

## 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

|                            |   |
|----------------------------|---|
| <b>TITLE (PROVISIONAL)</b> | Online versus paper-based screening for depression and anxiety in adults with cystic fibrosis in Ireland: a cross-sectional exploratory study                               |
| <b>AUTHORS</b>             | Cronly, Jennifer; Duff, Alistair; Riekert, Kristin A.; Perry, Ivan; Fitzgerald, Anthony; Horgan, Aine; Lehane, Elaine; Howe, Barbara; Ni Chroinin, Muireann; Savage, Eileen |

### VERSION 1 – REVIEW

|                        |  |
|------------------------|--|
| <b>REVIEWER</b>        | Bradley Quon<br>University of British Columbia, Canada |
| <b>REVIEW RETURNED</b> | 25-Sep-2017  |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | <p>This is the first study to examine paper-based vs. online depression and anxiety screening in CF. The study was not initially designed to compare the two as the study started with paper-based and then converted to online through the CF Ireland community network when recruitment with paper-based was deemed poor. As a result, there's self-selection bias as identified by the authors which limits the generalizability. Furthermore, their sample only represented 28% of the adult CF population in Ireland. The authors carefully discuss the implications of online screening on clinic care delivery and the need for appropriate follow-up.</p> <p>The paper-based and online responses were not anonymous and therefore this should be mentioned earlier in the methods section (it is currently mentioned in the discussion section only)</p> <p>Why was the HADS administered and not the PHQ-9 and GAD-7? Rationale should be included much earlier in the introduction or methods.</p> <p>Did the patient need to be clinically stable when completing the questionnaire? If the patient is sick when filling out the questionnaire, responses may not be representative of their usual response.</p> <p>Page 6, line 16: It's not completely accurate to mention there were specific differences for categorical variables (i.e. those with &gt;2 categories) as the chi-square test examines the difference in the categorical distribution for the variable and not the statistical difference for an individual category between groups.</p> |
|-------------------------|---|

|  |  |
|--|--|
|  | <p>Page 7, line 23: It is unclear what the confidence intervals provided represent?</p> <p>Page 7, line 30-40: The text is redundant with the table. This can be more concise.</p> <p>Page 5, Line 30: FEV% should read FEV1%</p> <p>Page 5, Line 49: Fischer's should be Fisher's</p> <p>Page 6, line 15: repeat sentence</p> |
|--|--|

|                        |  |
|------------------------|--|
| <b>REVIEWER</b>        | Wouter van Ballegooijen<br>Vrije Universiteit Amsterdam, Netherlands |
| <b>REVIEW RETURNED</b> | 08-Oct-2017  |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | <p>This is a well-written manuscript and relevant to the field. I also commend the authors for writing this paper, because apparently it was not the original intention of their study to compare media for administering questionnaires. This paper practically legitimises further analyses of the main study. In my opinion that is a reason to publish it. I have a major comment, however. So far, research on equivalence of paper and online questionnaires, especially concerning the HADS, has not focussed on measurement invariance. This applies to this study as well. Differences in mean total scores and prevalence may be due to differences in the sample and do not indicate different psychometric properties. A multiple-group measurement invariance test would test equivalence of factor structures and item intercepts, which would answer the question very well. Another comment is that instead of testing a different format, you are also testing a different setting, i.e. the circumstances of administering the questionnaire is different between conditions.</p> <p>For further explanation, see<br/><a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4766750/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4766750/</a></p> |
|-------------------------|---|

### VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Bradley Quon

Institution and Country: University of British Columbia, Canada Competing Interests: None declared

**COMMENT:** This is the first study to examine paper-based vs. online depression and anxiety screening in CF. The study was not initially designed to compare the two as the study started with paper-based and then converted to online through the CF Ireland community network when recruitment with paper-based was deemed poor. As a result, there's self-selection bias as identified by the authors which limits the generalizability. Furthermore, their sample only represented 28% of the adult CF population in Ireland. The authors carefully discuss the implications of online screening on clinic care delivery and the need for appropriate follow-up.

**RESPONSE:** Yes, we acknowledged the limits of generalisability in this study due to self-selection bias and the small sample size in the discussion section (in the 2nd paragraph on p.10)

COMMENT: The paper-based and online responses were not anonymous and therefore this should be mentioned earlier in the methods section (it is currently mentioned in the discussion section only)

RESPONSE: Screening can never be completed anonymously as services need to take action when depression and anxiety scores are elevated. This point is now more explicitly mentioned in the methods section (in the 2nd sentence under Data Collection Procedures on p.4). The referral process was mentioned in the original manuscript under the Methods section and has been left in (in the last sentence under Measures on p.5)

COMMENT: Why was the HADS administered and not the PHQ-9 and GAD-7? Rationale should be included much earlier in the introduction or methods.

RESPONSE: The HADS measure was the one selected by the International TIDES study protocol and which our study adopted. Our study had commenced including data collection before the publication of a consensus statement by the International Committee on Mental Health in CF (Quittner et al. 2016 published ahead of print in 2015) recommending the PHQ-9 and GAD-7 as screening measures in the CF population. We have noted this now in the last paragraph of section on Measures.

COMMENT: Did the patient need to be clinically stable when completing the questionnaire? If the patient is sick when filling out the questionnaire, responses may not be representative of their usual response.

RESPONSE: No. Our sample was cross-sectional and one of convenience at scheduled visits or online for some adults. None of the patients were hospitalised at the time of data collection and these would more likely have been sicker than those not hospitalised. That said, as a cross-sectional sample and given the nature of CF illness, we can expect that CF patients will always fluctuate between being well and less well or sick at the point of data collection.

COMMENT: Page 6, line 16: It's not completely accurate to mention there were specific differences for categorical variables (i.e. those with >2 categories) as the chi-square test examines the difference in the categorical distribution for the variable and not the statistical difference for an individual category between groups.

RESPONSE: Yes, this was incorrectly worded and has been changed (see 1st subheading under results section).

COMMENT: Page 7, line 23: It is unclear what the confidence intervals provided represent?

RESPONSE: Yes, this was not clear in the text. It has now been reworded to indicate that the confidence represents prevalence.

COMMENT: Page 7, line 30-40: The text is redundant with the table. This can be more concise.

RESPONSE: The redundant text has now been removed (see deletion in subheading: HADS scores and physical health outcome variables)

COMMENT: Page 5, Line 30: FEV% should read FEV1%

RESPONSE: FEV% has been changed to FEV1% in the text.

RESPONSE: Page 5, Line 49: Fischer's should be Fisher's Page 6, line 15: repeat sentence These typographical errors have been corrected in the text.

Reviewer: 2

Reviewer Name: Wouter van Ballegooijen

Institution and Country: Vrije Universiteit Amsterdam, Netherlands Competing Interests: None declared

COMMENT: This is a well-written manuscript and relevant to the field. I also commend the authors for writing this paper, because apparently it was not the original intention of their study to compare media for administering questionnaires. This paper practically legitimises further analyses of the main study. In my opinion that is a reason to publish it. I have a major comment, however. So far, research on equivalence of paper and online questionnaires, especially concerning the HADS, has not focussed on measurement invariance. This applies to this study as well. Differences in mean total scores and prevalence may be due to differences in the sample and do not indicate different psychometric properties. A multiple-group measurement invariance test would test equivalence of factor structures and item intercepts, which would answer the question very well.

RESPONSE: Yes, a multiple-group measurement invariance test conducted within the context of factor analysis would be needed to assess equivalence. Since the international guidelines for screening depression and anxiety are now recommending the PHQ-9 and GAD-7 as screening measures in the CF population (and already are being used in practice in some countries), a post-hoc assessment of HAD equivalence administered online vs paper based seems of little value. Also, the small sample size needed for measurement invariance was considered in our decisions. However, the reviewer's comment is important and so we have added the need for this in our manuscript when referring to the international recommendation to use the PHQ-9 and GAD-7 in the Cf population. (See page 10 2nd paragraph),

COMMENT: Another comment is that instead of testing a different format, you are also testing a different setting, i.e. the circumstances of administering the questionnaire is different between conditions.

RESPONSE The possibility that setting may have acted as a confounding variable is now included in the Discussion section under limitations (2nd paragraph on p.10). Also, we have includes settings within the context of measurement invariance as above.

#### VERSION 2 – REVIEW

|                        |  |
|------------------------|--|
| <b>REVIEWER</b>        | Bradley Quon<br>University of British Columbia, Canada |
| <b>REVIEW RETURNED</b> | 20-Nov-2017  |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | The authors have responded satisfactorily to my concerns. |
|-------------------------|---|

|                        |  |
|------------------------|--|
| <b>REVIEWER</b>        | Wouter van Ballegooijen<br>Vrije Universiteit Amsterdam, Netherlands |
| <b>REVIEW RETURNED</b> | 19-Nov-2017  |

|                         |                      |
|-------------------------|----------------------|
| <b>GENERAL COMMENTS</b> | No further comments. |
|-------------------------|----------------------|