PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	OCCURRENCE, MORTALITY, AND COST OF BRAIN DISORDERS IN DENMARK: A POPULATION-BASED COHORT STUDY
AUTHORS	Vestergaard, Søren; Rasmussen, Thomas; Stallknecht, Sandra; Olsen, Jens; Skipper, Niels; Sørensen, Henrik T.; Christiansen, Christian

VERSION 1 – REVIEW

REVIEWER	Peter Herbison University of Otago New Zealand
REVIEW RETURNED	21-Feb-2020

GENERAL COMMENTS	This is an interesting paper that is clearly reported. The one thing I found to comment on was that this was a matched study but it was not clear whether the analysis was matched or not. If it was not matched then I would have expected a reason for this decision.

REVIEWER	Sebastian Trautmann Medical School Hamburg
REVIEW RETURNED	11-Jun-2020

GENERAL COMMENTS	This study reports on estimates of the occurrence of mental and neurological disorders as well as associated mortality and costs.They provide potentially important data which could inform health care policies and funding schemes. I only have a few suggestions. 1. It is not exactly clear how the 25 groups of disorders were chosen. The considered groups are very heterogenous. I wonder whether it makes sense to treat headache (including forms of headache with relevant etiological mechanisms outside the brain) the same way as anxiety disorders. The same is true for the variety of infections of the central nervous system. How is this selection comparable to earlier studies? Moreover, the choice of included diagnoses per group seems arbitrary in parts. For example, why was alcohol abuse included, but not alcohol dependence? A better explanation of the rationale for the included groups and diagnoses is needed. 2. There are alternative methods to the human capital approach. The authors might add a short discussion of advantages and disadvantages using this one

	3. Why was the rather crude CCI score used to adjust for comorbidity and not the actual number of comorbid conditions? 4. As is always the case with registry data, the diagnoses are of unclear validity, probably far below high quality epidemiological data using diagnostic interviews. Even small amounts of measuring bias can accumulate heavily in such large samples. For example, it seems odd that depression and stress-related disorders occur with about the same probability, although depression is much more prevalent in epidemiological studies. I think this warrants further discussion. 5. I think more context is needed to interpret the findings, especially as the authors compare their data to earlier estimates for entire Europe. What is the expected prevalence given high quality epidemiological data in Denmark. How are mortality and health care expenditures compared to other regions.

REVIEWER	Poul Jennum Danish Center for Sleep Medicine Department of Clinical Neurophysiology Rigshospitalet Copenhagen Denmark
REVIEW RETURNED	27-Jun-2020

GENERAL COMMENTS	Current study seeks to clarify the overall national health costs of brain disease. This is done based on the use of national Danish registry databases. This is a previously published method. Thus, this is an important issue that has health policy significance. However, there are some conditions that make data difficult to understand data and their generalization. 1. data selection. Data is primarily selected from the Danish Hospital Diagnostic Register. In this way, there are no diagnoses that are primarily seen by specialists in the primary sector. Diagnoses such as Parkinson's disease, epilepsy, headaches are underestimated. 2. Some diagnoses are selected primarily for drug use; here you can doubt if you overlook untreated patients. 3. Selection of controls: it is stated that controls 10 are selected per diagnosis. This is probably realistic in the case of rare diagnoses, but in common diagnoses (apoplexy, dementia) comparative controls are unlikely to be found. This must be explained. 4. There are approximately 1 million people in the cohort, which corresponds to about 1/5 of the Danish population. This is undoubtedly the casemix of multi- or concomitant comorbidity. This gives a high probability of double counting. 5. All controls are completely free of neurological disease; in which case it is devoid of brain disease.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

This is an interesting paper that is clearly reported. The one thing I found to comment on was that this was a matched study but it was not clear whether the analysis was matched or not. If it was not matched then I would have expected a reason for this decision.

Response: We appreciate that the Reviewer considered the paper interesting and clearly reported. We agree with the reviewer, that matching persons with brain disorders to persons without these disorders is crucial to isolate the excess mortality and costs in patients with brain disorders. We now clarified in the abstract that the analyses were matched: *"Prevalence and incidence of hospital-diagnosed brain disorders, 1-year absolute and relative mortality, and attributable direct and indirect costs-of-illness compared with the corresponding matched cohorts."* (p. 52, ll. 39-42). We provide a more detailed description of the matched analyses in the methods section (p. 57, l. 9 - p. 60, l. 11)

Reviewer: 2

This study reports on estimates of the occurrence of mental and neurological disorders as well as associated mortality and costs. They provide potentially important data that could inform health care policies and funding schemes. I only have a few suggestions.

1. It is not exactly clear how the 25 groups of disorders were chosen. The considered groups are very heterogeneous. I wonder whether it makes sense to treat headache (including forms of headache with relevant etiological mechanisms outside the brain) the same way as anxiety disorders. The same is true for the variety of infections of the central nervous system. How is this selection comparable to earlier studies? Moreover, the choice of included diagnoses per group seems arbitrary in parts. For example, why was alcohol abuse included, but not alcohol dependence? A better explanation of the rationale for the included groups and diagnoses is needed.

Response: We are pleased, that the Reviewer considers our study to be potentially important. We agree, that groups of persons with brain disorders are heterogeneous. In this study, we focused on the major groups of neurological and mental disorders, and we prioritized including disorders previously reported to be important (common or severe conditions). We found inspiration for the grouping in a large-scale European study which included 19 different disorders: addictive disorders, affective disorders, anxiety disorders, brain tumour, childhood and adolescent disorders (developmental disorders), dementia, eating disorders, epilepsy, mental retardation, migraine, multiple sclerosis, neuromuscular disorders, Parkinson's disease, personality disorders, psychotic disorders, sleep disorders, somatoform disorders, stroke, and traumatic brain injury.[1]
However, with input from clinical experts in the field, we expanded the list of disorders, to encompass 25 disorders in all. All the clinical experts were invited to a face-to-face meeting to discuss the preliminary results and were acknowledged in the manuscript. As the aim of the study was to give an overview of a wide range of disorders, some degree of heterogeneity was necessary.

In order to compare estimates, we applied the same methods to all the included disorders, even though the nature of each group of disorder is different. Thus, our study is an overview, and detailed studies of each specific disorder are needed to further examine detailed aspects of their occurrence, mortality and costs.

Regarding "alcohol abuse" mentioned by the Reviewer. We named the group "alcohol abuse" and not "alcohol dependence" as dependence is a more narrow clinical diagnosis. We wanted to include also persons with diagnosed abuse who did not necessarily fulfil the criteria for dependence. Yet, we chose not include acute intoxication, as persons receiving a diagnosis of a single episode of alcohol intoxication do not necessarily have an alcohol abuse. Thus to minimize misclassification, we chose not to include this condition. This approach was supported by a professor and expert in alcohol abuse and alcohol-related disorders, who is acknowledged in the manuscript, with advise to include other relevant diagnoses reflecting alcohol abuse (ICD-10: F10 (excl. F10.0), E244, G312, G621, G721, I426, K292, K70, K852, K860, Q860).

After this pertinent advised from the Reviewer, we elaborated on our rationale for the group selection, an reorganized this section in the manuscript: "*The following 25 predefined groups of brain disorders were examined: alcohol abuse, anxiety disorders, bipolar disorder, brain tumours, cerebral palsy, dementia, depression, developmental and behavioural disorders, drug abuse, eating disorders, epilepsy, headache, infections of the central nervous system, intellectual disability, multiple sclerosis, neuromuscular disorders, other neurodegenerative disorders, Parkinson's disease, personality*

disorders, polyneuropathy, schizophrenia spectrum disorders, sleep disorders, stress-related disorders, stroke, and traumatic brain injury. Disorders were selected if expected to be common or critical, and we prioritized to select groups of disorders examined in previous studies to enable comparison of our results.[1, 2] For each of 25 specific brain disorders, we established two cohorts: a prevalent cohort of persons alive on 1 January 2015 who had a diagnosis of the specific brain disorder recorded during the 1995-2014 period and an incident cohort of persons with a first-time diagnosis recorded during the 1 January 2011 to 31 December 2015 period." (p. 56, ll. 23-45)

2. There are alternative methods to the human capital approach. The authors might add a short discussion of advantages and disadvantages using this one

Response: We agree that a discussion of advantages and disadvantages of the method is highly relevant, so we added following to the manuscript: "*In estimating the indirect costs (i.e. loss of productivity) of illness and premature death we applied the human capital approach. In the literature, this approach has been discussed and among others it has been argued that application of the human capital approach leads to an over-estimation of the productivity costs. Hence, alternative methods like the friction cost method have been proposed.[3] The idea behind the friction cost method is that the amount of production lost due to disease depends on the time (friction period) organizations need to restore the initial production level. Friction periods will differ by industry, type of work etc. and the challenge is to estimate relevant friction periods but application of the friction cost method leads to lower productivity cost estimates.[3] Taking these considerations into account, we present our indirect cost estimates separately (Figure 6) making it possible to assess the results without inclusion of the indirect costs.*" (p. 64, l. 42 – p. 65, l. 7)

3. Why was the rather crude CCI score used to adjust for comorbidity and not the actual number of comorbid conditions?

Response: We acknowledge that the CCI score is an aggregated measure of comorbidity, and adjusting for each of the disorders separately could minimise risk of residual confounding. Still, adjustment for the CCI score has been shown to account for most confounding by comorbidity in analyses of mortality.[4] The advantage of using an aggregated measure of comorbidity burden is the ability to adjust for comorbidity even when you have few outcomes (deaths), e.g., eating disorders, multiple sclerosis, cerebral palsy, and personality disorders (Figure 3). Adjustment for individual comorbid conditions when possible may reduce confounding, but to enable comparison of the mortality across the groups, we found it crucial to apply the same mortality analyses to all the included disorders,. As a sensitivity analysis, we have now repeated the Cox-regression adjusting for each of the 19 comorbid conditions separately for the *any brain disorder* cohort, which have sufficient number of outcomes. In this analysis, we found an adjusted HR of 4.9 (95% CI: 4.9 – 5.0), which was very close to the HR found our main analyses (HR=4.7 [95% CI: 4.7-4.8]). Thus, beside it being feasible, we consider it appropriate to adjust for CCI score when adjusting mortality for comorbid somatic conditions. To underline the limitation of adjusting for the CCI score in the manuscript, we now updated the section in the discussions: "*We described non-mental comorbidity as the proportions of persons in each cohort previously diagnosed with diseases from the CCI covering 19 groups of disorders. CCI score was used to adjust for confounding in our mortality analyses, and as CCI score is an aggregated measure of comorbidity we cannot rule out residual or unmeasured confounding in our estimates of HRs of death.*" (p. 64, ll. 16-26)

4. As is always the case with registry data, the diagnoses are of unclear validity, probably far below high quality epidemiological data using diagnostic interviews. Even small amounts of measuring bias can accumulate heavily in such large samples. For example, it seems odd that depression and stress-related disorders occur with about the same probability, although depression is much more prevalent in epidemiological studies. I think this warrants further discussion.

Response: We agree, that well-performed diagnostic interviews may decrease misclassification in our study for some disorders. On the other hand, the nationwide registry data allowed us to examine a sample of millions of individuals with minimal selection at inclusion (i.e., participants did not influence whether they participated) and minimal loss-to-follow up over time. We agree that some persons with mild events of some disorders e.g., depression, anxiety, and headache, were not included in the main analyses of our study if they did not seek hospital care for the condition.
A previous European study reported a 1-year prevalence of major depression of 6.9% and bipolar disorders of 0.9%.[1] Also, a Danish study used a validated questionnaire to screen for depression,

finding a prevalence of major depression of 3.3%,[5] close to the prevalence of depression of 3.2% in our main analyses. This indicates that we also primarily capture major depression, which is discussed in the manuscript (p. 63, l. 47 – p. 64, l. 7). However, in sensitivity analyses we found that 12.6% of Danes had either diagnosed depression or filled prescriptions of antidepressants with depression as prescription indicator within the prior 20 years (Suppl. Figure 4). Thus, when filled prescription of antidepressant is included as proxy for depression, the prevalence of depression was 4-fold higher than the prevalence of stress-related disorders (Suppl. Figure 4). We discussed this limitation in the manuscript: "*We may have underestimated the prevalence, incidence, and total cost of non-severe brain disorders, as some patients were treated solely in general practice or were undiagnosed. This is especially relevant for disorder that are mainly treated in primary care, or not treated at all, and therefore were not captured in our main analyses such as anxiety,[6] headache,[7] and sleep disorders.[8] We addressed this in a sensitivity analysis that also identified patient-based filled prescriptions for relevant medications.* " (p. 63, l. 47 – p. 64, l. 7)

Furthermore, we now addressed the prevalence estimates that were markedly different from previous ones in the largest European study of brain disorders "*Compared to the reported prevalence of separate disorders, we found markedly lower prevalence of anxiety (1.7% vs 12%), headache (1.7% vs 10%), depression (3.2% vs 6.5%), and sleep disorders (1.1% vs 8.7%), despite the longer lookback in our study (20-years vs. 1-year prevalence).[1] This may be explained by different data sources, as we in the main analyses only included persons with hospital-diagnosed disorders. Importantly, when we included persons with filled prescriptions in sensitivity analyses prevalence increased considerably*." (p. 65, ll. 27-41).

5. I think more context is needed to interpret the findings, especially as the authors compare their data to earlier estimates for entire Europe. What is the expected prevalence given high quality epidemiological data in Denmark. How are mortality and health care expenditures compared to other regions.

Response: We agree that a more detailed discussion of the context and perspectives of our findings would be interesting. Within the allowed word limits of original articles, we now elaborated on the perspectives focusing on dementia as an example "*While we had detailed data on direct costs, we lacked information on municipally supported rehabilitation, assistance supplies, and transportation costs related to treatment and rehabilitation. Similarly, our cost analyses did not include intangible costs (e.g., due to decreased quality of life) and costs of informal care provided by relatives, which may be considerable in conditions such as dementia.[9] Yet, we included costs of nursing homes, sheltered accommodation, and home nursing, and we found the annual cost per person with dementia (€30K) similar to that previously reported in developed countries.[10, 11]*" (p. 64, ll.28-40)
We prioritized describing the methods we used in detail, and to report and discuss the main findings in the manuscript. Moreover, we provided all the detailed results in the supplementary to enable the readers of BMJ Open to put our findings into perspective. Importantly, to further enable comparison of costs in Denmark with costs other countries, we estimated the annual cost per persons with brain disorders and provided these figures both graphically (Figure 5 and Suppl. Figure 3) and tabulated (Suppl. Table 4). To keep the manuscript readable, we did not find it feasible to discuss the prevalence, incidence, mortality and costs of each of the 25 specific disorders in the relation to previous findings. We hope that the readers will find it both accessible and informative.

Reviewer: 3

Current study seeks to clarify the overall national health costs of brain disease. This is done based on the use of national Danish registry databases. This is a previously published method. Thus, this is an important issue that has health policy significance. However, there are some conditions that make data difficult to understand data and their generalization.

1. Data selection. Data is primarily selected from the Danish Hospital Diagnostic Register. In this way, there are no diagnoses that are primarily seen by specialists in the primary sector. Diagnoses such as Parkinson's disease, epilepsy, headaches are underestimated.

Response: We are grateful, that the Reviewer consider the issues we addressed to have health policy significance. We agree on this insightful comment, that occurrence and costs of some disorders may be underestimated more than others, and we aimed to address this appropriately in the discussion section "*We may have underestimated the prevalence, incidence, and total cost of non-*

severe brain disorders, as some patients were treated solely in general practice or were undiagnosed. This is especially relevant for disorder that are mainly treated in primary care, or not treated at all, and therefore were not captured in our main analyses such as anxiety,[6] headache,[7] and sleep disorders.[8] We addressed this in a sensitivity analysis that also identified patient-based filled prescriptions for relevant medications." (p. 63, l. 49 – p. 64 l.7) and "*Compared to the reported prevalence of separate disorders, we found markedly lower prevalence of anxiety (1.7% vs 12%), headache (1.7% vs 10%), depression (3.2% vs 6.5%), and sleep disorders (1.1% vs 8.7%), despite the longer lookback in our study (20-years vs. 1-year prevalence).[1] This may be explained by different data sources, as we in the main analyses only included persons with hospital-diagnosed disorders. Importantly, when we included persons with filled prescriptions in sensitivity analyses, the prevalence increased considerably indicating that we underestimated the occurrence of these disorders in our main analyses*" (p. 65, ll. 27-41).

2. Some diagnoses are selected primarily for drug use; here you can doubt if you overlook untreated patients.

Response: It is a highly relevant point, that not only do we underestimate some disorders due to lack of primary care data, but additionally for some disorders such as drug abuse and alcohol abuse a considerable proportion of patients are not treated at all. We now emphasized this in the discussions section "*We may have underestimated the prevalence, incidence, and total cost of non-severe brain disorders, as some patients were treated solely in general practice, or were undiagnosed or untreated. This is especially relevant for disorder that are mainly treated in primary care, or not treated at all…*" (p. 63, ll. 49-56)

3. Selection of controls: it is stated that controls 10 are selected per diagnosis. This is probably realistic in the case of rare diagnoses, but in common diagnoses (apoplexy, dementia) comparative controls are unlikely to be found. This must be explained.

Response: Certainly, when disorders are common, it can be difficult to sample enough controls matched on age and sex. In this study, we sampled controls with replacement from the entire Danish population as recommended,[12] which enabled us to sample 10 comparisons of same age and sex per diagnosed individual requiring that comparisons were free of that specific disorder at the matching date. We now clarified that we sampled with replacement in the manuscript "*Each person in the brain disorder cohorts was matched to 10 living persons from the general population on birth year and sex (sampled with replacement).[12]*" (p. 57, ll. 14-16).

4. There are approximately 1 million people in the cohort, which corresponds to about 1/5 of the Danish population. This is undoubtedly the casemix of multi- or concomitant comorbidity. This gives a high probability of double counting.

Response: It is correct that many persons in our study had concurrent/comorbid brain disorders, as shown in Suppl. Figure 2 and mentioned in the results and discussion sections. We were indeed aware of the risk of double counting single individuals when estimating occurrence and costs, and we addressed this by only including each individual in the *any brain disorder* cohort. We now clarified this important point in the abstract "*We identified 1,075,081 persons with at least one prevalent brain disorders (any brain disorder) on 1 January 2015, corresponding to 18.9% of the Danish population*" (p. 52, ll. 46-49). Also, we now underlined it in the methods sections: "*We also identified every Danish resident with any brain disorder, i.e., each person with any of the specific 25 disorders was identified on the date of his or her first diagnosis. To avoid double-counting, every person could only be included once in the any brain disorder cohort*" (p. 57, ll. 4-10) As we included only unique individuals in the *any brain disorder* cohort, the number of persons in this cohort is not equal to the sum of persons in the cohorts with the 25 specific disorders. Likewise we counted each cost component only once in the *any brain disorder* cohort, and if combining the cost of each of the 25 specific brain disorders, the sum by far exceeded the costs of persons with *any brain disorders*.
We agree with the Reviewer, that it is remarkable that on in five Danes have been diagnosed in hospitals with at least one brain disorder during the prior 20 years, and this finding is consistent previous findings from Denmark and Europe.

5. All controls are completely free of neurological disease; in which case it is devoid of brain disease.

Response: As correctly mentioned by the Reviewer, the matched comparisons to the *any brain disorder* cohort were required not to have had any brain disorder at the date of matching. In the any brain disorder cohort it was only possibly to sample 10 comparisons for each person with a disorder because we sampled with replacement as recommended.[12] However, when matching comparisons for one of the 25 specific disorders, we restricted to persons without that specific disorder (i.e. we did not restrict to persons without any brain disorder in these 25 comparison cohorts). We now clarified this important aspect in the manuscript: "*Matched persons could not have the brain disorder of interest as of the index date of the person with the brain disorder.*" (p. 57, ll. 16-19)

REFERENCES

1. Gustavsson A, Svensson M, Jacobi F, et al. Cost of disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol* 2011;21(10):718-79. doi: 10.1016/j.euroneuro.2011.08.008

2. Whiteford HA, Ferrari AJ, Degenhardt L, et al. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. *PLoS One* 2015;10(2):e0116820. doi: 10.1371/journal.pone.0116820

3. Drummond M, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. 3. ed. ed. Oxford: Oxford University Press 2005.

4. Schneeweiss S, Seeger JD, Maclure M, et al. Performance of Comorbidity Scores to Control for Confounding in Epidemiologic Studies using Claims Data. *American Journal of Epidemiology* 2001;154(9):854-64. doi: 10.1093/aje/154.9.854

5. Olsen LR, Mortensen EL, Bech P. Prevalence of major depression and stress indicators in the Danish general population. *Acta Psychiatrica Scandinavica* 2004;109(2):96-103. doi: 10.1046/j.0001-690X.2003.00231.x

6. Alonso J, Angermeyer MC, Bernert S, et al. Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatr Scand Suppl* 2004(420):21-7. doi: 10.1111/j.1600-0047.2004.00327.x

7. Stovner L, Andrée C. Prevalence of headache in Europe: A review for the Eurolight project. *The journal of headache and pain* 2010;11:289-99. doi: 10.1007/s10194-010-0217-0

8. Kerkhof GA. Epidemiology of sleep and sleep disorders in The Netherlands. *Sleep Med* 2017;30:229-39. doi: 10.1016/j.sleep.2016.09.015

9. Frahm-Falkenberg S, Ibsen R, Kjellberg J, et al. Health, social and economic consequences of dementias: a comparative national cohort study. *European journal of neurology* 2016;23(9):1400-07. doi: 10.1111/ene.13043

10. Schaller S, Mauskopf J, Kriza C, et al. The main cost drivers in dementia: a systematic review. *Int J Geriatr Psychiatry* 2015;30(2):111-29. doi: 10.1002/gps.4198

11. Cantarero-Prieto D, Leon PL, Blazquez-Fernandez C, et al. The economic cost of dementia: A systematic review. *Dementia (London)* 2019:1471301219837776. doi: 10.1177/1471301219837776

12. Heide-Jørgensen U, Adelborg K, Kahlert J, et al. Sampling strategies for selecting general population comparison cohorts. *Clin Epidemiol* 2018;10:1325-37. doi: 10.2147/clep.S164456

VERSION 2 – REVIEW

REVIEWER	Peter Herbison University of Otago New Zealand
REVIEW RETURNED	12-Aug-2020

GENERAL COMMENTS	I think it is still unclear whether a matched or unmatched analysis was done in this study. This is an individually matched case control study. In these studies it is normal to perform a matched analysis when comparing cases and controls. For binary outcomes (such as mortality) instead of logistic regression you carry out conditional logistic regression. Now in SAS conditional logistic regression uses the proportional hazards regression command (PHREG). The authors say that this command is what they used but they included age and sex as co-variates. These were the variables used in the matching process, so carry no information, which makes it unclear to me whether conditional, or unconditional methods were used(i.e. matched or unmatched). This needs to be clarified and if unmatched analyses were carried out, this needs to be explained. SAS provides information on matched analyses at https://support.sas.com/resources/papers/proceedings/ proceedings/sugi29/208-29.pdf. Similarly for continuous outcomes a matched analysis should be done, using something like PROC MIXED with a variable for each matched set used as a random effect. To understand the results it is necessary to clearly specify what analysis was done.

REVIEWER	Sebastian Trautmann Medical School Hamburg, Germany
REVIEW RETURNED	04-Aug-2020

GENERAL COMMENTS	The authors were very responsive to the reviewer comments. I have nothing to add.

REVIEWER	Poul Jennum Danish Center for Sleep Medicine, Department of Clinical Neurophysiology Rigshospitalet Copenhagen Denmark
REVIEW RETURNED	03-Aug-2020

GENERAL COMMENTS	The author have addressed my comments.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

I think it is still unclear whether a matched or unmatched analysis was done in this study. This is an individually matched case control study. In these studies, it is normal to perform a matched analysis when comparing cases and controls. For binary outcomes (such as mortality) instead of logistic regression, you carry out conditional logistic regression. Now in SAS conditional logistic regression uses the proportional hazards regression command (PHREG). The authors say that this command is what they used but they included age and sex as co-variates. These were the variables used in the matching process, so carry no information, which makes it unclear to me whether conditional, or unconditional methods were used (i.e. matched or unmatched). This needs to be clarified and if unmatched analyses were carried out, this needs to be explained. SAS provides information on matched analyses at https://support.sas.com/resources/papers/proceedings/proceedings/sugi29/208-29.pdf.

Similarly for continuous outcomes a matched analysis should be done, using something like PROC MIXED with a variable for each matched set used as a random effect.

To understand the results it is necessary to clearly specify what analysis was done.

Response: We thank for this pertinent advice, and we have now specified that we used "…*an unstratified Cox regression model adjusted for age, sex, and CCI score.*" (p. 58, ll. 49) Indeed the Reviewer is correct, that if our study was a matched case-control study, we needed to apply matched analyses (e.g. conditional logistic regression) to avoid bias. In the current matched cohort study, we could have chosen to do a conditional analysis accounting for the matching strata (e.g., stratified Cox regression) but given the study design, we would expect that our use of an unmatched analysis would give virtually the same estimates.[1, 2] We matched on birth year and sex only, and using an unconditional model is appropriate when matching on few variables (loose matching).[3] Of note, as we further adjusted for comorbidity (CCI score) which was not matched on, it was necessary to also adjust for the matching variables to avoid potential bias.[4]

To address the concern of the Reviewer, we wanted to ensure that the use of a conditional model would not change the results in our study. Therefore, we repeated the mortality analysis for the *Any brain disorder* cohort including hospital-diagnosed as in the main analysis. Using a stratified Cox regression (PHREG with a strata statement including the matching variables) yielded a HR effect estimate of 4.76 (95% CI: 4.70-4.83). This is very close to the HR estimated by the unstratified Cox regression (PHREG without a strata statement) in our main analysis (HR of 4.74 [95% CI: 4.68-4.81]). This suggests that the difference between the stratified and unstratified models is minimal in this study, and for the above stated reasons, we find it appropriate not to change the current analyses.

In the sensitivity analyses, the ordinary least squares (OLS) regression was not performed on the matched populations. Rather it included every Danish resident and their total direct health care costs, with age, sex, and each of the 25 brain disorders as explanatory variables. After this relevant review comment, we have now explained this in the manuscript: "*In addition, we performed an ordinary least squares (OLS) regression to compute attributable direct costs, in which we modelled the average annual costs per person for each group of brain disorders. This OLS regression included every Danish resident (not only the matched cohorts) with age, sex, and each of the 25 brain disorders as explanatory variables, and thus, costs of every brain disorder was adjusted for costs of comorbid brain disorders."* (p. 59, l. 53 – p. 60, l. 9)

Reviewer 2
The authors were very responsive to the reviewer comments. I have nothing to add.

Response: We are pleased, that our response satisfied the Reviewer.

Reviewer 3
The author have addressed my comments.

Response: We are pleased, that our response satisfied the Reviewer.

References

1. Cummings P, McKnight B, Greenland S. Matched Cohort Methods for Injury Research. Epidemiologic Reviews. 2003;25(1):43-50.

2. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. 758 p.

3. David WH, Stanley L, Susanne M. Applied survival analysis : regression modeling of time-to-event data. 2. ed. ed. Hosmer DW, Lemeshow S, May S, editors. Hoboken, N.J: Wiley-Interscience; 2008.

4. Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. International journal of epidemiology. 2013;42(3):860-9.

VERSION 3 – REVIEW

REVIEWER	Peter Herbison University of Otago New Zealand
REVIEW RETURNED	25-Aug-2020

GENERAL COMMENTS	I am now happy with this paper. With a ample this size it is appropriate to do an unmatched analysis adjusting for the matching factors. But it must be clear that that was what was done.