PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Depression among Adult Patients with Primary Brain Tumor: A Cross-Sectional Study of Risk Factors in a Low-Middle Income Country
AUTHORS	Pidani, Anum; Siddiqui, Amna Rehana; Azam, Iqbal; shamim, Muhammad Shahzad; Jabbar, Adnan; Khan, Shameel

VERSION 1 – REVIEW

REVIEWER	Adomas Bunevicius University of Virginia, USA
REVIEW RETURNED	21-Jul-2019

GENERAL COMMENTS	In this study authors evaluated predictors of self reported depression (PHQ-9) in heterogenous sample of brain tumor patients in terms of tumor diagnoses and stage of therapy. They found that depression symptoms were associated with worse functional status. KPS score was associated with employment status, treatment stage, and tumor recurrence. While this is the first study of this kind performed in Pakistan, however results largely represent clinical scenario in other countries. Aside from insurance status, it is not clearly described why risk factors of depression should be different in Low-middle income countries vs. other countries. Abstract: authors say “biopsy proven”. This is not clear as most patients undergo tumor resection with subsequent pathology report that establish brain tumor diagnosis. Introduction, para 2: authors describe definition of depression according to WHO. This is not acceptable definition of depression because it has clear criteria as defined by the DSM and is used for clinical practice and research studies. Were there any restriction with regards to tumor diagnosis. It is important specify that because prognosis and patient symptoms can be very different in gliomas (intrinsic and incurable brain tumor) vs. meningiomas or pituitary adenomas. Indeed, the sample the majority pf patients had benign tumors, such as meningiomas and pituitary adenomas, and only 21 had high grade glioma This high variability of different types of brain tumor in the samples challenge generalizability of the study findings Why did authors selected p-value threshold of =<0.25?

	Looking into Table 2, authors should provide p values for analyses performed, because 95% values of some predictors cross 1, thus questioning if they were indeed significantly associated with outcome of interest. This needs to be checked.

REVIEWER	Alasdair G. Rooney University of Edinburgh Scotland (UK)
REVIEW RETURNED	31-Jul-2019

GENERAL COMMENTS	The authors sought to identify clinical and demographic characteristics associated with depression in Pakistani brain tumour patients. They conducted a single-centre cross-sectional study (n=132) using the PHQ-9 >10 to identify a positive screening prevalence of 39%. Poorer functional status (KPS) was independently associated with higher depression scores. The main strength of the paper is its novel focus on 'non-Western' patients, which are well worth studying given the social and healthcare differences between first world and low/middle-income countries such as Pakistan. Other strengths include prospective design, a relatively large sample size, interesting data on strategies used to handle stress, and the use in the PHQ-9 of a partially-validated outcome scale in this clinical population. The use of the STROBE criteria checklist is also noted. General comments: Minor grammatical errors are scattered throughout which could be addressed in editorial. "Depression" is a term which can be used to encompass conditions ranging from the clinical syndrome of DSM-V Major Depressive Disorder, to general lay sadness. Here it appears to imply a clinical diagnosis, which the PHQ-9 alone cannot make. The PHQ-9 just measures 'depressive symptoms'. This should be addressed – such as by replacing 'depression' with the term 'depressive symptoms' throughout, or something similar. Abstract: P1 Line 49. The prevalence figure for high depressive symptoms (39%) should have 95% confidence intervals added. P2 Line 5. Under the heading "Strengths and limitations of this study", it's not immediately clear which of the given statements are strengths and which are meant as limitations. P2 Line 7. The authors state, "The major strength of this study is its ability to analyze data using robust statistical techniques." It's true that attempts are made to control for confounders, but even 132 patients give only limited power for regression analyses with over 30 measured variables. Nor – as an observational study – can one ever be sure that every important confounder has been controlled. I think it's questionable as to how 'robust' statistics can be in this limited context and suggest the authors amend or delete this statement. Introduction:

	There are several problems with the references that would benefit from careful review. A few examples: - The World Health Organisation definition of depression is referenced to [5], which is in fact a paper by Mainio et al that (unless I missed it) does not mention the World Health Organisation's definition of depression. - The authors state: "The worldwide prevalence of depression in cancer patients is 25% with higher rates among Asian countries [8]." The cited reference (Ostrom et al) is a paper describing the aetiology of glioma that does not appear to contain any statements about the prevalence of depression in cancer. - The quoted prevalence range of 10-40% is referenced to [10], which is a technical and theoretical review by Starkweather et al. of how we should conceptualise depression in these patients, and (again unless missed) does not present a simple prevalence range. - References [7] and [12] are the same paper. - Later in the text of the Methods section, Ref [23] is said to be a paper by Gholizadeh et al, but in the reference list, Ref [23] is the review on symptom clusters by Fox et al. For reasons of time I will leave it there but these examples hopefully illustrate that the references need line by line review. Methods P3 Line 20. "The exclusion criteria for study participants were as follows: diagnosis of depression prior to the diagnosis of brain tumor…" Can the authors please clarify how far back they looked (was this any lifetime diagnosis of depression, or just immediately prior to brain tumour diagnosis?) and also what was the reason for excluding these patients? Previous depression is one of the strongest predictors of future depression, so excluding it risks potential bias. How did the authors determine tumour location – did all patients have contrast enhanced brain imaging? How did they determine cognitive impairment? Statistical analysis P4 Line 6. What was the aim of the power calculation – power to do what? Sometimes in observational studies power calculations focus on narrowing 95% confidence intervals down to acceptable limits. But 95% CI are not mentioned. If they were the reason for the power calculation, what limits were the authors aiming for – and if they were not, what were you looking for enough power to do? P4 Line 13. Could the authors please reference a suitable precedent (and give the rationale) for calling significance at p=0.25 for univariate analyses? I've seen it done at p=0.1, but the approach taken here is even more permissive. P values aren't the be-all and end-all but there should be some precedent for the method, or alternatively they could re-do the analysis with a limit of p=0.1.

	P4 Line 16. “Propensity scores were computed to identify factors associated with functional status.” I don't know what propensity scores are (or how they are computed) – could the authors please give more detail? Results P7 Line 44. These propensity scores are then used in further analyses (beyond what is stated in Methods) as follows: “Propensity scores predicted from [the] above model were significantly associated with depression. Table 3 shows models to demonstrate [the] association of functional status (KPS) with depression and propensity scores for functional status (KPS) with depression.” These further analyses need to be described in Methods too – how they were done and why – because at the moment it is not clear what they mean. In particular I'm not clear on the difference between KPS and a propensity score for KPS. Discussion Limitations of the study should be adequately discussed. At the moment they are not mentioned. Table 1. A third column should ideally be added with p values for each univariate comparison, to allow readers to quickly scan all the associations. - Travel cost: please state the units of the numbers given (is it rupees?) - Overall treatment cost: I am not sure what "lac" rupees means. -‘Strategies to handle stress’, ‘Brain structures involved’ and ‘First symptoms involved before brain tumour diagnosis’ add up to >132. I realise why but these variables should be marked with an asterisk or something and the reason denoted in the table legend, for clarity. -‘Tumour grade’ only totals 131 – one patient is missing? Table 2. The presentation of this Table is a little unclear. I can see that for a given number of levels of a variable there are “n-1” entries in the column “PR + 95% CI”, but I am wondering if those for Factor 1 (Current Employment Status) are placed on the wrong rows. Could the authors please review that this table is laid out exactly as they intended? Also please expand slightly on what the relevance of the “Reference Category” asterisk is, in the table legend. (Is it meant with reference from, or with reference to?) Table 3. The methodology behind this analysis needs clarified in Methods as noted above.

	Other It would be good to see a Table summarising the average scores / frequencies for each PHQ-9 symptom. Often brain tumour patients may report significant somatic symptoms (sleep, appetite, fatigue etc) without experiencing the "core" symptoms of depressed mood or anhedonia. Indeed, it would be good to know specifically how many patients scored >10 on PHQ-9 without endorsing either low mood or anhedonia. The frequency of high scoring on the suicide item 9 would be good to know. Were any patients taking antidepressants?

REVIEWER	Robert Greevy Vanderbilt University Medical Center
REVIEW RETURNED	17-Dec-2019

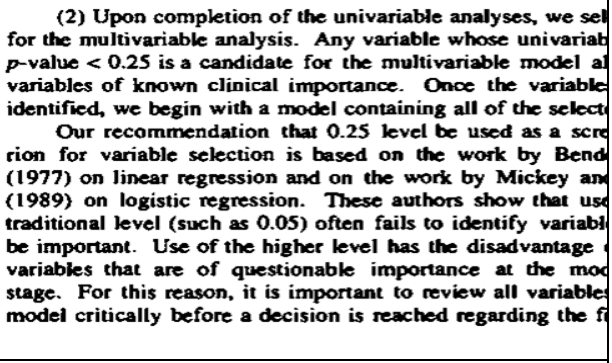
GENERAL COMMENTS	In this review of “Depression among Adult Patients with Primary Brain Tumor: A Cross-Sectional Study of Risk Factors in a Low-Middle Income Country”, I will focus primarily on the methodology. Major comments: Please clarify what is meant by “Cox algorithm regression” in the abstract and “logistic regression cox algorithm” in the methods. Explain how and why the Cox algorithm is being used. Please clarify how and why propensity scores (PS) are being used. As I’m reading the paper, it appears PS are being used primarily for modeling: Depression ~ KPS + PS, where PS comes from KPS ~ employment status + treatment stage + tumor recurrence If that’s correct, the justification is something along the lines of the following. Having 51 patients with PHQ-9 ≥ 10, we restricted our model for Depression to have 5 degrees of freedom (df) at most, i.e. 10 cases per df. Given you have 5 df to work with, did you consider using a spline or polynomial on PS? Depression ~ KPS + spline(PS) or Depression ~ KPS + PS + PS^2 + PS^3 Please explain the creation of the PS model further. For example, it is not clear if the PS consists only of (referencing Table 2) KPS ~ employment status + treatment stage + tumor recurrence or whether additional variables are in the model. The PS model does not need to be as restricted as the Depression model but some concern of overfitting remains. The goal of the PS analysis is to adjust for important confounding between Depression and KPS. It is not to find statistically significant predictors of KPS. Finding meaningful predictors of KPS is a useful goal by itself. But mixing that goal into the PS analysis can lead to bad practices such as dropping an important but statistically non-significant variable from

	the PS model or including a statistically significant variable that is not associated with the Depression. That said, with n=30 patients having KPS scores ≤ 70, it is reasonable to limit what can go into the PS. It will be important to avoid the misinterpretation that every variable that was considered for the PS model is now being adjusted for. Only the variables in the final PS model are being adjusted for in the model of Depression ~ KPS + PS. The authors can avoid this by saying "controlling for employment status, treatment stage, and tumor recurrence" instead of saying "controlling for covariates". Because the two main findings are built around understanding factors associated with Depression and KPS, I would find it useful to have Table 1 to have three columns instead of two, i.e. PHQ-9 ≥ 10 (n=51), PHQ-9 < 10 (n=81), and total (n=132). If one of those columns had to be dropped due to space constraints, the total column should be the one to drop. Additionally, I would be useful to see Table 2 contain all of the variables in Table 1 with the columns determined by KPS status. The authors could mark the variables contained in the PS with an asterisk and a comment in the table's footnotes. If I'm understanding the PS analysis correctly, I'm confused why Table 3 is referring to two models. Isn't there just one model here, i.e. Depression ~ KPS + PS? The abbreviation "PR" should be included when prevalence ratio is first mentioned in the body of the paper. A brief sentence defining the PR could be helpful as it is used less frequently than risk ratio (RR). Alternatively, RR could be used depending on the target audience. Please clarify how PR are being estimated. Logistic regression would give prevalence odd ratios (POR). I found it unusual to refer to p-values <0.25 as "significant" in the univariate analysis. This particularly stood out when followed by "After adjusting for the effect of other variables in multivariable model, functional status (KPS) remained the only significant variable [which now used a 0.05 level threshold] with P-value <0.001." If "significant" is being used as a marker for having a strong signal, it would make more sense to use a threshold of 0.001 throughout. In this case, the univariate and multivariate analyses are consistent with only KPS meeting that threshold. If instead, p-values are being used to select groups of variables to discuss or to include in a modeling procedure, I suggest simply dropping the word "significant" throughout and referring specifically to each of the various p-value thresholds being used. Minor comments: The stepwise model building is not ideal, but for this dataset where the signals are pretty distinct, I think it is okay. As described above, the univariate and multivariate analyses told the same story, i.e. KPS has a strong association with depression. Treating PHQ-9 as an ordinal variable and using a proportional odds model could be beneficial, but again, I suspect the current analysis is capturing the big signals in the data.

	The participants section implies no one refused consent to participate. Is this correct? Were records kept on the number of patients excluded due to prior diagnosis of depression, etc.? Could this be included? It is not essential. In the statistical analysis section, the description of power is unnecessary given all of the estimates have 95% confidence intervals presented. The precision of the study is what it is, and the confidence interval width fully captures that information. The power description is not hurting anything, but if the authors need to cut something to help make room for all of my other requests, I suggest they cut that. "We also checked multicollinearity between all the predictor variables." How was this done, e.g. via variance inflation factors? Was anything done in response to multicollinearity? I found that there were several minor grammatical errors that did not interfere with interpretation but should be corrected before publication.

VERSION 1 – AUTHOR RESPONSE

0.0 **S#**	**Reviewer comments and feedback**	**Response**
A.	Reviewer 1: Adomas Bunevicius	Thank you Dr. Bunevicius for your feedback.
1	While this is the first study of this kind performed in Pakistan, however results largely represent clinical scenario in other countries.	This is true. Our results are pretty much similar to what is concluded in other similar studies conducted at western part of world. But, our purpose of conducting this study was to obtain some evidence form LMICs and our part of world which can facilitate clinicians and policy makers to amend care policies for brain tumor patients by integrating psychosocial aspect of care earlier in the course of illness. We were successful in highlighting that a great no of primary brain tumor patients suffers psychological illness (Depression) during their disease process and thus this calls for focusing on both physical and psychosocial aspect of care for brain tumor patients.
2	Aside from insurance status, it is not clearly described why risk factors of depression should be different in Low-middle income countries vs. other countries.	This is what the purpose of implementing this study was. Few studies are done on similar topic but on western patients. There is no studies done on LMICs patient so we were not sure if there are any differences in terms of clinical, personal, socio-economic factors with regards to the development of depression among primary brain tumor patients. Also, literature related to this is

		scares and thus, this study gives us new insight about the topic and population residing in LMICs.
3.	Abstract: authors say “biopsy proven”. This is not clear as most patients undergo tumor resection with subsequent pathology report that establish brain tumor diagnosis.	Thank you for highlighting this. We have corrected it. All patients were enrolled based on their MRI contrast reports and underwent biopsy later.
4.	Introduction, para 2: authors describe definition of depression according to WHO. This is not acceptable definition of depression because it has clear criteria as defined by the DSM and is used for clinical practice and research studies.	Yes true. We have made corrections. Now we have defined depression based on DSM-V definition. Thank you for letting us know this.
5.	Were there any restriction with regards to tumor diagnosis? It is important specify that because prognosis and patient symptoms can be very different in glioma (intrinsic and incurable brain tumor) vs. meningioma or pituitary adenomas. Indeed, the sample the majority of patients had benign tumors, such as meningioma and pituitary adenomas, and only 21 had high grade glioma. This high variability of different types of brain tumor in the samples challenge generalizability of the study findings.	All the primary brain tumor patients had MRI contrast done to diagnose the tumor and we enroll only those patients who underwent surgery for primary brain tumor. So there was no restriction with regards to tumor diagnosis. This is true that our sample has high variability as both benign and malignant tumors were included in the study which challenges its generalizability of the study finding. But, this study explored that regardless of tumor type, depression is prevalent in all primary brain tumor patients
6.	Why did authors selected p-value threshold of =<0.25?	According to the book “Applied Logistic Regression” (second edition) by Hosmer and Lamshow (Page 95), selected 0.25 as cutoff at univraite will prevent losing important variable which are eliminated as a result of lower P-value cutoff.  (2) Upon completion of the univariable analyses, we sel for the multivariable analysis. Any variable whose univariab *p*-value < 0.25 is a candidate for the multivariable model al variables of known clinical importance. Once the variable identified, we begin with a model containing all of the selecte Our recommendation that 0.25 level be used as a scre rion for variable selection is based on the work by Bend (1977) on linear regression and on the work by Mickey an (1989) on logistic regression. These authors show that use traditional level (such as 0.05) often fails to identify variabl be important. Use of the higher level has the disadvantage variables that are of questionable importance at the mod stage. For this reason, it is important to review all variables model critically before a decision is reached regarding the f
7.	Looking into Table 2, authors should provide p values for analyses performed, because 95% values of some predictors cross 1, thus questioning if they were indeed significantly associated with outcome of interest. This needs to be checked.	Thank you for highlighting this. I have added the P-value now. Overall model P-value (F-test) is also included which is significant.
B.	Reviewer 1: Alasdair G. Rooney	Thank you Dr. Rooney for your feedback.

8	Minor grammatical errors are scattered throughout which could be addressed in editorial.	Thank you for emphasizing this point. I have tried to make all grammatical corrections.
9	"Depression" is a term which can be used to encompass conditions ranging from the clinical syndrome of DSM-V Major Depressive Disorder, to general lay sadness. Here it appears to imply a clinical diagnosis, which the PHQ-9 alone cannot make. The PHQ-9 just measures 'depressive symptoms'. This should be addressed – such as by replacing 'depression' with the term 'depressive symptoms' throughout, or something similar.	Yes very true. I have replaced depression with word "Depressive symptoms".
Abstract:		
1(	P1 Line 49. The prevalence figure for high depressive symptoms (39%) should have 95% confidence intervals added.	I have added 95% CI
1	P2 Line 5. Under the heading "Strengths and limitations of this study", it's not immediately clear which of the given statements are strengths and which are meant as limitations.	I have now added separate headings for both strengths and limitations
1	P2 Line 7. The authors state, "The major strength of this study is its ability to analyze data using robust statistical techniques." It's true that attempts are made to control for confounders, but even 132 patients give only limited power for regression analyses with over 30 measured variables. Nor – as an observational study – can one ever be sure that every important confounder has been controlled. I think it's questionable as to how 'robust' statistics can be in this limited context and suggest the authors amend or delete this statement.	Yes I have amended this statement
Introduction:		
1	There are several problems with the references that would benefit from careful review. A few examples: - The World Health Organisation definition of depression is referenced to [5], which is in fact a paper by Mainio et al that (unless I missed it) does not mention the	Thank you for highlighting this. I have rechecked and made amendments in referencing

	World Health Organisation's definition of depression. - The authors state: "The worldwide prevalence of depression in cancer patients is 25% with higher rates among Asian countries [8]." The cited reference (Ostrom et al) is a paper describing the aetiology of glioma that does not appear to contain any statements about the prevalence of depression in cancer. - The quoted prevalence range of 10-40% is referenced to [10], which is a technical and theoretical review by Starkweather et al. of how we should conceptualise depression in these patients, and (again unless missed) does not present a simple prevalence range. - References [7] and [12] are the same paper. - Later in the text of the Methods section, Ref [23] is said to be a paper by Gholizadeh et al, but in the reference list, Ref [23] is the review on symptom clusters by Fox et al. For reasons of time I will leave it there but these examples hopefully illustrate that the references need line by line review.	
Methods		
1	P3 Line 20. "The exclusion criteria for study participants were as follows: diagnosis of depression prior to the diagnosis of brain tumor…" Can the authors please clarify how far back they looked (was this any lifetime diagnosis of depression, or just immediately prior to brain tumour diagnosis?) and also what was the reason for excluding these patients? Previous depression is one of the strongest predictors of future depression, so excluding it risks potential bias.	All the patients who were currently (at the time of data recruitment) on anti-depressants or any other anti-psychotic drugs using before the diagnosis of primary brain tumor were excluded. However, patients with history of depression diagnosed and treated before primary brain tumor was diagnosed were not excluded from the study. Moreover, those patients who were diagnosed as having depression after the diagnosis or within one year of diagnosis of primary brain tumor were be included in the study. The purpose for such exclusion criteria was actually the question of the study. This study aimed to determine the factors associated with depression among primary brain tumor patients. Furthermore, depressive symptoms are often the first symptom of primary brain tumor and gradually with time other symptoms becomes apparent which eventually

		results in the diagnosis of primary brain tumor. Thus, patients who already had depression before the diagnosis of primary brain tumor especially for more than a year and were currently on treatment of depression will have different factors associated with depression and not primarily the brain tumor.
1	How did the authors determine tumour location – did all patients have contrast enhanced brain imaging?	Yes all patients had MRI contrast imaging done
1	P4 Line 6. What was the aim of the power calculation – power to do what? Sometimes in observational studies power calculations focus on narrowing 95% confidence intervals down to acceptable limits. But 95% CI are not mentioned. If they were the reason for the power calculation, what limits were the authors aiming for – and if they were not, what were you looking for enough power to do?	Thanks for identifying this. It wasn't power calculation instead it was sample size calculation. I have corrected this.
1	P4 Line 13. Could the authors please reference a suitable precedent (and give the rationale) for calling significance at p=0.25 for univariate analyses? I've seen it done at p=0.1, but the approach taken here is even more permissive. P values aren't the be-all and end-all but there should be some precedent for the method, or alternatively they could re-do the analysis with a limit of p=0.1.	According to the book "Applied Logistic Regression" (second edition) by Hosmer and Lamshow (Page 95), selected 0.25 as cutoff at univraite will prevent losing important variable which are eliminated as a result of lower P-value cutoff. **(2) Upon completion of the univariable analyses, we sel for the multivariable analysis. Any variable whose univariab *p*-value < 0.25 is a candidate for the multivariable model al variables of known clinical importance. Once the variable identified, we begin with a model containing all of the selecte Our recommendation that 0.25 level be used as a scre rion for variable selection is based on the work by Bend (1977) on linear regression and on the work by Mickey an (1989) on logistic regression. These authors show that use traditional level (such as 0.05) often fails to identify variabl be important. Use of the higher level has the disadvantage variables that are of questionable importance at the moc stage. For this reason, it is important to review all variables model critically before a decision is reached regarding the f**
1	P4 Line 16. "Propensity scores were computed to identify factors associated with functional status." I don't know what propensity scores are (or how they are computed) – could the authors please give more detail?	Thank you for asking this question. I have answer this in main manuscript as well. We calculated Propensity scores for the only significant variable left after performing multivariable model building (functional status). The purpose of computing propensity scores was to identify factor associated with the functional status and understand the viscous pathway of associations between explanatory variables and depression. To predict propensity scores, functional status was kept as dependent variable and was regress with other explanatory variables. After the final model was obtained for functional status, propensity scores

		were computed. At last, Propensity scores were regress against depression (dependent variable in the study) to see its association with depression. The cut-off for significance of propensity scores was ≤0.05.
Results		
1	P7 Line 44. These propensity scores are then used in further analyses (beyond what is stated in Methods) as follows: "Propensity scores predicted from [the] above model were significantly associated with depression. Table 3 shows models to demonstrate [the] association of functional status (KPS) with depression and propensity scores for functional status (KPS) with depression." These further analyses need to be described in Methods too – how they were done and why – because at the moment it is not clear what they mean. In particular I'm not clear on the difference between KPS and a propensity score for KPS.	Thank you for highlighting this. I have now added explanation for this in methodology section as well. KPS is a tool which was used to assess functional status. It was considered as one variable with two categories. On the other hand, Propensity scores for KPS was computed separately. To predict propensity scores, functional status (KPS) was kept as dependent variable and was regress with other explanatory variables (relapse, employment status, and treatment stage). After the final model was obtained for functional status, propensity scores were computed.
Discussion		
2	Limitations of the study should be adequately discussed. At the moment they are not mentioned.	Thank you for identifying this. I have added limitations in the discussion part
2	Table 1. A third column should ideally be added with p values for each univariate comparison, to allow readers to quickly scan all the associations. - Travel cost: please state the units of the numbers given (is it rupees?) - Overall treatment cost: I am not sure what "lac" rupees means. -'Strategies to handle stress', 'Brain structures involved' and 'First symptoms involved before brain tumour diagnosis' add	 I have added all the P-values of univariate in the descriptive format in results portion. It's in Rupees. I have made amendments as well. It's in Rupees. I have made amendments as well. Yes agreed! Done

	up to >132. I realise why but these variables should be marked with an asterisk or something and the reason denoted in the table legend, for clarity. -'Tumour grade' only totals 131 – one patient is missing?	Yes it was typo error. Corrected!
2.	Table 2. The presentation of this Table is a little unclear. I can see that for a given number of levels of a variable there are "n-1" entries in the column "PR + 95% CI", but I am wondering if those for Factor 1 (Current Employment Status) are placed on the wrong rows. Could the authors please review that this table is laid out exactly as they intended? Also please expand slightly on what the relevance of the "Reference Category" asterisk is, in the table legend. (Is it meant with reference from, or with reference to?)	Yes there were spacing error. Thanks for identifying. I have corrected it. Here reference category means the category of each variable which is kept as reference in analysis
2.	Table 3. The methodology behind this analysis needs clarified in Methods as noted above.	Done
2.	Other It would be good to see a Table summarizing the average scores / frequencies for each PHQ-9 symptom. Often brain tumour patients may report significant somatic symptoms (sleep, appetite, fatigue etc) without experiencing the "core" symptoms of depressed mood or anhedonia.	Thank you for highlighting these points. We can surely add scores/frequencies of each PHQ-9 symptom including last suicide item. However, this will take-up lot of space and manuscript is already exceeding word count. Moreover, in analysis, overall scores were included in a form of

	The frequency of high scoring on the suicide item 9 would be good to know. Indeed, it would be good to know specifically how many patients scored >10 on PHQ-9 without endorsing either low mood or anhedonia. Were any patients taking antidepressants?	dichotomous factor and thus each symptoms score is not required. It is already there in the top line of table 1. Altogether 51 (39%) patient had scored ≥10 on PHQ-9 No none of the patient were taking antidepressant.
C.	**Reviewer 3: Robert Greevy**	Thank you Dr. Greevy for your feedback.
2	Please clarify what is meant by "Cox algorithm regression" in the abstract and "logistic regression cox algorithm" in the methods. Explain how and why the Cox algorithm is being used.	Thank you for highlighting this. This is Cox Algorithm and logistic regression cox algorithm. I have also corrected this in manuscript. The purpose of using Cox Algorithm is the study design. We wanted to calculate Prevalence ratio (PR) as this was a cross-sectional study design. Our outcome was binary with two categories. We couldn't apply logistic regression as it will give us odds ratio which will actually overestimate prevalence ratio. Thus, we opted for Cox algorithm to compute crude and adjusted PRs.
2	Please clarify how and why propensity scores (PS) are being used. As I'm reading the paper, it appears PS are being used primarily for modeling: Depression ~ KPS + PS, where PS comes from KPS ~ employment status + treatment stage + tumor recurrence If that's correct, the justification is something along the lines of the following. Having 51 patients with PHQ-9 ≥ 10, we restricted our model for Depression to have 5 degrees of freedom (df) at most, i.e. 10 cases per df.	Yes PS was used in modeling. Yes this is correct justification. Thank you for clarifying this.

	Given you have 5 df to work with, did you consider using a spline or polynomial on PS? Depression ~ KPS + spline(PS) or Depression ~ KPS + PS + PS^2 + PS^3	Yes we applied polynomial on PS but it made model insignificant.
2	Please explain the creation of the PS model further. For example, it is not clear if the PS consists only of (referencing Table 2) KPS ~ employment status + treatment stage + tumor recurrence or whether additional variables are in the model. The PS model does not need to be as restricted as the Depression model but some concern of overfitting remains. The goal of the PS analysis is to adjust for important confounding between Depression and KPS. It is not to find statistically significant predictors of KPS. Finding meaningful predictors of KPS is a useful goal by itself. But mixing that goal into the PS analysis can lead to bad practices such as dropping an important but statistically non-significant variable from the PS model or including a statistically significant variable that is not associated with the Depression. That said, with n=30 patients having KPS scores ≤ 70, it is reasonable to limit what can go into the PS. It will be important to avoid the misinterpretation that every variable that was considered for the PS model is now being adjusted for. Only the variables in the final PS model are being adjusted for in the model of Depression ~ KPS + PS. The authors can avoid this by saying “controlling for employment status, treatment stage, and tumor recurrence” instead of saying “controlling for covariates”.	PS was created for only one significant variable in multivariable modeling which was KPS (functional status). To create PS model, first of all we kept KPS as outcome and regressed it will all the other explanatory variables. We found that when we regress employment status, tumor relapsed, and treatment stage with KPS, the overall P-value of model was significant. Then we computed propensity scores. That propensity scores were then regressed against main outcome (Depression) and found significantly associated with it. Yes we agree! We have now used “controlling for employment status, treatment stage, and tumor recurrence” instead of saying “controlling for covariates”.
2	Because the two main findings are built around understanding factors associated with Depression and KPS, I would find it useful to have Table 1 to have three columns instead of two, i.e. PHQ-9 ≥ 10 (n=51), PHQ-9 < 10 (n=81), and total (n=132). If one of those columns had to be dropped due to space constraints, the total column should be the one to drop. Additionally, I would be	Yes agreed! We have done that

	useful to see Table 2 contain all of the variables in Table 1 with the columns determined by KPS status. The authors could mark the variables contained in the PS with an asterisk and a comment in the table's footnotes.	
2	If I'm understanding the PS analysis correctly, I'm confused why Table 3 is referring to two models. Isn't there just one model here, i.e. Depression ~ KPS + PS?	Thanks for highlighting this. I have now separated both the models to make it clear for the readers
3	The abbreviation "PR" should be included when prevalence ratio is first mentioned in the body of the paper. A brief sentence defining the PR could be helpful as it is used less frequently than risk ratio (RR). Alternatively, RR could be used depending on the target audience.	Yes agreed! done
3	Please clarify how PR are being estimated. Logistic regression would give prevalence odd ratios (POR).	We calculated PR using cox algorithm
3	I found it unusual to refer to p-values <0.25 as "significant" in the univariate analysis. This particularly stood out when followed by "After adjusting for the effect of other variables in multivariable model, functional status (KPS) remained the only significant variable [which now used a 0.05 level threshold] with P-value <0.001." If "significant" is being used as a marker for having a strong signal, it would make more sense to use a threshold of 0.001 throughout. In this case, the univariate and multivariate analyses are consistent with only KPS meeting that threshold. If instead, p-values are being used to select groups of variables to discuss or to include in a modeling procedure, I suggest simply dropping the word "significant" throughout and referring specifically to each of the various p-value thresholds being used.	According to the book "Applied Logistic Regression" (second edition) by Hosmer and Lamshow (Page 95), selected 0.25 as cutoff at univraite will prevent losing important variable which are eliminated as a result of lower P-value cutoff. (2) Upon completion of the univariable analyses, we sel for the multivariable analysis. Any variable whose univariab *p*-value < 0.25 is a candidate for the multivariable model al variables of known clinical importance. Once the variable identified, we begin with a model containing all of the selecte Our recommendation that 0.25 level be used as a scre rion for variable selection is based on the work by Bend (1977) on linear regression and on the work by Mickey and (1989) on logistic regression. These authors show that use traditional level (such as 0.05) often fails to identify variabl be important. Use of the higher level has the disadvantage variables that are of questionable importance at the mod stage. For this reason, it is important to review all variables model critically before a decision is reached regarding the fi
3	Minor comments: The stepwise model building is not ideal, but for this dataset where the signals are pretty distinct, I think it is okay. As described above, the univariate and multivariate	Agreed! It was actually multivariable modeling in which each variable was added stepwise based on univariate P-values

	analyses told the same story, i.e. KPS has a strong association with depression.	
	Treating PHQ-9 as an ordinal variable and using a proportional odds model could be beneficial, but again, I suspect the current analysis is capturing the big signals in the data.	Yes Agreed
	The participants section implies no one refused consent to participate. Is this correct?	
	Were records kept on the number of patients excluded due to prior diagnosis of depression, etc.? Could this be included? It is not essential.	Yes none of the patients refused for the study
		Yes records are with research PI and team
	In the statistical analysis section, the description of power is unnecessary given all of the estimates have 95% confidence intervals presented. The precision of the study is what it is, and the confidence interval width fully captures that information. The power description is not hurting anything, but if the authors need to cut something to help make room for all of my other requests, I suggest they cut that.	Thanks for identifying this. It wasn't power calculation instead it was sample size calculation. I have corrected this.
	"We also checked multicollinearity between all the predictor variables." How was this done, e.g. via variance inflation factors? Was anything done in response to multicollinearity?	
	I found that there were several minor grammatical errors that did not interfere with interpretation but should be corrected before publication	To assess Multicollinearity, three different tests were used. Pearson's correlation was used for two normally distributed continuous variables, ETA was used for one qualitative and one quantitative variable whereas; Cramer's V was

		used for two qualitative variables. Moreover, the cut-off for Multicollinearity was 0.8. I have tried to correct all grammatical errors

VERSION 2 – REVIEW

REVIEWER	Adomas Bunevicius University of Virginia
REVIEW RETURNED	17-Feb-2020

GENERAL COMMENTS	I cannot see in the revised manuscript correction of inclusion criteria based on MRI criteria. Please be specific as this is important. That raises a question if any patients were found not to have primary brain tumor as sometimes GBM can look similar to brain metastases of lymphoma. Abstract: please specify the term “pre-structured questionnaire” Please specify what is “prevalence ratio” Limitations section” The study included cross-sectional data instead of prospective data which limits both temporality and direction of causation.” There is a confusion with terms. Cross sectional study can be prospective. In tables, use numerical value instead of using 0 before single digit numbers. Did any patients refused from participation in the study? My problem with this study is mixing different histological types of brain tumors together, but probably this study can help to address importance of depression in general neuro-oncology practice. Selection of liberal p value is another concern that increases the risk of false positive findings. Would results remain the same if authors used more traditional cut-off value of p value, ie. 0.05 or even 0.1 It seems like the main topic of the paper is depression, but they perform PSM analysis for functional status. Can you please also consider this analysis for PHQ-9 score?

REVIEWER	Alasdair Rooney University of Edinburgh Scotland UK
REVIEW RETURNED	26-Feb-2020

GENERAL COMMENTS	I am grateful to the authors for addressing most of my comments, or giving fair justification for not addressing them. I have a few remaining concerns:

	P3. In response to reviewer comments the authors have sought to clarify the exclusion criterion regarding previous depression (in the highlighted text on this page). However the added text is missing key information and consequently the sentence makes no grammatical sense. Diagnosis of depression for about one what? MRI diagnosis is mentioned in the abstract but should also be mentioned in the main text (Methods > Participants). All three reviewers queried the use of $p<0.25$ in the univariate analysis, so we can anticipate that readers will as well. The justifying source Hosman & Lamshow should be added as a reference. I still have concerns about the references. For instance Ref 6 is cited to justify the statement "depression affects about 350 million individuals worldwide and according to the Global Mental Health Survey (2014), nearly 1 in 20 individuals report having at least one episode of depression within a year [6]." Ref 6 is a review of HRQOL in glioma, which as far as I can tell from a quick read-through, does not mention the global prevalence of depression, or the Global Mental Health Survey. Meanwhile (for instance) Ref 9 is cited to justify the statement "The estimated prevalence of clinically diagnosed depression in Pakistan is approximately 6% out of which 3% are cancer patients [9]." But Ref 9 is a biobehavioural model of depressive symptoms in glioma by Starkweather et al. Did they discuss the prevalence of depression in Pakistan? Please double check ALL references again, not just the ones I highlighted last time or this time, one by one - the problem has been pointed out once and should not need repeated. The grammar is improved and mostly good, but continues to need correcting here and there. It is not for peer reviewers to sort this word by word. Careful line-by-line review within the study team by a native English speaker would be worthwhile seeking out if not done already.

VERSION 2 – AUTHOR RESPONSE

Responses to reviewer's feedback

S#	Feedback	Responses
	Reviewer 1: Dr. Adomas Bunevicius	Thank you for your feedback
	I cannot see in the revised manuscript correction of inclusion criteria based on MRI criteria. Please be specific as this is important. That raises a question if any patients were found not to have primary brain tumor as sometimes GBM can look similar to brain metastases of lymphoma.	Thank you for high lightening this. I agree that this is very important point and thus I have made corrections. All patients were diagnosed as having primary brain tumor using MRI. I have added this in methodology part (under heading "participants") as well

	Abstract: please specify the term “pre-structured questionnaire”	A pre-structured questionnaire is usually is a set of standardized questions with a fixed scheme, which specifies the exact wording and order of the questions, for gathering information from respondents. I have added the reference in the manuscript (Phellas, C. N., Bloch, A., & Seale, C. (2011). Structured methods: interviews, questionnaires and observation. Researching society and culture, 3, 181-205).
	Please specify what is “prevalence ratio”	Prevalence ratio is a measure of association often calculated to assess an association between predictor variables and outcome in cross-sectional studies. This is the ratio of the proportion of the persons with disease over the proportion with the exposure. We calculated Prevalence ratio (PR) as this was a cross-sectional study design. Our outcome was binary with two categories. We couldn’t apply logistic regression as it will give us odds ratio which will actually overestimate prevalence ratio. Thus, we opted for Cox algorithm to compute crude and adjusted PRs. I have added its reference too in the manuscript (Barros, A. J., & Hirakata, V. N. (2003). Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. BMC medical research methodology, 3(1), 21).
	Limitations section” The study included cross-sectional data instead of prospective data which limits both temporality and direction of causation.” There is a confusion with terms. Cross sectional study can be prospective	Thank you identifying this. This is very true. I have corrected this in the manuscript.
	In tables, use numerical value instead of using 0 before single digit numbers	Thank you for highlighting this. I have made changes in manuscript
	Did any patients refused from participation in the study?	No patients refused to participate in the study
	My problem with this study is mixing different histological types of brain tumors together, but probably this study can help to address importance of depression in general neuro-oncology practice.	Yes we agree. Thank you for mentioning this
	Selection of liberal p value is another concern that increases the risk of false positive findings. Would results remain	Thank you for raising this concern. The results would have been same even if we would have taken traditional P-values. This is because, P-

	the same if authors used more traditional cut-off value of p value, ie. 0.05 or even 0.1	value of ≤ 0.25 was taken only at univariate level to ensure that all clinically important variables are not missed. The P-value of multivariable analysis was kept ≤0.05. Those variables which had P-value of ≤0.25 at univariate were part of multivariable modeling (entered in model) but remained insignificant in multivariable analysis and thus only one variable with P-value of <0.001 (KPS) was significant after adjusting for other variables.
	It seems like the main topic of the paper is depression, but they perform PSM analysis for functional status. Can you please also consider this analysis for PHQ-9 score?	We only calculated PS (propensity scores) and did not do propensity score matching (PSM). PS was created for only one significant variable in multivariable modeling which was KPS (functional status). To create PS model, first of all we kept KPS (which was predictor variables) as outcome and regressed it with all the other explanatory variables. We found that when we regress employment status, tumor relapsed, and treatment stage with KPS, the overall P-value of model was significant. Then we computed propensity scores. That propensity scores were then regressed against main outcome (Depression) and found significantly associated with it. For PHQ-9 we cannot calculate PS as PHQ-9 scores were outcome of this study (measuring depression). Calculating PS for PHQ-9 means regressing PHQ-9 scores with all other explanatory variables which will yield same results as done via Cox regression method.
	Reviewer 2: Alasdair Rooney	Thank you for your feedback
	In response to reviewer comments the authors have sought to clarify the exclusion criterion regarding previous depression (in the highlighted text on this page). However the added text is missing key information and consequently the sentence makes no grammatical sense. Diagnosis of depression for about one what?	Thank you so much for helping us identifying this typo error. I have corrected this sentence “Diagnosis of depression for about one” to “diagnosis of depression for about one year prior to the diagnosis of brain tumor”
	MRI diagnosis is mentioned in the abstract but should also be mentioned in the main text (Methods > Participants).	Yes agreed! I have added this important point in methodology as well under heading “Participants”.
	All three reviewers queried the use of p<0.25 in the univariate analysis, so	Added! Thank you for this suggestion

	we can anticipate that readers will as well. The justifying source Hosman & Lamshow should be added as a reference.	
	I still have concerns about the references. For instance Ref 6 is cited to justify the statement “depression affects about 350 million individuals worldwide and according to the Global Mental Health Survey (2014), nearly 1 in 20 individuals report having at least one episode of depression within a year [6].” Ref 6 is a review of HRQOL in glioma, which as far as I can tell from a quick read-through, does not mention the global prevalence of depression, or the Global Mental Health Survey. Meanwhile (for instance) Ref 9 is cited to justify the statement “The estimated prevalence of clinically diagnosed depression in Pakistan is approximately 6% out of which 3% are cancer patients [9].” But Ref 9 is a biobehavioural model of depressive symptoms in glioma by Starkweather et al. Did they discuss the prevalence of depression in Pakistan? Please double check ALL references again, not just the ones I highlighted last time or this time, one by one - the problem has been pointed out once and should not need repeated.	Thank you for highlighting this. I have checked all the references and have made few corrections as well. Reference no 6 is changes: Tucci, V., & Moukaddam, N. (2017). We are the hollow men: The worldwide epidemic of mental illness, psychiatric and behavioral emergencies, and its impact on patients and providers. Journal of emergencies, trauma, and shock, 10(1), 4. Reference no 9 is also changes: Ahsan, J., et al., Spectrum of central nervous system tumours–a single center histopathological review of 761 cases over 5 years. Journal of Ayub Medical College Abbottabad, 2015. 27(1): p. 81-84. Rest I have also rechecked and made corrections where required.
	The grammar is improved and mostly good, but continues to need correcting here and there. It is not for peer reviewers to sort this word by word. Careful line-by-line review within the study team by a native English speaker would be worthwhile seeking out if not done already.	We have carefully reviewed the text and our team members have corrected the grammatical errors in the text Revise. We have tried our best to correct all the grammatical errors in the manuscript.