Elsevier Editorial System(tm) for Clinical Oncology

Manuscript Draft

Manuscript Number: 15-0924R1

Title: Do patient-reported outcome measures agree with clinical and photographic assessments of normal tissue effects after breast radiotherapy? The experience of the Standardisation of Breast Radiotherapy (START) Trials in early breast cancer

Article Type: Original Article

Keywords: breast radiotherapy, normal tissue effects, patient-reported outcomes

Corresponding Author: Prof. John Yarnold, FRCR

Corresponding Author's Institution: The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research

First Author: J S Haviland

Order of Authors: J S Haviland; P Hopwood; J Mills; M Sydenham; J M Bliss; John Yarnold, FRCR

Manuscript Region of Origin: UNITED KINGDOM

Abstract: Aims
In radiotherapy trials normal tissue effects (NTE) are important endpoints, and it is pertinent to ask whether patient reported outcome measures (PROMs) could replace clinical and/or photographic assessments. Data from the START breast radiotherapy trials are examined.

Materials and Methods
NTEs in the treated breast were recorded by i) annual clinical assessments, ii) photographs at 2 and 5 years, iii) PROMs at 6 months, 1, 2 and 5 years following radiotherapy. Hazard ratios for the radiotherapy schedules were compared. Measures of agreement of assessments at 2 and 5 years tested concordance.

Results
PROMs were available at 2 and/or 5 years for 1939 women, of whom 1870 had clinical and 1444 had photographic assessments. All methods were sensitive to the dose difference between schedules. Patients reported higher prevalence for all NTE endpoints than clinicians or photographs ($p<0.001$ for most NTEs). Concordance was generally poor; weighted kappa at 2 years ranged from 0.05 (telangiectasia) to 0.21 (shrinkage and oedema). Percentage agreement was lowest between PROMs and photographic assessments of change in breast appearance (38%).

Conclusions
All 3 methods produced similar conclusions for the comparison of trial schedules, despite low concordance between the methods on an individual patient basis. Careful consideration should be given to the different contributions of the measures of NTE in future radiotherapy trials.

Manuscript Number 15-0924

Article Title: Do patient-reported outcome measures agree with clinical and photographic assessments of normal tissue effects after breast radiotherapy? The experience of the Standardisation of Breast Radiotherapy (START) Trials in early breast cancer

In response to the editor's comment, we have added the following reference to the introduction (page 3):

"Patient-reported outcome measures in radiotherapy: clinical advances and research opportunities in measurement for survivorship" by S Faithfull, A Lemanska, T Chen. published in CO vol. 27 issue 11 Nov. 2015 pgs. 679-685.

Response to reviewers:

Reviewer #1: The paper from Haviland and colleagues reports on the relationships between the methods (patient recorded outcomes, clinical assessments and photographic assessments) used to determine normal tissue effects (NTEs) in the START A & B trials. These ground breaking studies have been critical in challenging previously accepted dogma (largely unsupported by substantive randomized evidence) concerning normal tissue sensitivity to fraction size in breast radiotherapy and have lead to the rationalization of radiotherapy treatment approaches at an international level.

The START triallists made serious efforts to collect information on NTEs - which determine the cosmetic consequences of breast radiotherapy - and provide here a useful comparison between these methods.

Their findings that the three methods produce similar 'trial level' estimates of NTEs, but that there is relatively poor 'individual patient-level' correlation are of critical importance for the conduct of future studies as well as the interpretation of the START trial results. The move towards patient reported outcomes in the anticipation that such outcomes are predominant (and may be less resource-demanding to collect) needs to be tempered with an understanding of how these endpoints relate to the more traditional endpoints conventionally reported in prospective studies.

We thank the reviewer for their supportive comments.

Reviewer #2: This is a timely paper with follow-up becoming ever my pressured in UK hospitals. The ability to consider PROMs and clinical photographs as the accepted measures of NTEs would be an important advance. I have minor comments only:

1. Abstract results page2, bracket should be after "(p<0.001" in the following sentence: Patients reported higher prevalence than clinicians or photographs (p<0.001 for most NTEs).

The text as written in our original manuscript is correct. Patient-reported prevalences were higher for all NTEs looked at in the analysis – "for most NTEs" refers to the p-value. The text has been revised to clarify this.

2. Acronym NTE incorrectly used in first sentence of introduction, "Traditional outcome measures of normal tissue responses (NTE) ....". Correctly explained later same paragraph.

This has been corrected on page 3 of the manuscript.

3. In the Introduction patient-reported outcome measures (PRO) in ….. should this be PROM? This is a question for the authors to consider.

We have changed PRO to PROM throughout the manuscript.

4. Ist line of materials and methods: The START-A and START-B trials recruited 4451 women between 1998 and 1993 from 35 …… I think you mean 2003 not 1993

It should read "from 1998 to 2002"; we have corrected this on page 4 of the manuscript.

5. Page 4: Trial-B patients were randomised to either 50 Gy in 25 fractions over 5 weeks (control) or 40 Gy in 15 fractions of 2.7 Gy over 3 weeks. Full details of the recruitment, and radiotherapy planning, delivery and verification protocols have been previously reported, as has the PRO study [11-13]. I think references should be 11-14

We have corrected this on page 4.

Reviewer #3: This study compares different methods of assessing changes in breast appearance using data from patients recruited from the START trials.

The data are presented satisfactorily, and agreement (concordance) between methods (using simple percentage agreement, weighted Kappa and Bowker's test) is analysed appropriately.

However, the choice of the statistical methodology used for the comparison of treatment effects is unclear, and the Abstract requires some amendments (see below).

1. Cox proportional hazards regression was used to analyse treatment effects (moderate/marked vs none/a little) between schedules (Figures 1 and 2). Why was this type of analysis used? Cox regression is conventionally used for censored time data (where study subjects have differing follow-up times). In this study, follow-up time (between assessments) is surely constant? If so, why wasn't logistic regression, which is used for binary outcome data measured at a specific follow-up point, used?

The results presented in Figures 1 and 2 use all available NTE data from the trials (as in the original trial publications) and not just the 2 and 5-year assessments used in the analyses of concordance of individual scores – i.e. at 6 months, 1, 2 and 5 years for PROMS, annually for the clinical assessments and at 2 and 5 years for photographs. These data were then analysed as time to first NTE event, hence why Cox regression was used. Text has been added to the statistical methods section on page 6 to clarify this.

2. Abstract. "Patients reported higher prevalence than clinicians or photographs…".

This statement is ambiguous as it stands. What outcome is being referred to here?

This has now been clarified in the text, as discussed in response to reviewer #2's first comment.

3. Abstract. "Concordance between the methods on an individual patient basis was low, but this does not prevent PRP and photographs being considered as the primary measure of NTE in future radiotherapy trials".

I am not convinced that this is a reasonable summary of the results. The Discussion appears to be saying that PRO and photographs complement the clinical findings, not that they should take priority?

The conclusions in the abstract have been revised.

Do patient-reported outcome measures agree with clinical and photographic assessments of normal tissue effects after breast radiotherapy? The experience of the Standardisation of Breast Radiotherapy (START) Trials in early breast cancer

JS Haviland[1,2], P Hopwood[2], J Mills[2], M Sydenham[2], JM Bliss*[2], JR Yarnold*[3] on behalf of the START Trialists' Group

[1] Faculty of Health Sciences, University of Southampton, Southampton, UK

[2] ICR-CTSU, Division of Clinical Studies, The Institute of Cancer Research, London, UK

[3] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK

* Joint last authors

Corresponding author:

Professor John R Yarnold

Division of Radiotherapy and Imaging

The Royal Marsden, Downs Road, Sutton SM2 5PT, UK

Tel: +44 208 661 3388, Fax: +44 208 661 3107, Email: john.yarnold@icr.ac.uk

Abstract

Aims

In radiotherapy trials normal tissue effects (NTE) are important endpoints, and it is pertinent to ask whether patient reported outcome measures (~~PRO~~PROMs) could replace clinical and/or photographic assessments. Data from the START breast radiotherapy trials are examined.

Materials and Methods

NTEs in the treated breast were recorded by i) annual clinical assessments, ii) photographs at 2 and 5 years, iii) ~~PRO~~PROMs at 6 months, 1, 2 and 5 years following radiotherapy. Hazard ratios for the radiotherapy schedules were compared. Measures of agreement of assessments at 2 and 5 years tested concordance.

Results

~~PRO~~PROMs were available at 2 and/or 5 years for 1939 women, of whom 1870 had clinical and 1444 had photographic assessments. All methods were sensitive to the dose difference between schedules. Patients reported higher prevalence for all NTE endpoints than clinicians or photographs ($p<0.001$ for most NTEs). Concordance was generally poor; weighted kappa at 2 years ranged from 0.05 (telangiectasia) to 0.21 (shrinkage and oedema). Percentage agreement was lowest between ~~PRO~~PROMs and photographic assessments of change in breast appearance (38%).

Conclusions

All 3 methods produced similar conclusions for the comparison of trial schedules, despite low ~~.~~ ~~C~~concordance between the methods on an individual patient basis~~ was low,~~. Careful consideration should be given to the different contributions of the measures of NTE ~~but this does not prevent PRO and photographs being considered as the primary measures of NTE~~ in future radiotherapy trials.

Keywords: breast radiotherapy, normal tissue effects, patient-reported outcomes

Introduction

Traditional outcome measures of normal tissue responses ~~(NTE)~~ to radiotherapy rely heavily, often exclusively, on clinical assessments using graded scales to score a wide range of early and late adverse effects [1-4]. Scoring systems, including Late Effects in Normal Tissues Subjective, Objective, Management and Analytic (LENT-SOMA), Radiation Therapy Oncology Group (RTOG) and Common Terminology Criteria for Adverse Events (CTCAE), feature symptomatology requiring health professionals to elicit and score responses to direct questions. Photographic assessments of change in breast appearance from a pre-radiotherapy baseline have become increasingly used in randomised trials of radiotherapy as they are usually scored by a small number of observers blinded to patient identity, treatment allocation and year of follow-up, unlike the clinical assessments which are scored by a large number of individuals in a multi-centre study [5]. In parallel, the use of carefully developed and validated quality of life instruments in psychosocial research and phase III cancer clinical trials has expanded considerably [6-8], together with the growing interest in use of PROMS in routine follow-up [9][ ]. With an increasing use of patient-reported outcome measures (~~PRO~~PROMs) in cancer clinical trials [10, 11], it is worth asking how comparable and interpretable are the different methods of assessment, and whether ~~PRO~~PROMs could become the primary means of scoring late normal tissue effects (NTE) of breast radiotherapy in trials. Against this background, the large-scale UK START randomised trials [12-15] of hypofractionated radiotherapy after primary surgery for early breast cancer were used to conduct exploratory analyses comparing different methods of assessment of late NTE after adjuvant breast radiotherapy with the primary aim of assessing if

~~PRO~~PROMs might take priority over, or replace, clinical and/or photographic assessments as outcome measures.

Materials and Methods

The START-A and START-B trials recruited 4451 women between 1998 and ~~1993~~ 2002 from 35 UK radiotherapy centres (ISRCTN59368779, MREC(1)98/86). Centres could opt to participate in the ~~PRO~~PROMs and photographic assessment studies, and if they participated, they were expected to invite every eligible trial patient to join. Thirty one (89%) centres opted to participate in the ~~PRO~~PROMs study and 29 (83%) in a photographic assessment study of change in breast appearance. Women with operable invasive breast cancer (International Union Against Cancer pT1-3a pN0-1 M0) requiring radiotherapy after surgery (breast-conserving surgery or mastectomy, with clear tumour margins ≥1 mm) were eligible for the trials if they were aged over 18 years, did not have an immediate surgical reconstruction, and were available for follow-up. Trial-A patients were randomised to either 50 Gy in 25 fractions (control) or 41.6 Gy in 13 fractions of 3.2 Gy or 39.0 Gy in 13 fractions of 3.0 Gy over 5 weeks. Trial-B patients were randomised to either 50 Gy in 25 fractions over 5 weeks (control) or 40 Gy in 15 fractions of 2.7 Gy over 3 weeks. Full details of the recruitment, and radiotherapy planning, delivery and verification protocols have been previously reported, as has the ~~PRO~~PROMs study [12-14].

Patients in the ~~PRO~~PROMs study completed baseline measures in clinic and were sent questionnaires to complete at home at 6 months, 1, 2 and 5 years following radiotherapy. Clinical assessments of NTE were collected at annual follow-up in all patients, and photographs were taken under standard conditions at post-surgical pre-radiotherapy baseline and at 2 and 5 years post-randomisation for patients who

had breast conserving surgery. The patient questionnaires included the i) EORTC QLQ-C30 core questionnaire and QLQ-BR23 breast-specific module [6, 16], from which the assessment of breast swelling over the previous 4 weeks (not at all, a little, quite a bit, very much) was used in this study of concordance, ii) Hospital Anxiety and Depression Scale [17], iii) 10-item Body Image Scale [18] and iv) 4 protocol-specific questions asking patients to score “change in breast appearance”, “breast hardness/firmness”, “reduction in size of breast” and “change in skin appearance” since radiotherapy; the first three questions applying only to patients with conserved breasts, and all items scored on a 4-point scale (none, a little, quite a bit, very much).

The annual clinical assessments of breast shrinkage, breast induration, telangiectasia and breast oedema were scored using the contralateral breast as a comparator and 4-point graded scales (none, a little, quite a bit, very much). Change in photographic breast appearance since radiotherapy was scored by a single team of 3 observers blind to patient identity, trial treatment allocation, year of follow-up and radiotherapy centre. The scoring method was validated in the START pilot trial [5]. Photographs at 2 and 5 years following radiotherapy were compared with a pre-radiotherapy (post-surgery) baseline and an overall score allocated for change in photographic breast appearance in the treated breast based on change in size, shrinkage and shape, on a 3-point scale (no change, mild change, marked change). Post-mastectomy patients were included in the ~~PRO~~PROMs and clinical assessments but not in the photographic assessments. Individual NTE were mapped between the different assessment methods in order to compare corresponding outcomes, as shown in Table 1.

Statistical methods

NTE assessments at all time-points in the trials were included in the comparison of radiotherapy schedules (i.e. from 6 months-5 years for the PROMSs, from 1-5 years for the clinical assessments, and at 2 and 5 years for the photographs). Time to first NTE event (defined as "quite a bit" or "very much" for the PROMs and clinical assessments, and any change (mild or marked) in photographic breast appearance) was calculated from date of randomisation, and survival analysis methods used to compare radiotherapy schedules. The hHazard ratios (HR) for the relative effects of the radiotherapy schedules in START-A were calculated for each NTE endpoint using Cox proportional hazards regression and compared between the different assessment methods using forest plots. Estimates of the α/β ratio for NTEs, which describes the sensitivity of normal tissues to fraction size, were obtained separately for the PROPROMs, clinician and photographic endpoints in START-A. Estimates of relative effects of the fractionation schedules in START-B are not presented in this paper as they do not contribute to the measurement of fraction sensitivity, only having two randomised groups in Trial B. HRs for the fractionation schedules in START-B have been published separately for the different NTE assessments, and showed consistent results [13-15].

For the concordance analyses, data from START Trials A and B were combined, and only 2 and 5-year assessments included as these were the time-points at which all three NTE assessment methods were used in the trials. For all PROPROMs and clinically-assessed endpoints there were few patients in the highest grade category, so moderate and marked categories were combined, resulting in 3-point scales corresponding to none, a little ("mild"), quite a bit / very much ("moderate / marked");

this also enabled comparison with the photographic assessments, which were scored on a similar 3-point scale. Corresponding NTE endpoints were matched between the ~~PRO~~PROMs, clinical and photographic assessments at each time point and compared on an individual patient basis using measures of concordance including percentage agreement (with 95% confidence interval, CI), weighted Kappa statistic (with 95%CI) and Bowker's test of symmetry [19]. Guidelines for interpreting the value of the weighted Kappa statistic in terms of the strength of agreement are <0.20: poor, 0.21-0.40: fair, 0.41-0.6: moderate, 0.61-0.8: good, 0.81-1.00: very good [20]. Bowker's test assesses the symmetry of a square table – i.e. whether there are more observations on one side of the diagonal than the other. The concordance analyses were also carried out stratifying on baseline patient characteristics such as age and quality of life scores (including anxiety and depression from the Hospital Anxiety and Depression Scale and body image from the Body Image Scale), to investigate whether these had any effect on the degree of concordance between NTE assessment methods.

Results

Of the 2208 women recruited into the overall START Trials ~~PRO~~PROMs study, self-assessments of NTEs were available at 2 and/or 5 years for 1939 (88%) patients, of whom 1870 also had clinical assessments at the same time-points (85% of all patients in ~~PRO~~PROMs study). Patient characteristics at baseline for the 1870 patients in this analysis are shown in Table 2, of whom 1574/1870 (84.2%) had breast conserving surgery and 1444/1574 (91%) had photographic assessments at 2 and/or 5 years.

Treatment effects on late NTE assessed by ~~PRO~~PROMs and by annual clinical assessment in START-A are shown side-by-side in Figure 1. Two test schedules (41.6 Gy and 39 Gy in 13 fractions) were compared with control (50 Gy in 25 fractions) in START-A. Comparing HR for corresponding endpoints, it can be seen that the treatment effects were of a similar size for ~~PRO~~PROMs and clinical assessments, with overlapping confidence intervals. Treatment effects on late NTE assessed by ~~PRO~~PROMs and by photographs for overall change in breast appearance were also similar (Figure 2). α/β estimates (adjusted for prognostic factors) for overall change in breast appearance were 2.9 Gy (95%CI 0.7-5.1 Gy) for ~~PRO~~PROMs and 2.6 Gy (95%CI 1.3-3.9 Gy) for photographic assessments. α/β estimates for individual NTE endpoints from clinical assessments have been reported [14] (there was no clinical assessment of overall cosmesis in the START Trials).

The comparison of overall rates of NTEs reported by ~~PRO~~PROMs and clinical assessments from START Trials A and B combined showed that patients reported a higher prevalence of breast changes (Figures 3a-d). Concordance between the assessments of corresponding NTEs on an individual patient basis was generally poor (Table 3). The lowest levels of percentage agreement between ~~PRO~~PROMs and clinicians were observed for breast induration / hardness (47% and 50% at 2 and 5 years, respectively), and breast shrinkage (53% and 47% at 2 and 5 years). The highest level of percentage agreement between ~~PRO~~PROMs and clinicians was for breast swelling/oedema (78% and 86% at 2 and 5 years), but the overall prevalence of oedema was very low (Figure 3c). Weighted kappa statistics also highlighted the low agreement between methods, ranging from 0.05 for

telangiectasia at 2 years (indicating poor agreement) to 0.21 for each of breast shrinkage and breast oedema at 2 years (indicating fair agreement). Results of Bowker's test of symmetry were highly statistically significant for all NTE endpoints, indicating a clear direction in the discordance of scoring between the different methods, with patients reporting more breast changes compared with clinical and photographic assessments (Table 3). There appeared to be no substantial differences in degree of concordance for individual NTE endpoints according to time since radiotherapy i.e. between 2 and 5 years (Table 3).

The comparison of ~~PRO~~PROMs and photographic assessments showed that patients reported a higher prevalence of overall change in breast appearance since radiotherapy and graded effects as more severe compared with the photographic assessments (Figure 3e). In testing concordance, agreement on an individual patient basis was low at 2 and 5 years (38% for each), with low weighted kappa values (0.09) and highly statistically significant discordance ($p<0.001$ for Bowker's test of symmetry); Table 3. Concordance of ~~PRO~~PROMs with clinical and photographic assessments of NTE appeared to be unaffected by patient factors including age, breast size, surgical deficit, baseline HADS anxiety and depression and body image scores (table in web appendix).

Discussion

Concordance between ~~PRO~~PROMs and NTE assessments as scored by clinicians and from photographs on an individual patient basis was poor. Percentage agreement between ~~PRO~~PROMs and clinical assessments of specific NTEs was around 50%, indicating that in only half of the patients the NTE was graded in the

same category of severity corresponding to none, mild, moderate/marked. Agreement was even lower between ~~PRO~~PROMs and photographs, where less than 40% graded NTEs the same. In our study, patients scored NTEs more frequently and more severely than results from clinicians or photographs. Concordance did not appear to be affected by patient characteristics including psychological measures (anxiety and depression), body image and factors associated with risk of NTEs (age, breast size and surgical deficit). It may not be surprising that concordance between the assessment methods on an individual patient basis was poor; this has been consistently reported in other studies [21-24]. These differences in ratings reflect the different paradigms in which symptoms are perceived and rated; these include variance in context, values, expectations and methodological influences as well as the different sociocultural backgrounds of subjects and doctors [25]. Published comparisons of clinician and patient self-assessments show considerable variability between ratings, especially for more subjective symptoms and often report, as in our study, a relative underestimate by clinicians compared with patients (e.g. Basch et al [26], Bruner et al [27], Fromme et al [23], Groenwold et al [28], Quinten et al [29], Stephens et al [30], Velikova et al [24]). However, the concordance analysis of NTE assessments in the Cambridge intensity-modulated breast radiotherapy trial found the opposite, with clinicians and photographic assessments reporting more NTEs compared with patients, possibly because the study was done in a single centre, with clinical ratings done by one person [31]. Others have shown more favourable rating of overall cosmesis following conservative treatment for breast cancer by patients compared with clinicians [32, 33], although these findings are not necessarily specific to late effects of radiotherapy. Kirchheiner et al [34] argued that some variation is “quite acceptable and comprehensible”, given the methodological

Formatted: French (France)
Field Code Changed
Formatted: French (France)
Formatted: French (France)
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed

differences between morbidity scoring by clinicians and patient-reported symptoms. Clinical and patient symptom ratings are typically not designed to be interchangeable, given that they often have different values and purposes, with patient assessments inherently encompassing impact on quality of life.

However, our study showed that despite the discordance between assessments on an individual basis, the three methods (~~PRO~~PROMs, clinical and photographs) generated similar estimates of relative treatment effects on NTE within the trials [12, 14, 15]. The discriminatory power of different assessments was equally good, in that ~~PRO~~PROMs generated the same estimates of α/β value for NTE in START-A (around 3 Gy) as photographs and clinical assessments (data for α/β values of clinical assessments of NTEs previously published [14]). From the trial outcome perspective, this consistency of treatment effects adds considerable weight to the overall interpretation and conclusions of the trial. However, the ~~PRO~~PROMs reported here were selected from a large number of multidimensional items assessed as part of the START quality of life sub-study, most of which would not be expected to discriminate so clearly between the schedules in the START trials, but are of value in understanding the experience of treatment effects over time. The ~~PRO~~PROMs items included in this analysis of concordance were those directly relevant to the hypothesis under test in the clinical trial, and therefore most likely to be sensitive to randomised differences in radiotherapy dose intensity. The ~~PRO~~PROMs needed to have a recognisable relationship with the pathophysiology (atrophy, fibrosis) of NTE, broadly corresponding to clinical scoring of change in size (atrophy), shape and texture (oedema, fibrosis) of the breast and change in photographic breast appearance (atrophy, distortion/fibrosis). This is in contrast with other clinically

relevant domains, such as physical and social functioning, that explore the impact on different aspects of quality of life [6, 16].

Clinicians are taught in training that symptomatology is the key to diagnosis, which they can only judge by listening to their patients and framing relevant questions. Clinicians act as surrogates for their patients in this context, so that if the relevant questions are known in advance (as they are in a clinical trial), there appears to be a good reason to prioritise the ~~PRO~~PROMs over the physical clinical assessments. Where physical signs are concerned, including breast size, shape and texture, this study suggests that patients are as sensitive as their doctors in scoring these changes too, provided the questions are framed appropriately. In this respect, it is possible to criticise our ~~PRO~~PROMs question, which asked patients to score changes since radiotherapy to the affected breast compared with the clinical assessment that compared the treated with the untreated breast at the time of the annual examination. Despite a variety of factors expected to influence how a woman responds to this question, the sensitivity to randomised dose indicates that the radiotherapy 'signal' was not lost. Doctors also develop their own frames of reference when assessing NTE, and the hundreds of clinical observers involved in scoring NTE in thousands of patients over a 10-year period, as in the START trials, necessarily contribute a lot of 'noise' in a scoring system. However, a disadvantage of reliance on ~~PRO~~PROMs in clinical trials is that they are traditionally labour-intensive to administer and generate large volumes of data, making heavy demands on trial management and statistical resources. Since modern data capture systems are increasingly able to collect outcome data directly from the patient (e.g. via an App), dispensing with clinical follow-up may appeal to patients as well as health

services operating under increasing pressures [35]. However, radiation effects are not viewed in isolation by patients and attention also needs to be paid to their concerns in the context of multi-modal treatments and adverse effects over time. Up to a third of patients report moderate or marked symptoms of the breast, arm and shoulder at 5 years, which may warrant engagement and advice from their clinical teams [13]. Thus more preparation and after care is needed for the success of patient self-management post-treatment and to improve quality of life [36]. Further, the acceptability of electronic symptom-reporting warrants evaluation in an aging population.

Despite adding to the administrative burden of clinical trials, the photographic assessments of NTEs provide valuable information, not least because they are scored generally by the same small team of observers who are blind to patient identity, randomised treatment allocation, year of follow-up and participating hospital. As it is generally not possible to blind treatment allocation in radiotherapy trials the photographic assessments provide the only unbiased comparison of normal tissue effects between randomised groups. In addition, as photographs provide a permanent record of breast effects at a fixed point in time, the assessments can be validated by repeat scoring from different teams of observers [5], thus making the scoring more standardised than ~~PRO~~PROMs or clinical assessments from physical examination. Photographs can also be filed and stored for use in future translational research investigating adverse effects of radiotherapy. There are some disadvantages to the use of photographic assessments in clinical trials, including financial and staff resources required, and they can be disliked by patients, but these

are outweighed by the benefits of retaining an unbiased comparison of NTEs within radiotherapy trials.

There is growing interest in investigating inherited risk factors for radiotherapy NTE, for which robust measures of NTE are needed that have a close relationship to the underlying pathophysiology [37], In this respect, the lack of concordance reported in this study is intriguing and potentially worrying. The prevalence and severity of NTEs reported by patients, clinicians and from photographs during follow-up were widely discordant in most cases. In trying to identify subgroups of patients with levels of NTE that are much more, or much less, severe than expected on the basis of known factors (breast size, radiotherapy dose etc.), it isn't possible to judge whether the clinical and photographic assessments of NTE severity are more or less valid than the ~~PRO~~PROMs, hence making identification of potential cases (and controls) for translational studies very difficult. Perhaps much depends on how the NTE assessment questions to patients and clinicians are posed, something that this study does not address.

In conclusion, the ~~PRO~~PROMs, clinical and photographic assessments of late NTE in the START trials generated consistent estimates of relative treatment effects between randomised groups, adding weight to the trials' overall findings. Discordance in the prevalence rates of NTE reported by the patients, clinicians and photographs could be expected for a number of well-established reasons, but this does not undermine an argument for prioritising ~~PRO~~PROMs and photographic assessments of NTEs in breast radiotherapy trials.

Conflict of Interest Statement

The authors have no conflicts of interest to disclose.

Role of the Funding Source

The START trials were funded by Cancer Research UK, Medical Research Council and the UK Department of Health (grant G9600656). The Cancer Research UK number for the START trials is CRUK/96/001. The funders had no role in study design or conduct, analysis or interpretation of data. The manuscript was approved by all authors and the corresponding author had the final decision to submit the manuscript for publication.

Acknowledgements

We thank all the patients who participated in the START trials, and the doctors, nurses, radiographers, physicists, and data managers at the participating centres. Continued data collection and analysis is made possible by a core grant from Cancer Research UK to the ICR-CTSU. We acknowledge NHS funding to the NIHR Biomedical Research Centre at The Royal Marsden and the ICR. The START Trialists' Group consists of the trial management group, consumers, trial steering committee, independent data monitoring committee, and the principal and main co-investigators at the participating centres (published previously [14]). We also acknowledge Claire Birch, who worked on preliminary concordance analyses of an earlier START dataset for her MSc dissertation.

References

[1] http://www.rtog.org/ResearchAssociates/AdverseEventReporting/RTOGEORTCLateRadiationMorbidityScoringSchema.aspx.

[2] http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf.

[3] Pavy JJ, Denekamp J, Letschert J, Littbrand B, Mornex F, Bernier J, et al. EORTC Late Effects Working Group. Late effects toxicity scoring: the SOMA scale. Radiother Oncol. 1995;35:11-5.

[4] Rubin P, Constine LS, Fajardo LF, Phillips TL, Wasserman TH. RTOG Late Effects Working Group. Overview. Late Effects of Normal Tissues (LENT) scoring system. Int J Radiat Oncol Biol Phys. 1995;31:1041-2.

[5] Haviland JS, Ashton A, Broad B, Gothard L, Owen JR, Tait D, et al. Evaluation of a method for grading late photographic change in breast appearance after radiotherapy for early breast cancer. Clin Oncol (R Coll Radiol). 2008;20:497-501.

[6] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85:365-76.

[7] Brady MJ, Cella DF, Mo F, Bonomi AE, Tulsky DS, Lloyd SR, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. J Clin Oncol. 1997;15:974-86.

[8] Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35:1095-108.

Formatted: Line spacing: Double
Formatted: Font: Arial, 12 pt
Field Code Changed
Formatted: Font: Arial, 12 pt, No underline, Font color: Auto
Formatted: Font: Arial, 12 pt
Formatted: Font: Arial, 12 pt, No underline, Font color: Auto
Formatted: Font: Arial, 12 pt
Formatted: Font: Arial, 12 pt, No underline, Font color: Auto
Formatted: Font: Arial, 12 pt
Formatted: Font: Arial, 12 pt, No underline, Font color: Auto
Formatted: Font: Arial, 12 pt
Formatted: Font: Arial, 12 pt, No underline, Font color: Auto
Formatted: Font: Arial, 12 pt
Formatted: Font: Arial, 12 pt, No underline, Font color: Auto
Formatted: Font: Arial, 12 pt
Formatted: Font: Arial, 12 pt, No underline, Font color: Auto
Formatted: Font: Arial, 12 pt
Formatted: Font: Arial, 12 pt, No underline, Font color: Auto
Formatted: Font: Arial, 12 pt

[9] Faithfull S, Lemanska A, Chen T. Patient-reported Outcome Measures in Radiotherapy: Clinical Advances and Research Opportunities in Measurement for Survivorship. Clin Oncol (R Coll Radiol). 2015;27:679-85.

[10] Basch E, Abernethy AP, Mullins CD, Reeve BB, Smith ML, Coons SJ, et al. Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. J Clin Oncol. 2012;30:4249-55.

[11] Rothman ML, Beltran P, Cappelleri JC, Lipscomb J, Teschendorf B, Mayo FDAP-ROCMG. Patient-reported outcomes: conceptual issues. Value Health. 2007;10 Suppl 2:S66-75.

[12] Bentzen SM, Agrawal RK, Aird EG, Barrett JM, Barrett-Lee PJ, Bliss JM, et al. The UK Standardisation of Breast Radiotherapy (START) Trial A of radiotherapy hypofractionation for treatment of early breast cancer: a randomised trial. Lancet Oncol. 2008;9:331-41.

[13] Hopwood P, Haviland JS, Sumo G, Mills J, Bliss JM, Yarnold JR. Comparison of patient-reported breast, arm, and shoulder symptoms and body image after radiotherapy for early breast cancer: 5-year follow-up in the randomised Standardisation of Breast Radiotherapy (START) trials. Lancet Oncol. 2010;11:231-40.

[14] Haviland JS, Owen JR, Dewar JA, Agrawal RK, Barrett J, Barrett-Lee PJ, et al. The UK Standardisation of Breast Radiotherapy (START) trials of radiotherapy hypofractionation for treatment of early breast cancer: 10-year follow-up results of two randomised controlled trials. Lancet Oncol. 2013;14:1086-94.

[15] Bentzen SM, Agrawal RK, Aird EG, Barrett JM, Barrett-Lee PJ, Bliss JM, et al. The UK Standardisation of Breast Radiotherapy (START) Trial B of radiotherapy

hypofractionation for treatment of early breast cancer: a randomised trial. Lancet. 2008;371:1098-107.

[16] Sprangers MA, Groenvold M, Arraras JI, Franklin J, te Velde A, Muller M, et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. J Clin Oncol. 1996;14:2756-68.

[17] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67:361-70.

[18] Hopwood P, Fletcher I, Lee A, Al Ghazal S. A body image scale for use with cancer patients. Eur J Cancer. 2001;37:189-97.

[19] Bowker AH. A test for symmetry in contingency tables. Journal of the American Statistical Association. 1948;43:572-4.

[20] Altman DG. Practical statistics for medical research. London: Chapman and Hall. 1991.

[21] Davidson SE, Trotti A, Ataman OU, Seong J, Lau FN, da Motta NW, et al. Improving the capture of adverse event data in clinical trials: the role of the International Atomic Energy Agency. Int J Radiat Oncol Biol Phys. 2007;69:1218-21.

[22] Vistad I, Cvancarova M, Fossa SD, Kristensen GB. Postradiotherapy morbidity in long-term survivors after locally advanced cervical cancer: how well do physicians' assessments agree with those of their patients? Int J Radiat Oncol Biol Phys. 2008;71:1335-42.

[23] Fromme EK, Eilers KM, Mori M, Hsieh YC, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. J Clin Oncol. 2004;22:3485-90.

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto, Italian (Italy)

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto, Italian (Italy)

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

[24] Velikova G, Wright P, Smith AB, Stark D, Perren T, Brown J, et al. Self-reported quality of life of individual cancer patients: concordance of results with disease course and medical records. J Clin Oncol. 2001;19:2064-73.

[25] Gotay C. Patient symptoms and clinician toxicity ratings: both have a role in cancer care. J Natl Cancer Inst. 2009;101:1602-3.

[26] Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst. 2009;101:1624-32.

[27] Bruner DW, Movsas B, Basch E. Capturing the patient perspective: patient-reported outcomes as clinical trial endpoints. American Society of Clinical Oncology educational book / ASCO American Society of Clinical Oncology Meeting. 2012:139-44.

[28] Groenvold M, Fayers PM, Petersen MA, Sprangers MA, Aaronson NK, Mouridsen HT. Breast cancer patients on adjuvant chemotherapy report a wide range of problems not identified by health-care staff. Breast Cancer Res Treat. 2007;103:185-95.

[29] Quinten C, Maringwa J, Gotay CC, Martinelli F, Coens C, Reeve BB, et al. Patient self-reports of symptoms and clinician ratings as predictors of overall cancer survival. J Natl Cancer Inst. 2011;103:1851-8.

[30] Stephens RJ, Hopwood P, Girling DJ, Machin D. Randomized trials with quality of life endpoints: are doctors' ratings of patients' physical symptoms interchangeable with patients' self-ratings? Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 1997;6:225-36.

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

[31] http://abstracts.ncri.org.uk/abstract/a-comparison-of-clinician-versus-patient-reported-outcomes-proms-for-late-normal-tissue-side-effects-following-breast-radiotherapy-results-of-the-cambridge-breast-intensity-modulated-radiotherapy-3/.

[32] Kaija H, Rauni S, Jorma I, Matti H. Consistency of patient- and doctor-assessed cosmetic outcome after conservative treatment of breast cancer. Breast Cancer Res Treat. 1997;45:225-8.

[33] Hoeller U, Kuhlmey A, Bajrovic A, Grader K, Berger J, Tribius S, et al. Cosmesis from the patient's and the doctor's view. Int J Radiat Oncol Biol Phys. 2003;57:345-54.

[34] Kirchheiner K, Nout R, Lindegaard J, Petric P, Limbergen EV, Jurgenliemk-Schulz IM, et al. Do clinicians and patients agree regarding symptoms? A comparison after definitive radiochemotherapy in 223 uterine cervical cancer patients. Strahlenther Onkol. 2012;188:933-9.

[35] Bruner DW, Hanisch LJ, Reeve BB, Trotti AM, Schrag D, Sit L, et al. Stakeholder perspectives on implementing the National Cancer Institute's patient-reported outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Translational behavioral medicine. 2011;1:110-22.

[36] Corner J, Wagland R, Glaser A, Richards SM. Qualitative analysis of patients' feedback from a PROMs survey of cancer patients in England. BMJ open. 2013;3.

[37] Andreassen CN, Barnett GC, Langendijk JA, Alsner J, De Ruysscher D, Krause M, et al. Conducting radiogenomic research--do not forget careful consideration of the clinical data. Radiother Oncol. 2012;105:337-40.

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto, German (Germany)

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Formatted: Font: Arial, 12 pt, No underline, Font color: Auto

Formatted: Font: Arial, 12 pt

Figure legends

Figure 1: Comparisons between randomised radiotherapy schedules in START Trial A for ~~PRO~~PROMs and clinical assessments of specific normal tissue effects

Figure 2: Comparisons between randomised radiotherapy schedules in START Trial A for ~~PRO~~PROMs and photographic assessments of overall change in breast appearance

Figure 3: Comparison of 5-year ~~PRO~~PROMs, clinical and photographic assessments of specific normal tissue effects in START Trials A and B

Figure 1: Comparisons between randomised radiotherapy schedules in START Trial A for ~~PRO~~PROMs and clinical assessments of specific normal tissue effects

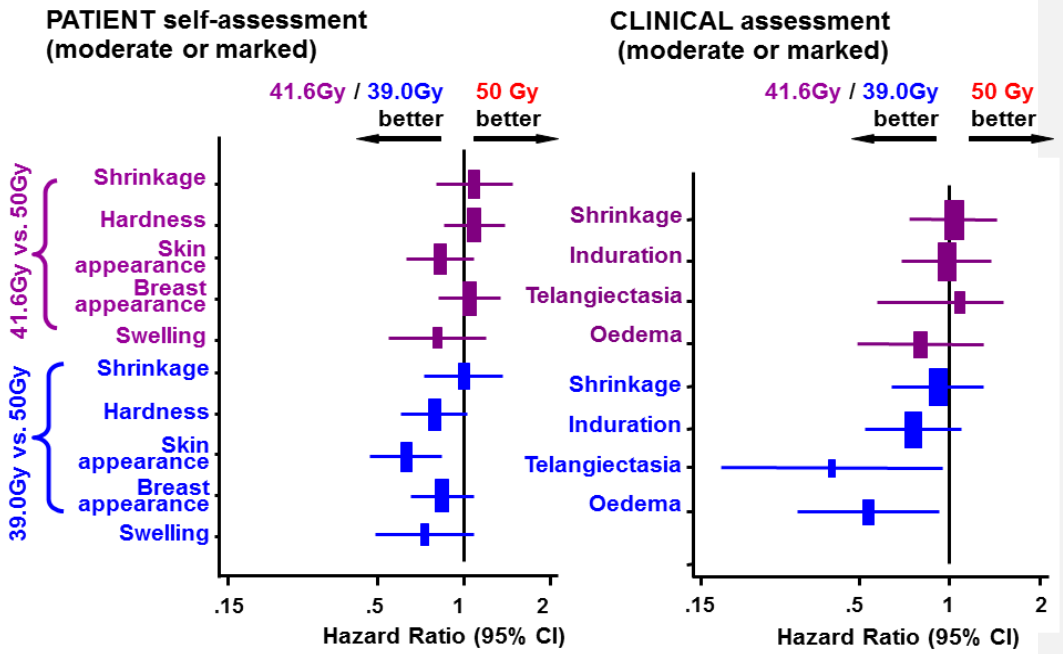


Figure 2: Comparisons between randomised radiotherapy schedules in START Trial A for ~~PRO~~PROMs and photographic assessments of overall change in breast appearance

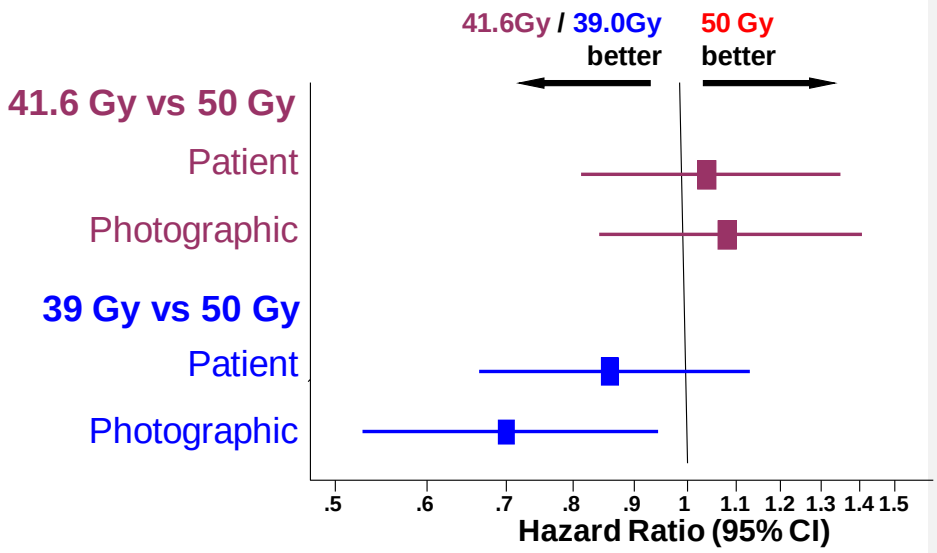
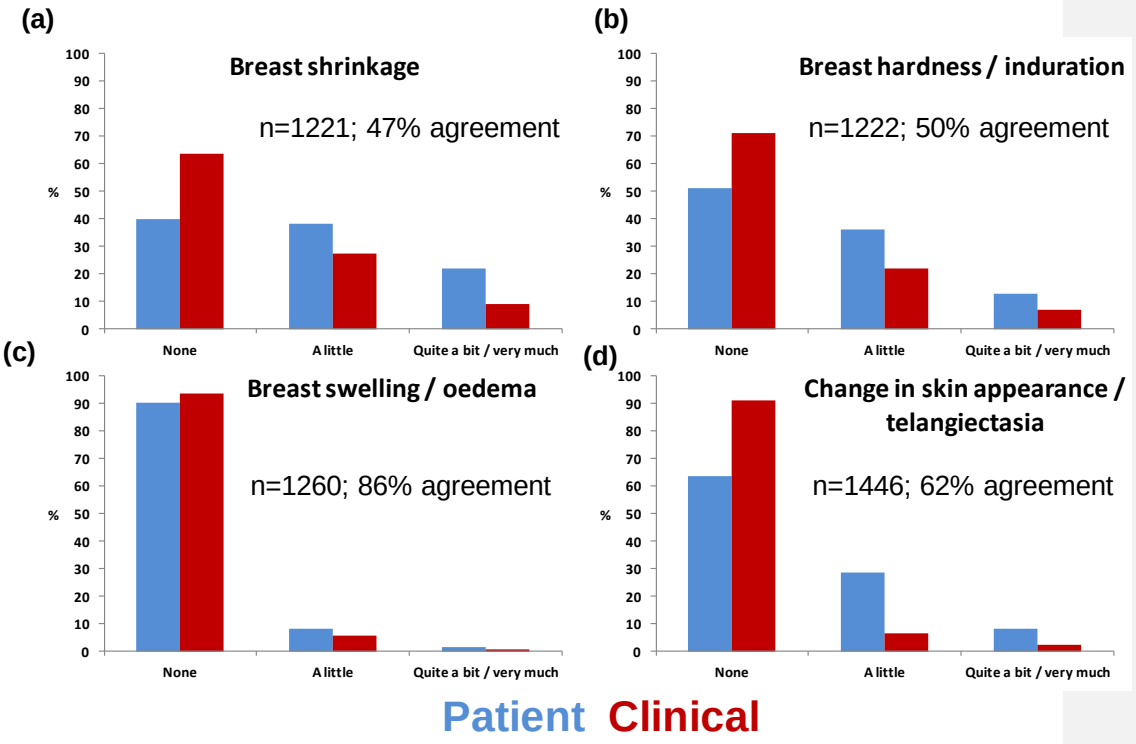


Figure 3: Comparison of 5-year ~~PRO~~PROMs, clinical and photographic assessments of specific normal tissue effects in START Trials A and B



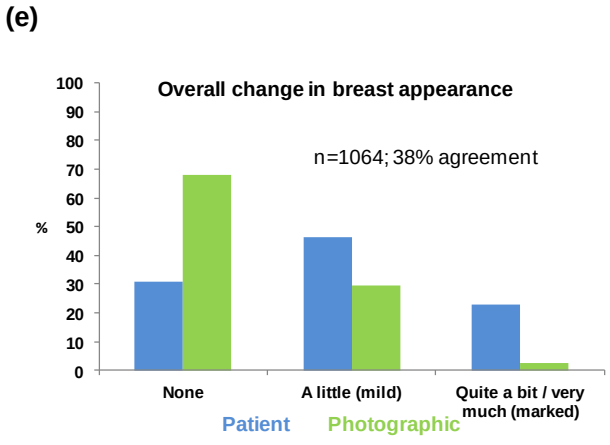


Table 1: Clinical and photographic outcome measures of specific late normal tissue effects in the breast and the corresponding ~~PRO~~PROM

Clinical assessment of late normal tissue effect in the treated breast	**Corresponding ~~PRO~~PROM used to test concordance with clinical or photographic assessment**[2]
Has the patient had any of the following adverse effects? Compare with contralateral breast[1]:	
Breast shrinkage	Has your affected breast become smaller as a result of your radiotherapy?[4]
Breast induration	Has your affected breast become harder/firmer to the touch since your radiotherapy?[4]
Breast oedema	During the past four weeks, was the area of your affected breast swollen?[5]
Telangiectasia	Has the appearance of the skin in the area of your affected breast changed since your radiotherapy?[4]
Has there been a change in photographic breast appearance compared with pre-radiotherapy baseline photograph?[3]	Has the overall appearance of your affected breast changed, compared with the other side, as a result of your radiotherapy?[4]

[1] Clinical assessments scored as none, a little, quite a bit, very much
[2] ~~PRO~~PROMs scored as not at all, a little, quite a bit, very much
[3] Photographic assessments scored as no change, mild change, marked change
[4] Protocol-specified items included in the patient questionnaire booklet under the heading “Since your breast radiotherapy”
[5] Question from the EORTC QLQ-BR23 breast cancer module

Table 2: Baseline characteristics of 1870 START Trial A and B patients with ~~PRO~~PROMs and clinical assessments of normal tissue effects at 2 and/or 5 years following radiotherapy

	Number of patients (%)
Age (years): mean (SD) [range]	57.0 (10.0) [27.1-86.0]
Type of primary surgery	
Breast conserving surgery	1574 (84.2)
Mastectomy	296 (15.8)
Axillary surgery	
None	55 (2.9)
Axillary clearance	1284 (68.7)
Axillary sampling	495 (26.5)
Sentinel node biopsy	36 (1.9)
Adjuvant chemotherapy	
No	1268 (67.8)
Yes	598 (32.0)
Unknown	4 (0.2)
Tamoxifen	
No	312 (16.7)
Yes	1554 (83.1)
Unknown	4 (0.2)
Breast size	
Small	154 (8.2)
Medium	1126 (60.2)
Large	228 (12.2)
Unknown – not in photographic study	362 (19.4)
Surgical deficit	
Small	872 (46.6)
Medium	496 (26.5)
Large	140 (7.5)
Unknown – not in photographic study	362 (19.4)
Hospital Anxiety and Depression Scale	
Anxiety	
Normal (0-7)	1287 (68.8)
Borderline (8-10)	322 (17.2)
Case (11+)	256 (13.7)
Unknown	5 (0.3)
Depression	
Normal (0-7)	1658 (88.7)
Borderline (8-10)	152 (8.1)
Case (11+)	52 (2.8)
Unknown	8 (0.4)
Body Image Scale (10-items): median (IQR) [range]	3 (0-8) [0-30]

SD = standard deviation; IQR = interquartile range
Breast size and surgical deficit assessed from baseline photographs
HADS scales range from 0-21
Body Image Scale ranges from 0-30, where a higher score indicates more concerns; unknown for 79 patients

Table 3: Concordance between ~~PRO~~PROMs and clinical or photographic assessments of specific normal tissue effects at 2 and 5 years in START Trials A and B

Clinicians	Patients None	Patients A little	Patients Quite a bit /very much	% agreement (95%CI)	Weighted Kappa (95%CI)	Bowker's test of symmetry, p-value
Breast shrinkage[1] – 2 years				755/1413;	0.21	<0.001
None	566	335	83	53.4%	(0.17-0.25)	
A little	107	158	70	(50.8-56.1%)		
Quite a bit / very much	18	45	31			
Breast shrinkage[1] – 5 years				579/1221;	0.19	<0.001
None	372	277	126	47.4%	(0.15-0.24)	
A little	96	151	87	(44.6-50.3%)		
Quite a bit / very much	18	38	56			
Breast induration / hardness[1] – 2 years				676/1439	0.12	<0.001
None	493	379	136	47.0%	(0.08-0.16)	
A little	112	152	73	(44.4-49.6%)		
Quite a bit / very much	31	32	31			
Breast induration / hardness[1] – 5 years				610/1222;	0.12	<0.001
None	482	295	94	49.9%	(0.07-0.16)	
A little	121	105	40	(47.1-52.8%)		
Quite a bit / very much	22	40	23			
Breast oedema / swelling[1] – 2 years				1144/1465;	0.21	0.017
None	1092	146	21	78.1%	(0.15-0.26)	
A little	109	51	9	(75.9-80.2%)		
Quite a bit / very much	16	20	1			
Breast oedema / swelling[1] – 5 years				1089/1260;	0.10	0.003
None	1076	86	19	86.4%	(0.04-0.17)	
A little	54	13	3	(84.4-88.2%)		
Quite a bit / very much	6	3	0			
Telangiectasia / change in skin appearance[2] – 2 years				959/1721;	0.05	<0.001
None	911	572	134	55.7%	(0.02-0.07)	
A little	32	42	11	(53.3-58.1%)		
Quite a bit / very much	6	7	6			
Telangiectasia / change in skin appearance[2] – 5 years				900/1446;	0.08	<0.001
None	859	369	90	62.2%	(0.04-0.12)	
A little	47	30	16	(59.7-64.7%)		
Quite a bit / very much	13	11	11			
Photographs						
Overall change in breast appearance[1] – 2 years				489/1290;	0.09	<0.001
None	331	525	130	37.9%	(0.06-0.11)	
Mild	56	141	78	(35.3-40.6%)		
Marked	4	8	17			
Overall change in breast appearance[1] – 5 years				409/1064;	0.09	<0.001
None	258	344	123	38.4%	(0.06-0.12)	
Mild	66	140	108	(35.5-41.4%)		
Marked	5	9	11			

CI = confidence interval
[1] breast conserving surgery patients only
[2] breast conserving surgery and mastectomy patients

Web Appendix: Concordance between ~~PRO~~PROMs and clinical or photographic assessments of specific normal tissue effects at 5 years stratified by baseline patient characteristics in START Trials A and B

	Breast shrinkage[1]		Breast induration/hardness[1]		Breast oedema/swelling[1]		Telangiectasia/change in skin appearance[2]		Overall change in breast appearance[1]	
	% agreement (95%CI)	Weighted Kappa (95%CI)	% agreement (95%CI)	Weighted Kappa (95%CI)	% agreement (95%CI)	Weighted Kappa (95%CI)	% agreement (95%CI)	Weighted Kappa (95%CI)	% agreement (95%CI)	Weighted Kappa (95%CI)
Age										
<50 years	43.7 (37.5-50.0)	0.22 (0.14-0.31)	47.4 (41.2-53.8)	0.09 (0.01-0.17)	N/A	N/A	56.7 (50.9-62.3)	0.06 (0.001-0.12)	37.6 (30.9-44.8)	0.05 (0-0.12)
≥50 years	48.4 (45.2-51.6)	0.20 (0.15-0.25)	50.6 (47.4-53.8)	0.13 (0.08-0.18)	86.9 (84.6-88.9)	0.12 (0.05-0.20)	63.7 (60.9-66.5)	0.09 (0.05-0.14)	39.2 (35.9-42.7)	0.11 (0.07-0.15)
Breast size										
Small	52.8 (43.7-61.8)	0.13 (0-0.26)	59.8 (50.5-68.5)	0.06 (0-0.19)	N/A	N/A	N/A	N/A	41.2 (32.2-50.8)	0.02 (0-0.06)
Medium	48.9 (45.5-52.2)	0.22 (0.17-0.27)	49.8 (46.4-53.1)	0.11 (0.05-0.16)	87.1 (84.6-89.2)	0.06 (0-0.13)	62.9 (59.6-66.1)	0.05 (0.01-0.10)	38.2 (34.7-41.7)	0.08 (0.05-0.12)
Large	37.8 (30.8-45.5)	0.10 (0-0.21)	44.6 (37.2-52.3)	0.10 (0-0.21)	80.5 (73.9-85.8)	0.20 (0.05-0.36)	48.7 (41.3-56.0)	0.07 (0-0.16)	36.9 (29.5-45.0)	0.06 (0-0.17)
Surgical deficit										
Small	50.9 (47.1-54.7)	0.21 (0.15-0.26)	50.1 (46.3-53.8)	0.10 (0.04-0.15)	84.9 (82.1-87.4)	0.06 (0-0.14)	62.6 (58.9-66.2)	0.08 (0.03-0.13)	40.5 (36.7-44.6)	0.10 (0.06-0.14)
Medium	43.9 (38.8-49.1)	0.16 (0.08-0.23)	53.0 (47.7-58.1)	0.20 (0.11-0.29)	90.8 (87.3-93.4)	0.28 (0.12-0.44)	60.8 (55.7-65.8)	0.03 (0-0.10)	36.9 (31.9-42.3)	0.08 (0.01-0.14)
Large	39.3 (30.3-49.0)	0.12 (0-0.24)	40.0 (30.9-49.8)	N/A	82.0 (73.3-88.4)	N/A	60.0 (50.2-69.1)	0.10 (0-0.23)	28.9 (20.3-39.1)	N/A
HADS anxiety										
0-7 (normal)	50.9 (47.5-54.3)	0.22 (0.17-0.27)	52.2 (48.8-55.6)	0.12 (0.07-0.18)	89.0 (86.7-90.9)	0.09 (0-0.17)	65.6 (62.6-68.5)	0.07 (0.03-0.12)	40.3 (36.8-43.9)	0.09 (0.05-0.13)
8-10 (borderline)	43.2 (36.4-50.3)	0.14 (0.05-0.22)	46.1 (39.2-53.2)	0.12 (0.03-0.22)	81.2 (75.2-86.1)	0.23 (0.08-0.38)	56.7 (50.3-63.0)	0.10 (0-0.20)	34.8 (28.0-42.2)	0.08 (0.01-0.15)
≥11 (case)	35.0 (27.7-43.0)	0.13 (0.04-0.22)	42.5 (34.8-50.1)	0.09 (0-0.19)	N/A	N/A	51.1 (43.7-58.4)	0.08 (0-0.17)	32.1 (24.3-40.9)	0.07 (0-0.15)
HADS depression										
0-7 (normal)	48.2 (45.2-51.2)	0.19 (0.15-0.24)	51.4 (48.4-54.4)	0.13 (0.08-0.18)	87.9 (85.8-89.7)	0.13 (0.06-0.21)	64.5 (61.9-67.1)	0.09 (0.04-0.13)	38.7 (35.6-41.9)	0.08 (0.05-0.12)
8-10 (borderline)	40.9 (30.7-51.9)	0.15 (0.01-0.29)	31.8 (22.5-42.7)	N/A	70.0 (59.3-79.0)	N/A	43.2 (34.0-53.0)	0.06 (0.02-0.11)	38.0 (27.5-49.6)	0.13 (0.04-0.23)
≥11 (case)	43.7 (26.8-62.1)	0.26 (0.07-0.46)	46.9 (29.5-65.0)	0.21 (0-0.48)	N/A	N/A	37.5 (23.2-54.2)	0.05 (0-0.19)	25.9 (11.9-46.6)	N/A
Body Image Scale[3]										
0-3	52.2 (48.4-56.1)	0.24 (0.18-0.30)	53.9 (50.1-57.7)	0.14 (0.07-0.20)	88.5 (85.9-90.8)	0.15 (0.05-0.25)	66.2 (62.7-69.9)	0.07 (0.01-0.13)	40.5 (36.6-44.6)	0.09 (0.04-0.14)
>3	41.3 (37.0-45.7)	0.14 (0.08-0.20)	43.6 (39.3-48.1)	0.08 (0.01-0.14)	83.7 (80.2-86.7)	0.05 (0-0.13)	57.8 (53.9-61.7)	0.08 (0.03-0.13)	35.0 (30.6-39.7)	0.09 (0.05-0.13)

CI = confidence interval; N/A = not available
[1] breast conserving surgery patients only
[2] breast conserving surgery and mastectomy patients
[3] 10-item Body Image Scale (possible range 0-30; median baseline score = 3)

***Title Page**

Do patient-reported outcome measures agree with clinical and photographic assessments of normal tissue effects after breast radiotherapy? The experience of the Standardisation of Breast Radiotherapy (START) Trials in early breast cancer

JS Haviland[1,2], P Hopwood[2], J Mills[2], M Sydenham[2], JM Bliss*[2], JR Yarnold*[3] on behalf of the START Trialists' Group

[1] Faculty of Health Sciences, University of Southampton, Southampton, UK

[2] ICR-CTSU, Division of Clinical Studies, The Institute of Cancer Research, London, UK

[3] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK

* Joint last authors

Corresponding author:

Professor John R Yarnold

Division of Radiotherapy and Imaging

The Royal Marsden, Downs Road, Sutton SM2 5PT, UK

Tel: +44 208 661 3388, Fax: +44 208 661 3107, Email: john.yarnold@icr.ac.uk

Do patient-reported outcome measures agree with clinical and photographic assessments of normal tissue effects after breast radiotherapy? The experience of the Standardisation of Breast Radiotherapy (START) Trials in early breast cancer

JS Haviland[1,2], P Hopwood[2], J Mills[2], M Sydenham[2], JM Bliss*[2], JR Yarnold*[3] on behalf of the START Trialists' Group

[1] Faculty of Health Sciences, University of Southampton, Southampton, UK

[2] ICR-CTSU, Division of Clinical Studies, The Institute of Cancer Research, London, UK

[3] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK

* Joint last authors

Corresponding author:

Professor John R Yarnold

Division of Radiotherapy and Imaging

The Royal Marsden, Downs Road, Sutton SM2 5PT, UK

Tel: +44 208 661 3388, Fax: +44 208 661 3107, Email: john.yarnold@icr.ac.uk

Abstract

Aims

In radiotherapy trials normal tissue effects (NTE) are important endpoints, and it is pertinent to ask whether patient reported outcome measures (PROMs) could replace clinical and/or photographic assessments. Data from the START breast radiotherapy trials are examined.

Materials and Methods

NTEs in the treated breast were recorded by i) annual clinical assessments, ii) photographs at 2 and 5 years, iii) PROMs at 6 months, 1, 2 and 5 years following radiotherapy. Hazard ratios for the radiotherapy schedules were compared. Measures of agreement of assessments at 2 and 5 years tested concordance.

Results

PROMs were available at 2 and/or 5 years for 1939 women, of whom 1870 had clinical and 1444 had photographic assessments. All methods were sensitive to the dose difference between schedules. Patients reported higher prevalence for all NTE endpoints than clinicians or photographs ($p<0.001$ for most NTEs). Concordance was generally poor; weighted kappa at 2 years ranged from 0.05 (telangiectasia) to 0.21 (shrinkage and oedema). Percentage agreement was lowest between PROMs and photographic assessments of change in breast appearance (38%).

Conclusions

All 3 methods produced similar conclusions for the comparison of trial schedules, despite low concordance between the methods on an individual patient basis. Careful consideration should be given to the different contributions of the measures of NTE in future radiotherapy trials.

Keywords: breast radiotherapy, normal tissue effects, patient-reported outcomes

Introduction

Traditional outcome measures of normal tissue responses to radiotherapy rely heavily, often exclusively, on clinical assessments using graded scales to score a wide range of early and late adverse effects [1-4]. Scoring systems, including Late Effects in Normal Tissues Subjective, Objective, Management and Analytic (LENT-SOMA), Radiation Therapy Oncology Group (RTOG) and Common Terminology Criteria for Adverse Events (CTCAE), feature symptomatology requiring health professionals to elicit and score responses to direct questions. Photographic assessments of change in breast appearance from a pre-radiotherapy baseline have become increasingly used in randomised trials of radiotherapy as they are usually scored by a small number of observers blinded to patient identity, treatment allocation and year of follow-up, unlike the clinical assessments which are scored by a large number of individuals in a multi-centre study [5]. In parallel, the use of carefully developed and validated quality of life instruments in psychosocial research and phase III cancer clinical trials has expanded considerably [6-8], together with the growing interest in use of PROMS in routine follow-up [9]. With an increasing use of patient-reported outcome measures (PROMs) in cancer clinical trials [10, 11], it is worth asking how comparable and interpretable are the different methods of assessment, and whether PROMs could become the primary means of scoring late normal tissue effects (NTE) of breast radiotherapy in trials. Against this background, the large-scale UK START randomised trials [12-15] of hypofractionated radiotherapy after primary surgery for early breast cancer were used to conduct exploratory analyses comparing different methods of assessment of late NTE after adjuvant breast radiotherapy with the primary aim of assessing if PROMs might take

priority over, or replace, clinical and/or photographic assessments as outcome measures.

Materials and Methods

The START-A and START-B trials recruited 4451 women between 1998 and 2002 from 35 UK radiotherapy centres (ISRCTN59368779, MREC(1)98/86). Centres could opt to participate in the PROMs and photographic assessment studies, and if they participated, they were expected to invite every eligible trial patient to join. Thirty one (89%) centres opted to participate in the PROMs study and 29 (83%) in a photographic assessment study of change in breast appearance. Women with operable invasive breast cancer (International Union Against Cancer pT1-3a pN0-1 M0) requiring radiotherapy after surgery (breast-conserving surgery or mastectomy, with clear tumour margins ≥1 mm) were eligible for the trials if they were aged over 18 years, did not have an immediate surgical reconstruction, and were available for follow-up. Trial-A patients were randomised to either 50 Gy in 25 fractions (control) or 41.6 Gy in 13 fractions of 3.2 Gy or 39.0 Gy in 13 fractions of 3.0 Gy over 5 weeks. Trial-B patients were randomised to either 50 Gy in 25 fractions over 5 weeks (control) or 40 Gy in 15 fractions of 2.7 Gy over 3 weeks. Full details of the recruitment, and radiotherapy planning, delivery and verification protocols have been previously reported, as has the PROMs study [12-14].

Patients in the PROMs study completed baseline measures in clinic and were sent questionnaires to complete at home at 6 months, 1, 2 and 5 years following radiotherapy. Clinical assessments of NTE were collected at annual follow-up in all patients, and photographs were taken under standard conditions at post-surgical

pre-radiotherapy baseline and at 2 and 5 years post-randomisation for patients who had breast conserving surgery. The patient questionnaires included the i) EORTC QLQ-C30 core questionnaire and QLQ-BR23 breast-specific module [6, 16], from which the assessment of breast swelling over the previous 4 weeks (not at all, a little, quite a bit, very much) was used in this study of concordance, ii) Hospital Anxiety and Depression Scale [17], iii) 10-item Body Image Scale [18] and iv) 4 protocol-specific questions asking patients to score “change in breast appearance”, “breast hardness/firmness”, “reduction in size of breast” and “change in skin appearance” since radiotherapy; the first three questions applying only to patients with conserved breasts, and all items scored on a 4-point scale (none, a little, quite a bit, very much).

The annual clinical assessments of breast shrinkage, breast induration, telangiectasia and breast oedema were scored using the contralateral breast as a comparator and 4-point graded scales (none, a little, quite a bit, very much). Change in photographic breast appearance since radiotherapy was scored by a single team of 3 observers blind to patient identity, trial treatment allocation, year of follow-up and radiotherapy centre. The scoring method was validated in the START pilot trial [5]. Photographs at 2 and 5 years following radiotherapy were compared with a pre-radiotherapy (post-surgery) baseline and an overall score allocated for change in photographic breast appearance in the treated breast based on change in size, shrinkage and shape, on a 3-point scale (no change, mild change, marked change). Post-mastectomy patients were included in the PROMs and clinical assessments but not in the photographic assessments. Individual NTE were mapped between the different assessment methods in order to compare corresponding outcomes, as shown in Table 1.

Statistical methods

NTE assessments at all time-points in the trials were included in the comparison of radiotherapy schedules (i.e. from 6 months-5 years for the PROMs, from 1-5 years for the clinical assessments, and at 2 and 5 years for the photographs). Time to first NTE event (defined as “quite a bit” or “very much” for the PROMs and clinical assessments, and any change (mild or marked) in photographic breast appearance) was calculated from date of randomisation, and survival analysis methods used to compare radiotherapy schedules. Hazard ratios (HR) for the relative effects of the radiotherapy schedules in START-A were calculated for each NTE endpoint using Cox proportional hazards regression and compared between the different assessment methods using forest plots. Estimates of the α/β ratio for NTEs, which describes the sensitivity of normal tissues to fraction size, were obtained separately for the PROMs, clinician and photographic endpoints in START-A. Estimates of relative effects of the fractionation schedules in START-B are not presented in this paper as they do not contribute to the measurement of fraction sensitivity, only having two randomised groups in Trial B. HRs for the fractionation schedules in START-B have been published separately for the different NTE assessments, and showed consistent results [13-15].

For the concordance analyses, data from START Trials A and B were combined, and only 2 and 5-year assessments included as these were the time-points at which all three NTE assessment methods were used in the trials. For all PROMs and clinically-assessed endpoints there were few patients in the highest grade category, so moderate and marked categories were combined, resulting in 3-point scales

corresponding to none, a little (“mild”), quite a bit / very much (“moderate / marked”); this also enabled comparison with the photographic assessments, which were scored on a similar 3-point scale. Corresponding NTE endpoints were matched between the PROMs, clinical and photographic assessments at each time point and compared on an individual patient basis using measures of concordance including percentage agreement (with 95% confidence interval, CI), weighted Kappa statistic (with 95%CI) and Bowker’s test of symmetry [19]. Guidelines for interpreting the value of the weighted Kappa statistic in terms of the strength of agreement are <0.20: poor, 0.21-0.40: fair, 0.41-0.6: moderate, 0.61-0.8: good, 0.81-1.00: very good [20]. Bowker’s test assesses the symmetry of a square table – i.e. whether there are more observations on one side of the diagonal than the other. The concordance analyses were also carried out stratifying on baseline patient characteristics such as age and quality of life scores (including anxiety and depression from the Hospital Anxiety and Depression Scale and body image from the Body Image Scale), to investigate whether these had any effect on the degree of concordance between NTE assessment methods.

Results

Of the 2208 women recruited into the overall START Trials PROMs study, self-assessments of NTEs were available at 2 and/or 5 years for 1939 (88%) patients, of whom 1870 also had clinical assessments at the same time-points (85% of all patients in PROMs study). Patient characteristics at baseline for the 1870 patients in this analysis are shown in Table 2, of whom 1574/1870 (84.2%) had breast conserving surgery and 1444/1574 (91%) had photographic assessments at 2 and/or 5 years.

Treatment effects on late NTE assessed by PROMs and by annual clinical assessment in START-A are shown side-by-side in Figure 1. Two test schedules (41.6 Gy and 39 Gy in 13 fractions) were compared with control (50 Gy in 25 fractions) in START-A. Comparing HR for corresponding endpoints, it can be seen that the treatment effects were of a similar size for PROMs and clinical assessments, with overlapping confidence intervals. Treatment effects on late NTE assessed by PROMs and by photographs for overall change in breast appearance were also similar (Figure 2). α/β estimates (adjusted for prognostic factors) for overall change in breast appearance were 2.9 Gy (95%CI 0.7-5.1 Gy) for PROMs and 2.6 Gy (95%CI 1.3-3.9 Gy) for photographic assessments. α/β estimates for individual NTE endpoints from clinical assessments have been reported [14] (there was no clinical assessment of overall cosmesis in the START Trials).

The comparison of overall rates of NTEs reported by PROMs and clinical assessments from START Trials A and B combined showed that patients reported a higher prevalence of breast changes (Figures 3a-d). Concordance between the assessments of corresponding NTEs on an individual patient basis was generally poor (Table 3). The lowest levels of percentage agreement between PROMs and clinicians were observed for breast induration / hardness (47% and 50% at 2 and 5 years, respectively), and breast shrinkage (53% and 47% at 2 and 5 years). The highest level of percentage agreement between PROMs and clinicians was for breast swelling/oedema (78% and 86% at 2 and 5 years), but the overall prevalence of oedema was very low (Figure 3c). Weighted kappa statistics also highlighted the low agreement between methods, ranging from 0.05 for telangiectasia at 2 years

(indicating poor agreement) to 0.21 for each of breast shrinkage and breast oedema at 2 years (indicating fair agreement). Results of Bowker's test of symmetry were highly statistically significant for all NTE endpoints, indicating a clear direction in the discordance of scoring between the different methods, with patients reporting more breast changes compared with clinical and photographic assessments (Table 3). There appeared to be no substantial differences in degree of concordance for individual NTE endpoints according to time since radiotherapy i.e. between 2 and 5 years (Table 3).

The comparison of PROMs and photographic assessments showed that patients reported a higher prevalence of overall change in breast appearance since radiotherapy and graded effects as more severe compared with the photographic assessments (Figure 3e). In testing concordance, agreement on an individual patient basis was low at 2 and 5 years (38% for each), with low weighted kappa values (0.09) and highly statistically significant discordance ($p<0.001$ for Bowker's test of symmetry); Table 3. Concordance of PROMs with clinical and photographic assessments of NTE appeared to be unaffected by patient factors including age, breast size, surgical deficit, baseline HADS anxiety and depression and body image scores (table in web appendix).

Discussion

Concordance between PROMs and NTE assessments as scored by clinicians and from photographs on an individual patient basis was poor. Percentage agreement between PROMs and clinical assessments of specific NTEs was around 50%, indicating that in only half of the patients the NTE was graded in the same category

of severity corresponding to none, mild, moderate/marked. Agreement was even lower between PROMs and photographs, where less than 40% graded NTEs the same. In our study, patients scored NTEs more frequently and more severely than results from clinicians or photographs. Concordance did not appear to be affected by patient characteristics including psychological measures (anxiety and depression), body image and factors associated with risk of NTEs (age, breast size and surgical deficit). It may not be surprising that concordance between the assessment methods on an individual patient basis was poor; this has been consistently reported in other studies [21-24]. These differences in ratings reflect the different paradigms in which symptoms are perceived and rated; these include variance in context, values, expectations and methodological influences as well as the different sociocultural backgrounds of subjects and doctors [25]. Published comparisons of clinician and patient self-assessments show considerable variability between ratings, especially for more subjective symptoms and often report, as in our study, a relative underestimate by clinicians compared with patients (e.g. Basch et al [26], Bruner et al [27], Fromme et al [23], Groenwold et al [28], Quinten et al [29], Stephens et al [30], Velikova et al [24]). However, the concordance analysis of NTE assessments in the Cambridge intensity-modulated breast radiotherapy trial found the opposite, with clinicians and photographic assessments reporting more NTEs compared with patients, possibly because the study was done in a single centre, with clinical ratings done by one person [31]. Others have shown more favourable rating of overall cosmesis following conservative treatment for breast cancer by patients compared with clinicians [32, 33], although these findings are not necessarily specific to late effects of radiotherapy. Kirchheiner et al [34] argued that some variation is “quite acceptable and comprehensible”, given the methodological differences

between morbidity scoring by clinicians and patient-reported symptoms. Clinical and patient symptom ratings are typically not designed to be interchangeable, given that they often have different values and purposes, with patient assessments inherently encompassing impact on quality of life.

However, our study showed that despite the discordance between assessments on an individual basis, the three methods (PROMs, clinical and photographs) generated similar estimates of relative treatment effects on NTE within the trials [12, 14, 15]. The discriminatory power of different assessments was equally good, in that PROMs generated the same estimates of α/β value for NTE in START-A (around 3 Gy) as photographs and clinical assessments (data for α/β values of clinical assessments of NTEs previously published [14]). From the trial outcome perspective, this consistency of treatment effects adds considerable weight to the overall interpretation and conclusions of the trial. However, the PROMs reported here were selected from a large number of multidimensional items assessed as part of the START quality of life sub-study, most of which would not be expected to discriminate so clearly between the schedules in the START trials, but are of value in understanding the experience of treatment effects over time. The PROMs items included in this analysis of concordance were those directly relevant to the hypothesis under test in the clinical trial, and therefore most likely to be sensitive to randomised differences in radiotherapy dose intensity. The PROMs needed to have a recognisable relationship with the pathophysiology (atrophy, fibrosis) of NTE, broadly corresponding to clinical scoring of change in size (atrophy), shape and texture (oedema, fibrosis) of the breast and change in photographic breast appearance (atrophy, distortion/fibrosis). This is in contrast with other clinically

relevant domains, such as physical and social functioning, that explore the impact on different aspects of quality of life [6, 16].

Clinicians are taught in training that symptomatology is the key to diagnosis, which they can only judge by listening to their patients and framing relevant questions. Clinicians act as surrogates for their patients in this context, so that if the relevant questions are known in advance (as they are in a clinical trial), there appears to be a good reason to prioritise the PROMs over the physical clinical assessments. Where physical signs are concerned, including breast size, shape and texture, this study suggests that patients are as sensitive as their doctors in scoring these changes too, provided the questions are framed appropriately. In this respect, it is possible to criticise our PROMs question, which asked patients to score changes since radiotherapy to the affected breast compared with the clinical assessment that compared the treated with the untreated breast at the time of the annual examination. Despite a variety of factors expected to influence how a woman responds to this question, the sensitivity to randomised dose indicates that the radiotherapy ‘signal’ was not lost. Doctors also develop their own frames of reference when assessing NTE, and the hundreds of clinical observers involved in scoring NTE in thousands of patients over a 10-year period, as in the START trials, necessarily contribute a lot of ‘noise’ in a scoring system. However, a disadvantage of reliance on PROMs in clinical trials is that they are traditionally labour-intensive to administer and generate large volumes of data, making heavy demands on trial management and statistical resources. Since modern data capture systems are increasingly able to collect outcome data directly from the patient (e.g. via an App), dispensing with clinical follow-up may appeal to patients as well as health services

operating under increasing pressures [35]. However, radiation effects are not viewed in isolation by patients and attention also needs to be paid to their concerns in the context of multi-modal treatments and adverse effects over time. Up to a third of patients report moderate or marked symptoms of the breast, arm and shoulder at 5 years, which may warrant engagement and advice from their clinical teams [13]. Thus more preparation and after care is needed for the success of patient self-management post-treatment and to improve quality of life [36]. Further, the acceptability of electronic symptom-reporting warrants evaluation in an aging population.

Despite adding to the administrative burden of clinical trials, the photographic assessments of NTEs provide valuable information, not least because they are scored generally by the same small team of observers who are blind to patient identity, randomised treatment allocation, year of follow-up and participating hospital. As it is generally not possible to blind treatment allocation in radiotherapy trials the photographic assessments provide the only unbiased comparison of normal tissue effects between randomised groups. In addition, as photographs provide a permanent record of breast effects at a fixed point in time, the assessments can be validated by repeat scoring from different teams of observers [5], thus making the scoring more standardised than PROMs or clinical assessments from physical examination. Photographs can also be filed and stored for use in future translational research investigating adverse effects of radiotherapy. There are some disadvantages to the use of photographic assessments in clinical trials, including financial and staff resources required, and they can be disliked by patients, but these

are outweighed by the benefits of retaining an unbiased comparison of NTEs within radiotherapy trials.

There is growing interest in investigating inherited risk factors for radiotherapy NTE, for which robust measures of NTE are needed that have a close relationship to the underlying pathophysiology [37], In this respect, the lack of concordance reported in this study is intriguing and potentially worrying. The prevalence and severity of NTEs reported by patients, clinicians and from photographs during follow-up were widely discordant in most cases. In trying to identify subgroups of patients with levels of NTE that are much more, or much less, severe than expected on the basis of known factors (breast size, radiotherapy dose etc.), it isn't possible to judge whether the clinical and photographic assessments of NTE severity are more or less valid than the PROMs, hence making identification of potential cases (and controls) for translational studies very difficult. Perhaps much depends on how the NTE assessment questions to patients and clinicians are posed, something that this study does not address.

In conclusion, the PROMs, clinical and photographic assessments of late NTE in the START trials generated consistent estimates of relative treatment effects between randomised groups, adding weight to the trials' overall findings. Discordance in the prevalence rates of NTE reported by the patients, clinicians and photographs could be expected for a number of well-established reasons, but this does not undermine an argument for prioritising PROMs and photographic assessments of NTEs in breast radiotherapy trials.

Conflict of Interest Statement

The authors have no conflicts of interest to disclose.

Role of the Funding Source

The START trials were funded by Cancer Research UK, Medical Research Council and the UK Department of Health (grant G9600656). The Cancer Research UK number for the START trials is CRUK/96/001. The funders had no role in study design or conduct, analysis or interpretation of data. The manuscript was approved by all authors and the corresponding author had the final decision to submit the manuscript for publication.

Acknowledgements

We thank all the patients who participated in the START trials, and the doctors, nurses, radiographers, physicists, and data managers at the participating centres. Continued data collection and analysis is made possible by a core grant from Cancer Research UK to the ICR-CTSU. We acknowledge NHS funding to the NIHR Biomedical Research Centre at The Royal Marsden and the ICR. The START Trialists' Group consists of the trial management group, consumers, trial steering committee, independent data monitoring committee, and the principal and main co-investigators at the participating centres (published previously [14]). We also acknowledge Claire Birch, who worked on preliminary concordance analyses of an earlier START dataset for her MSc dissertation.

References

[1] http://www.rtog.org/ResearchAssociates/AdverseEventReporting/RTOGEORTCLateRadiationMorbidityScoringSchema.aspx.

[2] http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf.

[3] Pavy JJ, Denekamp J, Letschert J, Littbrand B, Mornex F, Bernier J, et al. EORTC Late Effects Working Group. Late effects toxicity scoring: the SOMA scale. Radiother Oncol. 1995;35:11-5.

[4] Rubin P, Constine LS, Fajardo LF, Phillips TL, Wasserman TH. RTOG Late Effects Working Group. Overview. Late Effects of Normal Tissues (LENT) scoring system. Int J Radiat Oncol Biol Phys. 1995;31:1041-2.

[5] Haviland JS, Ashton A, Broad B, Gothard L, Owen JR, Tait D, et al. Evaluation of a method for grading late photographic change in breast appearance after radiotherapy for early breast cancer. Clin Oncol (R Coll Radiol). 2008;20:497-501.

[6] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85:365-76.

[7] Brady MJ, Cella DF, Mo F, Bonomi AE, Tulsky DS, Lloyd SR, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. J Clin Oncol. 1997;15:974-86.

[8] Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35:1095-108.

[9] Faithfull S, Lemanska A, Chen T. Patient-reported Outcome Measures in Radiotherapy: Clinical Advances and Research Opportunities in Measurement for Survivorship. Clin Oncol (R Coll Radiol). 2015;27:679-85.

[10] Basch E, Abernethy AP, Mullins CD, Reeve BB, Smith ML, Coons SJ, et al. Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. J Clin Oncol. 2012;30:4249-55.

[11] Rothman ML, Beltran P, Cappelleri JC, Lipscomb J, Teschendorf B, Mayo FDAP-ROCMG. Patient-reported outcomes: conceptual issues. Value Health. 2007;10 Suppl 2:S66-75.

[12] Bentzen SM, Agrawal RK, Aird EG, Barrett JM, Barrett-Lee PJ, Bliss JM, et al. The UK Standardisation of Breast Radiotherapy (START) Trial A of radiotherapy hypofractionation for treatment of early breast cancer: a randomised trial. Lancet Oncol. 2008;9:331-41.

[13] Hopwood P, Haviland JS, Sumo G, Mills J, Bliss JM, Yarnold JR. Comparison of patient-reported breast, arm, and shoulder symptoms and body image after radiotherapy for early breast cancer: 5-year follow-up in the randomised Standardisation of Breast Radiotherapy (START) trials. Lancet Oncol. 2010;11:231-40.

[14] Haviland JS, Owen JR, Dewar JA, Agrawal RK, Barrett J, Barrett-Lee PJ, et al. The UK Standardisation of Breast Radiotherapy (START) trials of radiotherapy hypofractionation for treatment of early breast cancer: 10-year follow-up results of two randomised controlled trials. Lancet Oncol. 2013;14:1086-94.

[15] Bentzen SM, Agrawal RK, Aird EG, Barrett JM, Barrett-Lee PJ, Bliss JM, et al. The UK Standardisation of Breast Radiotherapy (START) Trial B of radiotherapy

hypofractionation for treatment of early breast cancer: a randomised trial. Lancet. 2008;371:1098-107.

[16] Sprangers MA, Groenvold M, Arraras JI, Franklin J, te Velde A, Muller M, et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. J Clin Oncol. 1996;14:2756-68.

[17] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67:361-70.

[18] Hopwood P, Fletcher I, Lee A, Al Ghazal S. A body image scale for use with cancer patients. Eur J Cancer. 2001;37:189-97.

[19] Bowker AH. A test for symmetry in contingency tables. Journal of the American Statistical Association. 1948;43:572-4.

[20] Altman DG. Practical statistics for medical research. London: Chapman and Hall. 1991.

[21] Davidson SE, Trotti A, Ataman OU, Seong J, Lau FN, da Motta NW, et al. Improving the capture of adverse event data in clinical trials: the role of the International Atomic Energy Agency. Int J Radiat Oncol Biol Phys. 2007;69:1218-21.

[22] Vistad I, Cvancarova M, Fossa SD, Kristensen GB. Postradiotherapy morbidity in long-term survivors after locally advanced cervical cancer: how well do physicians' assessments agree with those of their patients? Int J Radiat Oncol Biol Phys. 2008;71:1335-42.

[23] Fromme EK, Eilers KM, Mori M, Hsieh YC, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. J Clin Oncol. 2004;22:3485-90.

[24] Velikova G, Wright P, Smith AB, Stark D, Perren T, Brown J, et al. Self-reported quality of life of individual cancer patients: concordance of results with disease course and medical records. J Clin Oncol. 2001;19:2064-73.

[25] Gotay C. Patient symptoms and clinician toxicity ratings: both have a role in cancer care. J Natl Cancer Inst. 2009;101:1602-3.

[26] Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst. 2009;101:1624-32.

[27] Bruner DW, Movsas B, Basch E. Capturing the patient perspective: patient-reported outcomes as clinical trial endpoints. American Society of Clinical Oncology educational book / ASCO American Society of Clinical Oncology Meeting. 2012:139-44.

[28] Groenvold M, Fayers PM, Petersen MA, Sprangers MA, Aaronson NK, Mouridsen HT. Breast cancer patients on adjuvant chemotherapy report a wide range of problems not identified by health-care staff. Breast Cancer Res Treat. 2007;103:185-95.

[29] Quinten C, Maringwa J, Gotay CC, Martinelli F, Coens C, Reeve BB, et al. Patient self-reports of symptoms and clinician ratings as predictors of overall cancer survival. J Natl Cancer Inst. 2011;103:1851-8.

[30] Stephens RJ, Hopwood P, Girling DJ, Machin D. Randomized trials with quality of life endpoints: are doctors' ratings of patients' physical symptoms interchangeable with patients' self-ratings? Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 1997;6:225-36.

[31] http://abstracts.ncri.org.uk/abstract/a-comparison-of-clinician-versus-patient-reported-outcomes-proms-for-late-normal-tissue-side-effects-following-breast-radiotherapy-results-of-the-cambridge-breast-intensity-modulated-radiotherapy-3/.

[32] Kaija H, Rauni S, Jorma I, Matti H. Consistency of patient- and doctor-assessed cosmetic outcome after conservative treatment of breast cancer. Breast Cancer Res Treat. 1997;45:225-8.

[33] Hoeller U, Kuhlmey A, Bajrovic A, Grader K, Berger J, Tribius S, et al. Cosmesis from the patient's and the doctor's view. Int J Radiat Oncol Biol Phys. 2003;57:345-54.

[34] Kirchheiner K, Nout R, Lindegaard J, Petric P, Limbergen EV, Jurgenliemk-Schulz IM, et al. Do clinicians and patients agree regarding symptoms? A comparison after definitive radiochemotherapy in 223 uterine cervical cancer patients. Strahlenther Onkol. 2012;188:933-9.

[35] Bruner DW, Hanisch LJ, Reeve BB, Trotti AM, Schrag D, Sit L, et al. Stakeholder perspectives on implementing the National Cancer Institute's patient-reported outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Translational behavioral medicine. 2011;1:110-22.

[36] Corner J, Wagland R, Glaser A, Richards SM. Qualitative analysis of patients' feedback from a PROMs survey of cancer patients in England. BMJ open. 2013;3.

[37] Andreassen CN, Barnett GC, Langendijk JA, Alsner J, De Ruysscher D, Krause M, et al. Conducting radiogenomic research--do not forget careful consideration of the clinical data. Radiother Oncol. 2012;105:337-40.

Figure legends

Figure 1: Comparisons between randomised radiotherapy schedules in START Trial A for PROMs and clinical assessments of specific normal tissue effects

Figure 2: Comparisons between randomised radiotherapy schedules in START Trial A for PROMs and photographic assessments of overall change in breast appearance

Figure 3: Comparison of 5-year PROMs, clinical and photographic assessments of specific normal tissue effects in START Trials A and B

Figure 1

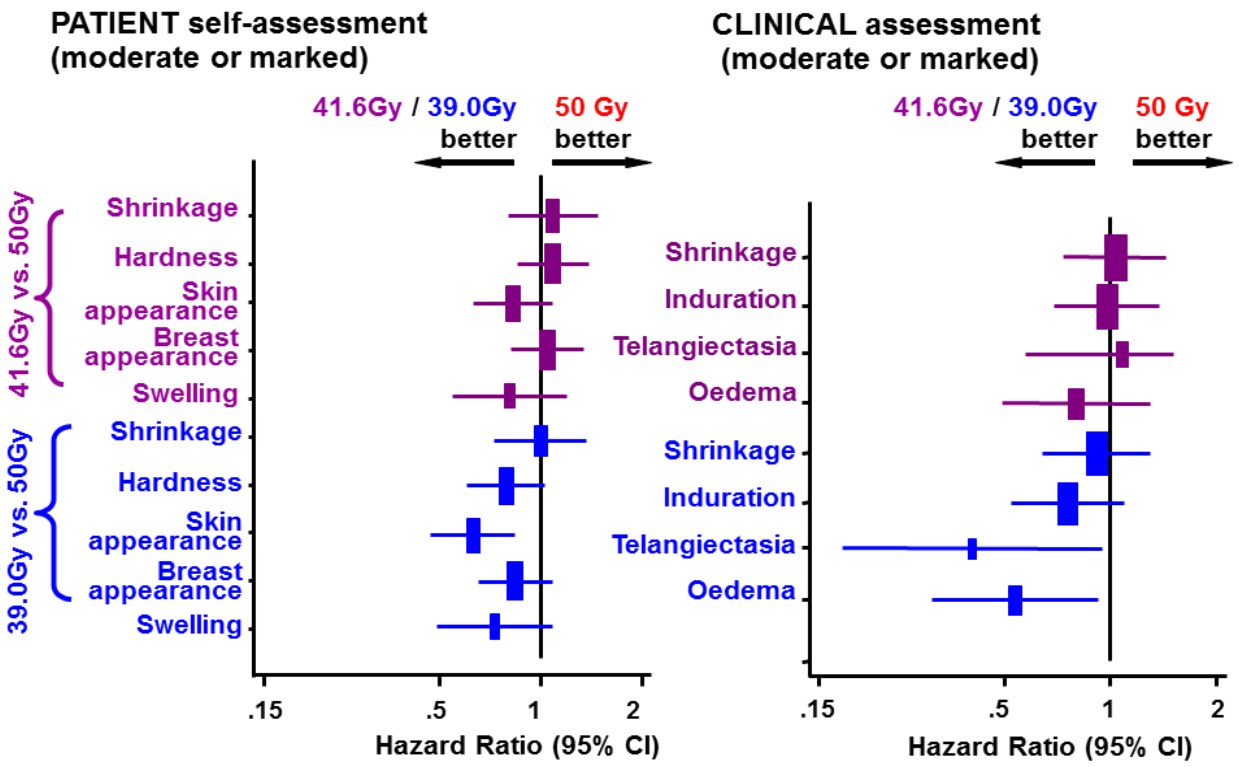

PATIENT self-assessment
(moderate or marked)
41.6Gy / 39.0Gy
better
50 Gy
better
41.6Gy vs. 50Gy
Shrinkage
Hardness
Skin appearance
Breast appearance
Swelling
39.0Gy vs. 50Gy
Shrinkage
Hardness
Skin appearance
Breast appearance
Swelling
.15
.5
1
2
Hazard Ratio (95% CI)
CLINICAL assessment
(moderate or marked)
41.6Gy / 39.0Gy
better
50 Gy
better
Shrinkage
Induration
Telangiectasia
Oedema
Shrinkage
Induration
Telangiectasia
Oedema
.15
.5
1
2
Hazard Ratio (95% CI)

Figure 2

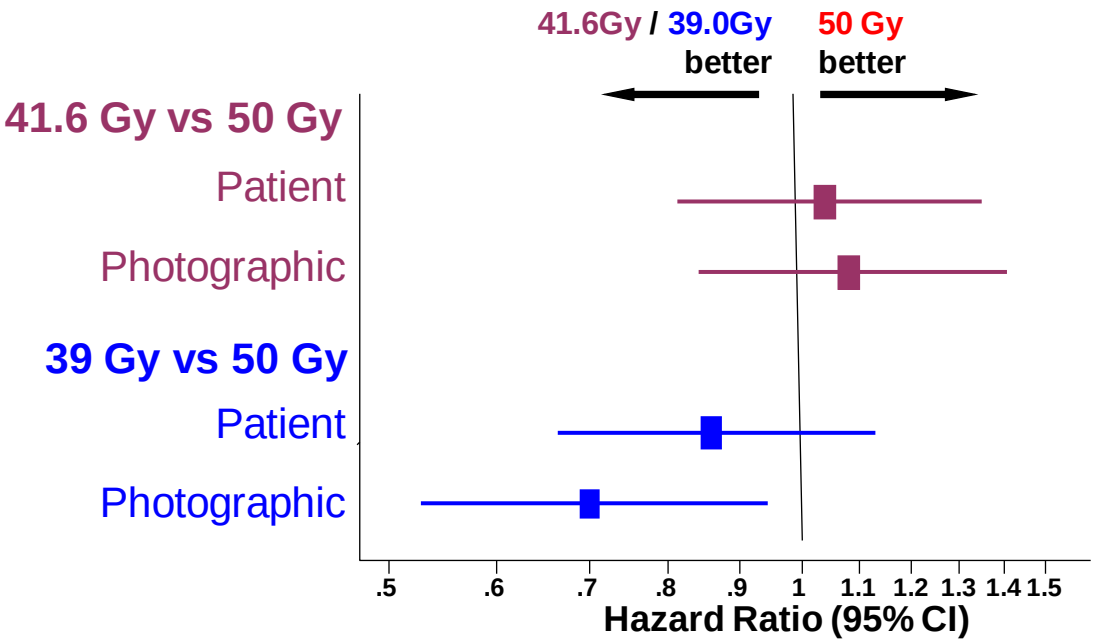

41.6Gy / 39.0Gy
better
50 Gy
better
41.6 Gy vs 50 Gy
Patient
Photographic
39 Gy vs 50 Gy
Patient
Photographic
.5
.6
.7
.8
.9
1
1.1
1.2
1.3
1.4
1.5
Hazard Ratio (95% CI)

Figure 3

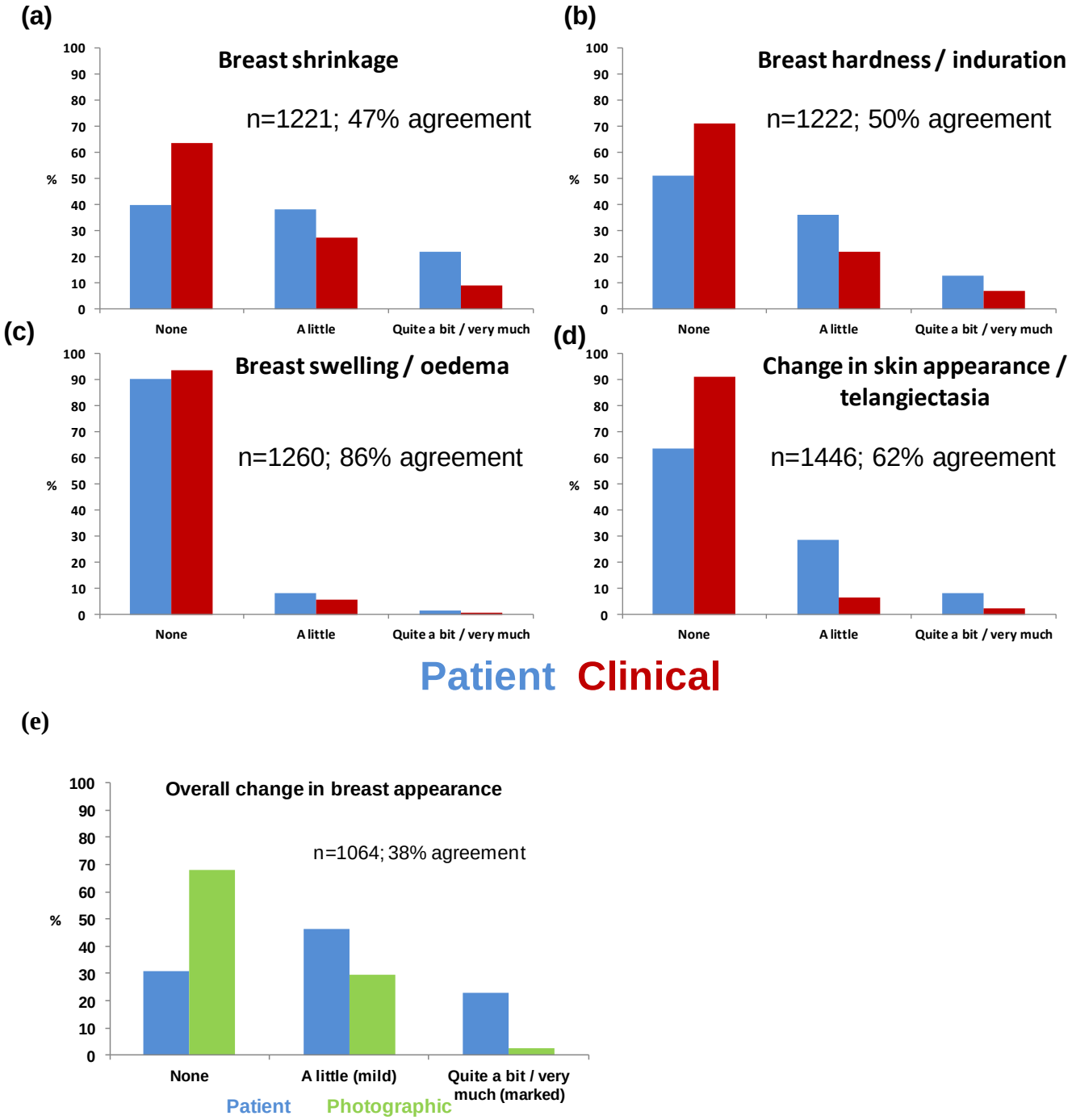

(a)
Breast shrinkage
n=1221; 47% agreement
(b)
Breast hardness / induration
n=1222; 50% agreement
(c)
Breast swelling / oedema
n=1260; 86% agreement
(d)
Change in skin appearance / telangiectasia
n=1446; 62% agreement
%
None
A little
Quite a bit / very much
Patient
Clinical
(e)
Overall change in breast appearance
n=1064; 38% agreement
None
A little (mild)
Quite a bit / very much (marked)
Patient
Photographic

Table 1: Clinical and photographic outcome measures of specific late normal tissue effects in the breast and the corresponding PROM

Clinical assessment of late normal tissue effect in the treated breast	**Corresponding PROM used to test concordance with clinical or photographic assessment**[2]
Has the patient had any of the following adverse effects? Compare with contralateral breast[1]:	
Breast shrinkage	Has your affected breast become smaller as a result of your radiotherapy?[4]
Breast induration	Has your affected breast become harder/firmer to the touch since your radiotherapy?[4]
Breast oedema	During the past four weeks, was the area of your affected breast swollen?[5]
Telangiectasia	Has the appearance of the skin in the area of your affected breast changed since your radiotherapy?[4]
Has there been a change in photographic breast appearance compared with pre-radiotherapy baseline photograph?[3]	Has the overall appearance of your affected breast changed, compared with the other side, as a result of your radiotherapy?[4]

[1] Clinical assessments scored as none, a little, quite a bit, very much
[2] PROMs scored as not at all, a little, quite a bit, very much
[3] Photographic assessments scored as no change, mild change, marked change
[4] Protocol-specified items included in the patient questionnaire booklet under the heading “Since your breast radiotherapy”
[5] Question from the EORTC QLQ-BR23 breast cancer module

Table 2: Baseline characteristics of 1870 START Trial A and B patients with PROMs and clinical assessments of normal tissue effects at 2 and/or 5 years following radiotherapy

	Number of patients (%)
Age (years): mean (SD) [range]	57.0 (10.0) [27.1-86.0]
Type of primary surgery	
Breast conserving surgery	1574 (84.2)
Mastectomy	296 (15.8)
Axillary surgery	
None	55 (2.9)
Axillary clearance	1284 (68.7)
Axillary sampling	495 (26.5)
Sentinel node biopsy	36 (1.9)
Adjuvant chemotherapy	
No	1268 (67.8)
Yes	598 (32.0)
Unknown	4 (0.2)
Tamoxifen	
No	312 (16.7)
Yes	1554 (83.1)
Unknown	4 (0.2)
Breast size	
Small	154 (8.2)
Medium	1126 (60.2)
Large	228 (12.2)
Unknown – not in photographic study	362 (19.4)
Surgical deficit	
Small	872 (46.6)
Medium	496 (26.5)
Large	140 (7.5)
Unknown – not in photographic study	362 (19.4)
Hospital Anxiety and Depression Scale	
Anxiety	
Normal (0-7)	1287 (68.8)
Borderline (8-10)	322 (17.2)
Case (11+)	256 (13.7)
Unknown	5 (0.3)
Depression	
Normal (0-7)	1658 (88.7)
Borderline (8-10)	152 (8.1)
Case (11+)	52 (2.8)
Unknown	8 (0.4)
Body Image Scale (10-items): median (IQR) [range]	3 (0-8) [0-30]

SD = standard deviation; IQR = interquartile range
Breast size and surgical deficit assessed from baseline photographs
HADS scales range from 0-21
Body Image Scale ranges from 0-30, where a higher score indicates more concerns; unknown for 79 patients

Table 3: Concordance between PROMs and clinical or photographic assessments of specific normal tissue effects at 2 and 5 years in START Trials A and B

Clinicians	Patients None	A little	Quite a bit /very much	% agreement (95%CI)	Weighted Kappa (95%CI)	Bowker's test of symmetry, p-value
Breast shrinkage[1] – 2 years				755/1413;	0.21	<0.001
None	566	335	83	53.4%	(0.17-0.25)	
A little	107	158	70	(50.8-56.1%)		
Quite a bit / very much	18	45	31			
Breast shrinkage[1] – 5 years				579/1221;	0.19	<0.001
None	372	277	126	47.4%	(0.15-0.24)	
A little	96	151	87	(44.6-50.3%)		
Quite a bit / very much	18	38	56			
Breast induration / hardness[1] – 2 years				676/1439	0.12	<0.001
None	493	379	136	47.0%	(0.08-0.16)	
A little	112	152	73	(44.4-49.6%)		
Quite a bit / very much	31	32	31			
Breast induration / hardness[1] – 5 years				610/1222;	0.12	<0.001
None	482	295	94	49.9%	(0.07-0.16)	
A little	121	105	40	(47.1-52.8%)		
Quite a bit / very much	22	40	23			
Breast oedema / swelling[1] – 2 years				1144/1465;	0.21	0.017
None	1092	146	21	78.1%	(0.15-0.26)	
A little	109	51	9	(75.9-80.2%)		
Quite a bit / very much	16	20	1			
Breast oedema / swelling[1] – 5 years				1089/1260;	0.10	0.003
None	1076	86	19	86.4%	(0.04-0.17)	
A little	54	13	3	(84.4-88.2%)		
Quite a bit / very much	6	3	0			
Telangiectasia / change in skin appearance[2] – 2 years				959/1721;	0.05	<0.001
None	911	572	134	55.7%	(0.02-0.07)	
A little	32	42	11	(53.3-58.1%)		
Quite a bit / very much	6	7	6			
Telangiectasia / change in skin appearance[2] – 5 years				900/1446;	0.08	<0.001
None	859	369	90	62.2%	(0.04-0.12)	
A little	47	30	16	(59.7-64.7%)		
Quite a bit / very much	13	11	11			
Photographs						
Overall change in breast appearance[1] – 2 years				489/1290;	0.09	<0.001
None	331	525	130	37.9%	(0.06-0.11)	
Mild	56	141	78	(35.3-40.6%)		
Marked	4	8	17			
Overall change in breast appearance[1] – 5 years				409/1064;	0.09	<0.001
None	258	344	123	38.4%	(0.06-0.12)	
Mild	66	140	108	(35.5-41.4%)		
Marked	5	9	11			

CI = confidence interval
[1] breast conserving surgery patients only
[2] breast conserving surgery and mastectomy patients

Illustrations

Web Appendix: Concordance between PROMs and clinical or photographic assessments of specific normal tissue effects at 5 years stratified by baseline patient characteristics in START Trials A and B

	Breast shrinkage[1]		Breast induration/hardness[1]		Breast oedema/swelling[1]		Telangiectasia/change in skin appearance[2]		Overall change in breast appearance[1]	
	% agreement (95%CI)	Weighted Kappa (95%CI)	% agreement (95%CI)	Weighted Kappa (95%CI)	% agreement (95%CI)	Weighted Kappa (95%CI)	% agreement (95%CI)	Weighted Kappa (95%CI)	% agreement (95%CI)	Weighted Kappa (95%CI)
Age										
<50 years	43.7 (37.5-50.0)	0.22 (0.14-0.31)	47.4 (41.2-53.8)	0.09 (0.01-0.17)	N/A	N/A	56.7 (50.9-62.3)	0.06 (0.001-0.12)	37.6 (30.9-44.8)	0.05 (0-0.12)
≥50 years	48.4 (45.2-51.6)	0.20 (0.15-0.25)	50.6 (47.4-53.8)	0.13 (0.08-0.18)	86.9 (84.6-88.9)	0.12 (0.05-0.20)	63.7 (60.9-66.5)	0.09 (0.05-0.14)	39.2 (35.9-42.7)	0.11 (0.07-0.15)
Breast size										
Small	52.8 (43.7-61.8)	0.13 (0-0.26)	59.8 (50.5-68.5)	0.06 (0-0.19)	N/A	N/A	N/A	N/A	41.2 (32.2-50.8)	0.02 (0-0.06)
Medium	48.9 (45.5-52.2)	0.22 (0.17-0.27)	49.8 (46.4-53.1)	0.11 (0.05-0.16)	87.1 (84.6-89.2)	0.06 (0-0.13)	62.9 (59.6-66.1)	0.05 (0.01-0.10)	38.2 (34.7-41.7)	0.08 (0.05-0.12)
Large	37.8 (30.8-45.5)	0.10 (0-0.21)	44.6 (37.2-52.3)	0.10 (0-0.21)	80.5 (73.9-85.8)	0.20 (0.05-0.36)	48.7 (41.3-56.0)	0.07 (0-0.16)	36.9 (29.5-45.0)	0.06 (0-0.17)
Surgical deficit										
Small	50.9 (47.1-54.7)	0.21 (0.15-0.26)	50.1 (46.3-53.8)	0.10 (0.04-0.15)	84.9 (82.1-87.4)	0.06 (0-0.14)	62.6 (58.9-66.2)	0.08 (0.03-0.13)	40.5 (36.7-44.6)	0.10 (0.06-0.14)
Medium	43.9 (38.8-49.1)	0.16 (0.08-0.23)	53.0 (47.7-58.1)	0.20 (0.11-0.29)	90.8 (87.3-93.4)	0.28 (0.12-0.44)	60.8 (55.7-65.8)	0.03 (0-0.10)	36.9 (31.9-42.3)	0.08 (0.01-0.14)
Large	39.3 (30.3-49.0)	0.12 (0-0.24)	40.0 (30.9-49.8)	N/A	82.0 (73.3-88.4)	N/A	60.0 (50.2-69.1)	0.10 (0-0.23)	28.9 (20.3-39.1)	N/A
HADS anxiety										
0-7 (normal)	50.9 (47.5-54.3)	0.22 (0.17-0.27)	52.2 (48.8-55.6)	0.12 (0.07-0.18)	89.0 (86.7-90.9)	0.09 (0-0.17)	65.6 (62.6-68.5)	0.07 (0.03-0.12)	40.3 (36.8-43.9)	0.09 (0.05-0.13)
8-10 (borderline)	43.2 (36.4-50.3)	0.14 (0.05-0.22)	46.1 (39.2-53.2)	0.12 (0.03-0.22)	81.2 (75.2-86.1)	0.23 (0.08-0.38)	56.7 (50.3-63.0)	0.10 (0-0.20)	34.8 (28.0-42.2)	0.08 (0.01-0.15)
≥11 (case)	35.0 (27.7-43.0)	0.13 (0.04-0.22)	42.5 (34.8-50.1)	0.09 (0-0.19)	N/A	N/A	51.1 (43.7-58.4)	0.08 (0-0.17)	32.1 (24.3-40.9)	0.07 (0-0.15)
HADS depression										
0-7 (normal)	48.2 (45.2-51.2)	0.19 (0.15-0.24)	51.4 (48.4-54.4)	0.13 (0.08-0.18)	87.9 (85.8-89.7)	0.13 (0.06-0.21)	64.5 (61.9-67.1)	0.09 (0.04-0.13)	38.7 (35.6-41.9)	0.08 (0.05-0.12)
8-10 (borderline)	40.9 (30.7-51.9)	0.15 (0.01-0.29)	31.8 (22.5-42.7)	N/A	70.0 (59.3-79.0)	N/A	43.2 (34.0-53.0)	0.06 (0.02-0.11)	38.0 (27.5-49.6)	0.13 (0.04-0.23)
≥11 (case)	43.7 (26.8-62.1)	0.26 (0.07-0.46)	46.9 (29.5-65.0)	0.21 (0-0.48)	N/A	N/A	37.5 (23.2-54.2)	0.05 (0-0.19)	25.9 (11.9-46.6)	N/A
Body Image Scale[3]										
0-3	52.2 (48.4-56.1)	0.24 (0.18-0.30)	53.9 (50.1-57.7)	0.14 (0.07-0.20)	88.5 (85.9-90.8)	0.15 (0.05-0.25)	66.2 (62.7-69.9)	0.07 (0.01-0.13)	40.5 (36.6-44.6)	0.09 (0.04-0.14)
>3	41.3 (37.0-45.7)	0.14 (0.08-0.20)	43.6 (39.3-48.1)	0.08 (0.01-0.14)	83.7 (80.2-86.7)	0.05 (0-0.13)	57.8 (53.9-61.7)	0.08 (0.03-0.13)	35.0 (30.6-39.7)	0.09 (0.05-0.13)

CI = confidence interval; N/A = not available

[1] breast conserving surgery patients only

[2] breast conserving surgery and mastectomy patients

[3] 10-item Body Image Scale (possible range 0-30; median baseline score = 3)

Covering Letter

Dr Rajesh Jena
Assistant Editor, Clinical Oncology

11 September 2015

The Institute of Cancer Research

John Yarnold
Professor of Clinical Oncology

The Royal Marsden, Downs Road
Sutton SM2 5PT
T +44 208 661 3388
E john.yarnold@icr.ac.uk

Registered office
The Institute of Cancer Research:
Royal Cancer Hospital
123 Old Brompton Road
London SW7 3RP

A Charity. Not for Profit.
Company Limited by Guarantee.
Registered in England No. 534147.
VAT Registration No. 849 0581 02

Dear Dr Jena

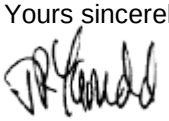
Do patient-reported outcome measures agree with clinical and photographic assessments of normal tissue effects after breast radiotherapy? The experience of the Standardisation of Breast Radiotherapy (START) Trials in early breast cancer

We would be grateful if you would please consider our manuscript for publication in Clinical Oncology, along with a related manuscript from Mukesh Mukesh and Charlotte Coles. We are submitting the manuscripts as a pair as the analyses were done in parallel, and each provides a different perspective on the measurement of normal tissue effects in breast radiotherapy trials. The manuscripts compare the assessments of normal tissue effects carried out within the START and Cambridge IMRT trials, and assess the concordance of patient-reported outcomes with clinical and photographic assessments, on an individual patient level as well as for overall treatment comparisons.

Although there is an extensive literature on the comparison of patient and doctor assessments in general, there is little available on assessments of long-term adverse effects of radiotherapy specifically. Given that the collection of data on late normal tissue effects forms a major part of data collection in radiotherapy trials, it is pertinent to investigate differences and similarities between assessment methods, and to question whether all are strictly necessary in future trials. The START Trials provide a large dataset enabling such concordance analyses, and we believe that our findings will provide valuable information to those designing and interpreting the findings of radiotherapy trials.

Kind regards.

Yours sincerely



John Yarnold
Professor of Clinical Oncology

Highlights

We are not submitting Highlights, and we are therefore not including a separate Highlights file in this submission.

CLINICAL ONCOLOGY

AUTHORSHIP RESPONSIBILITY, FINANCIAL DISCLOSURE & CONTRIBUTORSHIP

Manuscript No:

Do patient-reported outcome measures agree with clinical and photographic assessments of normal tissue effects after breast radiotherapy? The experience of the Standardisation of Breast Radiotherapy (START) Trials in early breast cancer

Corresponding author's name: John Yarnold
Address: The Royal Marsden, Downs Road, Sutton SM2 5PT
Tel: 020 8661 3388 Fax: 020 8661 3107 E-mail: john.yarnold@icr.ac.uk

By signing this form, each author affirms that he/she has read and agrees to all authorship responsibility, contributorship, and financial disclosure/conflict of interest matters.

1. **Authorship responsibility (*to be signed by all named authors at the bottom of this form*)**: My signature certifies that I have participated sufficiently in the conception and design of this work, the analysis of the data (where applicable) as well as the writing of the manuscript to take public responsibility. The manuscript represents valid original work. I attest that that the work described has not been published before (except in the form of an abstract or as part of a published lecture, review or thesis). It is not under consideration elsewhere, nor will it be submitted for publication elsewhere while under consideration by *Clinical Oncology*. I have reviewed this manuscript and approve its submission. The undersigned authors also warrant that this work contains no libellous or unlawful statements, does not infringe on the rights of others, or contains material or instructions that might cause harm or injury.

2. Financial Disclosure/Conflict of Interest: By signing below I certify that any affiliations with or involvement in any organisation with a direct financial interest in the subject matter or materials discussed in the paper are noted below. Otherwise my signature indicates that I have no such financial interest. All funding or project support is acknowledged in the manuscript.

Statement of Financial Interest: (please use additional sheet if necessary)

The authors have no conflicts of interest to disclose.

4. Contributorship: Please list each authors' contribution to the above manuscript. Please provide details such as, the original idea for the research, which individual(s) undertook the collection of data, proportion of patients entered by an individual, responsibility for the first draft of the manuscript, etc.

All authors were involved in the design of the study, analysis and interpretation of the data and contributed to the writing of the report.

Author 1 ___JS Haviland_____
_____
_____

Author 2 ___P Hopwood_____
_____
_____

Author 3 ___J Mills_____
_____
_____

Author 4 ___M Sydenham_____
_____
_____

Author 5 ___JM Bliss_____
_____
_____

Author 6 ___J Yarnold_____
_____
_____

Please also give details of persons (who may not be on the list of authors) who provided statistical advice for the data from the inception of the study and undertook statistical analyses.

-n/a----------

Signatures

Author 1 _____ Date __11.08.2015_____

Author 2 _____ Date ___18.08.2015_____

Author 3 _____ Date ___17.08.2015_____

Author 4 _____ Date ___18.08.2015_____

Author 5 _____ Date ___17/08/2015_____

Author 6 _____ Date __17.08.2015_____

*Conflict of Interest Statement

CLINICAL ONCOLOGY

AUTHORSHIP RESPONSIBILITY, FINANCIAL DISCLOSURE & CONTRIBUTORSHIP

Manuscript No:

Do patient-reported outcome measures agree with clinical and photographic assessments of normal tissue effects after breast radiotherapy? The experience of the Standardisation of Breast Radiotherapy (START) Trials in early breast cancer

Corresponding author's name: John Yarnold
Address: The Royal Marsden, Downs Road, Sutton SM2 5PT
Tel: 020 8661 3388 Fax: 020 8661 3107 E-mail: john.yarnold@icr.ac.uk

By signing this form, each author affirms that he/she has read and agrees to all authorship responsibility, contributorship, and financial disclosure/conflict of interest matters.

1. **Authorship responsibility (*to be signed by all named authors at the bottom of this form*)**: My signature certifies that I have participated sufficiently in the conception and design of this work, the analysis of the data (where applicable) as well as the writing of the manuscript to take public responsibility. The manuscript represents valid original work. I attest that that the work described has not been published before (except in the form of an abstract or as part of a published lecture, review or thesis). It is not under consideration elsewhere, nor will it be submitted for publication elsewhere while under consideration by *Clinical Oncology*. I have reviewed this manuscript and approve its submission. The undersigned authors also warrant that this work contains no libellous or unlawful statements, does not infringe on the rights of others, or contains material or instructions that might cause harm or injury.

2. Financial Disclosure/Conflict of Interest: By signing below I certify that any affiliations with or involvement in any organisation with a direct financial interest in the subject matter or materials discussed in the paper are noted below. Otherwise my signature indicates that I have no such financial interest. All funding or project support is acknowledged in the manuscript.

Statement of Financial Interest: (please use additional sheet if necessary)

The authors have no conflicts of interest to disclose.

4. Contributorship: Please list each authors' contribution to the above manuscript. Please provide details such as, the original idea for the research, which individual(s) undertook the collection of data, proportion of patients entered by an individual, responsibility for the first draft of the manuscript, etc.

All authors were involved in the design of the study, analysis and interpretation of the data and contributed to the writing of the report.

Author 1 ___JS Haviland_____
_____
_____

Author 2 ___P Hopwood_____
_____
_____

Author 3 ___J Mills_____
_____
_____

Author 4 ___M Sydenham_____
_____
_____

Author 5 ___JM Bliss_____
_____
_____

Author 6 ___J Yarnold_____
_____
_____

Please also give details of persons (who may not be on the list of authors) who provided statistical advice for the data from the inception of the study and undertook statistical analyses.

-n/a----------

Signatures

Author 1 _____ Date __11.08.2015_____

Author 2 _____ Date ___18.08.2015_____

Author 3 _____ Date ___17.08.2015_____

Author 4 _____ Date ___18.08.2015_____

Author 5 _____ Date ___17/08/2015_____

Author 6 _____ Date __17.08.2015_____