Author's response to reviews

Title: Feasibility, reliability and validity of a questionnaire on health care consumption and productivity loss in patients with a psychiatric disorder (TiC-P)

Authors:

Clazien Bouwmans (bouwmans@bmg.eur.nl)
Kim Jong, de (k.dejong.3@erasmusmc.nl)
Reinier Timman (r.timman@erasmusmc.nl)
Moniek Vlasveld (Mvlasveld@trimbos.nl)
Christina Feltz-Cornelis, van der (c.m.vdrfeltz@tilburguniversity.edu)
Siok Swan Tan (tan@bmg.eur.nl)
Leona Hakkaart- van R oijen (hakkaart@bmg.eur.nl)

Version: 2 Date: 3 May 2013

Author's response to reviews: see over
Reactions on reviewers comments on manuscript

Dear Mr. Ford,

First of all we would like to take the opportunity to thank you for the valuable and constructive comments on our manuscript ‘Feasibility, reliability and validity of a questionnaire on health care consumption and productivity loss in patients with a psychiatric disorder (TiC-P)’.

Please find below a point-by-point summary of our reaction to each reviewer’s comment. The original comments are marked in italics followed by our response.

Kind regards,

Clazien Bouwmans

Referee 1  Reviewer: Matthias Hunger

Major compulsory revisions:

1.) *The authors have assessed construct validity for two specific items only: the number of contacts with psychotherapists and long term absence from work. However, at several parts of the manuscript, results for the two items are generalized to the entire TiC-P questionnaire, e.g.:*

   Abstract: “The construct validity was assessed by comparing patient reported data with data derived from registries”.
   Abstract: “The comparison of patient-reported data and data from registrations indicate satisfactory construct validity of the TiC-P”.
   Abstract: “(...) results indicate that the TiC-P is a (...) valid instrument for collecting data on medical consumption and productivity losses”.

   Discussion, 1st paragraph: “Estimates on the number of contacts with healthcare providers in mental healthcare were in accordance to registries of health care providers”

   Discussion, 1st paragraph: “So the TiC-P, including SF-HLQ, seems a valid instrument for measuring healthcare utilization and productivity loss”. The authors should revise these statements making clear that their results are valid for two specific items only.

   We have adjusted/revised all statements concerning the construct validity of the TiC-P.

2.) *Similarly, the following statement in the discussion is speculative and not supported by the data. Therefore it should be removed:*

   “Assuming that the reported data of contacts with psychotherapists do not differ
from the reporting of other healthcare utilization, we may conclude that the TiC-P is a valid instrument to measure healthcare utilization.”

We have removed this sentence.

3.) Assessment of test-retest reliability requires that the construct being measured is relatively stable over time. As the recall periods in the authors’ test retest design do not or only partly overlap, the assessment of test retest reliability assumes that the frequency of healthcare consumption is constant over time. While this may be a reasonable assumption for weekly prescribed sessions with a psychotherapist, I think it is questionable for infrequent or irregular physician visits, like e.g. those to a social worker or hospital outpatient visits. As a consequence, reliability coefficients could indicate low agreement even if there was perfect reliability of the items. Given these concerns, the authors should think about removing the infrequent medical use categories in their reliability analyses in table 3.

We prefer to maintain the original version of table 3 to give a full presentation of the findings on all items related to health care consumption. Additionally, we have added comments on this subject in the discussion. However, if the reviewer deems it necessary we will remove some of the items.

Minor essential revisions:

4.) The intraclass correlation coefficient (ICC) is typically used to assess the consistency of multiple measurements on the same quantity in the case that the true score is not known. Therefore it is usually used as a measure of reliability. Is there a reason why the authors also used the ICC for the comparison between self-reports and register data (validity assessment), i.e. in a situation where the true healthcare utilization is actually known?

There was no specific reason why we did use the ICC to assess the construct validity. However, we agree with this comment. We have revised the manuscript on this part and we have provided information on absolute differences and absolute agreement and we have calculated spearman rank correlation coefficients for these items of the TiC-P.

5.) Discussion: “However, it can be assumed that the amount of short term absence from work (e.g. absence shorter than 2-4 weeks) is much easier to recall thus resulting in accurate information from self-reports.” I do not fully agree with this statement. In the Collaborative care study, all participants were on long term sick leave. To provide the true number of days on sick leave, these patients only have to recall the approximate day when they started their sick leave. In contrast, for patients to correctly report their short term absence from work, it is potentially necessary to recall more than one episode of sick leave and to recall the exact number of days for each of these episodes. The authors should remove the above statement from their manuscript unless they provide data supporting their hypothesis.

We don’t have information about differences in recalling short-term and long-term absence from work. Consequently, we have removed our statement on this point. However, based on the findings of the study of Severens et al we have concluded that applying a recall period of 2-4 weeks for short term absence from work will provide reliable patient reported data. This is added to the discussion.
6.) Results: There is no Table 5, so, the “Insert Table 5” statement should be deleted.

This statement is deleted.

7.) Results, last paragraph: “The difference between reported and registered date of absence from work was on average 1.5 days”.

Are these differences absolute differences? They should be! Why is the difference so small given that there was only an agreement of 70.9%, even after accepting a margin of 7 days?

We have checked the data. The absolute difference is 5.7 (SD 20.6). This is adjusted in the manuscript.

Discretionary revisions:
8.) Table 2: This table contains little information. I think the respective numbers could also be reported in the text.

The calculation of the cost of medication requires relatively comprehensive information on medication. We believe that the information in table 2 makes this more clear. Consequently, we have maintained this table in the manuscript.

9.) Introduction, last paragraph: “…is widely used in the Netherlands for economic evaluations in mental health”

Please give one or two references.

References are added

10.) Methods, third paragraph: “This resulted in a number of textual changes”:

Please give a reference on work on the development process of the TiC-P.

A reference is added

11.) Discussion, last paragraph: “In conclusion, the results from our study indicate that the TiC-P is a relatively good alternative for collecting resource use data”.

It is unclear what “relatively good” means and what the other alternatives are.

We have adjusted this conclusion.

Referee 2 Reviewer: Hildegard Seidl

Major compulsory Revisions
1. The main objective of this paper was to assess the validity of the TiC-P, but the applied validation only examined contacts with psychotherapists (one out of 14 questions) and long-term absence from work as part of the SF-HLQ. This needs to be clearly conveyed in the title, abstract (aims, methods and results), results, discussion and conclusions. The assumption that the validity of self-reported contacts with psychotherapists is comparable to the validity of items on other healthcare utilization is highly speculative and no evidence to demonstrate this could be provided. Therefore, this sentence has to be removed.
The text regarding the construct validity is adjusted in the abstract and throughout the whole manuscript. We kept the title of the manuscript as this serves as a more or less general indication of the content of the paper. Additionally, we have removed the statement concerning the generalisability of the validity of items on other healthcare utilization.

2. Construct validity was assessed by comparing the patient reported contacts with psychotherapists from the Monitoring Study with registered data. It is not clear why only data of 114 responders were analyzed, given that the Monitoring study included 631 participants and the reported number of contacts ranged from 0 to 8. Table 5 is missing.

We have deleted the statement concerning table 5.

A total of 129 therapists participated in the trial on a voluntary basis. Due to practical/logistical reasons, a total of 10 therapists were asked to provide registration data of contacts with patients participating in the Monitoring study. Each of these therapists was involved in the treatment of at least 10 patients. Since the registrations are not primarily aimed to extract contact data most data had to be derived manually. Seven therapists responded and provided registration data. The total number of measurements was 365. We believe that collecting more data from registries would not generate additional value for the analyses.

3. To assess feasibility, it was checked whether the items of the questionnaire were filled in completely. For medication intake, however, it is not clearly stated how large the random sample was and how the authors could detect whether a medication name was missing.

Data was derived from a random sample of 283 patients reporting the use of medication. This information is added to table 2.

The following is added to the method section: Completeness on medication name was checked manually and was defined missing if no name was reported or in cases that the medication was described in general terms (f.e. sleeping pill, antibiotic).

4. Discussion, paragraph 6, last sentence: “It is not expected that this may have major impact on the validity...”

The recall period is three times shorter than usual. This limitation should be discussed in more detail.

We have adjusted this statement: This (the longer recall period) may limit the generalizability of the results of this study for the current version of the TiC-P. Further research should indicate the impact of a longer recall period on the different items of the TiC-P.

Minor Essential Revisions

5. Methods, paragraph 5: “Additionally, depending on the relevance for the target population the questionnaire allows adding or leaving out specific items of resource utilisation.”

Why does the questionnaire allow adding or leaving out specific items?

Generally, patient questionnaires should not be longer than necessary. We assume that a short questionnaires is positive for the response rate. It is possible that specific items are not relevant for the target (study) population. F.e. contacts with a self-help group or CAD. If this is expected or known before the start of the
study the researcher may consider leaving out these items. However, since this is not relevant for the current study we did not explain this specifically.

6. Study design and analyses
How often was the TiC-P filled out in the Monitory Study? (How many follow-ups were there?)

The TiC-P was filled out at baseline and monthly during the first 3 months of the study, and every 3 months thereafter. The treatment duration of the patients differed. Follow-up ended at the moment that the therapy was completed.
This information is added to the study design.

7. Table 1: The diagnoses of only 76% of the participants are reported.

The diagnosis of the remaining patients was different including: disorders usually first diagnosed in infancy, childhood or adolescence, impulse control disorders, eating disorders, dissociative disorders, sexual disorders, substance-related disorders and psychotic disorders. This was added to table 1.

8. To assess reliability, a test-retest design was applied. How did the authors ensure that the questions on health care utilization at both measurements (test and retest) cover the same period?

The purpose of the re-test questionnaire was explained in a cover letter. A subsample was invited to participate within two weeks after submission of the original measurement. This was explained in the method section. Since the re-test was performed early at the beginning of the therapy it was assumed that medical consumption and absence from work will be relatively stable in this period. The implications for the results are discussed in the discussion section.

9. Discussion, paragraph 7: Please justify the statement: “It is plausible that the questionnaire is valid in other groups of patients.”

The statement is adjusted. We suppose that our findings can be generalized to other groups of patients. However, future research on this is desired.

Discretionary Revisions

10. Discussion, paragraph 7: “Agreement between the number of days of absence from work… and self-report of the patient was satisfactory” It should be mentioned that this agreement only is the result of accepting a margin of +/- 7 days.

We have adjusted this statement.