Author's response to reviews

Title: Satisfaction with care among patients with non-metastatic breast cancer: development and first steps of validation of the REPERES-60 questionnaire

Authors:

Gautier Defossez (gautier.defossez@univ-poitiers.fr)
Simone Mathoulin-Pelissier (mathoulin@bergonie.org)
Isabelle Ingrand (isabelle.ingrand@wanadoo.fr)
Isabelle Gasquet (isabelle.gasquet@sap.ap-hop-paris.fr)
Lynda Sifer (lynda.sifer@laposte.net)
Pierre Ingrand (pierre.ingrand@univ-poitiers.fr)
Roger Salamon (roger.salamon@isped.u-bordeaux2.fr)
Virginie Migeot (virginie.migeot@univ-poitiers.fr)

Version: 2 Date: 9 May 2007

Author's response to reviews: see over
Author’s response to reviews

Title: Satisfaction with care among patients with non-metastatic breast cancer: development and validation of the REPERES-60 questionnaire

Authors:
G Defossez, MD: gautier.defossez@univ-poitiers.fr
S Mathoulin-Pélissier, MD PhD: mathoulin@bergonie.org
I Ingrand, Research Engineer: isabelle.ingrand@wanadoo.fr
I Gasquet, MD: isabelle.gasquet@sap.ap-hop-paris.fr
L Sifer, sociologist: lynda.sifer@laposte.net
P Ingrand, MD PhD: pierre.ingrand@univ-poitiers.fr
R Salamon, MD PhD: roger.salamon@isped.u-bordeaux2.fr
V Migeot, MD PhD: virginie.migeot@univ-poitiers.fr

Version: 2    Date: Mai 2007

Poitiers, mai 7, 2007

Manuscript: 4752161561358770

Iratxe Puebla
Senior Assistant Editor
BMC-series journals

Dear Iratxe Puebla,

Thank you for considering the above paper and for the reviewers’ useful comments. I send you the manuscript “Satisfaction with care among patients with non-metastatic breast cancer: development and firsts steps of validation of the REPERES-60 questionnaire”, to your editorial consideration to publication in BMC Cancer.

The manuscript is now 4920-word long with 42 references (4 added), 1 figure and 4 tables. The abstract is 353-word long. We propose two additional files.

This manuscript has been modified according to the suggestions of the reviewers. Our point-by-point responses to their comments are attached.

The manuscript was translated and edited by our colleague Angela Swaine Verdier whose native language is English.

We hope that our responses have now made the manuscript acceptable for publication.

Thank you for your consideration.

Sincerely yours

Gautier Defossez, MD
University Institute Of Public Health
Rue de la Milétrie
86000 Poitiers
Email: gautier.defossez@univ-poitiers.fr
1. Reviewer: Anne Brédart

Reviewer's report

This paper describes the development and psychometric analysis of a questionnaire purported to assess breast cancer patient satisfaction with the care received during the overall process of breast cancer treatment and follow-up, in a French population. This work rests and expands on an initial French validation of Davies & Ware Consumer Satisfaction Survey developed in the United States (manual published in 1991) (CSS-VF). The CSS-VF comprises 79 items to which 42 new items are added as a result of two breast cancer patient focus groups. The resulting 121-item questionnaire has been administered to 975 patients (one month after the end of treatment) and, for test-retest reliability testing, to a random sub-sample of 172 (one year later and one week apart).

This work is relevant. My main concern regards the reporting of this study.

Major Compulsory revisions

Different research questions are addressed which confuses the understanding of the work performed. This paper is long while missing important information. I would suggest splitting this paper in two, focus first on the qualitative part and second on the quantitative part of the study. This would allow for detailing aspects, for example, of the methodology and results of the focus groups (prompts for discussions, patients' exact reports, …). For example (p8), it seems that patients had to give their priorities on issues raised; this does not appear clearly in the paper. Patients' input does not appear clearly as such. The methodological originality of the focus groups for developing more specific patient satisfaction items/sections should be better highlighted.

[1] Without these details many aspects of the results presented have to be taken for granted. In the second paper focusing on psychometric results, the qualitative study (identification/selection of new questionnaire issues) has then just to be summarised. The readers should clearly understand how the new issues have been produced.

We understand the suggestion by this reviewer to split the article into two distinct papers, in view of the scope afforded by the two phases. However, since the Editor (and other reviewer) has not made this request, we propose to establish a better balance between the two parts of the article. On the one hand we present a more detailed description of the two focus group phases, and we are including information on item drafting and elements in connection with the new dimensions, as suggested. On the other, this necessary addition is counterbalanced by the reduction of the results section for the validation phase, these being synthesised following the remarks made by the different reviewers (reduction in text in favour of an organisation diagram and synthetic table of results).

We hold a record of the material derived from the verbatim following the first focus group. It is possible to include it in the annexes to the article, so that the reader can visualise the links between the verbatim material, the existing dimensions of the CSS-VF and the new dimensions.
The development of this questionnaire is mainly based on a questionnaire developed in the US and for primary care. A significant number of studies using specific patient satisfaction questionnaires are now available in the cancer field; in their literature review, authors do not refer to this literature on which they could have based/compared their own work.

In the discussion we have added this comparison of the REPERES-60 questionnaire with the main satisfaction questionnaires in the field of cancer. The literature mainly reports on instruments developed in hospital environments, the most commonly used today being the EORTC IN-PATSAT32 questionnaire.

In a hospital environment, the staffs involved are solely on the one hand cancer specialists and on the other the hospital healthcare staff. In a complete care itinerary, all the doctors encountered over the itinerary are involved. This explains why:

1. the enumeration of staff involved in the two questionnaires differs: doctors and nurses as opposed to all the doctors over the full itinerary (primary and secondary care);
2. if the majority of the dimensions explored by the REPERES-60 questionnaire are also found in the IN-PATSAT32 (interpersonal skills, technical skills, information provision, availability, access, exchange of information, comfort, global satisfaction) psychological support, mentioned by the patients as necessary for the announcement of the diagnosis and on leaving the medical environment, is not explored in the IN-PATSAT32.

Some paragraphs are wrongly placed in the paper. E.g., page 5 (2nd paragraph): "In view of the repeated development..." refers to the rationale of this study rather than to a methodological aspect thus should be place in introduction; page 13, the rationale for not exploring the reproducibility of specific scales should be placed in the methodology rather than the result section.

The first paragraph cited has indeed been repositioned in the introduction to the article.

On account of the return to single items for items Q1 to Q63 (thus corresponding to the dimensions of primary and secondary care in the questionnaire), the Kappa coefficients were intentionally removed from table 2 of the previous version of the article, so as not to encumber the results. Reproducibility at dimension level was however necessarily assessed for all 13 dimensions identified, using the intra-class correlation coefficient (table 3).

Methodological details should be justified: why a one-week lapse for test-retest (good results may reflect recall bias; after the end of treatment patients’ health may be stable allowing for a longer time lapse between assessments)?

The time interval for assessing reproducibility of the questionnaire was fixed at a minimum of one week. As the reviewer underlines, too short a time lapse involves a risk of recall bias. However time lapses over three months can lead, in particular in the area of cancer, to variations in interpretation of the experience of care by the patient. The time lapse chosen is thus a compromise between these two arguments. For instance, response rates for an interval set at 2 weeks appear better than for a re-assessment at three months; in this respect we have added a reference (Anne Brédart et al, Timing of patient satisfaction assessment: effect on questionnaire acceptability, completeness of data, reliability and variability of scores. Patient Education and Counselling, 2002, p131-6).
[5] Criteria for item selection, statistical estimate threshold choices need also to been given good reasons (e.g.: 0.80 for redundancy) or referenced.

A strong correlation observed between two items indicates, according to several authors, redundancy in the exploration of information. A threshold of 0.80 cannot be a consensus value for concluding that two items provide the same information. Nevertheless, it does make it possible to tend towards a reduction in the number of items, without necessarily altering the measure. Three references are in support of this interpretation, and the two English-language references have been added:

2. Boyle GJ. Does item homogeneity indicate internal consistency or item redundancy in psychometric scales. Person Individ Diff. 1994, 12: 291-4;

[6] Result section: place of tables/annexes (or results not shown) should be indicated to allow explicit reading of the results. Results could be formulated more clearly and meaningfully (providing item contents, grouping positive or negative results, extracting essential information). A table listing items, whether they are eliminating/maintained and why, could be helpful.

The results tables in association with the text of the article were integrated as an additional file at the time of the electronic submission, and did not appear in the body of the article. This error has been corrected so that the link with the text is easier.

However, as noted earlier, the results have been altered according to the recommendations of the various reviewers. Table 2, providing information on the properties for item selection, has been completed and reshaped to integrate detail of what became of each item.

[7] Discussion section: item scores for each medical specialist were highly correlated, how authors explain this and what is the implication for care improvement?

We have integrate in discussion this reviewer's remarks. The strong correlations between items can be explained by the way in which cancer is cared for in the French healthcare system. This care is classically provided in one or several health facilities, but with collegiate decisions among the specialists concerned (surgeons and oncologists). A patient when interviewed appears to recall global aspects of care (there is therefore a correlation among specialist scores) rather than aspects concerning each specialist. This tendency is today even more marked in the French "cancer plan" which seeks to develop this collaboration by proposing the establishment of cancer coordination centres.

[8] Tables/annexes: it is not clear why all item numbers for annex 2 do not correspond with the final questionnaire in appendix. Some items/sections do not appear logically (chronological/content) ordered, e.g.: the last section content is closer to sections 2, 3, 4; overall satisfaction could be assessed at the end; whatever the choice, it should be argued.

The link between the items in the questionnaire derived from the focus group and the final REPERES-60 questionnaire is now included in the revised results table 2.

Concerning the order in which the items appear in the questionnaire, the new dimensions were simply added following on the existing dimensions of the CSS-VF. We have no objection to placing the global satisfaction dimension at the end of the questionnaire; we merely reflected the chronology of the questions as asked.
Minor essential revisions

[9] Authors should pay attention to the inappropriate use of some terms: e.g.: complementary treatment refers to adjuvant treatment whereas complementary may be associated to CAM (Complementary Alternative Medicine).

The terminology has been altered to integrate this reviewer's remarks (adjuvant treatment is appropriate).
Reviewer's report

Title: Satisfaction with care among patients with non-metastatic breast cancer: development and validation of the REPERES-60 questionnaire

2. Reviewer: Giovanni Apolone

Reviewer's report

The paper by Gautier et al describes the process of development and preliminary test of a new satisfaction with care questionnaire to be used on (French) patients with non-metastatic breast cancer. The rationale of starting this long and complex process comes from the fact that, so far, no satisfaction with care questionnaires that are appropriate for ambulatory breast cancer patients (after the first primary treatment) are available (at least in French), thus a new tool is required to be applied in quality, outcome or evaluative studies.

Methods used were standard multi-step studies and evaluations that implied the review of literature, the identification of pre-existing instruments to be used as basis, the identification of a multi-dimensional a priori conceptual and measurement model, the conduction of (qualitative) focus groups to identify candidate items and test those extracted from other tools, a first quantitative (cross-sectional) evaluation on a large sample of patients. This phase of the “validation” process implied either to identify the best set of items using psychometric and other statistics methods (level of completion/compliance, item distribution, redundancy, factorial structure, reproducibility, item-scale and scale-scale correlations, scale consistency and reproducibility, etc) or the analysis of the association/correlation with other external indicators, and the assessment of the known-group validity of the new tools (the 60-item REPERES questionnaire).

The paper is well organized, the methods are agreeable, data and findings are fairly reported and discussed. Nevertheless, a few points deserve attention.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

[10] The process of development and validation is well described throughout the paper. Nevertheless, there is some confusion between the steps (phases), the methods (qualitative, quantitative), the intentions (questionnaire development, validation), the techniques (psychometric, statistics), the analysis (test-retest, factorial, multi-trait, etc). At the end of the “Background” or at the beginning of the “Method” a clear description of the entire “validation” process could be inserted, including also the steps/phases planned but not yet implemented. A summary figure/synopsis may help (See ref 24 for an example)

A synthetic description of the two development and validation phases of the questionnaire has been added at the start of the "methods" section. The terminology has been altered and a synopsis diagram has indeed made it possible to represent the various phases in development.

[11] “Validity” is not a yes/not attribute but the results of the accumulation of evidence produced through a long process where hypotheses are formulated and then sequentially challenged with several empirical testings. What so far done is promising, but some further tests are required before claiming that this instrument is “validated”. I
recommend to amend from the paper all these types of “over-interpretation” (see for example the first three lines in the discussion!)

The validation of an instrument is certainly a long ongoing process. We have amended the "over-interpretation".

[12] When testing external validity (page 15), p-values are used to judge the presence or absence of discrimination. This is not fully correct as a statistical significance may be assent (higher than 0.05) either for the (small) size of the effect under evaluation (a correlation or a difference between mean groups) or for the small sample size of the groups compared. I suggest to use other indicators/index of relative capability to pick-up differences between groups (as the estimation of confidence intervals or, when appropriate, the ratio between T or F tests).

We suggest integrating the results that were initially in the annexes into the body of the article. This provides the reader with the different means and SD for the groups compared. A difference of 5 points between two scores can be considered as significant discrimination. Further to this, the samples are relatively large (at least 218 per group). The methodology has been completed and the p-values removed from the text.

[13] In the Discussion some limitations of the current results should be mentioned and discussed to warn reader about the preliminary nature of findings. For example, I am sure that what done and what presented support the psychometric validity of the new questionnaire. I am also pretty convinced that items and scales behave in a sound way either internally or with some external indicators. In addition, some of correlations are easily interpretable, from the clinical point-of-view. I have also appreciated the Authors honesty when suggesting that “...the original CSS-VF can be used its own, ...” while “...the 4 new dimensions are complementary ...” (page 17, first lines). Nevertheless, some warnings about the need of further independent tests, the need to also test other dimensions of “validity” such as the responsiveness (in longitudinal studies, eventually with a before/after design), and the evaluation of its robustness in other settings are required. Finally, readers should be strongly advised to be cautious in generalizing present results to other linguistic, cultural and health care settings.

In a similar manner as for the response to remark 11, the preliminary nature of the work has been stated in the article, in particular in relation to the perspectives for validation and properties that have as yet not been explored. These remarks have been put at the end of the discussion.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

[14] The Title should be changed to warn readers about the preliminary nature of findings (the questionnaire is promising but not yet validated

We agree with this remark, and propose to leave the editor the choice of the title: « Satisfaction with care among patients with non-metastatic breast cancer: development and first steps of validation of the REPERES-60 questionnaire”.
Reviewer's report

Title: Satisfaction with care among patients with non-metastatic breast cancer: development and validation of the REPERES-60 questionnaire

3. Reviewer: Chih-Hung H Chang

Reviewer's report

General
This paper reports the development and validation of the REPERES-60 questionnaire to assess satisfaction with care among patients with non-metastatic breast cancer. However, the authors did not provide strong justification of the needs of developing such a new instrument. It is also not clear how the “different phases of treatment” issue has been incorporating into the item generation process. Although the manuscript seems to be written in a straightforward fashion, it would benefit greatly from some editorial assistance (e.g., spelling, grammar, tense, etc.) to make it more reader friendly. Some important data collection procedures seem to be missing in the “Method” section. Re-organizing some of the information presented might also be helpful.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Page 1:

[15] It would be helpful to describe a bit more on: 1) how the 45 new items were created; 2) how the patients were selected; 3) the demographic and clinical characteristics of the participants; 4) how the divergent validity and discriminant validity were tested.

The abstract has been altered to integrate this reviewer’s remarks.

Page 2:
[16] What are the 4 new dimensions and what does the REPERES-60 stand for?

Taking account of the remarks on this article, in particular on the balance between the qualitative development phase and the quantitative validation phase, the abstract has been altered.

Page 4:
[17] It would be helpful to summarize the development focus or results from some of the existing satisfaction instruments (references 10-17) and discuss why they do not address the issue of not incorporating “patient opinion in different phases of treatment for breast cancer”.

These remarks were also made by the first reviewer.
A more detailed description of the two focus groups and elements used in item drafting and the development of the new dimensions has been included.
We are ready, if required, to include the list of elements obtained from the verbatim from the first focus group in the annexes to this paper.
[18] Consider moving the “This study was conducted in ....” Paragraph to “Method” section.
This paragraph has been positioned at the start of the "methods" section.

Pages 5-6:
[19] How the new items were written and by whom?
A list of the elements derived from the verbatim was developed by the REPERES team (sociologists working on content analysis), using transcripts of the interviews with the first focus group. On reading this material, the patients in the second focus group "validated" the items in the CSS-VF, and suggested replicating them according to the type of health professional. As for the new items, they were discussed and drafted one by one by the patients (with methodological support from a sociologist).

Page 6:
[20] It is confusing, at least to this reviewer, that why there were 79 items derived from the original CSS-VF?
We propose to integrate table 2 of the results, which enables clarification of this point. The adaptation into French of the CSS led to distinguishing the dimensions of primary care (GPs) and secondary care (specialists), unlike the American questionnaire, this being on account of differences in the way the two health systems are run. These alterations related to the dimensions access to care, communication and competence of doctors, and possibilities for choice among doctors. The suggestion made by the focus group of integrating all the different health professionals led to a replication of each item for each type of professional: GP, gynaecologist, radiologist for primary care (diagnosis and follow-up), and surgeon and oncologist for secondary care (treatment).

Page 7:
[21] The data collection procedures in three different time period would merit some clarifications. It is somewhat ambiguous at the present form (The first paragraph on page 7). What were the response rates?
The assessment of reproducibility of the questionnaire required interviewing patients at a stable time in their illness, which is why the test-retest was performed one year after the end of initial treatment. This time lapse, moreover, enabled us to obtain data to assess sensitivity to change of the instrument.
The response rates are respectively 87% (one month after the end of treatment and 95% (one year afterwards for test-retest). This is detailed on page 11.

Page 9:
[22] It is not clear why the “Varimax” (orthogonal) rotation was chosen as the dimensions were thought to be correlated.
There was no preconceived idea about the correlation between dimensions in the questionnaire. "Varimax" orthogonal rotation is the method that is most widely used in the process of questionnaire validation. The results derived from the factor analysis after orthogonal rotation are very convincing.
[23] The use of the EORTC QLQ-C30 for examining the validity of the REPERES-60 seems problematic. Why the authors expected that the correlations between the two instruments should be low (r < .40).

As the second reviewer remarked, questionnaire validation is a lengthy ongoing process, for which evidence needs to be gathered before concluding on the validity of the instrument. The results that we set out in the article correspond to the first stages of questionnaire validation. To hypothesise that satisfaction and quality of life concepts are independent (and therefore not strongly correlated) was a way of supporting the external validity of the instrument, and of verifying the coherence of the definition given for the concept of satisfaction. If the correlations observed between the two concepts had proved to be high, the results would have been in favour of proximity, rather than independence, between the two tools.

[24] Why the authors expected those patients who were younger patients, with higher education, having problems of communicating, and with poorer health would have lower scores (less satisfaction). Any literature or their own work to support this? Also how the “experienced problems of communicating in the announcement of the diagnosis” was determined?

These hypotheses are based on reports in the literature concerning satisfaction in cancer. The references are cited page 10 [34, 35, 36] to support the relevance of the hypotheses.

Other data on the cohort of REPERES patients was collected alongside the satisfaction data, in response to other research needs:
- Quality of life of patients was assessed using the EORTC QLQC30 questionnaire so as to gain knowledge on patients' perceived health;
- A questionnaire on the patients' care itinerary was sent to each, responses providing information on communication problems experienced by the patients at the time of the announcement of the diagnosis.

[25] Please provide relevant citations for Wilcoxon test.


Page 11:
[26] Had there any other clinical information been collected in addition to the basic demographics?

The clinical data was collected by a clinical research assistant for each patient in the REPERES cohort. All patients had non-metastatic breast cancer (cohort inclusion criteria). Thus the majority had surgery and radiotherapy (plus chemotherapy according to prognostic factors). The clinical aspects of patient care in this cohort have been the subject of another article.

[27] Consider using abbreviated item contents in Table 2 (full content can be seen in the Appendix).

The items were in full in table 2 to avoid the reader losing information on their content. This approach makes it possible to view all the items removed after the selection phase. The revision of the results following the remark of one of the reviewers leads us to think that the complete table 2 is required.
[28] The process to examine the reproducibility issue seems contradictory to what the authors pointed out (*Reproducibility was studied at dimensional level using the intra-class correlation coefficient...*).

The reproducibility of the measure was studied item by item using a weighted kappa coefficient (item selection, “methods” section, page 9), and latter for the different dimensions using an intra-class correlation coefficient (properties of the final instrument, “methods” section, page 11).

[29] How the 12-factor solution was reached? Eigenvalues (not reported) or Scree plot (not shown)?

The number of factors retained was determined from eigenvalues of one or more.

We suggest adding the results in the annexes if the Editor agrees.

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Difference</th>
<th>Proportion</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23.445308</td>
<td>0.3907</td>
<td>0.3907</td>
</tr>
<tr>
<td>2</td>
<td>3.8796988</td>
<td>0.793244</td>
<td>0.0647</td>
</tr>
<tr>
<td>3</td>
<td>3.0863744</td>
<td>0.826021</td>
<td>0.0514</td>
</tr>
<tr>
<td>4</td>
<td>2.2637721</td>
<td>0.559358</td>
<td>0.0377</td>
</tr>
<tr>
<td>5</td>
<td>2.2073365</td>
<td>0.840902</td>
<td>0.0364</td>
</tr>
<tr>
<td>6</td>
<td>2.0228562</td>
<td>0.309782</td>
<td>0.0337</td>
</tr>
<tr>
<td>7</td>
<td>1.7136780</td>
<td>0.159316</td>
<td>0.0286</td>
</tr>
<tr>
<td>8</td>
<td>1.5543263</td>
<td>0.121764</td>
<td>0.0259</td>
</tr>
<tr>
<td>9</td>
<td>1.4225600</td>
<td>0.172739</td>
<td>0.0239</td>
</tr>
<tr>
<td>10</td>
<td>1.2598206</td>
<td>0.140348</td>
<td>0.0210</td>
</tr>
<tr>
<td>11</td>
<td>1.1167722</td>
<td>0.100164</td>
<td>0.0186</td>
</tr>
<tr>
<td>12</td>
<td>1.0166076</td>
<td>0.045637</td>
<td>0.0169</td>
</tr>
</tbody>
</table>

[30] What are the correlations between the derived factors?

An analysis of the inter-score correlation matrix was conducted to ascertain independence of dimensions for correlation coefficients below 0.70. Just one high inter-scale correlation was observed between listening abilities and provision of information by doctors and psychosocial support ($r=0.73$). The inter-scale correlation coefficients range from 0.16 (global satisfaction and insurance cover) to 0.73 (listening abilities and provision of information by doctors and psychosocial support).

[31] No data were reported in this paper to support the statement (… made it possible to reduce the time required to reach a finalized, validated version.).

Since we started from the CSS-VF questionnaire, we gained time that was put towards the development of the complementary dimensions to the CSS-VF. This work resulted, via original methods, in the early availability of a questionnaire that is relevant to the field of study. This approach finds support in two papers:


*Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)*