Clinical Oncology xxx (2016) 1–9

Contents lists available at ScienceDirect

Clinical Oncology

journal homepage: www.clinicaloncologyonline.net

Original Article

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

J.S. Haviland [*†], P. Hopwood [†], J. Mills [†], M. Sydenham [†], J.M. Bliss [†1], J.R. Yarnold [‡1] on behalf of the START Trialists' Group

[*] Faculty of Health Sciences, University of Southampton, Southampton, UK
[†] ICR-CTSU, Division of Clinical Studies, The Institute of Cancer Research, London, UK
[‡] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK

Received 11 September 2015; received in revised form 1 December 2015; accepted 15 December 2015

Abstract

Aims: In radiotherapy trials, normal tissue effects (NTE) are important end points and it is pertinent to ask whether patient-reported outcome measures (PROMs) could replace clinical and/or photographic assessments. Data from the Standardisation of Breast Radiotherapy (START) 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 after 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 a higher prevalence for all NTE end points 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). The percentage agreement was lowest between PROMs and photographic assessments of change in breast appearance (38%).
Conclusions: All three 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.

Key words: 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 a 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 blind 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 multicentre 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

Author for correspondence: J.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.
E-mail address: john.yarnold@icr.ac.uk (J.R. Yarnold).
[1] Joint last authors.

http://dx.doi.org/10.1016/j.clon.2016.01.011

with a growing interest in the use of patient-reported outcome measures (PROMS) in routine follow-up [9]. With an increasing use of 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 Standardisation of Breast Radiotherapy (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 the 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 $\geq$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 after radiotherapy. Clinical assessments of NTE were collected at annual follow-up in all patients, and photographs were taken under standard conditions at a post-surgical pre-radiotherapy baseline and at 2 and 5 years after randomisation for patients who had breast-conserving surgery. The patient questionnaires included the (i) European Organization for Research and Treatment of Cancer (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) four 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 four-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 four-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 three 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 after 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 three-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 to 5 years for the PROMs, from 1 to 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 the date of randomisation, and survival analysis methods used to compare radiotherapy schedules. Hazard ratios for the relative effects of the radiotherapy schedules in START-A were calculated for each NTE end point 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 end points 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. Hazard ratios 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 end points there were few patients in the highest grade category, so moderate and marked categories were combined, resulting in three-point

Please cite this article in press as: Haviland JS, et al., 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, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.01.011

Table 1
Clinical and photographic outcome measures of specific late normal tissue effects in the breast and the corresponding patient-reported outcome measure (PROM)

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

* Clinical assessments scored as none, a little, quite a bit, very much.
† Photographic assessments scored as no change, mild change, marked change.
‡ PROMs scored as not at all, a little, quite a bit, very much.
§ Protocol-specified items included in the patient questionnaire booklet under the heading 'Since your breast radiotherapy'.
¶ Question from the EORTC QLQ-BR23 breast cancer module.

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 three-point scale. Corresponding NTE end points 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), weighted kappa statistic (with 95% confidence interval) 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 the 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 the control (50 Gy in 25 fractions) in START-A. Comparing hazard ratios for corresponding end points, 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% confidence interval 0.7–5.1 Gy) for PROMs and 2.6 Gy (95% confidence interval 1.3–3.9 Gy) for photographic assessments. α/β estimates for individual NTE end points 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 (Figure 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

Please cite this article in press as: Haviland JS, et al., 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, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.01.011

Table 2
Baseline characteristics of 1870 Standardisation of Breast Radiotherapy (START) Trial A and B patients with patient-reported outcome measures (PROMs) and clinical assessments of normal tissue effects at 2 and/or 5 years after radiotherapy

	Number of patients (%)
Age (years): mean (standard deviation) [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 (interquartile range) [range]	3 (0–8) [0–30]

* Breast size and surgical deficit assessed from baseline photographs.
† Hospital Anxiety and Depression Scale ranges from 0 to 21.
‡ Body Image Scale ranges from 0 to 30, where a higher score indicates more concerns; unknown for 79 patients.

end points, 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 seemed to be no substantial differences in degree of concordance for individual NTE end points 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 seem to be unaffected by patient factors, including age, breast size, surgical deficit, baseline Hospital Anxiety and Depression Scale anxiety and depression and body image scores (see 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. The percentage agreement between PROMs and clinical assessments of specific NTEs was around 50%, indicating that in only half 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 seem 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. [23,24,26–30]). 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 carried out in a single centre, with clinical ratings carried out by one person [31]. Others have shown a more favourable rating of overall cosmesis after 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

Please cite this article in press as: Haviland JS, et al., 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, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.01.011

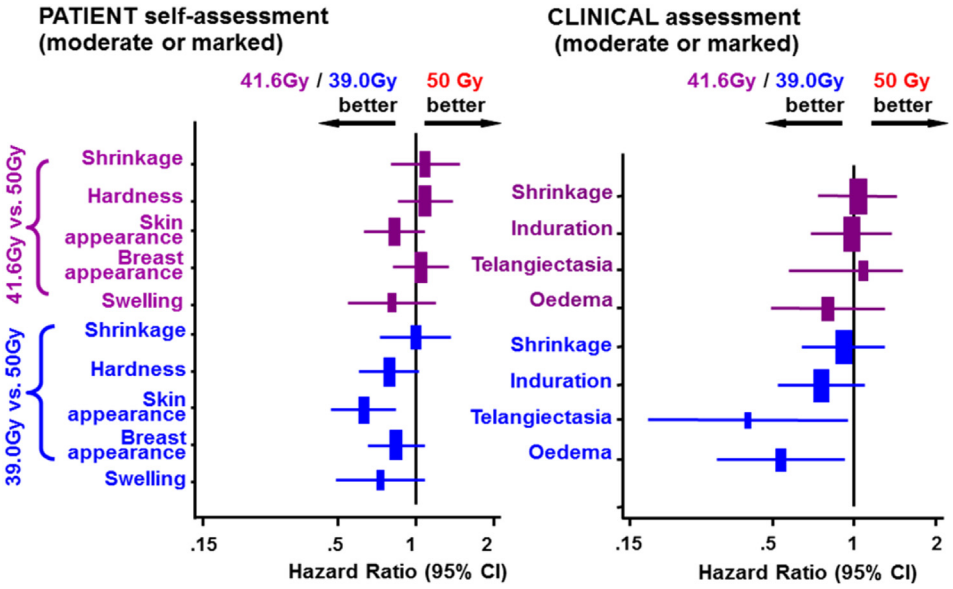

Fig 1. Comparisons between randomised radiotherapy schedules in Standardisation of Breast Radiotherapy (START) Trial A for patient-reported outcome measures (PROMs) and clinical assessments of specific normal tissue effects.

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 probably 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 effect on different aspects of quality of life [6,16].

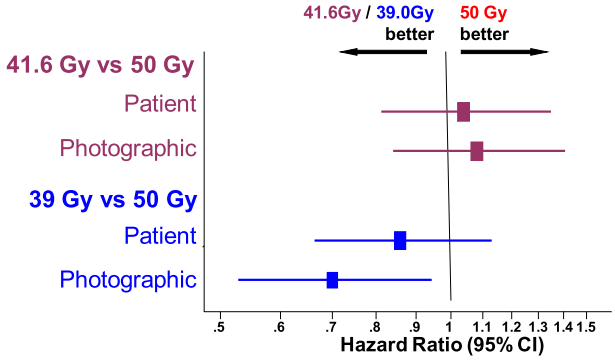

Fig 2. Comparisons between randomised radiotherapy schedules in Standardisation of Breast Radiotherapy (START) Trial A for patient-reported outcome measures (PROMs) and photographic assessments of overall change in breast appearance.

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 seems 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

Please cite this article in press as: Haviland JS, et al., 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, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.01.011

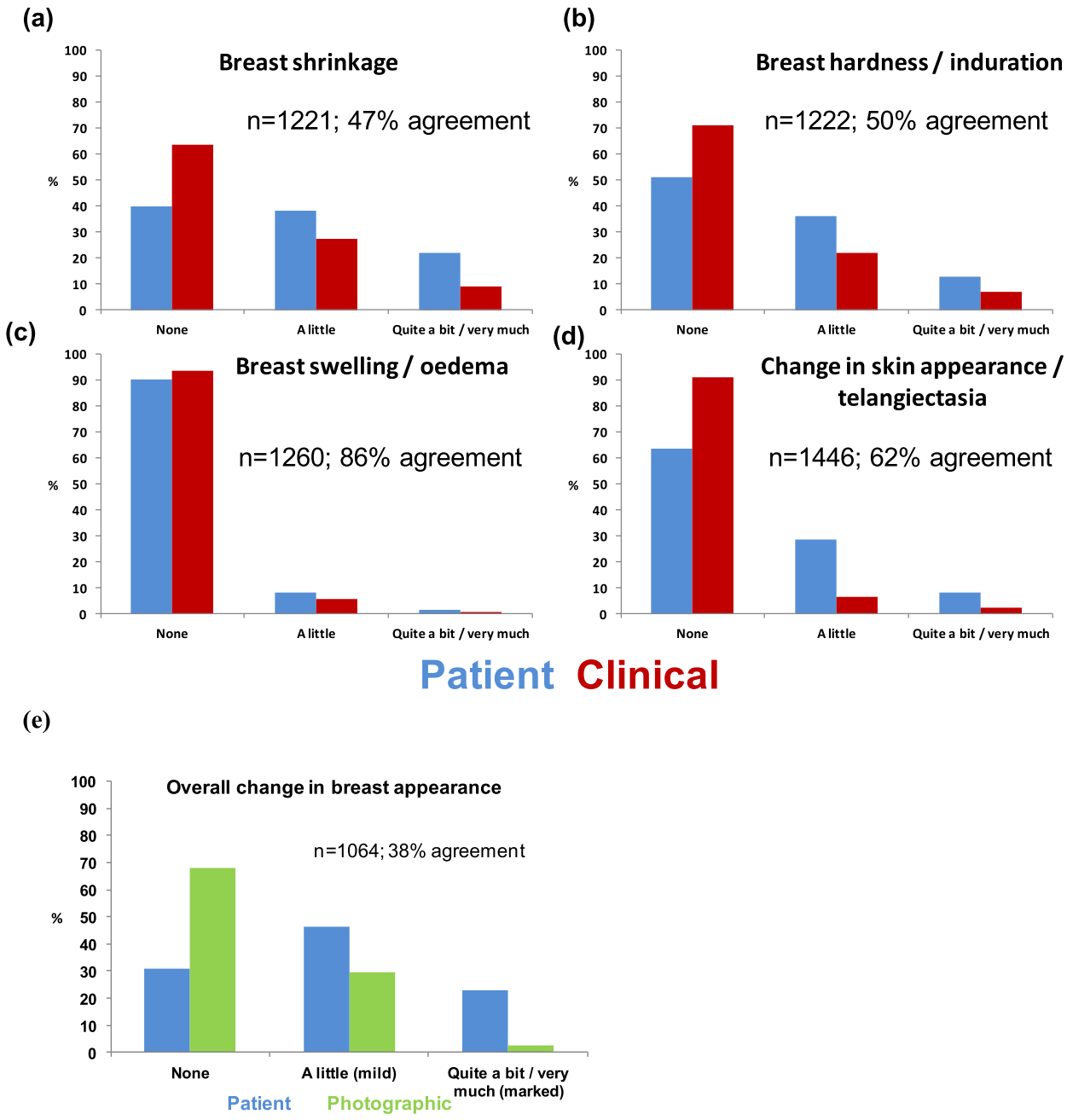

Fig 3. Comparison of 5 year patient-reported outcome measures (PROMs), clinical and photographic assessments of specific normal tissue effects in Standardisation of Breast Radiotherapy (START) Trials A and B.

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 multimodal 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]. Furthermore, 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

Please cite this article in press as: Haviland JS, et al., 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, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.01.011

Table 3
Concordance between patient-reported outcome measures (PROMs) and clinical or photographic assessments of specific normal tissue effects at 2 and 5 years in Standardisation of Breast Radiotherapy (START) Trials A and B

Clinicians	Patients			% agreement (95% confidence interval)	Weighted kappa (95% confidence interval)	Bowker's test of symmetry, *P*-value
	None	A little	Quite a bit /very much			
Breast shrinkage* – 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* – 5 years				579/1221; 47.4%	0.19	<0.001
None	372	277	126	(44.6–50.3%)	(0.15–0.24)	
A little	96	151	87			
Quite a bit/very much	18	38	56			
Breast induration/hardness* – 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* – 5 years				610/1222; 49.9%	0.12	<0.001
None	482	295	94	(47.1–52.8%)	(0.07–0.16)	
A little	121	105	40			
Quite a bit/very much	22	40	23			
Breast oedema/swelling* – 2 years				1144/1465; 78.1%	0.21	0.017
None	1092	146	21	(75.9–80.2%)	(0.15–0.26)	
A little	109	51	9			
Quite a bit/very much	16	20	1			
Breast oedema/swelling* – 5 years				1089/1260; 86.4%	0.10	0.003
None	1076	86	19	(84.4–88.2%)	(0.04–0.17)	
A little	54	13	3			
Quite a bit/very much	6	3	0			
Telangiectasia/change in skin appearance† – 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† – 5 years				900/1446; 62.2%	0.08	<0.001
None	859	369	90	(59.7–64.7%)	(0.04–0.12)	
A little	47	30	16			
Quite a bit/very much	13	11	11			
Photographs						
Overall change in breast appearance* – 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* – 5 years				409/1064; 38.4%	0.09	<0.001
None	258	344	123	(35.5–41.4%)	(0.06–0.12)	
Mild	66	140	108			
Marked	5	9	11			

* Breast-conserving surgery patients only.
† Breast-conserving surgery and mastectomy patients.

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 NTEs 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

Please cite this article in press as: Haviland JS, et al., 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, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.01.011

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 is not 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.

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. The START trials were funded by Cancer Research UK (CRUK/96/001), Medical Research Council and the UK Department of Health (grant G9600656). The funders had no role in study design or conduct, analysis or interpretation of data.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.clon.2016.01.011.

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, *et al*. EORTC Late Effects Working Group. Late effects toxicity scoring: the SOMA scale. *Radiother Oncol* 1995;35:11–15.

[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;31:1041–1042.

[5] Haviland JS, Ashton A, Broad B, *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, *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;85:365–376.

[7] Brady MJ, Cella DF, Mo F, *et al*. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. *J Clin Oncol* 1997;15:974–986.

[8] Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35:1095–1108.

[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;27:679–685.

[10] Basch E, Abernethy AP, Mullins CD, *et al*. Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. *J Clin Oncol* 2012;30:4249–4255.

[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–S75.

[12] Bentzen SM, Agrawal RK, Aird EG, *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;9:331–341.

[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;11:231–240.

[14] Haviland JS, Owen JR, Dewar JA, *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;14:1086–1094.

[15] Bentzen SM, Agrawal RK, Aird EG, *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;371:1098–1107.

[16] Sprangers MA, Groenvold M, Arraras JI, *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;14:2756–2768.

[17] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–370.

[18] Hopwood P, Fletcher I, Lee A, Al Ghazal S. A body image scale for use with cancer patients. *Eur J Cancer* 2001;37:189–197.

[19] Bowker AH. A test for symmetry in contingency tables. *J Am Stat Assoc* 1948;43:572–574.

Please cite this article in press as: Haviland JS, et al., 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, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.01.011

[20] Altman DG. *Practical statistics for medical research*. London: Chapman and Hall; 1991.
[21] Davidson SE, Trotti A, Ataman OU, *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;69:1218–1221.
[22] Vistad I, Cvancarova M, Fossa SD, Kristensen GB. Post-radiotherapy 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;71:1335–1342.
[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;22:3485–3490.
[24] Velikova G, Wright P, Smith AB, *et al*. Self-reported quality of life of individual cancer patients: concordance of results with disease course and medical records. *J Clin Oncol* 2001;19:2064–2073.
[25] Gotay C. Patient symptoms and clinician toxicity ratings: both have a role in cancer care. *J Natl Cancer Inst* 2009;101:1602–1603.
[26] Basch E, Jia X, Heller G, *et al*. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J Natl Cancer Inst* 2009;101:1624–1632.
[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. p. 139–144.
[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;103:185–195.
[29] Quinten C, Maringwa J, Gotay CC, *et al*. Patient self-reports of symptoms and clinician ratings as predictors of overall cancer survival. *J Natl Cancer Inst* 2011;103:1851–1858.
[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? *Qual Life Res* 1997;6:225–236.
[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;45:225–228.
[33] Hoeller U, Kuhlmey A, Bajrovic A, *et al*. Cosmesis from the patient's and the doctor's view. *Int J Radiat Oncol Biol Phys* 2003;57:345–354.
[34] Kirchheiner K, Nout R, Lindegaard J, *et al*. Do clinicians and patients agree regarding symptoms? A comparison after definitive radiochemotherapy in 223 uterine cervical cancer patients. *Strahlenther Onkol* 2012;188:933–939.
[35] Bruner DW, Hanisch LJ, Reeve BB, *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). *Trans Behav Med* 2011;1:110–122.
[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, *et al*. Conducting radiogenomic research – do not forget careful consideration of the clinical data. *Radiother Oncol* 2012;105:337–340.

Please cite this article in press as: Haviland JS, et al., 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, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.01.011