

Months and Severity Score (MOSES) in a Phase III trial (PARCER): A new comprehensive method for reporting adverse events in oncology clinical trials

Nilesh Ranjan,^{a,1} Supriya Chopra,^{b,1*} Akshay Mangaj,^a Pallavi Rane,^c Mayuri Charnalia,^b Sadhana Kannan,^c Tapas Dora,^a Reena Engineer,^a Umesh Mahantshetty,^a Lavanya Gurram,^a Prachi Mittal,^a Jaya Ghosh,^f Amita Maheshwari,^d TS Shylasree,^d Sudeep Gupta,^e and SK Shrivastava^a

^aDepartment of Radiation Oncology, Tata Memorial Hospital, Homi Bhabha National Institute, Mumbai, Maharashtra, India

^bDepartment of Radiation Oncology, ACTREC, Tata Memorial Centre, Homi Bhabha National Institute, Navi Mumbai, Maharashtra, India

^cDepartment of Epidemiology and Clinical Trials Unit, ACTREC, Tata Memorial Centre, Homi Bhabha National Institute, Navi Mumbai, Maharashtra, India

^dDepartment of Gynae Surgical Oncology, Tata Memorial Hospital, Homi Bhabha National Institute, Mumbai, Maharashtra, India

^eDepartment of Medical Oncology, ACTREC, Tata Memorial Centre, Homi Bhabha National Institute, Navi Mumbai, Maharashtra, India

^fDepartment of Medical Oncology, Tata Memorial Hospital, Homi Bhabha National Institute, Mumbai, Maharashtra, India

Summary

Background Adverse event reporting in oncology trials lacks temporal description. We propose a toxicity summarizing method that incorporates time.

Methods Patients recruited in a phase III trial (NCT01279135) that compared three-dimensional conformal radiotherapy (3DCRT) and intensity modulated radiotherapy (IMRT) for late toxicity in cervical cancer were included. Adverse events were reported using Common Terminology Criteria for Adverse Events (CTCAE) v3.0 and quality of life (QoL) with EORTC QLQ-C30 and CX24. A total of six symptoms with a related QoL question (diarrhoea, abdominal pain, anorexia, urinary incontinence, frequency and fatigue) were included. Month and severity score [MOSES = \sum (CTCAE grade x proportionate time)] was calculated. Cumulative-MOSES (C-MOSES) was calculated by summing these 6 individual MOSES. QoL was categorized as "substantially symptomatic" or "not". Receiver operator curve analysis was performed to determine the MOSES cut off that predicts for substantial QoL symptoms. CTCAE and MOSES were tested for accurately categorizing QoL impact.

Findings In the construction dataset, 201/300 patients had symptoms. MOSES > 0.20 had higher accuracy than CTCAE for predicting impact on QoL related to diarrhoea (85% vs. 69%), anorexia (61% vs 51%), abdominal pain (71% vs. 57%), urinary incontinence (72% vs. 61%) and frequency (62% vs. 59%). C-MOSES > 0.70 correlated with reduction in role functioning and global QoL. While no difference was seen in CTCAE grade ≥ 1 Gastrointestinal (GI) toxicity between 3DCRT or IMRT arm, 3DCRT had higher C-MOSES than IMRT (HR=0.64; 95% CI 0.41–0.99, $p = 0.04$).

Interpretation MOSES has higher accuracy than CTCAE in categorizing symptom specific and functional QoL. These results require further external validation.

Funding None.

Copyright © 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: MOSES; Late toxicity; Quality of life; CTCAE; Cervix cancer

*Corresponding author.

E-mail addresses: schopra@actrec.gov.in (N. Ranjan), schopra@actrec.gov.in (S. Chopra).

¹ Joint First Authors

Introduction

The last two decades have witnessed a rapid evolution of radiation technology, systemic therapies including targeted agents and immunotherapy, and increasing

eClinicalMedicine

2022;47: 101390

Published online xxx

<https://doi.org/10.1016/j.eclinm.2022.101390>

eclinm.2022.101390

Research in Context

Evidence before this study

We searched MEDLINE, Embase, the JBI Evidence-based Practice Database, and the Cochrane Database of Systematic Reviews from 1990 to 2021. Search terms included [MeSH Terms] "late toxicity reporting", "summarizing late toxicity", "toxicity in oncology trials" and "alternative toxicity reporting methods". We observed that late toxicity reporting in most of the oncology trials was based on reporting the moderate to severe grade of toxicity in different organ systems with limited information on evolution of adverse effects and their temporal course. Furthermore, lower grades of toxicity (persistent or otherwise) are seldom reported. Adverse event reporting systems like TAME, Tox-T and LAPERS have been developed and used for a more comprehensive description sometime also reflecting impact on QOL.

Added value of this study

We developed a new method of summarizing physician-reported toxicity to predict patient-reported QOL. MOSES allows for summation of time weighted toxicity scores across organ systems leading to a cumulative numerical toxicity burden score (C-MOSES). New scores like these, which can summate multisystem events, also have the possibility of providing a more comprehensive differentiation of treatment interventions and can possibly be complementary to CTCAE.

Implications of all the available evidence

MOSES incorporates the dimension of time, the multiplicity of events and provides weightage to persistent lower grades of toxicity and hence, provides a more complete, longitudinal numerical depiction of adverse events than CTCAE including the possibility to summate overall symptom burden into a single score. The initial methodological development and internal validation has been performed. However, further external validation is needed to further test its robustness and performance.

combination treatments.¹⁻⁷ Also, in recent years, there has been an increased focus on radiation dose escalation and an increase in the use of abbreviated or hypo-fractionated schedules.⁸⁻¹⁰ Strategies to improve disease outcomes or reduce the treatment duration are likely to be associated with an increase in treatment-related toxicity burden. When different treatment options are available, the choice of treatment is based on the intervention effect that also includes risk-benefit (or therapeutic) ratio. Therefore, it is important to summarize overall adverse events to represent the overall morbidity that any treatment entails. While most randomized trials report adverse events using common terminology criteria for adverse events (CTCAE), it's

often summarized as the maximum grade of toxicity across an organ system. It does not take into consideration the multiplicity of events within an organ system and the temporal course of events. This is important for adverse events that persist and could therefore impact the patients long term quality of life (QOL). While late adverse events are observed with many oncological interventions their occurrence is best reported in patients receiving radiation treatment. However, currently, there is no recommended method to account for persistence or change in the severity of toxicity during long term follow up. Also, when patients experience multi-organ symptoms (e.g. diarrhoea and urinary incontinence) or multiple symptoms within an organ system (e.g. nausea, diarrhoea and bleeding per rectum), there is no straightforward way of summarizing the combined impact of adverse events. Another important aspect of adverse event reporting is the need to incorporate patient-reported outcome measures in the final adverse events summary. Multiple studies have demonstrated the discordance between physician and patient-reported events.^{8,9} The current methods of QOL reporting focus on reporting QOL as "before-after" or at "predefined multiple time point based analysis" for the entire cohort.⁸⁻¹² While the existing clinical reporting methods may provide crude incidence and snapshots of toxicity related endpoints, comprehensive adverse event reporting that takes temporal trends of multi-organ systems into account is desirable. This may possibly be more representative of the patient's QOL and help in better understanding of impact of any treatment intervention on morbidity.

Innovative methods for summarizing toxicity have also been reported in the past. TAME method¹³ integrates time and multiplicity with the severity of toxicity. It summarizes adverse-event data by assigning three different scores- T-score for acute toxicity (T), A-score for late toxicity (A), M score for mortality risk (M) due to toxicity, and finally combining all of them to generate results (E). TAME method concluded that CTCAE maximum grade method excluded 29–70% of acute adverse events. There was a 500% increase in the relative toxicity burden when different treatment regimens were compared, which standard methods of reporting did not encompass. The Toxicity over Time (ToxT)¹⁴ approach used area under the curve (AUC) analyses to summarize adverse event profiles over the entire course of a study.⁸ However, limitations of this methodology were statistical complexity, heterogeneity, and absence of interpretability. More recently, for summarizing late toxicity in gynaecological cancers LAPERS¹⁵ (Late, persistent, substantial, treatment-related symptoms after radiotherapy) system was developed. LAPERS is a binary scoring system that categorizes patients as LAPERS or non-LAPERS depending upon the deterioration of symptoms from baseline. It is assessed in each patient and for each symptom separately and can be

used for both physicians or patients reported symptoms. It takes the duration of symptoms into account by defining the persistence of substantial symptoms in half of the follow-ups and therefore rare adverse events. Limitations of LAPERS include the binary categorization, and as may be expected, all LAPERS or non-LAPERS patients do not have the same burden of toxicity. Also, it is limited in reflecting the lower grade persistent, isolated severe symptoms and temporal course of toxicity.

To overcome the limitations of the above toxicity reporting systems, we developed the **month** and **severity** score (MOSES) system that imputes time weightage to the CTCAE score. We hypothesize that such a method of summarizing adverse events in addition to CTCAE could predict with higher accuracy the impact of therapeutic intervention on a patient's QOL. In this paper, we report the methodological development and initial validation of MOSES.

Methods

Study design

For the purpose of the study patients recruited into the PARCER trial, a phase III RCT of postoperative image-guided intensity-modulated radiotherapy (IG-IMRT) vs three-dimensional conformal radiation therapy (3D-CRT) for late toxicity reduction (NCT01279135) were included.⁴ The primary endpoint of the PARCER trial was physician reported grade ≥ 2 late gastrointestinal (GI) toxicity.⁴ A total of 300 patients were randomized to 3D-CRT and IG-IMRT. Patients received after surgery 50 Gy/25#/5weeks along with concurrent weekly cisplatin (40 mg/m²) followed by 2 # of high dose rate brachytherapy of 6 Gy each. Prospectively data on acute and late adverse events were recorded for GI, Genitourinary (GU) and haematolymphoid system using CTCAE reporting. In addition, lymphedema, fatigue and constitutional symptoms were assessed at time points of pre-treatment, during radiation treatment and on each scheduled follow-up (every 3 monthly for the first 2 years and 6 monthly thereafter). GI adverse event assessment included diarrhoea, perforation, obstruction, distension, anorexia, malabsorption, nausea, vomiting, necrosis, ulcer, haemorrhage, pain and constipation. GU toxicities included cystitis, urinary fistula, frequency, incontinence and bladder spasm. EORTC QLQ-C30 and QLQ-CX24 modules were used for QOL assessment at corresponding time points. The final study analyses that used CTCAE maximum grade method demonstrated the lower late GI grade ≥ 2 toxicity with IG-IMRT and were recently reported.⁴

Data collection

Institutional ethics committee approval was obtained to access the study database for development of MOSES

system. Patients consented for participation in the index trial and provided permission of use of toxicity data. Additional consent was not needed for this study. The methodological development did not employ intervention attribution and included all patients with adverse events irrespective of allocated intervention in the index trial. Out of all the available adverse events, we selected a total of six different late toxicities - three GI toxicities (diarrhoea, abdominal pain and anorexia), two GU toxicities (urinary incontinence and urinary frequency) and fatigue (Figure 1). These also constituted the most common late adverse events and had a corresponding QOL question in EORTC QLQ-C30 or QLQ-CX24.

For inclusion in the present study, patients should have been symptomatic either on physician assessment (CTCAE grade >0) or on patient-reported assessment (QOL) that was performed > 90 days of treatment completion.¹¹ The last date of adverse events data collection for this study corresponded to the most recent clinical follow up in the absence of disease relapse. In patients with disease relapse, the adverse event and QOL data at the time of relapse was excluded therefore making adverse event and QOL available only until the preceding follow up (at least 90 days before disease relapse). This was done to ensure that disease and adverse events symptoms do not overlap for assessment. Patients who had no adverse events on either physician or patient-reported assessment or had clinical follow up of <12 months or disease relapse before 12 months were excluded. For the present study, the maximum CTCAE grade for GI, GU and fatigue subscales was recorded from the trial database.

MOSES calculation

In the proposed methodology, two types of scores were allocated to each adverse event-proportional time score (P) for the duration of time spent in a particular grade of adverse event (proportion of the number of months of follow up) and severity score (S) for the grade of toxicity (CTCAE grade). For allocating a score to an individual patient, the total duration of follow up was first calculated in months, and then proportional time score, i.e. the proportion of duration spent in different grades of toxicity, was calculated. For each proportional duration of follow up, a severity score (S) was assigned for each symptom based on the corresponding CTCAE grade in the database. The final score was then calculated as $\Sigma (P \times S)$. For example, a patient who spends 12 out of 24 months in Grade 1 toxicity and 12 months without any toxicity will have a proportionate score of $12/24 = 0.50$ and severity score of 1 corresponding to CTCAE grade 1, which will assign a final score of $0.50 \times 1 = 0.50$. Another patient who spends two out of 24 months in Grade 3, two out of 24 months in grade 2 and one out of 24 months in grade 1 toxicity will have a final score of $(2/24 \times 3 + 2/24 \times 2 + 1/24 \times 1) = 0.45$.

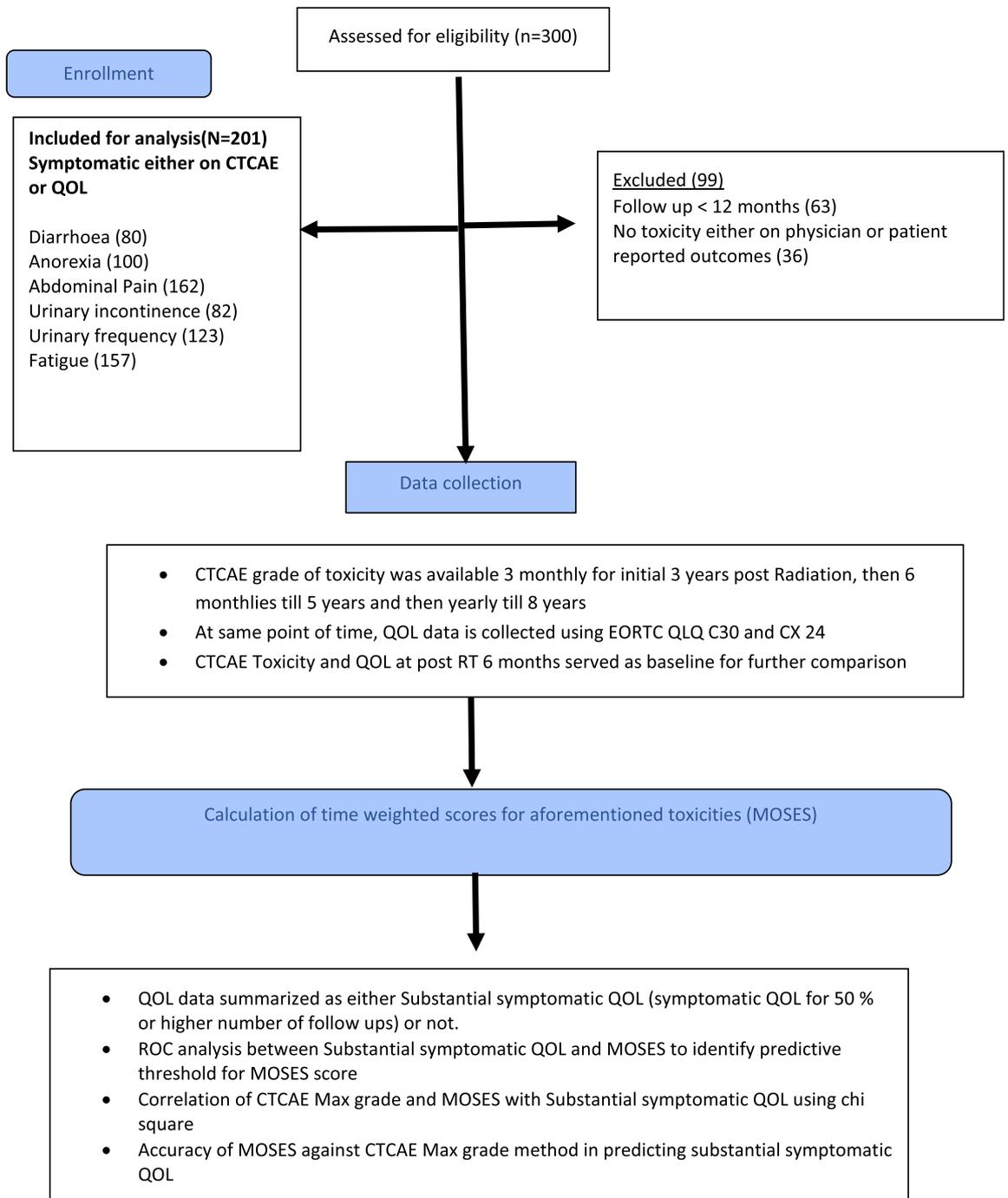


Figure 1. Flow diagram describing selection of patients and process for the development of MOSES reporting system.

Examples of calculation of MOSES is summarised in supplementary Table 1. Using similar weightage for each of the six toxicities, a cumulative MOSES (C-MOSES) score was calculated by summing MOSES of individual adverse event items [C-MOSES= MOSES (Diarrhoea+ Anorexia+ Pain+ Urinary Frequency+ Urinary Incontinence+ Fatigue)].

QOL scoring

QOL scoring was performed as per the standard recommendations using the corresponding questions from EORTC QLQ-C30 and QLQ-CX24 questionnaires for specific toxicities.¹¹ Higher scores in the functional domain suggested a higher level of functioning, and a higher score in the symptom’s domain suggest more

symptoms. Traditional QOL scoring evaluates symptom changes over time, and patients QOL at different predefined time points is assessed to understand the impact of an intervention.^{11,12} However, in between these predefined time-points, there could be an improvement or deterioration in QOL. Therefore, to evaluate the performance of MOSES, we categorized QOL as “substantially symptomatic on QOL reporting” and “insignificant change in QOL”, where substantially symptomatic on QOL reporting was defined as “symptomatic with any score for at least 50% of follow-ups.”

For the purpose of this study, missing CTCAE and QOL data was considered as missing at random. The missing data was imputed only if the duration of missing consecutive follow-ups was ≤ 12 months. In such a case the scores of preceding and subsequent follow up was assumed for each half of the missing follow up. If follow up was not attended for >12 consecutive months, data collection was censored at last such time point of consecutive follow up to avoid over imputation. For example, if patient attended 36 months of follow up and then directly came after 24 months for regular follow up. In such a case only 36 months of CTCAE and QOL data was considered for this study.

Analysis of intervention effects using CTCAE and MOSES

Though the MOSES development did not assume intervention effect or attribution, we reanalysed the gastrointestinal adverse events index trial (PARCER) to internally validate MOSES within the construction data set. The PARCER trial⁴ was designed to determine difference in late GI grade ≥ 2 toxicities between 3DCRT and IG-IMRT. As MOSES development included all grades of toxicity therefore for this exploratory validation, we used all grades of CTCAE GI toxicity therefore categorizing CTCAE as grade ≥ 1 vs. not and GI C-MOSES as ≥ 0.70 and < 0.70 .

Statistical analysis

Considering patient-reported QOL as the gold standard, both MOSES and CTCAE system were tested for identifying patients who were “substantially symptomatic on QOL reporting”. With MOSES as the “test factor” and $> 50\%$ symptomatic QOL status as the “event of interest”, the cut-off score of MOSES to predict deterioration in QOL was assessed using receiver operator characteristics (ROC) analysis. All patients, including those with MOSES score of zero (i.e. no physician-reported adverse events but with patient-reported symptoms on QOL), were included. A cut off was selected that provided a balance between sensitivity and specificity. All adverse event MOSES were categorized across the selected threshold for further analysis. While there was a minor variation in this cut off for various toxicity endpoints,

the score was rounded off to a uniform value of 0.20 for ease of applicability. For the cumulative MOSES (C-MOSES), the chosen cut off was 0.70 as a vast majority of symptomatic patients have 3–4 symptoms related to GI. To assess stability of the chosen cut off sensitivity analysis for the choice of cut-offs was performed using MOSES cut off between a 0.15–0.25 and QOL categorization definition of $>25\%$, $>50\%$, and $>75\%$.

CTCAE was categorized Grade ≥ 1 or grade 0. Chi-square test was used to test the discriminatory ability of CTCAE and MOSES in categorizing patients as “substantially symptomatic on QOL reporting or not”. p -value < 0.05 was considered statistically significant. Similarly, C-MOSES and CTCAE maximum grade was tested for their ability to predict substantial change in QOL (symptom scales and functional scales). Accuracy of CTCAE was calculated as the number of patients with [CTCAE Grade ≥ 1 with substantial QOL change + CTCAE Grade = 0 with insignificant QOL Change]/ total number of patients. Similarly, the accuracy of MOSES was calculated as =number of patients with [MOSES ≥ 0.2 with substantial QOL change + MOSES < 0.2 with insignificant QOL Change]/ total number of patients.

The GI adverse events in the index trial were reanalysed using cumulative incidence (1-Kaplan Meier) method both with CTCAE and C-MOSES. The interventions (3DCRT or IG-IMRT) were compared for their impact on late GI toxicity for each of the toxicity reporting methods.

Role of the funding source

No specific funding was obtained for this study. The primary data set was acquired from the database of the phase III trial, and its funders had no role in this study design, analysis, interpretation or writing of this report. The principal investigator (SC) and designated study staff had access to the complete dataset along with the final responsibility to submit this manuscript for publication.

¹Results

Out of the 300 patients randomized in the phase III trial, 201 patients who had any of the six listed long-term adverse events either in physician or patient-reported outcomes were included [supplementary Table 2]. Table 1 shows the baseline characteristics of the study population. As per the CTCAE method, a total of 371, 51, 8 and 1 grade 1,2,3,4 adverse events [Table 2] were reported with a total of 431 events [median of

¹ TAH+BSO: Total abdominal hysterectomy with bilateral salpingoophorectomy; CTRT: Chemoradiation; RT: Radiation; 3D: Three dimensional; IMRT: Intensity Modulated Radiation Therapy; VMAT: Volumetric Modulated Arc Therapy.

Age, median in years (range)	50 years (31–83)
BMI Mean (range)	24 (13.7–37.1)
≥ 24	100 (49.8%)
<24	101 (50.2%)
Comorbidities at diagnosis	33 (16.4%)
Hypertension	24 (12%)
Diabetes Mellitus	15 (7.5%)
Histology	
Squamous cell carcinoma	78.1%
Adenocarcinoma	19.4%
Adeno-squamous cell carcinoma	2.5%
Type of surgery	
Simple hysterectomy	40%
Wertheim hysterectomy	45%
TAH+BSO	7%
Adjuvant treatment	
CTRT	78%
RT alone	22%
EBRT technique	
3D conformal	49%
IMRT/VMAT	51%
Overall Treatment time, Mean(Range, in days)	51 (38–77)

Table 1: Patient, disease and treatment characteristics (n = 201).

^a BMI: Body Mass Index; TAH+BSO: Total abdominal hysterectomy and bilateral salpingoophorectomy; CTRT: Chemoradiation; RT: Radiation; 3D: Three Dimensional; IMRT: Intensity Modulated Radiotherapy; VMAT: Volumetric Modulated Arc Therapy.

2 events/patient (range 0–6)] . Additionally, many patients reported impaired QOL, whereas, on CTCAE, grade 0 was allocated by the physician. [Table 2]

Symptom specific MOSES

GI toxicity. Table 3 summarizes the mean MOSES across different CTCAE grades of GI toxicity. As a first example, the temporal course (evolution and resolution) of toxicity in patients with CTCAE grade 1–3 diarrhoea is summarized in Figure 2a-b. For each of these patients, CTCAE grade and their corresponding MOSES is depicted in Figure 3. As seen in Table 3, MOSES for CTCAE grade 3 diarrhoea ranges from 0.23 to 2.13, which is almost a 10-fold change within the same CTCAE grade that can be attributed to the duration spent with symptoms of toxicity (Figure 3). Similarly, in patients with CTCAE grade 2 diarrhoea, MOSES had a range of 0.18 to 0.63, almost a 3-fold change. MOSES predicted with higher accuracy diarrhoea related impact on QOL than CTCAE (85% vs. 69%, p value= 0.001).

For abdominal pain, MOSES across different CTCAE grades are shown in Table 3 and supplementary figure 1a-d, in which the variation and overlap of scores can be observed. Both CTCAE and MOSES correlated significantly with QOL (Table 4), but the accuracy of CTCAE in predicting impact on QOL was lower than MOSES (57% and 71%, respectively). For anorexia, 100 patients with symptoms were included of which 67% had grade 0 allocated by physicians on CTCAE and 33% had grade

CTCAEToxicity	No. of patients ^a (out of 300)	Grade 0 ^b	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Diarrhoea	80	55	17	4	4	0	0
Anorexia	100	67	33	0	0	0	0
Abdominal Pain	162	29	107	24	2	0	0
Urinary frequency	123	59	57	6	0	1	0
Urinary incontinence	82	42	37	2	1	0	0
Fatigue	157	21	120	15	1	0	0

Table 2: Distribution of CTCAE grades for various adverse events amongst study population.

^a These are patients who either had physician or patient reported toxicity in the entire study cohort. The median number of symptoms per patient on CTCAE recording were 2 (0–6).

^b The grade 0 patients are those who were scored grade 0 by the physician but have reported any symptom on patient reported QOL questionnaire.

Symptom	Diarrhoea	Abdominal pain	Anorexia	Urinary incontinence	Urinary frequency	Fatigue
CTCAE Grade 1	0.14(0.04–0.32)	0.19 (0.03–0.61)	0.20(0.03–0.65)	0.16 (0.03–0.56)	0.17 (0.04–0.56)	0.27 (0.04–1)
CTCAE Grade 2	0.46(0.18–0.63)	0.43 (0.14–0.91)	NA	0.40 (0.39–0.41)	0.40 (0.13–1.04)	0.54 (0.14–1.17)
CTCAE Grade 3	0.85(0.23–2.13)	1.23 (1.04–1.43)	NA	0.67 (0.67)	NA	0.71 (0.71)
CTCAE Grade 4	NA	NA	NA	NA	0.45 (0.45)	NA

Table 3: Mean MOSES of each grade of all adverse event items tested for the study.

NA=Not applicable.

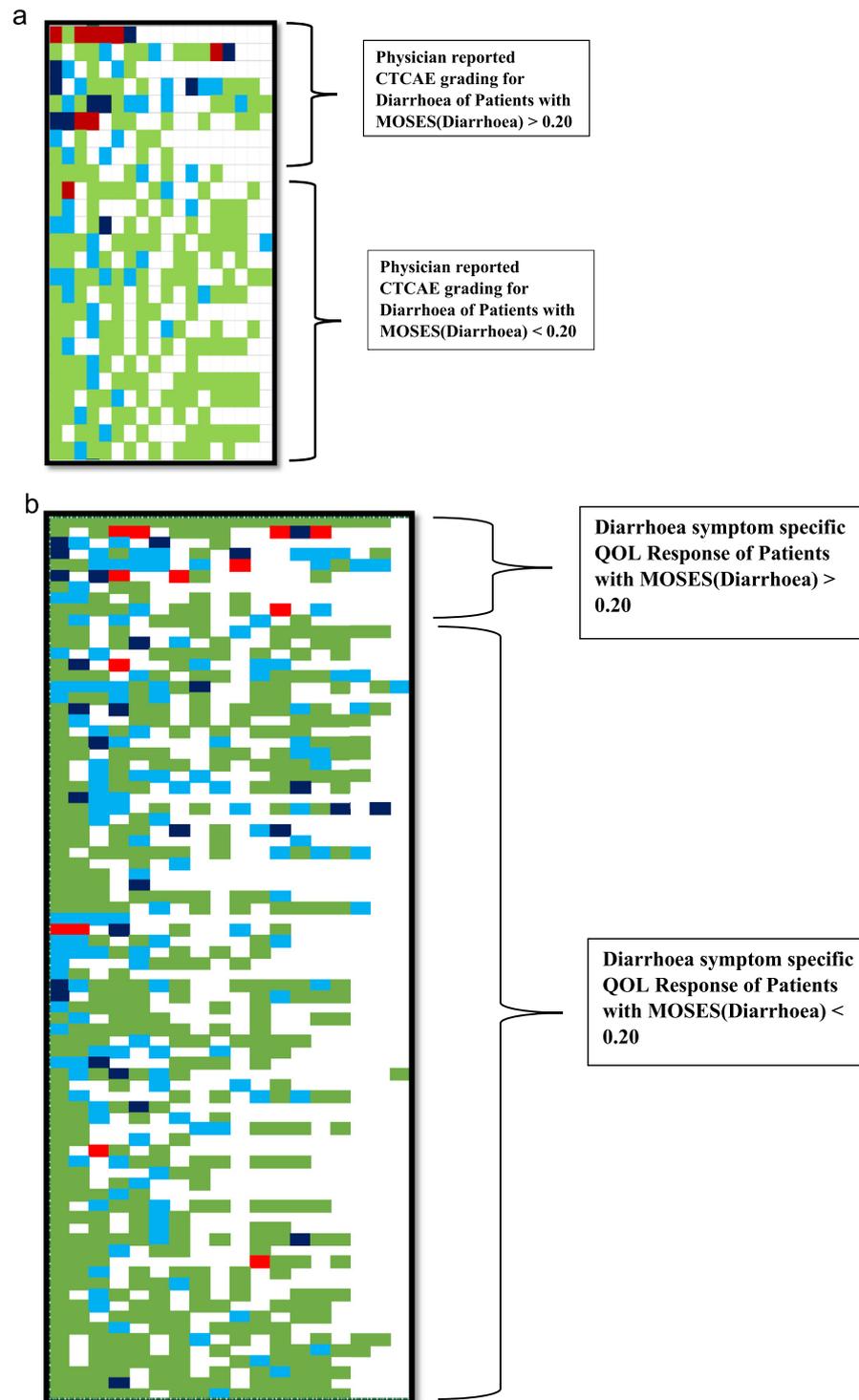


Figure 2. (a) Longitudinal profile of CTCAE grading for patients with diarrhoea on physician reporting for the duration of their scheduled clinical follow up. Colours reflect the physician reported CTCAE grade 0 (light green), grade 1 (light blue), grade 2 (dark blue), and grade 3 (dark red). Missing colour blocks represent missed clinical follow up. (b) Longitudinal profile of QOL responses of patients with diarrhoea as per scheduled clinical follow up. Colours reflect the EORTC answers "not at all" (light green), "a little" (light blue), "quite a bit" (dark blue), "very much" (dark red). Missing blocks represent missing clinical follow up or missing QOL scoring.

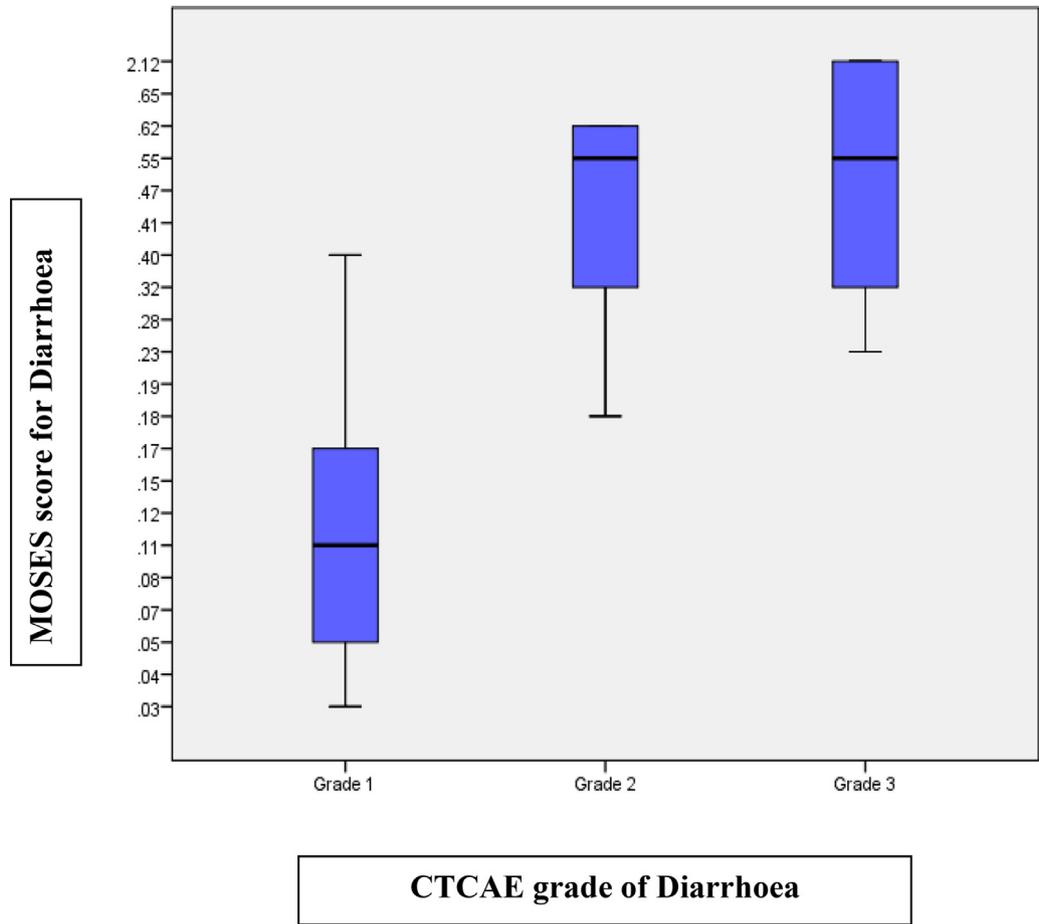


Figure 3. Distribution of MOSES across different grades of diarrhoea. For grade 1, MOSES range was from 0.04 to 0.32, for Grade 2 MOSES ranges from 0.18 to 0.63 and MOSES for Grade 3 ranges from 0.23 to 2.13.

QOL Symptoms	CTCAE maximum grade method				MOSES Method				
	Sensitivity	Specificity	Accuracy	p-value (φ^*)	Sensitivity	Specificity	Accuracy	p-value (φ^*)	AUC
Diarrhoea	50%	73%	69%	0.096 (0.186)	43%	94%	85%	0.001 (0.42)	0.67
Anorexia	25%	63%	51%	0.24 (0.11)	9%	85%	61%	0.4 (0.02)	0.45
Abdominal Pain	88%	24%	57%	0.046 (0.156)	58%	85%	71%	0.001 (0.45)	0.76
Urinary incontinence	65%	59%	61%	0.04 (0.226)	30%	91%	72%	0.01 (0.27)	0.65
Urinary frequency	63%	56%	59%	0.045 (0.18)	21%	91%	62%	0.06 (0.16)	0.63
Fatigue	90%	24%	76%	0.03 (0.165)	63%	70%	64%	0.001 (0.27)	0.71
QOL functional Domains	CTCAE maximum grade				Cumulative-MOSES				
Physical function	87%	13%	70%	0.982 (0.002)	56%	65%	58%	0.003 (0.213)	0.67
Role function	86%	13%	31%	0.892 (0.10)	46%	80%	72%	0.001 (0.256)	0.68
Social function	85%	10%	67%	0.088 (0.153)	27%	100%	28%	0.22 (0.085)	0.77
Global health status/QOL	22%	90%	38%	0.446 (0.54)	32%	92%	47%	0.001 (0.235)	0.59

Table 4: Sensitivity, Specificity and Accuracy of CTCAE, MOSES and C-MOSES in predicting substantial QOL symptoms and functional scales.

* φ – phi value is used to measure effect size where 0.2, 0.5 and 0.8 represents small medium and large effect size (Cohen,1988), area under curve comparison not feasible for both methods as CTCAE uses categorical and MOSES uses continuous data.

1 anorexia. Both CTCAE and MOSES did not correlate with QOL ($p = 0.24$ and 0.90 respectively). The accuracy of CTCAE and MOSES in predicting anorexia was 51% and 61% respectively (Table 4).

Fatigue. Out of the 157 patients included for fatigue, grade 1, 2 and 3 fatigue was reported by 120, 15 and 1 patient respectively. MOSES for the grade 2 patients ranges from 0.14 to 1.17 while for grade 1 patients, it ranges from 0.04 to 1, which is almost 25-fold. In this case, the accuracy of CTCAE was higher than MOSES, however, the AUC of MOSES was also high (0.71) and was acceptable for reporting (Table 4).

GU toxicity. For Genito-urinary toxicity assessment, urinary incontinence and urinary frequency were included. Out of a total of 82 patients with urinary incontinence, 37 patients had grade 1 CTCAE events for which MOSES varied from 0.03 to 0.56. The distribution of the MOSES score for urinary frequency and incontinence is depicted in supplementary figures 5 and 6. Only two patients had grade 2 toxicity, and one patient had grade 3 toxicity. The accuracy of MOSES was better than CTCAE for predicting QOL impact of urinary incontinence (61% for CTCAE and 72% for MOSES). For urinary frequency, 57 patients had CTCAE grade 1 toxicity, with MOSES ranging from 0.04 to 0.56, a 14-fold difference. The accuracy of CTCAE and MOSES in predicting QOL impact was 59% and 62%, respectively (Table 4). Area under curve (AUC) of MOSES for urinary incontinence and frequency was 0.65 and 0.63 respectively (Table 4).

Cumulative moles (C-MOSES). To estimate the cumulative burden of toxicity among patients, we first determined the number of symptoms per patient and cumulative MOSES was estimated for all 201 patients for these 6 most common toxicities. The summary of C-MOSES scores across CTCAE grades is summarised in supplementary Table 3. On investigating the correlation between C-MOSES and the number of adverse events/patients, we recognized that though patients with multiple adverse events (>2) should have a greater probability of higher C-MOSES, however a substantial proportion of patients (40%) with multiple adverse events did not necessarily have C-MOSES >0.70 . This observation supports the need for summing both grade and temporal duration for estimating impact. Accuracy of CTCAE in predicting role functioning and global QOL domains was 31% and 38% respectively whereas for C-MOSES it was 72% and 47% respectively (Table 4).

Sensitivity analysis

As the results could be influenced by choice of either MOSES cut off, or the definition of "substantial QOL

impact," we performed sensitivity analysis using different cut-offs from ROC analysis for MOSES. The use of different cut-offs for MOSES between 0.15 and 0.25 did not improve the accuracy of predicting QOL impact hence a cut off of 0.20 was retained and recommended for individual CTCAE symptom items. We also evaluated the accuracy of MOSES by excluding patients who did not have a MOSES grade (i.e. CTCAE grade 0) but had reported symptoms on QOL. Though this approach improved sensitivity (50 to 85%) and accuracy (68.7% to 80%) for reporting diarrhoea and some other symptoms, we continued to include all 201 patients to minimize the optimistic bias.

We also performed a sensitivity analysis for the cut off for QOL assessment, i.e. symptomatic on patient-reported outcomes $>25%$, $>50%$ or $>75%$ of follow up visits. The highest accuracy of MOSES was observed when patients reported symptoms in $>75%$ to follow up visits. Only a minor difference was observed for functional scales at the cut off of 50 and 75% (supplementary Table 4.5). However, as choice of "symptomatic $>75%$ follow up time" could exclude many patients with symptoms; hence, for QOL correlation, the original definition and cut-offs were retained.

Reanalysis of GI events with MOSES

While CTCAE method reported no difference in cumulative incidence of late GI grade ≥ 1 adverse events between the 3DCRT and IG-IMRT arms (HR=0.98; 95% CI 0.75–1.28, $p = 0.88$), C-MOSES ≥ 0.70 reported statistically significant difference in GI adverse events in two arms (HR=0.64; 95% CI 0.41–0.99, $p = 0.04$) [Figure 4].

Discussion

MOSES is a new toxicity summarising method that is developed from longitudinal CTCAE data. MOSES incorporates the dimension of time, severity and multiplicity of events. It also takes into account persistence of low-grade events, thereby providing a longitudinal depiction of physician-reported late adverse events using CTCAE. MOSES is summarised as a mathematical score that describes the burden of late adverse events. In this methodological development study, six most commonly reported adverse events that were subjective (anorexia, fatigue, pain) and objective (diarrhoea, urinary frequency and incontinence) were included to test the performance of MOSES across the range of most common late adverse events after pelvic radiotherapy. MOSES and CTCAE were further compared for accurately identifying "substantially symptomatic patients" on QOL assessment.

We observed a wide variation of MOSES across each allocated CTCAE grade, and at the same time, there was a significant overlap in the MOSES scores of patients

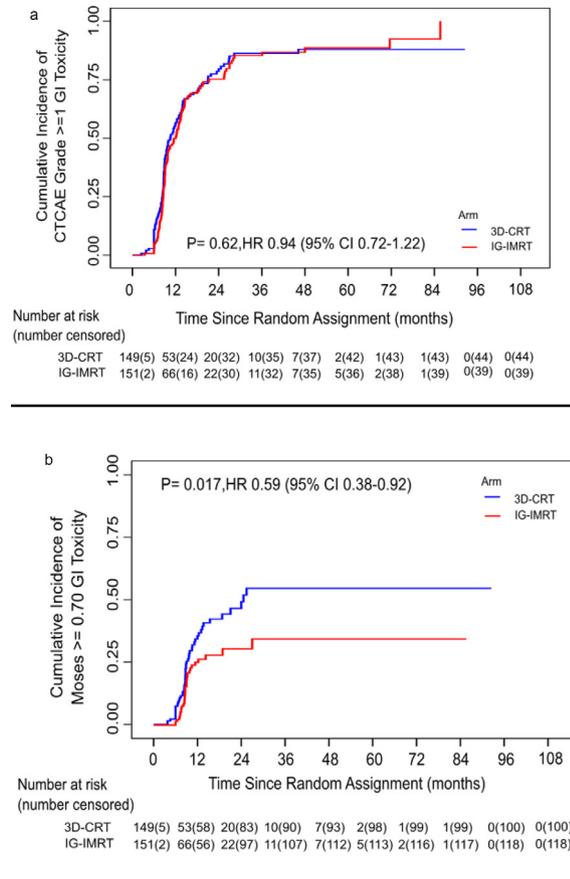


Figure 4. (a) Figure depicting cumulative incidence of grade ≥ 1 Late GI toxicity by treatment arm, (b) Figure depicting cumulative MOSES for GI toxicity by treatment arm.

with different CTCAE grades. Many patients with CTCAE grade 1 had MOSES higher than in a patient who was allocated CTCAE grade 2 or 3. This suggests that there could be a significant impact of lower grade toxicities if they persist for a substantial duration and one-time point CTCAE grade allocation may not completely describe the spectrum of adverse events. These limitations of CTCAE reporting using maximum grade also possibly contributes to the lack of agreement between physician and patient-reported outcomes. While we observed a relatively higher accuracy of MOSES to predict the impact on QOL, we observed that the sensitivity of MOSES in detecting symptomatic patients on QOL was low as compared to its' specificity. This could be attributed to the inclusion of patients that had MOSES score of zero (but had a corresponding QOL symptom reported). However, to minimize the optimistic bias, we report the sensitivity, specificity and accuracy after including all patients.

Similar to the limitations in the CTCAE reporting, the methodology of reporting and summarizing QOL hinges on reporting changes in QOL from baseline to predefined times points during follow up. For example,

a patient who reports "not at all" for QOL item of pain for up to 24 months of follow up but "very much" at 36 months and after that "not at all" could be labelled to have significant deterioration of QOL in the before-after analysis if limited to baseline and 36 months. In contrast, in real life the deterioration has been only for a fraction of follow up. To overcome these limitations, we categorized patients QOL scores as "substantially symptomatic or not" on follow up based on whether patients reported symptoms in at least 50% of their follow up duration. In addition to providing a summary of longitudinal QOL, this method also compares physician and patient-reported outcomes per patient rather than at the study level. Almost a similar approach of categorizing CTCAE and QOL is previously reported.¹⁵

Overall, MOSES provided higher accuracy for identifying QOL impact for a vast majority of common adverse events with good AUC, suggesting that imputing temporal trends and severity weightage can improve the congruence of physician and patient assessment. Also, C-MOSES was more accurate than CTCAE in predicting functional and global QOL (Table 4). MOSES cut-offs that we utilized and proposed in this study

could be different in a separate cohort of patients based on how patients report their QOL in different socio-cultural environments. In an attempt to internally validate MOSES, we reanalysed all gastrointestinal toxicities in PARCER study by defining CTCAE grade ≥ 1 toxicity as an event. The difference between two treatment arms i.e., 3DCRT and IG-IMRT were not observed if all CTCAE grades of GI toxicity were included. However, using C-MOSES difference in treatment intervention was very clear and favoured IG-IMRT. This suggests that IG-IMRT not only reduce moderate to severe toxicity (as in index trial) but also the duration of toxicity which was not necessarily captured using CTCAE method (Figure 4). External validation of MOSES is therefore needed to test its ability in discriminating intervention effects of advanced radiation techniques as compared to CTCAE.

MOSES has certain limitations. It is developed from physician reported CTCAE grading, and inter-observer variation in grading is well known. Therefore, MOSES has an inherent limitation of physician scoring bias which can be minimized by asking two physicians to provide scoring. Unfortunately, multi-physician scoring was not available in this study. Though MOSES accounts for the duration spent in all toxicity grades, it does not directly reflect the interventions and their impact on adverse event severity and duration. It also does not reflect severity of CTCAE grade 5 events accurately. Furthermore, if a patient has a small period of a serious adverse event that resolves after the intervention, it is likely that MOSES will allocate a lower score to such a patient. Therefore, MOSES should be used in addition to CTCAE to understand the severity and temporal course of the toxicity. Though we report higher accuracy of MOSES than CTCAE it's important to recognise that QOL is a multidimensional evaluation that may not be governed by physician's assessment of time and severity. An additional parameter of weightage of symptom burden could likely further improve the performance of MOSES. Nevertheless, the differences in the patient and the physician-reported outcome would continue to be there as objective scoring systems may not have a perfect method of imputing the patient's adaptability to the new circumstances. Secondly, MOSES is derived from patients receiving postoperative (chemo) radiation. It's relevance as compared to CTCAE needs to be further tested in patients undergoing interventions that have higher frequency of acute or short term adverse events rather than late adverse events (e.g. surgery or chemotherapy when administered as a stand-alone treatment). Missing follow up data can also pose a challenge in the calculation of MOSES. Also, during patient's follow-up, the start and end date of adverse events can be unclear and can have a bearing on the calculation of the score. While the performance of this metric could be improved by making follow up calls to assess the resolution or evolution of adverse effects in between planned follow-ups, the trial database currently

does not have such an information. Also an important limitation of MOSES is its lack of direct applicability in routine day to day care of patients where decisions are often made on an individual case basis. However, it could be very effectively used as a supplement to CTCAE to understand the symptom burden in clinical trials or deployed in survivorship clinics.

In summary, we developed a method of summing and reporting toxicity that has potential to provide a better correlation with patient-reported symptoms and can be a valuable complement to CTCAE. External validation is needed in the future to test the robustness and applicability in summarising data in clinical trials.

Contributors

NR participated in project concept, design, analysis, manuscript preparation and approval of the study. SC participated in project concept, design, analysis, manuscript preparation and final approval. SC is also the Principal Investigator for PARCER Trial. Both NR and SC have accessed and verified the data and responsible for decision to submit manuscript. AM participated in the design and analysis of the study. MC was responsible for data collection, data assembly and preparation of the manuscript. SK was responsible for the statistical analysis of the PARCER trial and PR performed the statistical analysis of the current work. TD, UM, RE, LG and PM participated in manuscript draft review and final approval of draft. JG, SG, AM and TSS participated in data acquisition, revising project for intellectual content and final approval of the draft. SKS participated in the final manuscript draft review and approval and is the Co-Principal Investigator of the PARCER study.

Declaration of interests

We declare no conflict of interests

Acknowledgements

We thank Kathrin Kirchheiner, Medical University of Vienna, for her critical discussions during the development of the MOSES.

Data sharing statement

Reader can access extra data (individual patient data) through the principal investigators after due institutional and additional regulatory approvals. Automated calculation of MOSES is feasible through an excel sheet developed by the authors and corresponding author may be contacted for this information.

Funding

None.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2022.101390.

References

- 1 Pötter R, Tanderup K, Schmid MP, et al. EMBRACE Collaborative Group. MRI-guided adaptive brachytherapy in locally advanced cervical cancer (EMBRACE-I): a multicentre prospective cohort study. *Lancet Oncol*. 2021;22:538–547.
- 2 Klopp AH, Yeung AR, Deshmukh S, et al. Patient-reported toxicity during pelvic intensity-modulated radiation therapy: NRG oncology-RTOG 1203. *J Clin Oncol*. 2019;36:2538–2544.
- 3 Yeung AR, Pugh SL, Klopp AH, et al. Improvement in patient-reported outcomes with intensity-modulated radiotherapy (RT) compared with standard RT: a report from the NRG Oncology RTOG 1203 study. *J Clin Oncol*. 2020;38:1685–1692.
- 4 Chopra S, Gupta S, Kannan S, et al. Late toxicity after adjuvant conventional radiation versus image-guided intensity-modulated radiotherapy for cervical cancer (PARCER): a randomized controlled trial. *J Clin Oncol*. 2021;39:3682–3692.
- 5 Mayadev JS, Enserro D, Lin YG, et al. Sequential ipilimumab after chemoradiotherapy in curative-intent treatment of patients with node-positive cervical cancer. *JAMA Oncol*. 2019;6:92–99.
- 6 Duska LR, Scalici JM, Temkin SM, et al. Results of an early safety analysis of a study of the combination of pembrolizumab and pelvic chemoradiation in locally advanced cervical cancer. *Cancer*. 2020;126:4948–5956.
- 7 Study of chemoradiotherapy with or without pembrolizumab (MK-3475) for the treatment of locally advanced cervical cancer (MK-3475-A18/KEYNOTE-A18/ENGOT-cx11). <https://clinicaltrials.gov/ct2/show/NCT04221945> Site accessed September 4th 2020.
- 8 Widmark A, Gunnlaugsson A, Beckman L, et al. Ultra-hypofractionated versus conventionally fractionated radiotherapy for prostate cancer: 5-year outcomes of the HYPO-RT-PC randomized, non-inferiority, phase 3 trial. *Lancet*. 2019;394:385–395.
- 9 Trialists' Group TS. The UK Standardisation of Breast Radiotherapy (START) Trial B of radiotherapy hypofractionation for treatment of early breast cancer: a randomized trial. *Lancet*. 2008;371:1098–1107.
- 10 Nout RA, Putter H, Jargenliemk-Schulz IM, et al. Five-year quality of life of endometrial cancer patients treated in the randomized Postoperative Radiation Therapy in Endometrial Cancer (PORTEC-2) trial and comparison with norm data. *Eur J Cancer*. 2012;48:1638–1648.
- 11 Fayers P, Aaronson NK, Bjordal K, Sullivan M. *EORTC QLQ-C30 Scoring Manual*. European Organisation for Research and Treatment of Cancer; 1995.
- 12 Cocks K, King MT, Velikova G, et al. Evidence-based guidelines for interpreting change scores for the European organisation for the research and treatment of cancer quality of life questionnaire core 30. *Eur J Cancer*. 2012;48:1713–1721.
- 13 Trotti A, Pajak TF, Gwede CK, et al. TAME: development of a new method for summarising adverse events of cancer treatment by the radiation therapy oncology group. *Lancet Oncol*. 2007;8:613–624.
- 14 Thanarajasingam G, Atherton PJ, Novotny PJ, Loprinzi CL, Sloan JA, Grothey A. Longitudinal adverse event assessment in oncology clinical trials: the Toxicity over Time (ToxT) analysis of Alliance trials NCCTG N9741 and 979254. *Lancet Oncol*. 2016;17:663–670.
- 15 Kirchheiner K, Pötter R, Nout RA, et al. Late, Persistent, Substantial, Treatment-Related Symptoms After Radiation Therapy (LAPERS): a new method for longitudinal analysis of late morbidity—applied in the EMBRACE study. *Int J Radiat Oncol Biol Phys*. 2020;106:300–309.