

## 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>	A cross-sectional study to develop and describe psychometric characteristics of a patient-reported instrument (PROFFIT) for measuring financial toxicity of cancer within a public healthcare system
<b>AUTHORS</b>	Riva, Silvia; Arenare, Laura; Di Maio, Massimo; Efficace, Fabio; Montesarchio, Vincenzo; Frontini, Luciano; Giannarelli, Diana; Bryce, Jane; Del Campo, Laura; De Lorenzo, Francesco; Iannelli, Elisabetta; Tracò, Francesca; Gitto, Lara; Jommi, Claudio; Vaccaro, Concetta Maria; Barberio, Daniela; Cinieri, Saverio; Porta, Camillo; Del Mastro, Lucia; Zagonel, Vittorina; Cogoni, Alessio Aligi; Bordonaro, Roberto; Gimigliano, Anna; Piccirillo, Maria Carmela; Guizzaro, Lorenzo; Gallo, Ciro; Perrone, Francesco

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Jeremy P Harris University of California Irvine
<b>REVIEW RETURNED</b>	14-Apr-2021

<b>GENERAL COMMENTS</b>	<p>The authors report a quantitative analysis of a novel survey created using an Italian population. The interest in capturing and testing financial toxicity is commendable. In general, the methodology is expertly done and one of the strengths of the paper.</p> <p>The main critique has to do with generalizability. It should be made clear that generalizability of the Determinants has to be limited to countries with a national health service or public option. However, the dichotomy is not as straightforward as the authors sometimes imply (it is not just a private-payer system like in the US versus having an available universal National Health Service). There are also grey area systems in regards to the payer systems (for example Canada's healthcare system, for which about 2/3 of the population has supplementary private insurance). In fact, even in the US systems like Medicaid and Medicare complicate the picture of financial toxicity in the US. A greater amount of detail of the uniqueness of the Italian healthcare system, how many patients have private care, and the average amount of out-of-pocket expenditures would help. Specifically, I am concerned about the following questions from the survey when generalizing outside of Italy:</p> <ul style="list-style-type: none"><li>- The NHS covers all health costs related to my illness</li><li>- I have paid for one or more private medical examinations for my illness</li><li>- I have paid for additional medicines or supplements related to my illness</li></ul> <p>It should be clarified that the Financial Toxicity (7-question) portion is most-easily transferable among systems.</p>
-------------------------	---

	<p>I would like to see some recommendations about generalizing the questionnaire beyond Italian and English. Should the original Italian be used? The committee approach from this group was well-conducted, but it could certainly limit the use of the questionnaire if comparable methods were recommended for other languages.</p> <p>It would be very useful to see the time from initial diagnosis to the time the survey was conducted in table 1. In the United States the length of cancer therapy is highly correlated to financial toxicity (development of bankruptcy, loss of private insurance). In a study that is meant to cover all varieties of cancer and across all levels of treatment, the time from diagnosis to survey administration could be an important surrogate for the amount of therapy and out-of-pocket expenses.</p> <p>Minor</p> <ol style="list-style-type: none"> <li>1. Worth having another English speaker look through this. In general the English is excellent, but I am able to tell there were parts written by a non-native speaker (beyond what I think is just a difference between American and British English). For example, Figure 1 is awkward (“travels”, “long way hospital”, “money for travels”). Or on page 9 line 5-8, “These items were tested as for comprehensibility...”.</li> <li>2. Imputation of the mean was utilized to account for missing information. That does not seem to be the best approach, since, for example, as the authors point out, 37% of information was missing for the five items related to job, from patients who declared themselves retired or jobless. It would be misleading to assume that the patients who were without jobs had the same financial issues as those with jobs (hence mean imputation would not be accurate).</li> <li>3. It would be useful to see the items that were excluded for intraclass correlation and communality as well as the results of these tests in the supplement (seeing as how these are the primary tests leading to the final survey). How close were these additional items from meeting the criteria for inclusion?</li> <li>4. The scoring procedure is a little confusing. I would suggest adding one or two examples to make clear that the FT score is separate from the determinants score. How the determinants score is summed is a little confusing since there are two separate scoring procedures.</li> </ol>
--	--

<b>REVIEWER</b>	Salene Jones Fred Hutchinson Cancer Research Center
<b>REVIEW RETURNED</b>	14-Apr-2021

<b>GENERAL COMMENTS</b>	<p>This paper reports on the development of a financial toxicity measure in cancer among Italian patients. Financial toxicity is a substantial problem in cancer care and better measures are needed to assess this crucial aspect of quality of life. This is a very well-written paper. The following specific comments on the reported study are provided.</p> <ol style="list-style-type: none"> <li>1. This paper would be greatly enhanced by construct validity analyses. Without construct validity analyses, this measure cannot be considered usable by other research teams or in clinical care. Even a cross-sectional construct validity analysis would be helpful if the authors are waiting on longitudinal data.</li> <li>2. The use of varimax rotation is difficult to justify. It is also unclear if principal components analysis was used and if so, an alternative method of estimation would be more appropriate for determining the actual latent structure. Also, scree plots tend to be unreliable.</li> </ol>
-------------------------	--

	<p>3. The sample mean imputation for the employment items seems difficult to justify.</p> <p>4. The outcomes and determinants subscales would be an interesting feature of this scale but it is difficult to judge without the item content reported.</p> <p>5. If the authors have time since diagnosis data, that should be reported. Financial toxicity can change and compound over time so knowing the time since diagnosis for the sample will inform whether this measure is better for those in or recently in treatment or for long-term survivors.</p> <p>6. Financial toxicity literature is increasingly showing that this is a multidimensional construct. Given that there are already general measures such as the COST in cancer and general financial hardship measures, it is unclear what an additional general financial hardship measure will add. Some of this would be addressed by providing the item content in an appendix.</p> <p>7. The authors should be commended for providing this instrument free of charge.</p>
--	---

<b>REVIEWER</b>	Fumiko Chino Memorial Sloan Kettering Cancer Center, Radiation Oncology
<b>REVIEW RETURNED</b>	22-Apr-2021

<b>GENERAL COMMENTS</b>	<p>Authors present on the develop a patient reported tool for assessing financial toxicity in a country with a public/nationalized healthcare system. This manuscript is important and useful as the bulk of financial toxicity research has been done in the US with fee for service insurance and high copays; patients receiving care in another healthcare system with different access and affordability concerns may require a different screening/assessment tool. Overall the methodology appears strong.</p> <p>Major concerns:</p> <p>1) Please be more explicit about the limitations of this analysis and tool development. For example, length of treatment and stage of disease does not appear to be captured and would anticipate that this may factor greatly into financial toxicity. Having a representative sample for this development project is important. Additionally race/ethnicity may be a limiting factor with using this instrument in other racially/ethnically diverse countries with nationalized healthcare when Italy is relatively homogenous [so much so that race/ethnicity is not even reported here].</p> <p>2) As many readers will not necessarily be familiar with the Italian health care system, would briefly explaining the healthcare system in Italy and compare/contrast it to other nationalized healthcare systems vs mixed public/private systems like the US.</p> <p>3) In figure 1 it is not immediately clear to this reviewer how #14-16 “Health workers” fit into the framework. Do you mean the lack of training for these parties in considering patient costs/financial burden?</p> <p>Minor concerns:</p> <p>1. Overall well written however that are some instances where the phrasing a little odd or inefficient: For example: “While developing PROFFIT, a complex matter derived from management of items related to job activities. Indeed, around one third of patients did not respond to these items.” Additionally some text is actually not well understood, example: “We decided not to define a fixed temporal frame to which refer the response, differently from what is frequently done in patient reported</p>
-------------------------	---

	<p>outcomes.” I understand what you mean after reading the full paragraph but the first line was initially a bit puzzling.  Figure 1: “Long way hospital” should be probably be rephrased as “Distance from Hospital” or similar  2. page 10, line 3, typo? “Consequently”?  3. page 11, line 40: missing words? “In principle, such deviation might reduce Therefore,”</p>
--	---

**VERSION 1 – AUTHOR RESPONSE**

Reviewer: 1  
Dr. Jeremy P Harris, University of California Irvine

Comments to the Author:  
The authors report a quantitative analysis of a novel survey created using an Italian population. The interest in capturing and testing financial toxicity is commendable. In general, the methodology is expertly done and one of the strengths of the paper.

**Response: thanks a lot.**

1. The main critique has to do with generalizability. It should be made clear that generalizability of the Determinants has to be limited to countries with a national health service or public option. However, the dichotomy is not as straightforward as the authors sometimes imply (it is not just a private-payer system like in the US versus having an available universal National Health Service). There are also grey area systems in regards to the payer systems (for example Canada’s healthcare system, for which about 2/3 of the population has supplementary private insurance). In fact, even in the US systems like Medicaid and Medicare complicate the picture of financial toxicity in the US. A greater amount of detail of the uniqueness of the Italian healthcare system, how many patients have private care, and the average amount of out-of-pocket expenditures would help. Specifically, I am concerned about the following questions from the survey when generalizing outside of Italy:

- The NHS covers all health costs related to my illness
- I have paid for one or more private medical examinations for my illness
- I have paid for additional medicines or supplements related to my illness

It should be clarified that the Financial Toxicity (7-question) portion is most-easily transferable among systems.

**Response: we perfectly agree with this comment. Out of the two parts of the PROFFIT questionnaire, in principle, the 7-item FT-score could be immediately generalizable to every system, once validity has been confirmed, while the 9 single-item determinants are strictly dependent on the healthcare system. The latter ones were acknowledged by patients in the cognitive interviews and should be the variable part of the questionnaire to be assessed in the various frameworks. Possibly, as it seems from other comments too, we were confusing in reporting and discussing the two parts. Therefore; (1) we have clarified the paragraph on limitations in terms of generalizability outside Italy in the discussion, even explaining the “grey zone” and that caution should be paid to determinants; (2) we have added a description of the Italian NHS and data on the extension of the coverage and the amount of out-of-pocket annual expenditure.**

2. I would like to see some recommendations about generalizing the questionnaire beyond Italian and English. Should the original Italian be used? The committee approach from this group was well-conducted, but it could certainly limit the use of the questionnaire if comparable methods were recommended for other languages.

**Response: methodology of translation and cross-cultural validation of PROs has been outlined by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the manuscript contains three papers where some indication can be found on this point (references #10, #11, and #15). As for the PROFFIT questionnaire we translated the Italian version for better understanding by the scientific community, and, as we said above, in principle, the 7-item FT-score could be easily transferable to the English-speaking countries while waiting for a complete cross-cultural validation of the whole questionnaire.**

3. It would be very useful to see the time from initial diagnosis to the time the survey was conducted in table 1. In the United States the length of cancer therapy is highly correlated to financial toxicity (development of bankruptcy, loss of private insurance). In a study that is meant to cover all varieties of cancer and across all levels of treatment, the time from diagnosis to survey administration could be an important surrogate for the amount of therapy and out-of-pocket expenses.

**Response: we have added these data in table 1. As you can see, it seems that, in Italy, time from cancer diagnosis is not related to financial toxicity; however, together with other clinical questions, it will be better assessed in a larger sample in the ongoing step 4 of the project.**

Minor

1. Worth having another English speaker look through this. In general, the English is excellent, but I am able to tell there were parts written by a non-native speaker (beyond what I think is just a difference between American and British English). For example, Figure 1 is awkward (“travels”, “long way hospital”, “money for travels”). Or on page 9 line 5-8, “These items were tested as for comprehensibility...”.

**Response: we have reviewed the language at our best. The figure has been modified.**

2. Imputation of the mean was utilized to account for missing information. That does not seem to be the best approach, since, for example, as the authors point out, 37% of information was missing for the five items related to job, from patients who declared themselves retired or jobless. It would be misleading to assume that the patients who were without jobs had the same financial issues as those with jobs (hence mean imputation would not be accurate).

**Response: we understand that some confusion arose by the way we reported methods and results; particularly, it seems that an excessive use of mean imputation was perceived by this Reviewer and by Reviewer #2. However, we believe this was not the case and we try to clarify it here and in the paper. A more detailed description of the analysis has been reported in the appendix that has been extensively modified.**

**Once we ascertained that 68/184 (37%) patients, who declared themselves retired or jobless (i.e. househusbands, housewives or individuals in search of employment), did not respond to items related to job, we decided to perform between-item correlation analysis separately for job items, with a denominator of 116 cases, and for all the other items, with a denominator of 184 cases. From this analysis, three job-related items were excluded and two remained (one among the outcome items and one among the determinants). Therefore, no use of mean imputation was done in the between-item correlation analysis.**

**Thereafter, we decided to perform EFA within two parallel data-sets: the restricted sample (where patients with missing information were excluded) and the full sample (where the mean value was imputed). However, within the analysis of determinants, the job-related item (Q102) was the first one removed in the restricted sample due to the lowest communality; therefore, the analysis in the remaining 14 items was only performed in the full sample; and, again, no use of mean imputation was done in the EFA of determinants.**

**Within the analysis of the Financial score (7 items, one of which was job-related – Q99) results are reported in details for both data-sets and do substantially agree.**

**We chose to input the average score rather than the minimum score because the latter could be reasonable for retired people, but not for younger people without job. Thus, imputing the minimum score would definitely bias the score toward the null, while imputing the average could possibly only slightly overestimate the financial issues. Further, this choice is consistent with the calculus of the score, where the missing items are not considered in the denominator. This question will be further dealt with in the next validation steps. Actually, this was the only step of the analysis where we used mean imputation and it was done in parallel with the analysis without mean imputation.**

**We understand that this pathway was not clearly reported, and we have modified the result section. In addition, we have slightly modified the paragraph of the discussion, where we also explained the direction of the potential biases arising in the two approaches.**

3. It would be useful to see the items that were excluded for intraclass correlation and communality as well as the results of these tests in the supplement (seeing as how these are the primary tests leading to the final survey). How close were these additional items from meeting the criteria for inclusion?

**Response: A more detailed description of the questionnaire development has been added in the appendix, reporting findings at each analysis step.**

**Further, following this comment, we have added in the appendix the complete list of 30 Italian items that were the starting point for analyses reported in this paper, although we have the formal English translation only for the items of the final questionnaire.**

4. The scoring procedure is a little confusing. I would suggest adding one or two examples to make clear that the FT score is separate from the determinants score. How the determinants score is summed is a little confusing since there are two separate scoring procedures.

**Response. ok, we added some examples for the calculation of the FT score and of single determinants. For determinants there is no summary score; they have to be used as single items.**

Reviewer: 2

Dr. Salene Jones, Fred Hutchinson Cancer Research Center

Comments to the Author:

This paper reports on the development of a financial toxicity measure in cancer among Italian patients. Financial toxicity is a substantial problem in cancer care and better measures are needed to assess this crucial aspect of quality of life. This is a very well-written paper.

**Response: thanks.**

The following specific comments on the reported study are provided.

1. This paper would be greatly enhanced by construct validity analyses. Without construct validity analyses, this measure cannot be considered usable by other research teams or in clinical care. Even a cross-sectional construct validity analysis would be helpful if the authors are waiting on longitudinal data.

**Response: as outlined in the protocol (Riva et al., BMJ, 2019), construct and criterion validity, sensitivity, reliability, generalizability will be assessed in a larger independent sample in step 4 of our project (recruitment is ongoing). To date we can see that the PROFFIT score is sensitive to differences between groups of patients who could be anticipated to be different (Italian macro-regions, age, education level and burden disease beyond cancer). We added information in the paper (Results and Discussion) and in the Appendix (Table S5); however, we may anticipate that these findings could suffer from an overfitting bias and need to be independently validated.**

2. The use of varimax rotation is difficult to justify. It is also unclear if principal components analysis was used and if so, an alternative method of estimation would be more appropriate for determining the actual latent structure. Also, scree plots tend to be unreliable.

**Response: For EFA we used the Principal Axis Factor (PAF) extraction, with Varimax rotation, that is the default option of SPSS and the most used approach. We were looking for a single factor, and orthogonal rotation allowed us to remove items with low communalities and factor loadings. The same results were found when oblique Promax rotation was applied. Only minimal differences were found in the restricted sample, where the similar interpretation of the two axes was emphasized. Of course, no difference can be found in the full sample where only one factor is retained.**

3. The sample mean imputation for the employment items seems difficult to justify.

**Response: we report here the same response given to Reviewer #1. We understand that some confusion arose by the way we reported methods and results; particularly, it seems that an excessive use of mean imputation was perceived by this Reviewer and by Reviewer #1. However, we believe this was not the case and we try to clarify it here and in the paper. A more detailed description of the analysis has been reported in the appendix that has been extensively modified.**

**Once we ascertained that 68/184 (37%) patients, who declared themselves retired or jobless (i.e. househusbands, housewives or individuals in search of employment), did not respond to**

items related to job, we decided to perform between-item correlation analysis separately for job items, with a denominator of 116 cases, and for all the other items, with a denominator of 184 cases. From this analysis, three job-related items were excluded and two remained (one among the outcome items and one among the determinants). Therefore, no use of mean imputation was done in the between-item correlation analysis.

Thereafter, we decided to perform EFA within two parallel data-sets: the restricted sample (where patients with missing information were excluded) and the full sample (where the mean value was imputed). However, within the analysis of determinants, the job-related item (Q102) was the first one removed in the restricted sample due to the lowest communality; therefore, the analysis in the remaining 14 items was only performed in the full sample; and, again, no use of mean imputation was done in the EFA of determinants.

Within the analysis of the Financial score (7 items, one of which was job-related – Q99) results are reported in details for both data-sets and do substantially agree.

We chose to input the average score rather than the minimum score because the latter could be reasonable for retired people, but not for younger people without job. Thus, imputing the minimum score would definitely bias the score toward the null, while imputing the average could possibly only slightly overestimate the financial issues. Further, this choice is consistent with the calculus of the score, where the missing items are not considered in the denominator. This question will be further dealt with in the next validation steps. Actually, this was the only step of the analysis where we used mean imputation and it was done in parallel with the analysis without mean imputation.

We understand that this pathway was not clearly reported, and we have modified the result section. In addition, we have slightly modified the paragraph of the discussion, where we also explained the direction of the potential biases arising in the two approaches.

4. The outcomes and determinants subscales would be an interesting feature of this scale but it is difficult to judge without the item content reported.

**Response: item content is reported in table 2 both in Italian and English translation to make comprehension easier. However, we point out that single items are used for determinants.**

5. If the authors have time since diagnosis data, that should be reported. Financial toxicity can change and compound over time so knowing the time since diagnosis for the sample will inform whether this measure is better for those in or recently in treatment or for long-term survivors.

**Response: we report here the same response given to Reviewer #1. We have added these data in table 1. As you can see, it seems that, in Italy, time from cancer diagnosis is not related to financial toxicity; however, together with other clinical questions, it will be better assessed in a larger sample in the ongoing step 4 of the project.**

6. Financial toxicity literature is increasingly showing that this is a multidimensional construct. Given that there are already general measures such as the COST in cancer and general financial hardship measures, it is unclear what an additional general financial hardship measure will add. Some of this would be addressed by providing the item content in an appendix.

**Response: item content was already reported in table 2, both in Italian and English translation to make comprehension easier. The opportunity of having context specific instruments to measure financial toxicity has been previously underlined by other Authors (ref #21 Rotter J, Spencer JC, Wheeler SB. Financial Toxicity in Advanced and Metastatic Cancer: Overburdened and Underprepared. J Oncol Pract 2019; JOP1800518) and by us (ref #20). In principle, an attempt to develop instruments within a given context should always been done when other available instruments have been developed in structurally different contexts. Of course strong structural differences exist between the US and Italian health system that might drive different determinants and estimators of FT, and this is why we started our project. Tools developed in different contexts might not necessarily be different; however, this is the case of COST and PROFFIT. As we discuss in the manuscript, PROFFIT is much more focused than COST on determinants of FT, and this seems very obvious in a system where patients feel that they should pay nothing for their disease. Rather, PROFFIT and COST are similar if we consider the FT score of PROFFIT, that is the score estimating the effect of FT at the single-patient level. Based on these data, it is clear that using COST in Italy would probably estimate quite well the degree of FT but could not generate hypotheses for contrasting it. On the contrary, PROFFIT does produce at the same time an estimate of the problem and gives indication on its determinant allowing to plan corrective actions.**

7. The authors should be commended for providing this instrument free of charge.

**Response: thank you.**

Reviewer: 3

Dr. Fumiko Chino, Memorial Sloan Kettering Cancer Center

Comments to the Author:

Authors present on the develop a patient reported tool for assessing financial toxicity in a country with a public/nationalized healthcare system. This manuscript is important and useful as the bulk of financial toxicity research has been done in the US with fee for service insurance and high copays; patients receiving care in another healthcare system with different access and affordability concerns may require a different screening/assessment tool. Overall the methodology appears strong.

Major concerns:

1) Please be more explicit about the limitations of this analysis and tool development. For example, length of treatment and stage of disease does not appear to be captured and would anticipate that this may factor greatly into financial toxicity. Having a representative sample for this development project is important. Additionally, race/ethnicity may be a limiting factor with using this instrument in other racially/ethnically diverse countries with nationalized healthcare when Italy is relatively homogenous [so much so that race/ethnicity is not even reported here].

**Response: unfortunately, we did not collect specific information on stage of disease, but it is clear from the data on time from initial diagnosis (added to table 1) that more than half of the patients were at advanced stage. We are further studying this within the ongoing external validation study. The comment on the race is absolutely correct. We did not report details because there were only three patients with ethnicity other than Caucasian (two from India and one from an unknown site).**

2) As many readers will not necessarily be familiar with the Italian health care system, would briefly explaining the healthcare system in Italy and compare/contrast it to other nationalized healthcare systems vs mixed public/private systems like the US.

**Response: we added the following paragraph in the discussion, also dealing with a comment pf Reviewer #1. The Italian health care system was shaped, since 1978, as a National Health Service (NHS) model, where the State is the most important financier, via general tax levies. [22] The NHS model prevails in Northern and Southern European Countries, whereas Central Europe is mostly characterized by social insurance-based model, funded by payroll taxes. Regardless the model, the European health care systems are characterized by a high proportion of healthcare expenditure covered by compulsory public programs, ranging from 66% in Spain to 78% in UK, compared to 49% in the USA. [19] The Italian NHS is decentralised, since regions are responsible for healthcare budget. [22] In Europe decentralisation does not depend on the healthcare system model: both NHS-shaped models (e.g. UK vs Spain) and social-insurance models (e.g. France vs Germany) are centralised vs decentralised respectively. Italy shows a lower intermediation of private expenditure than the other major European countries: in 2018 out-of-pocket expenditure accounted for 89% of private expenditure in Italy, compared to 40%, 55% and 75% in Germany, France and UK/Spain respectively. [23] The mean yearly amount of out-of-pocket expenses for cancer patients was estimated in the same year to be 1841 euros within a survey conducted by the Federazione italiana delle Associazioni di Volontariato in Oncologia – FAVO. [24]**

3) In figure 1 it is not immediately clear to this reviewer how #14-16 “Health workers” fit into the framework. Do you mean the lack of training for these parties in considering patient costs/financial burden?

**Response: nice comment. Lack of training may be the original cause and might be a target for intervention and corrective actions. Within the framework, patients clearly attribute economic damage to malpractice or low-level practice of health workers. And, within importance analyses, items related to these determinants were the top scorer. The phrasing in the figure has been changed to be more explicit (Support from health staff) and hopefully clearer.**



Minor concerns:

1. Overall well written however that are some instances where the phrasing a little odd or inefficient:

For example: “While developing PROFFIT, a complex matter derived from management of items related to job activities. Indeed, around one third of patients did not respond to these items.”

Additionally some text is actually not well understood, example: “We decided not to define a fixed temporal frame to which refer the response, differently from what is frequently done in patient reported outcomes.” I understand what you mean after reading the full paragraph but the first line was initially a bit puzzling.

Figure 1: “Long way hospital” should be probably be rephrased as “Distance from Hospital” or similar

**Response: thanks for these suggestions. The first phrase was removed, the second was rephrased and the figure was changed.**

2. page 10, line 3, typo? “Consequently”?

**Response: yes, corrected.**

3. page 11, line 40: missing words? “In principle, such deviation might reduce Therefore,”

**Response: yes, missing words and wrong sense. Rephrased.**

### VERSION 2 – REVIEW

<b>REVIEWER</b>	Salene Jones Fred Hutchinson Cancer Research Center
<b>REVIEW RETURNED</b>	29-Jun-2021

<b>GENERAL COMMENTS</b>	<p>This paper is a revision and reports on the development of financial toxicity questionnaire. The revision is only partially responsive to the previous review. The following changes are suggested.</p> <p>1) While it is good that the authors conducted a Promax factor analysis, this should replace the varimax rotation analysis. An uncorrelated factor structure is very difficult to justify and the authors should either switch to the Promax analysis or provide justification for the varimax rotation as requested in the previous review.</p> <p>2) Construct validity analyses are not justified. There should be a reason for examining those specific factors and there needs to an actual interpretation of the differences if demographic factors are used.</p> <p>3) While the authors have added a small amount of text about the multidimensional nature of financial hardship, the discussion still does not speak to how the new measure either does or does not address these issues.</p>
-------------------------	---

<b>REVIEWER</b>	Jeremy P Harris University of California Irvine
<b>REVIEW RETURNED</b>	06-Jul-2021

<b>GENERAL COMMENTS</b>	<p>The authors have adequately addressed my concerns and comments. I believe that COST v2 is now in use, so the discussion should focus on differences with PROFFIT and COST v2 rather than v1.</p>
-------------------------	---

### VERSION 2 – AUTHOR RESPONSE

*Reviewer: 1*

*Dr. Jeremy P Harris, University of California Irvine*

*Comments to the Author:*

*The authors have adequately addressed my concerns and comments. I believe that COST v2 is now*

*in use, so the discussion should focus on differences with PROFFIT and COST v2 rather than v1.*

**Response:** We have reported on page 16 that recently, the COST-FACIT version 2 has been developed. In this version, an additional item was added to reflect overall financial wellbeing. However, this additional item was not included in the calculation of the summary score in the original validation study and this makes difficult to make any comparisons with the US context, at the present time.

*Reviewer: 2*

*Dr. Salene Jones, Fred Hutchinson Cancer Research Center*

*Comments to the Author:*

*This paper is a revision and reports on the development of financial toxicity questionnaire. The revision is only partially responsive to the previous review. The following changes are suggested.*

1) *While it is good that the authors conducted a Promax factor analysis, this should replace the varimax rotation analysis. An uncorrelated factor structure is very difficult to justify and the authors should either switch to the Promax analysis or provide justification for the varimax rotation as requested in the previous review.*

**Response:** We partially disagree with the Reviewer because we think we must report what we did, and we first used Varimax rotation. Although we acknowledge that oblique rotation eventually better caught relationships between axes, as we clarified in the appendix, checking for orthogonal axes is the primary analysis within an exploratory approach, where the underlying structure is scrutinized without preliminary hypotheses. It is true that we hoped to find a unique FT scale, but we were working within the framework of a universal health system, and we could not rule out the option that items were only sparsely correlated and no definite structure was found. Actually, two axes were found with many items cross loading on them, leading to the Promax rotation. It is reassuring that similar findings were found with orthogonal and oblique rotations. We think that the whole path should be reported to readers, but we would not dismiss the Reviewer's request and we modified the paper, adding Promax in the Methods (page 7), while it was already reported in the Results (page 12) section of the paper.

2) *Construct validity analyses are not justified. There should be a reason for examining those specific factors and there needs to an actual interpretation of the differences if demographic factors are used.*

**Response:** We added a preliminary construct validity analysis following the requests of the Reviewers, and we used information available in the sample, that was defined a priori, according to main demographic and clinical differences (Tab S5). It is true that we did not

**clarify enough the limitations of this approach in the paper, and we did it now (page 8), emphasizing that findings are only suggestive and need to be independently validated in a larger independent sample, whose recruitment is ongoing (as for the protocol). We understand that the Reviewer is not requiring an unplanned analysis in a new independent sample, that at this time would possibly be underpowered. We will perform it as soon as we reach the planned sample size.**

3) *While the authors have added a small amount of text about the multidimensional nature of financial hardship, the discussion still does not speak to how the new measure either does or does not address these issues.*

**Response: Hopefully, we modified the Discussion section, according to the Reviewer's requests (page 18).**