertainment of dementia outcomes such as that used in our study, thus forcing the "new user" study to rely on a less optimal outcome definition such as diagnosis codes from administrative or automated data, which have been shown to have limited sensitivity. Additionally, benzodiazepines may be used sporadically over long periods of time on an as-needed basis which could make it challenging to identify true "new users", particularly in an elderly cohort."

4. **What is the rationale for treating only stroke and coronary heart disease as time-varying in the models? Updated information was also available for other covariates.**

Response: Given our 10 year exposure window, it is challenging to determine the right time during the window to measure potential confounders. We sought to balance adequately adjusting for confounding but also not over-correcting as brought up by the Committee in comment 6 below. Covariates in models were age at study entry, sex, education, hypertension, diabetes, smoking, stroke, coronary heart disease (CHD), BMI, exercise, self-rated health and depressive symptoms. Stroke and CHD were treated as time-varying covariates because these are major life events that are quite likely to happen after the age of 65 (and by extension, after enrollment in the ACT study) and may have a significant impact on dementia risk. We also felt it was very unlikely that their occurrence would be influenced by prior benzodiazepine use, which made us more comfortable treating them as time-varying. We conducted a sensitivity analysis in which we additionally used time-varying depressive symptoms (reported in the manuscript) and found no appreciable change in risk estimates. Age at study entry, sex, and education are not time varying which leaves six additional variables that could have been modelled as time varying for which we instead chose to use baseline values. In the case of some of these (e.g. physical activity), we were concerned about over-correction that might occur from adjusting in a time varying way for factors that may be influenced by benzodiazepine use, and so we sought to use the earliest measure available to us. In the case of others (e.g. BMI), we thought the baseline measure was likely a sufficient capture of the target health characteristic, especially in terms of capturing its potential relationship with dementia risk, and thus did not need updating over course of the study (especially if it could lead to an increase in observations excluded due to missing measures at later study visits).

5. **About 3.7% appear to be excluded because of missing covariates. What information was missing? Is there any indication that this subgroup may be biased in any way associated with outcome? Does a sensitivity analysis using multiple imputation alter the findings?**

Response: We quantified the missing data for each variable and have added this information as a footnote in Table 2 (footnote a). The missing variables are distributed as follows: education (n=1), body mass index (n=75), smoking (n=7), exercise (n=8), self-rated health (n=5), depressive symptoms (n=56). The two ways that missing data could affect our results are by introducing bias in the HRs and by decreasing their precision. With only about 3.7% missing data it is unlikely that either bias or

precision would be meaningfully affected. Since the standard error of log(HR) is a function of the square root of sample size, decreasing the sample size by 3.7% increases the standard errors of log(HR) by about 2%. An increase in standard errors of this magnitude is insufficient to alter any of our study's conclusions. Similarly, with such a small proportion of missing data, it is unlikely that our results would be biased to the extent necessary to alter our findings.

6. **Re the covariates adjusted for – could these be associated with outcome and exposure, leading to over-correction. For example, is it valid to adjust for self-rated health and depressive symptoms?**

   Response: We are not entirely sure what is meant by this comment but suspect it may refer to concerns about adjusting for variables that could be intermediates in the causal pathway – that is, they could be consequences of benzodiazepine use and in turn increase the risk of dementia. If that were the case, then adjusting for them might mask a true association between benzodiazepine use and dementia. We considered this issue at length in designing our analysis, and it contributed to our decision not to treat these variables as time-varying (see response to comment #4 above.) Based on the literature and our own work, we selected the variables that we thought would be confounders of the relationship between benzodiazepine use and dementia a priori. A strength of ACT is that we have access to potential confounders that are not available in administrative data. A concern is that those who use benzodiazepines have poorer health and have other health conditions or characteristics that are associated with the outcome. These characteristics can be difficult to measure and it is very likely that measuring only specific comorbid conditions such as CHD is not adequate to capture health status more broadly. Variables adjusted for in our analysis, such as self-rated health and depressive symptoms, serve to help capture overall health status that may be associated with both risk of dementia and use of benzodiazepines. We believe the threat of bias due to not adjusting for such confounding factors is a larger threat to validity than potential bias due to over-correction.

   We would be happy to add supplementary material or additional text to the manuscript to more fully describe our decisions regarding confounders if the Committee advises (Comments 4 and 6).

7. **Whilst information is given pre-adjustment for dementia and AD (figure 3), for the CASI modelling there is none (table 6).**

   Response: We have added the age-adjusted results for the CASI. We have created a second table (new tables 6 and 7) and presented both age-adjusted and multivariable results for the mean change in CASI and CASI-IRT and for the rate of change for both outcomes.

REFEREES COMMENTS

Reviewer: 1

8. The main issue that I believe requires clarification stems from the very low exposure of the population over a long period, which makes it difficult to determine meaningful comparison groups. If I understand the data correctly, the authors are comparing three relatively low levels of exposure over a long period to people with no exposure. The 'high exposure group' of 121 or more standard doses over 10 years, would equate to over about 4 months of daily exposure over 10 years. Therefore, it is not very surprising that there is not a dose response relationship with the 'low', 'medium' and 'high' exposures described. I wonder if the clinical message of this paper is

predominantly about the risks of low exposure? Please justify the dose cut-offs investigated here further.

Response: Please see reply to Comment 2 above. We agree that benzodiazepine use in this cohort may be lower than what has been reported in studies from other countries. Some key studies of this topic did not have information about dose or duration and so were not able to describe intensity of use in their subjects (Billioti de Gage 2012). The "high use" group in the Canadian study was people with 180 days of use or more in 5 years; however, more description about the extent of use in this highest group was not provided (Billioti de Gage 2014). In our study, the median level of use in our highest use category was 375 TSDD over 10 years – equivalent to slightly more than one year of daily use. We have added information regarding the level of use in our highest use group in the results section (page 10, last line):

"Within the highest benzodiazepine category (>120 TSDD), the median level of use was 375 TSDD (equivalent to slightly over one year of daily use)."

We also believe that this reviewer's concern is addressed by the additional analyses we have conducted examining benzodiazepine use as a continuous variable, as suggested by the Committee; these results are consistent with our primary analysis (Comment 2 above).

Reviewer: 2

9. The authors acknowledge the limitations of the CASI, and employ an IRT approach. A minor suggestion is to add a sentence or two explaining how that approach might absolve a concern about ceiling effects or relative insensitivity of the CASI to detect cognitive decline in this relatively well-educated population. A comment in the Discussion related to this issue is also suggested.

Response: The reviewer has raised a good point that the CASI-IRT addresses a limitation, the CASI's imprecision in estimating cognitive decline in the higher range of cognitive function. As suggested, we have added more information in the methods (page 6, last para) and in the discussion as suggested (page 15, last para):

"Because of this, there is a nonlinear relationship between CASI score and underlying cognitive ability resulting in imprecision at the higher end of the scale.[22] Thus, we performed secondary analyses after applying item response theory (IRT) methods to generate CASI-IRT scores which have linear scaling properties (Parscale, Scientific Software International Inc, Chicago, Illinois).[23] This addresses the relative insensitivity of the CASI to detect cognitive decline in people with high cognitive functioning."

(page 16, line 13): *"Furthermore, the CASI is a screening tool, making it relatively insensitive to detect cognitive decline in people with high cognitive functioning. However, we still did not find an association between benzodiazepine use and cognitive decline when using statistical methods addressing this limitation"*

Reviewer: 3

The manuscript is quite well designed but should be altered and clarified since some parts of the design could be confusing. Moreover, some methodological choices, particularly those concerning the definition and measurement of exposure, require more precise explanations and/or supplementary analyses, since they may have led to underestimating the association. For example, the possibility of a

depletion of susceptible bias should be discussed as well as the fact that the long-term user group, as defined by the authors, could have included a significant number of short-term users (see below). These points are crucial since the study concludes that there is no association between benzodiazepines and dementia in chronic users (the 30% increase in risk being explained by short-term users only), and therefore challenges the conclusions of the studies conducted so far on the same topic. Since the potential impact on public health is major, the appropriateness of the message delivered to prescribers is crucial. Indeed, some conclusions drawn from the study could go against current international recommendations. For example, one could conclude that no excess risk is to be feared with long-term treatments while short-term use is associated with a 30% increased risk. Finally, the authors conclude that their results "may alleviate concerns raised by prior studies", a claim that should be mitigated given the level of proof of the results of what is an observational study, and the high public health relevance of the topic.

MAIN COMMENTS

10. The study design should be clarified or more precisely explained. I assume the authors considered cumulative benzodiazepine exposure in the 1-to-10-year period prior to enrolment in the ACT cohort. However, several sentences in the manuscript are confusing and could suggest that the observation period for ascertaining exposure started ten years before the event (and not before enrolment in the ATC cohort; see abstract and Figure 2) or lasted after enrolment in the ACT cohort (see the legend of Table 3). For example:

    Response: We apologize that this was not communicated clearly. In fact, we did not focus only on exposure prior to study enrollment; rather, we measured use in relation to events (dementia onset) occurring during follow-up. We chose this approach because we anticipated that substantial additional exposure could accumulate during follow-up and that it could potentially affect dementia risk. Thus we calculated a measure of time-varying exposure that rolled forward in time. In the Cox models, we recalculated the 10 year exposure for all participants at risk at the time of each new event (incident dementia). We have clarified this design issue in several places as described below:

    a. Abstract (page 2, lines 34-39) "Benzodiazepine exposure was defined as the total standardized daily doses dispensing in a rolling exposure window covering the prior 10 years…" It is unclear what "the prior 10 years" refers to: (1) the date of dementia (more evocative of a retrospective design) or (2) the date of enrolment in the cohort (more appropriate for a prospective design).

    Response: We have made the following changes to the abstract as suggested:

    "*Benzodiazepine exposure was defined from computerized pharmacy data and consisted of the total standardized daily doses (TSDDs) dispensed over a 10 year period (a rolling window that moved forward in time during follow-up). We excluded the most recent 1 year because of possible prodromal symptoms. Multivariable Cox proportional hazard models were used to examine time-varying benzodiazepine use and dementia risk."*

    b. In order to avoid this ambiguity, Figure 2 could provide the mean and maximum follow-up time after the observation period (during which the presence of dementia is investigated).

Response: We do not understand what the reviewer is suggesting. Figure 2 is a schematic of how we defined our exposure for the two different outcomes (dementia and cognitive trajectory) and the sensitivity analysis. It is not meant to include specific data. The total person years, mean and SD of our participants' follow-up time are included in the results section (page 11, under subheading "Dementia and AD").

"The 3,434 participants included in these analyses accrued 25,068 person-years of follow-up, with a mean (SD) of 7.3 (4.8) years."

c. Page 7, line 20 to 23: "... we summed the SDD for all benzodiazepine pharmacy fills during the exposure period (described below)". Again, in this part ("below"), the starting point for measuring exposure (ten year before dementia or enrolment in the cohort) is unclear.

Response: We have clarified this section as suggested (Methods, page 7, para 3):

"For each participant, we summed the SDD for all benzodiazepine pharmacy fills during the 10-year exposure window to create a cumulative total standardized daily dose (TSDD). We constructed a time-varying measure defined as the TSDD dispensed over a 10-year window after excluding dispensings in the most recent 1 year, which could have been for prodromal dementia symptoms. Figure 2 illustrates how exposure windows were defined. The 10 year window was first calculated using data from prior to ACT enrollment and then moved forward in time throughout follow-up."

d. Even if I am not fully grasping what the authors really did, the methodological choices should be justified more precisely.

Response: We apologize for not being clear about our methodological choices in our manuscript. Because we hypothesized that cumulative medication exposure, in particular heavy total exposure as might be accumulated over a long time period (e.g. several years), would be the most harmful, we measured exposure over a 10 year period. To ensure equal opportunity to assess exposure across participants, we limited our analyses to people who as of ACT enrollment had at least 10 years of prior enrollment in this health care system (which consisted of 86% of the original sample). We also felt that new exposures accumulating during follow up might contribute to dementia risk, so we allowed the window to move forward through time. We have included a rationale for the selection of our exposure window on page 7, para 2:

"Studies have not delineated what pattern of benzodiazepine exposure might be important for increasing risk of dementia (e.g. long-term sustained use versus several episodes of periodic use). We hypothesized that cumulative medication exposure, particularly exposure that might accumulate over a long time period, was the most plausible causal mechanism by which benzodiazepine use could impact dementia risk. Therefore, we selected a 10 year window based on this hypothesis and on methodologic and practical considerations."

11. Appropriateness of assessment period and definition of exposure. The 10-year event-free period used by the authors for measuring exposure before the index date (i.e. date of enrolment in the cohort, or the date of dementia) could be too conservative. Indeed, prevalent users, mostly those exposed at the beginning of the observation period (i.e. 5 to 10 years before the enrolment in the cohort), may have developed dementia before enrolment in the ACT cohort. This could result in a

depletion of susceptible bias when estimating the association between benzodiazepines and dementia. The “prevalent user design” makes this bias more likely since it is likely that a significant proportion of users started their treatment before this observation period. It would make more sense to shorten the observation period (5 years before enrolment could be satisfactory) and to introduce a “new initiator design”. This would partly reduce the likelihood of this bias. Moreover, doing so would probably reduce both the proportion of persons excluded for having fewer than 10 years before enrolment in the ACT cohort (N=674/4724) and the risk of selection bias.

Response: Please see Comment #3 under Committee Review for our discussion about susceptibility bias and the additional text that we added to the discussion section under the limitations section. As this reviewer suggests, a new-user design is one way to handle this issue. However, there are limitations to this approach as well. We did not use a new user design for several reasons. Most importantly, a new user design would not be well suited to examine our principal hypothesis: that cumulative medication exposure, as might accumulate over a long time period, would be associated with dementia. To our knowledge, no prior studies have been able to look back many years or decades to identify true “new users” of benzodiazepines and then follow these subjects forward for many years using rigorous methods to identify incident dementia—an approach that is not likely to be feasible for several reasons. In a typical "new user" design, an extremely large database is often needed to find new users who are then followed forward in time for the outcome of interest, here, incident dementia. This design would make it very difficult to use the rigorous process for ascertainment of dementia outcomes as used in our study, thus forcing the "new user" study to rely on a less optimal outcome definition such as diagnosis codes from administrative or automated data. It is unlikely that a large population based cohort study such as ours or others that followed 5000 people with careful assessments for incident dementia every 2 yrs would be large enough to support a nested "new user" type of study within it. Also, in most cases, studies lack the data needed to truly identify lifetime “new users” of a medication. Instead they are simply identifying people who have not used a medication in a recent, fairly brief time period such as 6 to 12 months. They typically are not able to look back many years to rule out past use. In the case of benzodiazepines, it is very likely that in a “new user” study, many of the apparent “new users” would have been exposed in the more distant past, and thus susceptibility bias could still be present. Lastly, if we were to try to implement a “new user” design within the existing ACT dataset, we would greatly reduce our sample size, affecting precision of estimates. We understand that there are trade-offs between the desirable features of a new user design and other methodologic challenges including the need for high-quality outcome measures.

With regard to shortening the observation period to 5 years, this would create other challenges that could undermine the validity of findings. It would not add that many more people to the sample (we would regain about 14% of the sample or 674 people), and based on our examinations of our pharmacy data, a moderate number of people with a history of benzodiazepine use prior to this 5 year window would be inaccurately classified as “new users.” Furthermore, it would limit our ability to study heavier use, a concern raised by other reviewers (e.g. Reviewer 1; see Comment 8 above.) We undertook some descriptive analyses to illustrate some of our concerns about implementing a new user design to address our study question:

- Among 3434 ACT participants, 2717 (79.12%) had no benzodiazepine fills in the 5 years prior to their ACT baseline visit (and thus would constitute the sample for a new user analysis such as the one suggested by Reviewer 3). Among these 2717 apparent “new users”, 301 (11.1%) had 1+ benzodiazepine fills in the 5- 10 years prior to baseline visit, and therefore are not truly new

users. If we were to look back farther in time, we would undoubtedly find additional participants with a history of more remote use, including some with substantial prior use. This highlights the challenge in trying to apply a "new user" design to study medications that are widely used over the lifespan, and often used episodically, such as benzodiazepines.

12. Definition of chronic users does not seem optimal. The cut-off chosen to define exposure might not make it possible to highlight exposure profiles actually at risk of developing dementia. Indeed, the cut-off chosen by the authors to define chronic use (i.e. >120 TSDD cumulative use during the 10-year observation period) is likely to have mixed chronic users (supposed to be at risk) and sporadic users (not supposed to be at risk). This cut-off was adequately chosen by Olfson et al.1 in their recent study to define chronic use within a one-year observation period. However, keeping the same threshold for a 10-year period is questionable since >120 TSDD may also correspond to occasional uses.

Response: Our intention was not to study "chronic use" per se; we are sorry if that was unclear. In fact our primary focus was on cumulative use over a relatively long time period, which could result from various patterns of use: prolonged use; repeated episodes of use; or short episodes of high dose use. We are not sure what the basis is for the reviewer's statement that sporadic users are 'not supposed to be at risk;' we did not mean to imply this in our manuscript. In fact we are not aware of studies (either epidemiologic, animal or basic science studies) that can provide guidance about what specific pattern of benzodiazepine exposure might be important for increasing dementia risk. Past studies have either based exposure ascertainment on self-report of use in the past 2 weeks (which does not capture chronic use particularly well)(Billioti de Gage 2012) or using a summary of dispensing data such as we have (with no explicit definition of long-term use or requirement that use be continuous).(Billioti de Gage 2014; Imfeld 2015) We intentionally did not describe our highest use group as "long-term" or "chronic" users as we did not characterize usage according to duration or pattern of use. We believe that we share this limitation with the other studies that have been published in this area. We have tried to clarify this issue in a few places throughout the manuscript.

- Please refer to Response to 10.d for inclusion of information in the methods regarding the rationale for our exposure.
- We have clarified the extent of use within the highest benzodiazepine category to give readers an idea of the amount of use in this exposure group (please see response to comment 8 above)
- We have clarified that people could achieve a certain threshold of use by different patterns of benzodiazepine use. Please see Methods (page 8, para 1):

"As examples, a person would reach the highest level of exposure in a 10 year period if he/she took any of the following for a total of 121 days or longer: temazepam 15 mg, triazolam 0.125 mg or lorazepam 2 mg. This level of use could be achieved by daily use for 121 days, or could represent episodic use over several years."

With regard to the concern with the cut-off of 120 used in our study, we also believe that this reviewer's concern is addressed by the additional analyses we have conducted examining benzodiazepine use as a continuous variable, as suggested by the Committee; these results are consistent with our primary analysis (Comment 2 above).

13. Time measurement of confounders may be questionable. The period for measuring putative confounders should be more clearly mentioned in the text (this information seems only provided for age). Since exposure measurement started 10 years before study entry, measuring confounders at this date or one year before does not seem appropriate. The measurement should be made earlier, ideally at the start of exposure.

    Response: We agree that ideally the measurement of confounders should be made at the start of exposure. As discussed above (see response to Comments #3 and 11), it is very difficult to obtain covariate measures prior to exposure to medications like benzodiazepines that are very widely used and also used episodically over a lifetime. In addition, some of our measures were collected specifically for this study (e.g. depressive symptoms, self rated health status) and thus are not available prior to study enrollment. We believe that we share this limitation with the other high quality studies conducted in this area that also have been unable to assess confounders at the time the benzodiazepine was first used.(Billioti de Gage 2012; Billioti de Gage 2014; Imfeld 2015) Given our 10 year exposure window, it is challenging to determine the right time during the window to measure potential confounding factors. Please see #4 under Committee Review for our discussion on this issue. Information about the inclusion of covariates (time-varying versus baseline values) is provided on page 9:

    "*We adjusted for age at study entry, sex, educational level, hypertension, diabetes mellitus, current smoking, stroke, coronary heart disease, body mass index, regular exercise, self-rated health, and depressive symptoms (variables defined in Table 2). We included time-varying measures for coronary heart disease and stroke and the values from the ACT baseline visit for all other covariates*."

    As stated in our response to comment 6 above, we would be happy to add supplementary material or additional text to the manuscript to more fully describe our decisions regarding confounders if the Committee advises.

14. Points related to the discussion section:

    - Comments about the representativeness of the ACT cohort are lacking

    Response: We have clarified this in the discussion section (Page 16, end of 1st para)

    "Lastly, most participants were white and relatively well-educated, and so our results may not be generalizable to other groups".

    - The long-term group as defined by the authors could include heterogeneous exposure patterns, so the ACT cohort might not make it possible to assess the risk of dementia in actual chronic users of benzodiazepines. This putative limitation should be mentioned in the discussion section since it could partly explain the absence of association found in the long-term group.

    Response: See our response to Comment 12 above. We agree with the reviewer that our study cannot draw conclusions about the impact of "chronic" use per se. As mentioned in response to comment 12 above, the prior well-conducted studies in this area also did not examine risk according to pattern of use. We have addressed this concern in the limitations section as suggested (page 16):

    "We defined our exposure groups using a 10 year exposure window and therefore our highest group is likely to include heterogeneous exposure patterns (mix of chronic and episodic users). Likewise,

prior studies have not described or examined pattern of benzodiazpine use. {Billioti de Gage, 2014 #309;Billioti de Gage, 2015 #357} Other designs would be necessary to try to address whether pattern of benzodiazepine use was important for dementia risk."

- Page 13 line 53-56: "This may alleviate concerns raised by prior studies for those people who have a long history of benzodiazepine use". As mentioned above, this sentence should be altered since the design was apparently not optimal for assessing the association in chronic users of benzodiazepines.

Response: As suggested, we have deleted this sentence from the manuscript.

SPECIFIC COMMENTS

15. Abstract (Page 2, line 30). The sentence "The association between 10-year cumulative benzodiazepine exposure" introduces a confusion about the duration of benzodiazepine use in the exposed group and should be altered.

Response: We have altered the abstract to improve the clarity as suggested (See response to comment 10 above).

16. Introduction (Page 4, line 9-10). The authors should mention that the 9-12% benzodiazepine prevalence refers to the US community-dwelling elderly population. Indeed, the prevalence could be much higher in other developed countries.

Response: We have made the alteration as suggested (page 4, first sentence).

"Benzodiazepines are widely prescribed to treat insomnia and anxiety with approximately 9-12% of older adults in the United States reporting use."

17. Introduction (Page 5, line 6-11). The wording of the objective is somewhat confusing. One could deduce that the authors assessed the effect of benzodiazepine use over the 10 subsequent years. In fact, 10 years was the period used for observing exposure, which is quite different.

Response: As suggested, we have made the following edits to clarify the objective (page 5, line 3):
"to evaluate the association between cumulative benzodiazepine use and the risk for dementia and cognitive decline. We hypothesized that higher cumulative use would be associated with increased risk"

18. Methods. The ACT cohort focuses on dementia and relies on a robust diagnosis for dementia cases, which could be more emphasized in the text. For example, the authors should mention who actually made the diagnosis (a GP, a clinician, a neuropsychiatrist, etc.) and provide the criteria used.

Response: As suggested we have added more detail to the description of the diagnosis of dementia cases on page 6, line 10:

"The results, along with clinical data from participants' medical records, were then reviewed in a multidisciplinary consensus conference including the examining physician, a neuropsychologist, another study physician, and the study nurse. The diagnoses of dementia and AD were made using standard research criteria."

Reviewer: 4

19. There is potential for variability in the exposure calculation between this study and that of Billioti de Gage (2014), the most closely related literature describing the relationship between benzodiazepines and dementia. Suggest the authors add a comment about such differences between their approach and Billioti de Gage in the discussion (page 12, line 37-57). The authors note credible reasons for differences found between studies, but this particular difference should emphasized as the measurement of the exposure variable is critically important and this study represents a new method. The use of pharmacy dispensing/claims data is indeed novel, but with conflicting results published in using these data, future work will need to determine merits of the exposure definitions to reproduce or refute these results.

Response: We agree that the Billioti de Gage (2014) study is the most closely related study to ours. Both studies used automated pharmacy data for medication exposure with slight differences in the calculation of the prescribed daily dose (or TSSD in our study) and categorization. The primary difference was the what the total number of mg/prescription was divided by to arrive at the daily dose; in our study it was the minimum effective dose, in the Billioti de Gage study (2014) it was the average daily dose of the source population, which to our knowledge, was not provided in the manuscript to allow comparison with ours. However, the major difference between these studies is how the outcome was ascertained, and we believe that this difference may be more important in explaining the discrepant findings. We have addressed this issue in the discussion section of the manuscript. As suggested we have added the following sentence about differences in exposure measurement in the discussion (page 14, line 7)

"Although the method to calculate and categorize the cumulative daily dose differed in these studies, it is unlikely that these differences explain our discrepant results. Our analysis using a continuous measure for benzodiazepine exposure supported our primary findings of no association with higher doses. We explored extending the lag period used to exclude prescriptions because of prodromal symptoms to more closely match the Canadian study and continued to find no association."

20. The authors approach to CASI scores with sensitivity analysis using IRT is commendable. However, CASI, as with other cognitive screening tests, is less reliable among those with higher cognitive function (sensitivity and specificity are established for diagnosis of dementia with very acceptable characteristics). The authors report in this study a population without cognitive impairment at baseline and use an outcome of dementia, with little validity to compare the intermediate phases of mild cognitive impairment. Certainly the CASI analysis is a credible attempt to address such changes, however the discussion section could be improved by an additional comment addressing limitations of cognitive screening measures in the transition from (and between) normal cognition, MCI, and mild dementia.

Response**:** As the reviewer states nicely, the CASI on its own is not able to reliably identify cases of mild cognitive impairment, so we cannot comment on whether benzodiazepine use may be associated with risk of MCI. Please see response to Comment #9 above.