Table 1. Internal reliability of the subscales at both time points

	Baseline		Time 2			
Subscale	Cronbach's alpha (raw)	Cronbach's alpha (standardised)	Cronbach's alpha (raw)	Cronbach's alpha (standardised)		
Bowel symptoms (non-stoma)	0.78*	0.78*	0.80*	0.80*		
Bowel symptoms (stoma)	0.64	0.64	0.81*	0.81*		
Pain/discomfort	0.89*	0.89*	0.93*	0.93*		
Stoma care	0.65	0.65	0.87*	0.89*		
Vaginal symptoms	0.88*	0.88*	0.89*	0.89*		

<sup>\*</sup>Acceptable consistency in responses to items within the scales, Cronbach's alpha coefficient >0.70

Table 2. Test-retest analyses for clinically stable participants

Subscale / Item	Mean (SD) paired difference (Baseline minus time 2)	95% CI	ICC (2,1)	n
Bowel symptoms (non-stoma)	2.16 (13.38)	-0.94 to 5.26	0.81*	74
Bowel symptoms (stoma)	2.38 (11.5)	-8.26 to 13.02	0.96*	7
Pain/discomfort	1.69 (12.91)	-1.14 to 4.53	0.87*	82
Stoma care	-4.76 (28.59)	-31.20 to 21.68	0.59	7
Vaginal symptoms	4.56 (19.85)	-0.75 to 9.88	0.79*	56
Frequent urination	0.00 (22.55)	-4.89 to 4.89	0.71*	84
Keeping clean	-2.41 (23.15)	-7.47 to 2.65	0.68	83
Proximity to toilet	2.78 (23.26)	-2.27 to 7.83	0.69	84
Lower limb oedema	2.78 (18.08)	-1.15 to 6.70	0.72*	84
Planning activities	-1.98 (24.46)	-7.29 to 3.32	0.72*	84

Subscale / Item	Mean (SD) paired difference (Baseline minus time 2)	95% CI	ICC (2,1)	n
Interest in sex		-2.21 to 4.74	0.84*	79
Sex life	0.88 (18.84)	-3.43 to 5.18	0.90*	76
Pain during intercourse	0.93 (18.53)	-3.43 to 5.28	0.81*	72
Erectile problems	0.00 (14.91)	-6.79 to 6.79	0.92*	21

<sup>\*</sup>Acceptable consistency in responses to items on two repeated assessments, ICC <u>></u>0.70

Table 3. Multi-trait scaling

	Convergent validity (Items correlate >0.40 with own scale)	Discriminant validity (items do not correlate as highly with another scale) <sup>1</sup>
Bowel symptoms (non-stoma)	0.66-0.81	0.16-0.51
Bowel symptoms (stoma)	0.86-0.87	0.16-0.51
Pain/discomfort	0.75-0.86	0.07-0.53
Stoma care	0.74-0.81	0.12-0.79
Vaginal symptoms	0.88-0.91	0.10-0.50

<sup>&</sup>lt;sup>1</sup> Correlations between scales with overlapping items not produced (Bowel symptoms (non-stoma) and Bowel symptoms (stoma))

Table 4. QLQ-ANL27 scale correlations with conceptually related QLQ-C30 scales

Conceptually related QLQ-C30 subscale	EORTC QLQ ANL27 Bowel symptoms (non-stoma)	EORTC QLQ ANL27 Bowel symptoms (stoma)	EORTC QLQ ANL27 Pain/discomfort
QLQ-C30 Constipation	0.21	-0.34	N/A
QLQ-C30 Diarrhea	0.61*	0.23	N/A
QLQ-C30 Pain	N/A	N/A	0.70*

Conceptually related QLQ-C30 subscale	EORTC QLQ ANL27 Bowel symptoms (non-stoma)	EORTC QLQ ANL27 Bowel symptoms (stoma)	EORTC QLQ ANL27 Pain/discomfort	
QLQ-C30 Physical functioning	-0.36	-0.15	N/A	
QLQ-C30 Social functioning	-0.40	-0.02	N/A	

<sup>\*</sup>Scales are conceptually related to an acceptable level, r>0.40

Table 5. Known group comparisons

Subscale	Treatment	Treatment	Localised or	Radiotherapy	Performance
	phase (acute	status (on or	locoregional	type IMRT or	status (0 or
	vs. early vs.	off)		CRT)	<b>1)</b> <sup>1</sup>
	late)				
Bowel	ES=0.28	ES=0.17	ES=0.12	ES=0.06	ES=0.25*
symptoms	P=0.057	P=0.365	P=0.327	P=0.645	P=0.055
(non stoma)					
Pain /	ES=0.92***	ES= 0.20*	ES=0.03	ES=0.04	ES=0.38*
discomfort	P<0.0001	P=0.270	P=0.818	P=0.761	P=0.002
Vaginal	ES=0.44*	ES=0.05	ES=0.01	ES=0.21*	ES=0.07
symptoms	P=0.003	P=0.831	P=0.950	P=0.226	P=0.649

 $<sup>^{1}</sup>$ Given that only 14 participants recorded their performance status as 2 or 3, comparisons were only performed for those scoring 0 or 1.

Table 6. Responsiveness to deterioration over time

<sup>\*</sup>Small effect size  $\geq$ 0.20

<sup>\*\*\*</sup>Large effect size <a>0.80</a>

Subscale <sup>1</sup>	N	Paired mean difference (Baseline minus second)	Lower CI	Upper CI	Baseline SD	Effect Size		Effect size category
Bowel symptoms (non- stoma)	22	-17.9	-27.7	-8.1	26.6	-0.7	0.0011	Moderate deterioration
Pain / discomfort	29	-34.9	-47.0	-22.8	26.8	-1.3	<.0001	Large deterioration
Vaginal symptoms	20	-31.7	-48.7	-14.6	20.2	-1.6	0.0010	Large deterioration

<sup>&</sup>lt;sup>1</sup>For symptom scales a negative difference means that second timepoint had higher (worse) scores than baseline, while a positive difference means that second timepoint had lower (better) scores than baseline.