PEER REVIEW HISTORY

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	A one year longitudinal study of psychological distress and self-assessed health in survivors of out-of-hospital cardiac arrest.
AUTHORS	Viktorisson, Adam; Sunnerhagen, Katharina; Johansson, Dongni; Herlitz, Johan; Axelsson, Åsa

VERSION 1 - REVIEW

REVIEWER	Andreas Dinkel Department of Psychosomatic Medicine and Psychotherapy, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany
REVIEW RETURNED	15-Mar-2019

GENERAL COMMENTS	This study deals with the psychological adaptation of people who survived out-of-hospital cardiac arrest. Surviving this life-threatening event is associated with cognitive and emotional deficits, even in survivors with good neurological outcomes. However, longitudinal studies investigating psychological adaptation over time are sparse, as are studies on predictors of adaptational outcomes. The authors conducted a longitudinal study with survivors of out-of-hospital cardiac arrest. Recruitment took place over 4 years. 298 out of 3082 cases were alive at 30 days, and 150 fulfilled the inclusion criteria. 94 of those patients responded at the 3 month follow-up, and 74 after 12 months. The results showed a decline of anxiety and depressive symptoms and an increase in health-related quality of life from 3 to 12 months post resuscitation. Physical comorbidity, age, and gender showed predictive associations with distress and HRQoL. This is an interesting and relevant study. However, there are some critical aspects I would like to address: Major 1. For use in the multivariable logistic regression, the outcomes – HADS scores for anxiety and depression and EQ 5D score – were dichotomized. Unfortunately, the authors do not report how many survivors showed clinically significant symptoms of anxiety and depression or diminished HRQQoL. First, the prevalence rate would be interesting to know in itself. Second, and more important, number of cases is essential for the evaluation of the results of the logistic regression. As only 70 survivors are included in the

	regression, it makes a difference if there are, e.g., 50 % cases or 20 % cases. Without these data, the number of events per variable is unknown. Therefore, it is not possible to evaluate the adequacy of the logistic regression. – In light of the inconsistent results, which are hard to interpret in a concise way, it seems that the results are compromised by methodological problems. 2. A related issue: There are cut-off scores available for the HADS. However, the authors used comparison data from a Swedish community sample. As the cut-off scores of the HADS has not been validated in a sample of out-of-hospital cardiac arrest survivors (but in many other samples), it seems viable to use reference scores from a community sample. – However, it is currently not fully clear how caseness was defined. Did the authors use the mean of the community sample as cut-off? This would lead to a substantial number of false positive clinical cases. Usually, one would apply 1 or even 2 SD as threshold for clinical levels of distress/reduced HRQoL. 3. In table 3, the authors report the final results of the multivariable logistic regression. – I did not understand which variables were available for the univariate analysis. All of the variables listed in table 1? How did the authors re-code the variables, as sample size was quite low for some variables, e.g. 5 for CABG or3 for back pain. 4. Moreover, why did the authors apply univariate screening of potential predictors before conducting the multivariable analysis? Methodologists often criticize the practice of bivariate screening before conducting multivariable analysis. Here, the authors state that they selected predictors on the basis of clinical reasoning. So, this is a deductive approach. The authors state that they are interested in a special set of predictors, but at the same time they exclude predictors they are interested in based on bivariate screening. One argument might be that they were interested in too many predictors, given the small sample size, and therefore tried to identify the most promising predictors. However, it might be more appropriate to reduce the number of predictors a priori and to refrain from bivariate screening. 5. The authors report the percentage of participants who showed an increase or decrease in the symptom level of anxiety and depression. As far as I understood, the authors did not apply a critical change score, which seems quite problematic. Clearly, an increase of 1 point in the summary score of the anxiety or depression subscale represents an increase, but the importance of this increase is unclear. Thus, the authors might want to apply RCI for the determination of reliable change (or even recovery/deterioration). Minor: 6. Abstract, methods/results: Please present sample size. 7. Abstract, results: "psychological distress" – Please make clear that psychological distress equals the assessment of anxiety and depression. – Furthermore, please state the prevalence rate of clinical anxiety and depression. 8. Introduction, line 32: "little is known about factors associated with post cardiac arrest improvement". Then, 4 referenes are given. Please summarize what is actually known on factors associated with outcome. 9. Introduction, line 39. Why do the authors explicitly mention Sweden? The authors refer to ref 7 and 14 which, as far as I can see, do not deal with the situation in Sweden.

	10. Procedure and questionnaire, p. 5, line 7: EQ-5D "is designed to measure health" – please change to subjective health or health-related quality of life. 11. Results, p. 6, line 55. The authors present data on gender differences. However, they did not pose a research question pertaining to sex differences. It is unclear why the authors highlight this variable – they might also choose other variables to present subgroup results. 12. Results, p. 7, line 8: Logistic regression – the authors speak of correlations, but they investigated odds ratios. Please correct. 13. Limitations, p. 9. Please discuss the small sample size as limiting factor. 14. Figure 2: Really necessary? – I found this figure a little bit hard to understand.

REVIEWER	Sarah Voss University of the West of England Bristol
REVIEW RETURNED	29-Mar-2019

GENERAL COMMENTS	Thank you for the invitation to review this paper. This is an interesting topic and the research adds to the evidence base of longer term well-being following OHCA. It is a well written paper that appropriately builds on work previously reported by the authors. I have only minor suggestions for improvement. Patient and public involvement: Has any work been carried out by the study team to investigate how important or relevant the research question is to members of the public? I would like to see detail on how patients were approached and consented. Did any patients decline to consent? This is different from not responding. Is damage to the brain anoxic (as currently written) or hypoxic? I think that it would be preferable to use the term hypoxic. VAS scale is described as a thermometer; this should be changed to something like 'a measurement instrument that uses a continuum of values' I would suggest that the authors use 'self-assessed poor health', rather than 'poor self-assessed health' throughout. I would argue that the use of the word 'emotion' is not appropriate. The research uses measures designed to assess psychological rather than emotional constructs and is therefore describing psychological well-being, or psychological distress, rather than emotion; consider revising this. It is not possible to differentiate between patients with CPC1 and CPC2; this should be described in the limitations of the study. Are there any limitations associated with the data being collected 7-9 years ago?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Thank you for your comments. They have helped us to improve the manuscript and we have tried to follow all your recommendations. Below are the responses to your comments.

1. For use in the multivariable logistic regression, the outcomes – HADS scores for anxiety and depression and EQ 5D score – were dichotomized. Unfortunately, the authors do not report how many survivors showed clinically significant symptoms of anxiety and depression or diminished HRQQoL. First, the prevalence rate would be interesting to know in itself. Second, and more important, number of cases is essential for the evaluation of the results of the logistic regression. As only 70 survivors are included in the regression, it makes a difference if there are, e.g., 50 % cases or 20 % cases. Without these data, the number of events per variable is unknown. Therefore, it is not possible to evaluate the adequacy of the logistic regression. – In light of the inconsistent results, which are hard to interpret in a concise way, it seems that the results are compromised by methodological problems.

Individual scores in HADS and EQ-5D-3L are presented in Figure 2. We have added the following paragraph to the results section “Mild to moderate anxiety (>7 in HADS-A) was reported by 22 survivors at the three months follow-up and by 17 at 12 months. Mild to moderate depression (>7 in HADS-D) was reported by 10 survivors at three months and by four at 12 months. A score indicating severe anxiety (>10 in HADS-A) was reported by 12 survivors at three months and by eight at 12 months, while severe depression (>10 in HADS-D) was reported by three at three months and one at 12 months.” However, for a logistic regression, the class imbalance has previously shown little impact on the prediction accuracy of the output model.1 The EQ-5D-3L has no cut-off and the mean scores are presented in Table 2.

2. A related issue: There are cut-off scores available for the HADS. However, the authors used comparison data from a Swedish community sample. As the cut-off scores of the HADS has not been validated in a sample of out-of-hospital cardiac arrest survivors (but in many other samples), it seems viable to use reference scores from a community sample. – However, it is currently not fully clear how caseness was defined. Did the authors use the mean of the community sample as cut-off? This would lead to a substantial number of false positive clinical cases. Usually, one would apply 1 or even 2 SD as threshold for clinical levels of distress/reduced HRQoL.

The outcome was dichotomized as good or bad with regard to mean scores drawn from the Swedish population-samples referred to. 2 SD below the mean, matched to age (within a 10 year span) and gender, was considered a bad outcome. This approach was chosen partly in order to compare the HADS and EQ-5D-3L alike (as EQ-5D-3L has no cut-off value for good HRQoL). Furthermore, survivors of OHCA are known to report low levels of psychological distress why population averages were considered comparable. We have added the following to the methods section (Statistics and data analysis): “A bad outcome was considered a score two standard deviations below the mean of the reference population.”

3. In table 3, the authors report the final results of the multivariable logistic regression. – I did not understand which variables were available for the univariate analysis. All of the variables listed in table 1? How did the authors re-code the variables, as sample size was quite low for some variables, e.g. 5 for CABG or 3 for back pain.

Only variables marked with a raised a (a) in Table 1 were included as predictors in the univariate regression analyses. Comorbidity was included as “any treatment requiring comorbidity” or “no comorbidity” (60 of 81).

4. Moreover, why did the authors apply univariate screening of potential predictors before conducting the multivariable analysis? Methodologists often criticize the practice of bivariate screening before conducting multivariable analysis. Here, the authors state that they selected predictors on the basis of clinical reasoning. So, this is a deductive approach. The authors state that they are interested in a special set of predictors, but at the same time they exclude predictors they are interested in based on bivariate screening. One argument might be that they were interested in too many predictors, given the small sample size, and therefore tried to identify the most promising predictors. However, it might be more appropriate to reduce the number of predictors a priori and to refrain from bivariate screening.

We appreciate this reflection. Variables included in the univariate regression were chosen based on clinical reasoning, in total six different variables as seen in Table 1. We chose to include no more than six due to the sample size. We have added “With regard to the sample size, only six variables were evaluated”

Statistical models are used as a method to provide clinically relevant results in relation to the study aim. Predictors that are input into the model should carry some extent of theory base rather than only derived on the statistical significance. We believe therefore it is important to include variables of clinical interest into the model to improve clinical interpretability of the results. A “purposeful selection” described by Hosmer and Lemeshow2 is used as the univariate analyses for screening variables to avoid a potential overfitting in the multivariate analysis due to many predictors in a given small sample size. As was already mentioned by the reviewers, we do aware some discussions regarding the use of univariate analyses before the multivariate analysis. The small sample size limits the use of other available methods of variable selection that can be applied to the data sets. As a limitation, it has been added in the Discussion.

5. The authors report the percentage of participants who showed an increase or decrease in the symptom level of anxiety and depression. As far as I understood, the authors did not apply a critical change score, which seems quite problematic. Clearly, an increase of 1 point in the summary score of the anxiety or depression subscale represents an increase, but the importance of this increase is unclear. Thus, the authors might want to apply RCI for the determination of reliable change (or even recovery/deterioration).

Thank you for this suggestion. We agree that this is of great importance for the interpretation of individual results. However, we believe that the addition of a critical change score may complicate the understanding of the results and is better left to the Discussion section. As stated in the second paragraph of the discussion, estimates of the MCID from other patient populations indicate that the improvements found between three and 12 months in this study might not be clinically noticeable in the group at large.

6. Abstract, methods/results: Please present sample size.

We have added the following to the abstract (Results): “Of 298 survivors, 85 were eligible for this study and 74 responded.”

7. Abstract, results: "psychological distress" – Please make clear that psychological distress equals the assessment of anxiety and depression. – Furthermore, please state the prevalence rate of clinical anxiety and depression.

We have done as suggested, please see the Objectives and Results sections of the abstract.

8. Introduction, line 32: "little is known about factors associated with post cardiac arrest improvement". Then, 4 references are given. Please summarize what is actually known on factors associated with outcome.

We have rearranged the sentence and added “Time to awakening, gender, age, role of bystander, use of hypothermia and percutaneous coronary intervention (PCI) have previously been related to the outcome after cardiac arrest” to the Introduction.

9. Introduction, line 39. Why do the authors explicitly mention Sweden? The authors refer to ref 7 and 14 which, as far as I can see, do not deal with the situation in Sweden.

Thank you for pointing this out. “In Sweden” has been removed from the text

10. Procedure and questionnaire, p. 5, line 7: EQ-5D "is designed to measure health" – please change to subjective health or health-related quality of life.

We have changed the description in this part to “subjective health” as suggested.

11. Results, p. 6, line 55. The authors present data on gender differences. However, they did not pose a research question pertaining to sex differences. It is unclear why the authors highlight this variable – they might also choose other variables to present subgroup results.

As was stated in the Introduction, gender is an important predictor associated with the outcome of cardiac arrest. We believe that it is important to provide results on gender difference. Therefore, we would like to keep this in the results.

12. Results, p. 7, line 8: Logistic regression – the authors speak of correlations, but they investigated odds ratios. Please correct.

Thank you for point this out. It has been corrected to odds ratios (p.7 line 8).

13. Limitations, p. 9. Please discuss the small sample size as limiting factor.

We have added "... the relatively small sample size which further limits the generalisability of the results. " to the Strengths and Limitations section.

14. Figure 2: Really necessary? – I found this figure a little bit hard to understand.

We appreciate this reflection. We argue that Figure 2 contribute with interesting information as it depict the development of each individual survivor and visualize the heterogenicity of the group. We have altered the Figure legend to make the interpretation easier.

Reviewer: 2

We thank you for your comments and have tried to follow all your suggestions. It is our belief that they have helped to improve the manuscript.

1. Patient and public involvement: Has any work been carried out by the study team to investigate how important or relevant the research question is to members of the public?

At the time when this study was preformed we were yet to start working with patients as partners in research. Therefore, unfortunately we have done no such work. We have added "...and no patient opinion regarding the subject has been obtained." To the Patient and public involvement section.

2. I would like to see detail on how patients were approached and consented. Did any patients decline to consent? This is different from not responding.

A written consent was sent out and signed by each participant. We have added “Survivors who responded to the questionnaires and gave a written consent were included.” Details of study recruitment are presented in Figure 1.

3. Is damage to the brain anoxic (as currently written) or hypoxic? I think that it would be preferable to use the term hypoxic.

We have done as suggested and changed to “hypoxic”.

4. VAS scale is described as a thermometer; this should be changed to something like ‘a measurement instrument that uses a continuum of values’

We have changed the description to “EQ VAS is visualized as a continuum of value scale where the patient assesses his or her current health on a scale from zero to 100, with endpoints labelled ‘best imaginable health state’ and ‘worst imaginable health state’.”

5. I would suggest that the authors use ‘self-assessed poor health’, rather than ‘poor self-assessed health’ throughout.

We have done as suggested.

6. I would argue that the use of the word ‘emotion’ is not appropriate. The research uses measures designed to assess psychological rather than emotional constructs and is therefore describing psychological well-being, or psychological distress, rather than emotion; consider revising this.

We agree and have changed “emotional” to “psychological” throughout.

7. It is not possible to differentiate between patients with CPC1 and CPC2; this should be described in the limitations of the study.

We have added “Furthermore, it is not possible to differentiate between patients with a CPC-score of one and two.” To the Strengths and Limitations section.

8. Are there any limitations associated with the data being collected 7-9 years ago?

The organization around resuscitation and the post cardiac arrest follow-up guidelines in Sweden has not undergone any large changes during this time, therefore we believe that the data is still relevant.

Reference

1. Crone SF, Finlay S. Instance sampling in credit scoring: An empirical study of sample size and balancing. International Journal of Forecasting 2012; 28(1): 224-38.

2. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons; 2013.

VERSION 2 – REVIEW

REVIEWER	Andreas Dinkel Department of Psychosomatic Medicine and Psychotherapy, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany
REVIEW RETURNED	03-May-2019

GENERAL COMMENTS	Thank you for your efforts in revising the manuscript. In my view, the manuscript has improved, and I have just a few final comments. 1. The authors added the number of cases for anxiety and depression. Then, they argue that class imbalance has shown little impact on the prediction accuracy of the output model. The authors are right that the common rule of thumb of 10 events per 1 variable (EPV) has been criticized, and there is ongoing debate about the relevance of EPV. However, those who critizise the EPV criterion stress that other variables should be taken into account when evaluating prediction accuracy of logistic regression, like overall sample size and correlational structure. Thus, I think the authors should briefly refer to EPV and other criteria for the evaluation of prediction accuracy, and state that due to the sample size and the low EPV the current results of the logistic regression should be interpreted cautiously. Some further comments: 2. Abstract, results: 85 survivors were eligible, 78 responded – please add percentages relating to the total number of 298 survivors.

	3. Abstract, results, lines 24/25: Please change to "Clinically relevant anxiety"…"while clinical depression". Furthermore, please present percentages in addition to the absolute number of cases. 4. The authors might want to include a statement on gender in the last para of the introduction, as they were especially interested in gender differences. 5. Results, p. 7, line 13: Please add percentages to the absolute number of cases for anxiety and depression. 6. Discussion, p. 9, lines 6, 12, 22, 25: The authors speak of correlations – "associations" would be more appropriate.

REVIEWER	Sarah Voss The University of the West of England, Bristol
REVIEW RETURNED	25-Apr-2019

GENERAL COMMENTS	Comments adressed.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Thank you for your comments. It is our belief that we have been able to change the manuscript in accordance your recommendations.

1. The authors added the number of cases for anxiety and depression. Then, they argue that class imbalance has shown little impact on the prediction accuracy of the output model. The authors are right that the common rule of thumb of 10 events per 1 variable (EPV) has been criticized, and there is ongoing debate about the relevance of EPV. However, those who critizise the EPV criterion stress that other variables should be taken into account when evaluating prediction accuracy of logistic regression, like overall sample size and correlational structure. Thus, I think the authors should briefly refer to EPV and other criteria for the evaluation of prediction accuracy, and state that due to the sample size and the low EPV the current results of the logistic regression should be interpreted cautiously.

We have added "… the relatively small sample size, resulting in a low count of events per variable in the logistic regression (five variables were evaluated in 70 participants), why the significance of these results should be interpreted with caution." to the Strengths and limitation section of the manuscript.

2. Abstract, results: 85 survivors were eligible, 78 responded – please add percentages relating to the total number of 298 survivors.

We appreciate this comment. Unfortunately we cannot add this due to the word restriction of the abstract. Furthermore, we argue that nothing else can be removed in order to make this space.

3. Abstract, results, lines 24/25: Please change to "Clinically relevant anxiety"…"while clinical depression". Furthermore, please present percentages in addition to the absolute number of cases.

Once again, we appreciate this comment. Unfortunately we cannot add this due to the word restriction of the abstract.

4. The authors might want to include a statement on gender in the last para of the introduction, as they were especially interested in gender differences.

We have changed the phrasing to “Further, we will also evaluate gender differences as well as predictors of psychological distress and self-assessed poor health at these time points.” In the last sentence of the last paragraph of the introduction.

5. Results, p. 7, line 13: Please add percentages to the absolute number of cases for anxiety and depression.

We have done as suggested.

6. Discussion, p. 9, lines 6, 12, 22, 25: The authors speak of correlations – "associations" would be more appropriate.

We have changed “correlations” to “associations” as suggested.