

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form ([see an example](#)) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	The Sydney Multisite Intervention of LaughterBosses and ElderClowns (SMILE) study: A cluster randomised trial of humour therapy in nursing homes
<b>AUTHORS</b>	Low, Lee-Fay; Brodaty, Henry; Goodenough, Belinda; Spitzer, Peter; Bell, Jean-Paul; Fleming, Richard; Casey, Anne-Nicole; Liu, Zhixin; Chenoweth, Lynn

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Professor Bernie Warren PhD School of Dramatic Art, University of Windsor Windsor Ontario Canada
<b>REVIEW RETURNED</b>	28-Sep-2012

<b>THE STUDY</b>	This paper does not provide a definition or an explanation of what "Humour Therapy" is , what its limits are or how one can be trained to employ it . A footnote defining what we are talking about would help focus the reader!  There is not enough discussion concerning what the ElderClowns DID to give the reader a sense of the transferability of these results.
<b>RESULTS &amp; CONCLUSIONS</b>	There is no doubt in my mind that this is a significant study . However while the facts and figures are clear there is not enough narrative to give the reader a sense of what exactly they relate to.  Readers need to know more about the interactions of the ElderClown with the residents to get a sense of this work

<b>REVIEWER</b>	Dr Stephen-Mark Cooper Reader in Applied Statistics & Data Analysis Cardiff School of Sport Cardiff Metropolitan University Cyncoed Campus Cyncoed Road Cyncoed CARDIFF, CF24 6XD Wales, UK  There are no competing interests
<b>REVIEW RETURNED</b>	24-Oct-2012

<b>GENERAL COMMENTS</b>	In this manuscript (MS) the authors apply novel and appropriate statistical methods to examine the SMILE study which is a cluster randomised controlled trial that attempts to evaluate the effect of humour therapy on depression, agitation, behavioural disturbances,
-------------------------	--

	<p>social engagement and quality of life in nursing home residents in Sydney, Australia. The study is conducted on a large sample of subjects.</p> <p>As I am not an expert in mental health, or the body of research therein, I have limited myself to making general comments about the MS and its presentation, and more specifically the research design and the statistical analyses of the collected scores. The study is well-designed, the research is well-organised, the data seem to have been collected appropriately and the data are expertly handled and analysed. As far as the methodology is concerned I am convinced that I could replicate the study based on the description the authors provide in the MS. The authors communicate their arguments in a clear and concise manner in a paper that is well crafted and skilfully written.</p> <p>If I have any criticisms of the MS at all it is to do with the lack of any consideration of confirmation of underlying assumptions related to the choice of statistical tests, and no consideration of inferential indices such as effect sizes and power related to stated hypotheses (other than when justifying the sample size). Sometimes, consistency in the expression of terms; for example the use of hyphenation throughout, the superscripted numbers identifying reference citations and when multiple citations are used – the use of <i>et al</i> – this is an abbreviation and should be presented in italics (it is a Latin term) thus (<i>et al.</i>). But, I guess this is likely to be corrected at the proof reading stage of the MS. I also found the amount of information, and the manner in which it was being expressed, in Tables 1 and 2 confusing. I would also like to see all statistical indices expressed in italics where appropriate (e.g. <i>P</i>, <i>r</i>, <i>n</i>, <i>t</i>, <i>F</i>, <i>U</i>, <math>X^2</math> etc.) with the relevant degrees of freedom subscripted as appropriate (e.g. <math>r_{df}</math>, <math>t_{df}</math>, <math>F_{df}</math>, <math>U_{df}</math> etc.).</p> <p>I realise that such points might seem pedantic but then ... one man's pedantry is another man's precision! Seriously, this was a really good paper, and even though I have no background in the academic area, it was one which I thoroughly enjoyed reading.</p> <p>My specific comments are as follows.</p> <p><b>Abstract (page 3)</b></p>
--	---

	<p><b>Line 27:</b> ... and seven did not provide ...</p> <p><b>Line 50:</b> ... and self- and proxy-rated quality ...</p> <p><b>Line 52:</b> what is DEMQOL? You've given the full names of inventories previously without the acronym.</p> <p><b>Page 4, line 5:</b> (and elsewhere in the MS) ... <math>P = 0.011</math> (spacing and italics)</p> <p><b>Line 10:</b> the same thing ... (95%CI/ 0.004 to 0.34, <math>P = 0.045</math>).</p> <p><b>Line 57:</b> in the copy of the MS that I downloaded for review purposes the side-heading 'Strengths and Limitations:' needs to be moved down so that it appears at the top of p.5.</p> <p><b>Introduction (page 6)</b></p> <p><b>Line 40:</b> (and throughout the MS) ... study (<math>n = 21</math>) of four ...</p> <p><b>Page 7, line 28:</b> I'm not sure that the sentence is full enough. I get the idea but would it be better as: 'Secondary outcomes of interest were to interrogate improvements in levels of agitation and other behavioural disturbances ....</p> <p><b>Methods (page 7)</b></p> <p><b>Line 38:</b> ... nursing homes. The study protocols ...</p> <p><b>Line 57:</b> .... Sydney within a one-hour ...</p> <p><b>Page 8, line 12:</b> ... were randomised to an intervention group or a control group. Eight ...</p> <p><b>Line 39:</b> ... within six months, experiencing ...</p> <p><b>Page 9, line 7:</b> '2) Between nine and 12 humour ...</p> <p><b>Lines 19 &amp; 21:</b> might have not may have</p> <p><b>Line 27:</b> LaughterBoss</p> <p><b>Line 34:</b> I'm not sure that the note about the purchase of the DVD is in keeping with the rubric of the journal!</p> <p><b>Line 46:</b> ... low care). A random number generator in Excel was used ....</p> <p><b>Page 10, line 22:</b> (and elsewhere in the MS) ... which is a clinician-rated depression scale.<sup>19</sup> ... The issue is to do with the consistency</p>
--	---

	<p>of hyphenation as I mentioned earlier. Also on page 10 there seems to be a problem with spacing and the superscripted reference citation numbers.</p> <p><b>Line 29:</b> ... behaviours (a mixed economy of spelling in the MS); the eight-item withdrawal ...</p> <p><b>Line 33:</b> what is DEMQOL? You've given the inventory names in full above followed by its acronym. What about DEMQOL?</p> <p><b>Line 57:</b> again, in the copy of the MS that I downloaded for review purposes the side-heading 'Analysis' needs to be moved down so that it appears at the top of p.11. However, should it not be 'Analyses' as there are more than one?</p> <p><b>Page 11, line 5:</b> ... effect size (Cohen's <math>d = 0.5</math>) difference</p> <p><b>Line 10:</b> what is a RACF study?</p> <p><b>Lines 12 &amp; 15:</b> ... intention-to-treat</p> <p><b>Line 15:</b> I think you need a reference for the SAS V9.2 software here</p> <p><b>Line 17:</b> <math>t</math>-test and <math>U</math> tests. Additionally, I guess the decision to use a parametric or non-parametric test of difference here at baseline was made on the nature of the scores collected (real number series – nominal, ordinal, interval or ratio)? But you deal with it simply by saying they were '... examined ... as appropriate'. I think you need to be more specific about how decisions were made – on what criteria. Also, the <math>t</math>-test tests the <math>H_0: \bar{X}_1 = \bar{X}_2</math>; the <math>X^2</math> test tests the <math>H_0: O = E</math>, and the Mann-Whitney test tests (probably here) <math>H_0: \eta_1 = \eta_2</math> yet no mention is made of underlying assumptions and hypotheses on the selection of the 'appropriate' tests and their outcomes. What level of statistical significance did you set throughout your study and why? From reading the MS I'm presuming alpha was set at <math>P \leq 0.05</math> but you need to justify why this was chosen in relation to the type of error you were willing to make.</p> <p><b>Line 24:</b> ... both within-resident correlation ...</p> <p><b>Line 28:</b> ... was adjusted using an approach recommended by Fitzmaurice <i>et al.</i><sup>30</sup> ...</p> <p><b>Line 46:</b> I think you need to add a reference for the Blom transformation here</p> <p><b>Line 53:</b> you say that the ratings were 'highly' correlated (<math>r = 0.863</math>). But this only gives a coefficient of determination of 74.5%. Presumably, this correlation was statistically significant and if so what was the <math>P</math>-value? Put the degrees of freedom for the zero-order correlation as a subscript here (<math>r_{df}</math>)</p> <p><b>Line 55:</b> ... to the primary analysis as outlined above.</p>
--	---

## Results (page 12)

**Line 12:** do you mean that the difference between groups was zero? Or do you mean that the difference between the groups was statistically non-significant? But we still do not know what your alpha level was and why you chose that level of significance.

**Line 22:** ... delivered (average ( $\pm$ SD) of  $11.24 \pm 0.97$  per facility). At least I'm assuming it is the SD and I question whether this sort of information can be summarised to this level of precision and certainly not to 2 decimal places. The same goes for the summary of sessions cited on line 24.

**Line 34:** it might be better to express this as '... over time, but the group by time interactions on depression, non-agitation behavioural disturbance, social engagement or resident- proxy-rated quality of life were non-significant ( $P > 0.05$ ).' That's assuming that the alpha was set at  $\leq 0.05$ .

**Line 38:** ... The group-by-time interaction was statistically significant for ... for covariates ( $P < 0.05$ ). That's assuming that the alpha was set at  $\leq 0.05$ .

**Line 45:** I think there is an error here as the 95%CI -0.004 to 0.34 includes zero and so cannot be statistically significant. In the abstract (p.4) you give this 95%CI as 0.004 to 0.34 which is probably correct. Also,  $P = 0.045$  (spacing, caps and italics) – this is the case for all  $P$ -values on page 12 and throughout the rest of the MS.

**Line 51:** ... was statistically significant with ...

**Page 13, line 3:** 'There were statistically significant engagement dose-by-time ... depression  $F_{df} = 6.72$ ,  $P = 0.000$  ... You need to subscript the degrees of freedom for all the  $F$ -ratios and caps and italics for all vales of  $P$  in this paragraph.

## Discussion (page 13)

**Line 33:** ... were 3.00 (95%CI 1.78 to 4.22)<sup>31</sup>. ...

**Line 38:** I think you need a reference here to endorse that risperidone is the most commonly used antipsychotic in Australia.

**Page 15, line 30:** might not may.

## Table 1 (page 20)

	<p>This is a 'busy' table. Italics for all <i>n</i> and italics and subscripted <i>dfs</i> for all values of <i>t</i>, <math>X^2</math> and <i>U</i>. Caps and italics for all <i>P</i>-values. Also check the spacing of all values – e.g. humour therapy/number with dementia = 145 (76.7%) and humour therapy/years lived in care = <math>2.8 \pm 3.1</math>.</p> <p><b>Table 2 (page 22)</b></p> <p>This is an even busier table! I think that the legend of the table could include more pertinent detail and the same issues as identified for Table 1 above are also relevant here. On p.23 'effect size' is mentioned but I don't recollect it being mentioned as part of your analysis methodology.</p>
--	--

## VERSION 1 – AUTHOR RESPONSE

Comments from Professor Bernie Warren

\*This paper does not provide a definition or an explanation of what "Humour Therapy" is , what its limits are or how one can be trained to employ it . A footnote defining what we are talking about would help focus the reader!

A sentence has been added to the discussion explaining what humour therapy is, and giving some examples. We believe that discussing how to deliver training in using humour may be outside the scope of this paper.

\*There is not enough discussion concerning what the ElderClowns DID to give the reader a sense of the transferability of these results.

\*Readers need to know more about the interactions of the ElderClown with the residents to get a sense of this work

We have added more detail to the examples of interactions between ElderClowns and residents. We have also added more information about some of the techniques used. We hope that this is enough detail for readers to understand the techniques and how they could be transferred.

Comments from Dr Stephen-Mark Cooper

\*If I have any criticisms of the MS at all it is to do with the lack of any consideration of confirmation of underlying assumptions related to the choice of statistical tests, and no consideration of inferential indices such as effect sizes and power related to stated hypotheses (other than when justifying the sample size). Sometimes, consistency in the expression of terms; for example the use of hyphenation throughout, the superscripted numbers identifying reference citations and when multiple citations are

used – the use of *et al* – this is an abbreviation and should be presented in italics (it is a Latin term) thus (*et al.*). But, I guess this is likely to be corrected at the proof reading stage of the MS. I also found the amount of information, and the manner in which it was being expressed, in Tables 1 and 2 confusing. I would also like to see all statistical indices expressed in italics where appropriate (e.g. *P*, *r*, *n*, *t*, *F*, *U*, *X*<sup>2</sup> etc.) with the relevant degrees of freedom subscripted as appropriate (e.g. *r*<sub>df</sub>, *t*<sub>df</sub>, *F*<sub>df</sub>, *U*<sub>df</sub> etc.).

clear there is not enough narrative to give the reader a sense of what exactly they relate to.

We have italicised '*et al*' however as per BMJ Open style we have not italicised statistical indices or subscripted degrees of freedom. We have added degrees of freedom to Table 1 and in the text where applicable.

\*Line 27: ... and seven did not provide ...

This has been changed as suggested.

\*Line 50: ... and self- and proxy-rated quality ...

This has been changed as suggested.

\*Line 52: what is DEMQOL? You've given the full names of inventories previously without the acronym.

The DEMQOL is a health related quality of life tool for dementia, however DEMQOL is the name of the tool, it is not an abbreviation. This has been clarified in the abstract and methods section.

\*Page 4, line 5: (and elsewhere in the MS) ... *P* = 0.011 (spacing and italics) Line 10: the same thing ... (95%CI 0.004 to 0.34, *P* = 0.045).

This has been corrected throughout.

\*Line 57: in the copy of the MS that I downloaded for review purposes the side-heading 'Strengths and Limitations:' needs to be moved down so that it appears at the top of p.5.

This has been changed as suggested.

\*Introduction (page 6)

\*Line 40: (and throughout the MS) ... study (*n* = 21) of four ...

This has been corrected throughout.

\*Page 7, line 28: I'm not sure that the sentence is full enough. I get the idea but would it be better as: 'Secondary outcomes of interest were to interrogate improvements in levels of agitation and other behavioural disturbances ....

The sentence has been corrected.

\*Methods (page 7)

\*Line 38: ... nursing homes. The study protocols ...

This has been changed as suggested.

\*Line 57: .... Sydney within a one-hour ...

This has been changed as suggested.

\*Page 8, line 12: ... were randomised to an intervention group or a control group. Eight ...

This has been changed as suggested.

\*Line 39: ... within six months, experiencing ...  
This has been changed as suggested.

\*Page 9, line 7: '2) Between nine and 12 humour ...  
This has been changed as suggested.

\*Lines 19 & 21: might have not may have  
These have been changed as suggested.

\*Line 27: LaughterBoss  
This has been changed as suggested.

\*Line 34: I'm not sure that the note about the purchase of the DVD is in keeping with the rubric of the journal!  
This has been deleted.

\*Line 46: ... low care). A random number generator in Excel was used ....  
This has been changed as suggested.

\*Page 10, line 22: (and elsewhere in the MS) ... which is a clinician-rated depression scale.<sup>19</sup> ... The issue is to do with the consistency of hyphenation as I mentioned earlier. Also on page 10 there seems to be a problem with spacing and the superscripted reference citation numbers.  
These have been changed as suggested.

\*Line 29: ... behaviours (a mixed economy of spelling in the MS); the eight-item withdrawal ...  
The spelling of behaviours has been corrected throughout as has randomised. Eight has been changed as suggested.

\*Line 33: what is DEMQOL? You've given the inventory names in full above followed by its acronym. What about DEMQOL?  
The DEMQOL is a health related quality of life tool for dementia, however DEMQOL is the name of the tool, it is not an abbreviation. This has been clarified in the abstract and methods section.

\*Line 57: again, in the copy of the MS that I downloaded for review purposes the side-heading 'Analysis' needs to be moved down so that it appears at the top of p.11. However, should it not be 'Analyses' as there are more than one?  
This has been changed as suggested.

\*Page 11, line 5: ... effect size (Cohen's  $d = 0.5$ ) difference Line 10: what is a RACF study?  
This has been changed as suggested. RACF stands for residential aged care facility; this has been changed to nursing home.

\*Lines 12 & 15: ... intention-to-treat  
This has been changed as suggested.

\*Line 15: I think you need a reference for the SAS V9.2 software here  
This reference has been added.

\*Line 17: t-test and U tests. Additionally, I guess the decision to use a parametric or non-parametric test of difference here at baseline was made on the nature of the scores collected (real number series – nominal, ordinal, interval or ratio)? But you deal with it simply by saying they were ' ... examined ...



as appropriate'. I think you need to be more specific about how decisions were made – on what criteria. Also, the t-test tests the  $H_0: \mu_1 = \mu_2$ ; the  $\chi^2$  test tests the  $H_0: O = E$ , and the Mann-Whitney test tests (probably here)  $H_0: \eta_1 = \eta_2$  yet no mention is made of underlying assumptions and hypotheses on the selection of the 'appropriate' tests and their outcomes.  
Detail on test selection has been added.

\*What level of statistical significance did you set throughout your study and why? From reading the MS I'm presuming alpha was set at  $P \leq 0.05$  but you need to justify why this was chosen in relation to the type of error you were willing to make.

The level and choice of alpha have been added.

\*Line 24: ... both within-resident correlation ...

This has been changed as suggested.

\*Line 28: ... was adjusted using an approach recommended by Fitzmaurice et al.<sup>30</sup> ...

This has been changed as suggested.

\*Line 46: I think you need to add a reference for the Blom transformation here

This reference has been added.

\*Line 53: you say that the ratings were 'highly' correlated ( $r = 0.863$ ). But this only gives a coefficient of determination of 74.5%. Presumably, this correlation was statistically significant and if so what was the P-value? Put the degrees of freedom for the zero-order correlation as a subscript here ( $r_{df}$ )

The p-value and degrees of freedom have been added.

\*Line 55: ... to the primary analysis as outlined above.

This has been changed as suggested.

## Results (page 12)

\*Line 12: do you mean that the difference between groups was zero? Or do you mean that the difference between the groups was statistically non-significant? But we still do not know what your alpha level was and why you chose that level of significance.

We have clarified that there were no significant differences between groups.

\*Line 22: ... delivered (average ( $\pm$ SD) of  $11.24 \pm 0.97$  per facility). At least I'm assuming it is the SD and I question whether this sort of information can be summarised to this level of precision and certainly not to 2 decimal places. The same goes for the summary of sessions cited on line 24.

We have clarified this statement, and reported in whole numbers.

\*Line 34: it might be better to express this as '... over time, but the group by time interactions on depression, non-agitation behavioural disturbance, social engagement or resident- proxy-rated quality of life were non-significant ( $P > 0.05$ ).' That's assuming that the alpha was set at  $\leq 0.05$ .

This has been changed as suggested.

\*Line 38: ... The group-by-time interaction was statistically significant for ... for covariates ( $P < 0.05$ ).'

That's assuming that the alpha was set at  $\leq 0.05$ .

This has been changed as suggested.

\*Line 45: I think there is an error here as the 95%CI -0.004 to 0.34 includes zero and so cannot be statistically significant. In the abstract (p.4) you give this 95%CI as 0.004 to 0.34 which is probably correct. Also,  $P = 0.045$  (spacing, caps and italics) – this is the case for all P-values on page 12 and

throughout the rest of the MS.  
Thank you for noticing this error. This has been corrected.

\*Line 51: ... was statistically significant with ...  
This has been changed as suggested.

\*Page 13, line 3: 'There were statistically significant engagement dose-by-time ... depression F df = 6.72, P = 0.000 ... You need to subscript the degrees of freedom for all the F-ratios and caps and italics for all values of P in this paragraph.  
This has been changed as suggested.

Discussion (page 13)

\*Line 33: ... were 3.00 (95%CI 1.78 to 4.22)<sup>31</sup>. ...  
This has been changed as suggested.

\*Line 38: I think you need a reference here to endorse that risperidone is the most commonly used antipsychotic in Australia.  
This reference has been added.

\*Page 15, line 30: might not may.  
This has been changed as suggested.

Table 1 (page 20)

\*This is a 'busy' table. Italics for all n and italics and subscripted dfs for all values of t, X<sup>2</sup> and U. Caps and italics for all P-values. Also check the spacing of all values – e.g. humour therapy/number with dementia = 145 (76.7%) and humour therapy/years lived in care = 2.8 ± 3.1.  
We have not italicised as this does not match BMJ Open style. We have checked spacing and added degrees of freedom.

Table 2 (page 22)

\*This is an even busier table! I think that the legend of the table could include more pertinent detail and the same issues as identified for Table 1 above are also relevant here. On p.23 'effect size' is mentioned but I don't recollect it being mentioned as part of your analysis methodology.  
The description of 'effect size' has been changed to 'adjusted mean difference', which is more accurate.

## VERSION 2 – REVIEW

<b>REVIEWER</b>	Cooper, Steve University of Wales Institute, Cardiff, Physiology and Health
<b>REVIEW RETURNED</b>	07-Dec-2012

<b>REPORTING &amp; ETHICS</b>	There are no competing interests in this research.
<b>GENERAL COMMENTS</b>	I am satisfied that the authors have adjusted the MS sufficiently well and have covered all the areas of concern that I raised in my initial review of the MS