The Cambridge Breast IMRT trial: comparison of clinician versus patient reported outcomes

Mukesh B. Mukesh¹,², Wendi Qian³, Charleen Chan Wah Hak⁴, Jenny S. Wilkinson¹, Gillian C. Barnett¹,⁵, Anne Margaret Moody¹, Charles Wilson¹, Charlotte E. Coles¹

¹ Oncology Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
² Department of Oncology, Colchester Hospital University NHS Foundation Trust, Essex, UK
³ Cambridge Cancer Trials Centre, Cambridge Clinical Trials Unit – Cancer Theme, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
⁴ School of Clinical Medicine, University of Cambridge, Cambridge, UK
⁵ Cancer Research-UK Centre for Genetic Epidemiology & Dept of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK

Author for correspondence:
Dr Mukesh B Mukesh
Oncology Centre, Box 193
Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge, CB2 0QQ;
Tel: 01223 348460, Fax: 01223 217094
Email: drmukesh12@doctors.net.uk

Running head: comparison of clinician versus patient reported outcomes post breast radiotherapy
Key words: breast cancer, radiotherapy late treatment toxicity, concordance

19 pages, 3 tables, 2 figures, appendix (4 tables)
Abstract

Background: Breast radiotherapy associated toxicity is often reported using clinical and photographic assessments. The addition of patient-reported outcome measures (PROMs) is becoming more common. This study investigates the concordance between clinician and patient reported outcomes.

Methods: The Cambridge Breast IMRT trial prospectively collected data on clinician assessment and PROMs at 2 and 5 years after breast radiotherapy. Clinician assessment included physical examination and photographic assessment. PROMs included EORTC BR23 questionnaire and four breast radiotherapy specific questions. The correlation between patient and clinician scores were analysed on an independent patient basis using percentage agreement, the Cohen’s kappa co-efficient (k) and Bowker’s test of symmetry. The analysis was repeated after stratifying patients based on age, baseline HADS and baseline body image score.

Results: At 2 and 5 years, a weak level of concordance was seen between the clinician-based assessment and PROMS for all the five toxicity endpoints (k = 0.05 to 0.21), with individual patient based agreement of 32.9%-78.3% and a highly discordant Bowker’s test of symmetry (p<0.001). The most frequently reported moderate-severe toxicity by patients was change in breast appearance (14% at both 2 and 5 years), while it was breast induration (36% and 25% at 2 and 5 years respectively) by the clinicians. The lack of concordance was not affected by patient’s age, baseline HADS score and baseline BIS score.

Conclusions: This study has found that moderate-severe toxicity reported by patients is low and the overall concordance between clinicians and patients is low. This could be due to methodological limitations or alternatively reflects the subjective nature of
PROMs. Incorporation of patient’s perception on treatment related toxicity will have important implications for treatment decisions and follow-up care.
Introduction

Early cancer detection and the use of adjuvant therapies including chemotherapy and radiotherapy have significantly improved breast cancer survival rates. As the number of breast cancer survivors’ increase, efforts to reduce the long term treatment related morbidity is paramount. While addressing this issue, it is imperative that the methodology of toxicity assessment is standardised.

Traditionally, breast radiotherapy associated toxicity has been reported using clinician-based assessment tools including physical examination and/or photographic assessment [1-3]. More recently, patient-reported outcome measures (PROMs) have been introduced as they provide patient perception of their own health condition and treatment toxicity. This raises the concept of PROMs replacing clinician-based assessments in clinical trials and therefore minimising the number of trial follow-up visits for patients.

The large Cambridge Breast IMRT trial demonstrated superior overall cosmesis and reduced skin telangiectasia with IMRT using clinician-based assessment, though the benefits of IMRT could not be demonstrated using PROMs [4, 5]. This report analysed the concordance between clinician- and patient-based assessment at 2 and 5 years post breast radiotherapy and investigate factors which may explain difference in outcome between clinicians and patients.
Materials and Methods

Patient population

The Cambridge Breast IMRT trial opened in April 2003 and was closed to recruitment in June 2007 (n=1145). The Cambridge Research Ethics Committee provided ethical approval to the study. Eligibility criteria included women with operable unilateral histologically confirmed invasive breast cancer (T1-3, N0-1, M0) or ductal carcinoma \textit{in situ} requiring post-operative radiotherapy (RT) after tumour excision by breast-conserving surgery. The full details of the trial including patient characteristics and treatment details have previously been published [6].

Clinician assessment

The late breast tissue toxicity post RT was assessed by clinicians using serial photographs and clinical examination.

Photographic assessment

Frontal photographs of both breasts were taken after primary surgery and before RT (baseline) and repeated at 2 and 5 years post RT. Two photographs were taken, one with the hands resting on the hips, the other with the arms raised above the head. These photographs were scored by a multi-disciplinary panel of clinicians (3 at any one time). The 2 and 5 year photographs were compared with post-operative baseline photograph for RT associated breast shrinkage and scored on a validated 3-point scale (none/minimal = 1, mild = 2, marked = 3). The panel also assessed overall cosmesis on photographs taken at 2 and 5 years and scored them using a 3 point score (good, moderate and poor cosmesis). This method of scoring has been validated and shown
to be quicker than using three independent scorers with re-scoring of discrepancies and final resolution through discussion [1]. The inter-observer variability of this assessment has previously been assessed [7].

**Clinical assessment**

One clinician assessed the treated breast 2 and 5 years post RT for breast oedema, skin telangiectasia and palpable induration. Each of these endpoints were graded 0 to 3 (none, a little, quite a bit, very much) on the scale used in the START trials [2, 3]. An oncologist (CEC) assessed the first 70 patients at 2 years and trained the breast research radiographer (JW), who then assessed the remaining patients.

**Patient reported outcome measure (PROM)**

All patients enrolled in the trial were offered participation in the PROMs study. Patients completed validated quality of life questionnaires (EORTC QLQ-C30, BR23 and Body Image Scale [8]) prior to starting radiotherapy and then at 6 months, 2 years and 5 years after RT completion. Four additional questions were added to the questionnaire to record patient’s assessment of breast tissue toxicity and graded 1-4 (none, a little, quite a bit, very much): change in skin appearance of affected breast, overall change in breast appearance, breast shrinkage and breast hardness/firmness.

**Comparison of clinician versus patient reported outcomes**
Each clinician-assessed toxicity endpoint was paired with a PROM for comparison (Table 1 for toxicity items and Table 2 for score system). Both assessments were performed at the same time point. The agreement between patient and clinician scores were analysed on an individual patient basis using measures of agreement including percentage agreement, the Cohen’s kappa co-efficient ($k$) and Bowker’s test of symmetry. A zero-weight row or column was used in calculating the kappa statistics (by using the ZEROS option in SAS) when no patients were in a particularly grouped category. A $k$ value of $<0.4$ was considered weak, $0.4-0.75$ as fair to good and $>0.75$ as excellent concordance [9].

The analysis was also carried out by stratifying patients based on age (<50 or ≥50 years), baseline Hospital Anxiety and Depression score (HADS) {normal, baseline & cases} and baseline body image score (BIS) {≤3 or >3} to investigate whether these factors have an effect on the degree of concordance.
Results

Clinician-based toxicity assessments were available for 80.4% (921/1145) and 52.2% (598/1145) patients at 2 and 5 years respectively. The PROMs were available for 78.7% (901/1145) and 56.3% (645/1145) patients at 2 and 5 years respectively.

At 2 years, a weak level of concordance was seen between the clinician-based assessment and PROMS for all the five toxicity endpoints ($k = 0.05$ to $0.21$), with individual patient based agreement of $32.9\%-58.6\%$ and a highly discordant Bowker’s test of symmetry ($p<0.001$) (table 3). Clinicians consistently reported higher treatment toxicity as compared to patients, except for breast shrinkage (figure 1). Clinicians scored $23\%$ patients with moderate/severe breast oedema, whereas only $2\%$ patients reported moderate/severe breast swelling at 2 years. $36\%$ patients were rated to have developed moderate/severe breast fibrosis, whereas on PROMs, $11\%$ patients reported moderate/severe breast firmness. A higher proportion of patients reported moderate/severe smaller breast as compared to clinicians at 2 years ($10\%$ versus $6\%$).

The concordance between the clinicians and patients rating of toxicity remained weak at 5 years ($k = 0.10$ to $0.18$). The individual patient-based agreement was $39.2\%-78.3\%$ and Bowker’s test of symmetry showed significant discordance ($p<0.01$) except for overall change in breast appearance ($p=0.28$) (table 3). Patients under-reported on breast firmness and breast oedema and over-reported on breast shrinkage, as compared to clinical and photographic assessments at 5 years (figure 2).

The lack of concordance between clinician-based assessment and PROMs were not affected by patient’s age, baseline HADS score and baseline BIS score (appendix table 4-7).
Discussion

This study compares the toxicity assessment by patients and clinicians at 2 and 5 years after breast radiotherapy. The level of concordance between patients and clinicians measured by kappa co-efficient was fairly poor at both 2 and 5 years, with patients’ often under-reporting toxicity as compared to clinicians ($k$ ranging from 0.05-0.20).

The most frequent patient-reported moderate-severe equivalent to ‘quite a bit/very much’ toxicity was change in breast appearance (14% at both 2 and 5 years) followed by smaller breast post-RT (10% and 13% at 2 and 5 years respectively). Conversely, the most frequently reported moderate-severe toxicity on clinician-based assessment was breast induration (36% and 25% at 2 and 5 years respectively). Patients reported the toxicity endpoints of ‘breast swelling’ and ‘firmness to the breast’ more favourably, whereas ‘skin changes’ and smaller breast’ were reported more critically as compared to the clinicians. One would expect the PROMs and clinician assessment to complement each other. However, our study shows poor concordance between PROMs and clinician assessments. Other studies have also previously reported on the lack of agreement between clinician-based assessment and PROMs after breast radiotherapy [10-13]. Most have suggested that patients rate their treatment toxicity outcome more positively as compared to clinicians. Conversely, the large multi-centre START trial suggest that patients score their treatment toxicity more severely as compared to clinicians [14].

The lack of concordance may reflect a real difference in opinion between clinicians and patients regarding toxicity. On the other hand, it is possible that this degree of
disagreement is related to the methodological limitations and/or patient factors. These factors will now be discussed in detail.

Methodological factors affecting concordance

In the current study, clinicians and patients may have measured slightly different paired endpoints. For example, clinician endpoint ‘telangiectasia’ was paired with the less specific PROMs endpoint ‘skin changes post-RT’, which includes other changes like pigmentation and erythema. At five years, 30% patients reported some skin changes as compared to 17% telangiectasia rate by the clinicians. Similarly for breast shrinkage, clinician assessment compared only the ipsilateral breast using paired baseline post-surgery and post-RT photographs. In contrast, patients may have compared the ipsilateral breast with the contralateral breast, which would not differentiate between breast shrinkage caused by surgery and RT respectively. At 5 years, 13% patients had reported quite a bit/very much smaller breast shrinkage as compared to 7% breast shrinkage on clinician assessment.

One clinician (JW) performed most of the clinician-based assessment and it is possible that the assessor was particularly critical during toxicity scoring as compared with patients, leading to significant disagreement. However, poor concordance was also seen for the photographic assessment (performed by a group of clinicians), suggesting that these differences are not fully accounted by the possible overcritical scoring from the clinician JW.

In addition to the above methodological limitation, there are global issues with the assessment of PROMs. Even though patients evaluated their treatment toxicity prospective, their ratings may be influenced by the pre-RT counselling, and some
patients may accept changes in the treated breast as expected. It is also possible that a patient’s overall healthcare experience influences their PROMs.

PROMs tools are a subjective measure of patients’ satisfaction post treatment as compared to the true rating of treatment toxicity. One would anticipate radiotherapy toxicity to increase with time. However, we found that many PROMs improved over time in the Cambridge trial [5]. It is likely that many patients learn to live with the radiotherapy changes with time and hence under-report on toxicity as compared to clinician assessments. In the current study, clinicians scored 81/531 patients with poor overall cosmesis at 5 years. Of these 81 patients, only 24 patients (30%) reported moderate/severe breast changes and 19 patients (23%) reported no change in breast appearance. Similarly, though clinicians scored 38 patients with moderate/severe breast shrinkage at 5 years, only 13 patients (34%) reported quite a bit/very much smaller breast on PROMs. Hoeller et al. [15] study of 287 patients found patients’ satisfaction for overall cosmesis was greater as compared to clinicians’ assessment, with 27% patients not judging severe changes of the breast as an unsatisfactory cosmetic result. Similar results have shown in another study, where 94% patients rated their breast appearance as excellent/good, even though 50% had rated moderate-very significant difference between the treated and untreated breast [16]. In the ACROSEIN trial, patients receiving concurrent chemo-RT were more likely to develop unsatisfactory cosmesis as compared to sequential chemo-RT (40% vs. 15%; p= 0.001) on clinicians assessment [17]. On the contrary, using PROMs, no significant difference was detected between the two arms (8% vs. 9%). However, a significant difference was reported for ‘difference between the two breasts’ (29% versus 14%) and ‘breast firmness’ (15% versus 6%) on PROMs, suggesting that for some patients these changes are acceptable and not likely to influence their cosmetic rating.
Patient related factors affecting concordance

As PROMs are considered as subjective in nature (i.e. how an objective toxicity is perceived and felt by the patient), it may be influenced by patient’s personal perspective on body image. Patients with positive body image may minimise the importance of treatment related changes, even if these changes are notable. Conversely, patients with poor body image may have difficulty in accepting bodily changes, even if they are small. Sneeuw et al. [18] studied the relationship between patients’ evaluation of overall cosmesis and psychosocial functions for 76 patients and found a strong association between cosmetic rating and body image (correlation coefficient =0.43). Other physical and psychosocial factors can also influence PROMs. Young age, ethnicity, and medical co-morbidities have previously been shown to influence patient’s rating of treatment toxicity [10, 19, 20]. Psychosocial state may also affect patient’s perception of cosmetic outcome/toxicity. In the ARCOSEIN trial, patients with probable cases of depression on HADS score perceived their breast to be larger, more deformed and with worse skin pigmentation as compared to the non-depressed patients after treatment [21]. Deshields et al. [22] evaluated if patients’ psycho-social status at the time of breast cancer diagnosis influenced patients’ rating of cosmesis. At a short follow up of 52 days, patients’ baseline QOL pre-RT was found to influence overall cosmesis rating, suggesting that psycho-social issues can affect patients rating of cosmetic endpoints.

In this study, the disagreement between patients and clinicians toxicity scoring could not be explained by patients age (<50 or ≥50 years), baseline HADS and baseline BIS (≤3 or >3), although the number of patients with moderate-severe toxicity was small in each group.
Most radiotherapy trials need long term data on efficacy and/or toxicity. The efficacy data can often be collected from cancer registry but patients need 5-10 years of clinical follow up for toxicity assessment. This can be inconvenient for patients and increase the cost of running the trials. Many breast cancer patients are now discharged early from clinical follow-up and may not wish to attend hospital appointments for trial follow-up alone. If PROMs could replace clinician assessments, clinical trials could be run more cost-effectively and it will also reduce the burden of trial follow-up on patients. However, the lack of concordance between PROMS and clinician assessments indicate that at present PROMs cannot replace well-established clinical methods. In addition, PROMs alone cannot be used to compare results from historical trials. One can argue that PROMs are more relevant but they still require validation before replacing clinician assessments. The clinical validity of PROMs tool may improve by involving patient advocates in the development of PROMs questionnaire. For example the IMPORT trials have included “Have you had a problem getting a bra to fit” in the PROMs questionnaire to get an objective assessment of post treatment changes [23].

Limitations of study

Despite this being one of the largest breast radiotherapy study investigating the issues of concordance between patients and clinicians, there are some limitations to the current study. The study was not purpose-designed to address the specific question of concordance between clinician-based assessment and PROMS. This is a single centre study with all patients from one region of the country. The report is limited to 5 year follow up, which may be considered as relatively short for some late RT effects. In addition, a significant number of patients were withdrawn from the 5 year analysis due to cancer related factors (death, recurrence, and new cancer), patient’s choice and
other social issues (unwell to travel, transport issues, moved out of area, etc) [4, 5].
The moderate-severe event rates for PROMs are low, which makes it difficult to tease out psycho-social factors affecting PROMs.

This study focuses on concordance between PROMs and clinician-based assessments, but concordance will never be perfect due to innate differences already discussed. Despite this, PROMs may still be very useful in detecting differences between treatment arms, as illustrated in the START trial. Haviland et al. [14] showed that PROMs are sensitive enough to distinguish changes in normal tissue toxicity between different whole breast radiotherapy fractionation arms. In comparison, PROMs did not show a significant reduction in normal tissue toxicity due to IMRT in the Cambridge breast IMRT trial. However, the absolute dose changes between standard radiotherapy and IMRT were minimal and to smaller volumes of breast tissue in the majority of cases [6]. Therefore, it is possible that PROMs were not sensitive enough to detect these subtle dosimetric changes.

It may be that clinician’s assessment and PROMs complement each other with both contributing different and important information. Patients are clearly the best judge of the impact of treatment on their physical and psychological wellbeing, therefore PROMs may be the most appropriate assessment method.
Conclusions

Using the currently validated toxicity tools, there is significant discordance between patients and clinicians in assessing radiotherapy toxicity. This could be due to current methodological limitations but more likely reflects the subjective nature of PROMs. PROMs are a complex concept, which incorporates physical, functional and psychosocial factors. Incorporation of patient’s perception on treatment related toxicity will have important implications for treatment decisions and follow-up care.

Conflict of Interest: None

Acknowledgements

All the women who participated in the study and the doctors, nurses, radiographers and physicists who were involved in their care.

Funding sources

Dr. Mukesh B Mukesh was funded by the Efficacy and Mechanism Evaluation programme, Medical Research Council, UK (Grant number: 09/150/16)
Dr. Gillian Barnett was funded by a fellowship from Cancer Research UK and The Royal College of Radiologists. She has also received funding from Addenbrooke’s Charitable Trust.
Miss Jennifer Wilkinson, breast research radiographer was funded by the Breast Cancer Campaign and is currently funded by the Comprehensive Local Research Network.
Dr. Charlotte Coles is supported by the Cambridge National Institute of Health Research Biomedical Research Centre.

**Disclaimer**

The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.