Abstract

Purpose/Objective:

To evaluate whether there is sufficient correlation between patient reported outcomes (PRO) and clinician reported outcomes (CRO) in bladder cancer follow up post radiotherapy to streamline data collection and to reduce trial follow up burden on patients, clinicians, and trial programmes.

Methods:

Patient reported outcome data were collected within the BC2001 trial using the Functional assessment of Cancer Therapy specific to bladder cancer (FACT-BL) questionnaire. Clinician reported outcome data were collected by clinicians using the LENT/SOM (Late effects in normal tissues subjective, objective and management). Data were collected at baseline, post-treatment, at 6- and 12-months post randomisation and then annually to 5 years.

Percentage agreement between clinician reported and patient reported outcome measures were evaluated at 2- and 5-years post randomisation. Concordance was tested using the weighted Kappa statistic with 95% confidence intervals.

Results

Correlation was evaluated between six categories of the FACT-BL and LENT/SOM scores. At 2 years the percentage agreement across these domains ranged from 45% to 78% with the weighted Kappa statistic between 0.07 and 0.35. Results were similar in year 5 with 48-83% agreement and kappa statistics between -0.02 and 0.21.

Conclusion

Correlation between clinician reported outcomes and patient reported outcomes in patients treated with radiotherapy for bladder cancer are generally poor. PROs appear more sensitive with higher grade events reported. Further work is needed to evaluate

whether PROs alone can be used to evaluate toxicity related outcomes in randomised controlled trials.

Keywords:
Bladder cancer, Patient reported outcomes, clinician reported outcomes, radiotherapy

Background

Bladder cancer is the 11th most common cancer in the UK(1). A significant proportion of patients are treated with bladder conserving therapy, with 48% of patients with stage II muscle invasive bladder cancer receiving radiotherapy in the first 12 months post diagnosis (2).

We have previously reported that long term normal tissue effects (NTE) in bladder cancer patients treated with radiotherapy are low (3)(4). Other papers presenting retrospective data on long term clinician reported outcomes (CRO) or patient reported outcomes (PRO) following bladder preservation treatment for muscle invasive bladder cancer also show low levels of NTE (5)(6)(7). Most studies present either PRO or CRO in isolation.

Whilst historically NTE data was collected using CRO, the use of PRO in clinical trials has been increasing (8) and both CRO and PRO are often collected in radiotherapy trials. Long term NTEs can occur up to ten years post treatment and current methods of data collection can be burdensome on patients, hospital clinical staff and clinical trials units. It is important to ascertain the best methods of collecting data to reduce this burden. Determining whether collection of both CRO and PRO could be rationalised or if PRO or CRO could be collected in isolation whilst still providing clinically relevant and valid information on NTE would reduce the resources spent on collecting data.

This analysis investigates, within the context of patients treated with radiotherapy for bladder cancer in a phase III randomised controlled trial, the degree of concordance on an individual patient level between PROs and CROs in a step towards understanding how best to collect data on NTE.

Methods

BC2001 is a phase III trial with a 2x2 partial factorial design that recruited 458 patients from 45 UK hospitals. Patients were randomised in a 1:1 ratio to receive radiotherapy with or without chemotherapy. Patients could also be randomised to receive standard

whole bladder radiotherapy or reduced high dose volume radiotherapy with a tumour boost. Recruitment to both the chemotherapy and the radiotherapy volume comparisons was encouraged but optional according to the eligibility and preference of the patient. All patients received conformal radiotherapy on consecutive weekdays according to the hospital site's standard regime (55Gy in 20 fractions or 64Gy in 32 fractions). Patients randomised to receive chemotherapy were given intravenous mitomycin C (12mg/m2) on day 1 of radiotherapy and a continuous infusion of 5-Flurouracil (5-FU) 500mg/m2/24 hours for 5 days during radiotherapy fraction 1-5 and 16-20 inclusive. Full details of the trial and efficacy and safety results have previously been reported (9)(3) and synchronous chemoradiotherapy is now a standard of care.

452/458 (99%) of BC2001 participants consented to the optional health-related quality of life (HRQoL) study. PROs were collected using the FACT-BL questionnaire completed on paper at baseline, end of treatment, 6 months post end of treatment and then annually to 5 years. The questionnaire includes 39 items scored on a five-point Likert scale and five subscales: physical well-being (PWB), social well-being (SWB), emotional well-being (EWB), functional well-being (FWB) and the bladder cancer subscale (BLCS). The FACT-BL total score (TOTAL, generated from the sum of the PWB, SWB, EWB, FWB and BLCS subscales) and Trial outcome index (TOI, generated as the sum of all of the items comprised in the PWB, FWB and BLCS sub scales) were calculated (10). Scoring and management of missing items were dealt with according to the FACIT administration and scoring guidelines (11). For questions within the FACT-BL that were phrased negatively, scoring was reversed, so high scores are indicative of a better quality of life. The randomised comparison of PROs has been previously reported, showing an initial fall in QoL immediately after treatment with a recovery to baseline after six months and no significant impact of addition of concomitant chemotherapy (4) .

CRO data was collected using the LENT/SOM (Late effects in normal tissues subjective, objective and management) score. Data was collected at baseline, post-treatment, at 6- and 12-months post randomisation and then annually to 5 years. The LENT/SOM scale is made up of questions in 3 domains (subjective, objective and management) graded 1 to 4 depending on severity.

Statistical analysis

PROs were paired with the relevant CRO with the FACT-BL and subjective score of the LENT/SOM being used. Six domains representing more common treatment related side effects were compared between the PRO and CRO: diarrhoea, bowel incontinence, urinary frequency, dysuria, urinary incontinence and sexual dysfunction. The exact pairing of the CRO and PRO was undertaken using clinical judgement of domains that were equivalent in meaning across the PRO and CRO and is presented in table 1. The primary endpoint for this exploratory analysis was the percentage agreement of PRO and CRO at 2- and 5-years post treatment in each domain.

Due to the low numbers of toxicity events recorded, the response levels in each tool were combined to make a 3-point overall scale (none, low/medium and high) for comparison. For the FACT-BL questionnaire responses “a little” or “somewhat” were classified as low/medium and “quite a bit” and “very much” were classified as high. In the LENT/SOM questions grades 1 and 2 were classified as low/medium and grades 3 and 4 as high. The groupings were selected similarly to previous work on PRO vs CRO (12) and to reflect the impact of toxicities on patients. The impact on patients with a “high” reported toxicity is likely to be clinically significant and require medical intervention, whilst those with reported toxicities falling into the low/medium grouping would be likely to resolve without medical intervention.

Agreement between the data collection method on an individual patient level was assessed using percentage agreement (with 95% confidence interval) and weighted kappa statistic (with 95% confidence interval). For interpretation of the strength of the agreement, weighted kappa statistics <0.20 were described as poor, 0.21-0.40 fair, 0.41-0.6 moderate, 0.61-0.8 good and 0.81-1.00 very good (13).

Results

At year 2, 180 patients had data available for at least one of the PRO domains analysed whilst 265 patients had CRO for at least one of the domains. 178 patients had data for both CRO and PRO and the analysis was performed on these patients.

At year 5, 113 patients had PRO data available with 174 patients with available CRO data. The analysis was performed on the 109 patients with available CRO and PRO data.

Overall prevalence of NTE

The majority of patients reported no long-term normal tissue effects at 2 and 5 years. The most commonly reported side effect on both PRO and CRO was sexual dysfunction where 58% of participants reported high level of sexual dysfunction on PRO and 22% on CRO at year 2. In general, there was a marked variation in prevalence of PRO and CRO reported NTEs. An example of this was when reporting dysuria where 26% of patients responding to the question; “it burns when I urinate” with “quite a bit” or “very much”, in comparison only 1% clinicians reported the patients to have grade 3 or 4 dysuria according to the LENT/SOM scoring.

Patients reported a higher prevalence in NTEs than clinicians in all six domains assessed at year 2 (figure 1) and 5 (figure 2). At year 2 the percentage agreement ranged from 34% for sexual dysfunction to 78% for diarrhoea (table 2). Concordance between PRO and CRO for the six domains was poor to fair as shown by the low weighted kappa.

Only one patient was reported to have a high grade of toxicity by clinicians which was scored as no toxicity on PRO. This was in the domain of urinary incontinence.

At year 5 the percentage agreement ranged from 24% for sexual dysfunction to 83% for diarrhoea. (Table 3.) Concordance using the weighted kappa statistic was poor in all domains apart from urinary frequency where it was fair (0.21).

Discussion

We have previously reported that the HRQoL and clinician reported toxicity of patients treated with radiotherapy for bladder cancer show a low level of marked or moderate side effects irrespective of the method of data collection. In this analysis results suggest a poor to fair correlation between PRO and CRO on an individual patient level

at both 2- and 5-years post randomisation. We have shown that patients tend to rate side effects more severely than clinicians, with only one incidence of a severe toxicity being reported within a CRO questionnaire that was not reported on a PRO questionnaire. This finding was consistent across the six domains analysed.

There are no previous data in published literature on the comparison of CRO and PRO in bladder cancer patients treated with radiotherapy. Within pelvic radiotherapy trials a previous study on pelvic radiotherapy for cervix cancer also showed poor correlation between PRO and CRO and higher reported toxicity levels in PRO when questionnaires were distributed at five years post treatment (14).

Previous studies in breast cancer radiotherapy have shown conflicting results. Patients within the START trial (15) scored toxicities more frequently and severely than clinicians. Similarly, Bhattacharya et al (12) investigated the correlation between PRO and CRO within the IMPORT LOW trial, showing discordance between PRO and CRO with a higher prevalence of toxicity reported by PROs. However, they were able to show that there was a similar risk ratio and effect size between the randomised radiotherapy treatment groups within the trial when evaluating the outcomes solely using CRO or PRO. This suggests that although there was discordance between PRO and CRO the clinical relevance in terms of effect of radiotherapy schedule was similar. Conversely in a further review in breast cancer patients, Mukesh et al (16) showed higher rates of toxicity being reported in CRO than PRO with weak concordance between them at 2 and 5 years. Limitations to this study were that one clinician carried out the majority of clinician-based assessments in this single centre study which may have reduced the expected variation in scoring between different clinicians. Overall, the radiotherapy technique showed improvement based on CROs but no improvement on PROs. Further studies of oncology patients following chemotherapy (17)(18)(19)(20) showed higher reported toxicities in PROs than CROs.

Our analysis showing largely poor concordance between the two methods of assessment of NTEs adds complexity to whether just one approach to data collection can be used in future bladder radiotherapy trials to collect accurate and relevant data for the trial. There may several reasons for the differences in reported frequency or severity of toxicities between PRO and CRO.

Clinicians may be able to evaluate toxicities in a more objective fashion than patients, bringing professional training and experience to their scoring. Similarly, clinicians can ensure that data are filled in for patients at key timepoints by assessing the toxicity in routine clinic follow ups rather than relying on the patient motivation to respond in PROs. However, particularly in multi-centre studies there is a risk of interobserver variability of clinician's reports with bias due to different levels of clinical experience. Furthermore, increasingly busy clinics may cause clinicians to rush responses to questions without appropriate probing of the patient to ensure they have the full picture of toxicity.

Alternatively, patients may be in a better position to communicate their subjective experience within the context of the question asked and this will correlate with the clinical relevance of that experience to the patient.

Our results show that patients report NTEs more frequently and severely than clinicians suggesting that PROs may be more sensitive to NTE than CROs. As radiotherapy delivery becomes more precise treatment related toxicity will be reduced. Picking up PRO data that is more relevant to patients may be a more impactful way of considering endpoints for clinical trials as PROs with the associated benefit that more frequently occurring events would improve statistical powering. Reassuringly, in our data PROs picked up the vast majority of toxicities in bladder cancer patients treated with radiotherapy. Only one high grade toxicity collect on CRO was not noted on the respective PRO.

One limitation to using solely PROs could be that some groups of patients might be more likely to fill in questionnaires than others, for example patients experiencing higher toxicity might be more motivated to respond and this could bias the results compared to CROs. It was reported in the main QoL paper (4) that at year 1, patients without HRQoL data were similar to those with data apart from a greater frequency of residual mass and incomplete resection at baseline. A further limitation would be that future trials using only PROs would not be able to be compared to previous trials using only CROs as the rates of NTE may appear higher in the PRO study even if no actual

difference existed due to increased reporting. This may make it hard to interpret patient toxicity of new treatments outside the context of a randomised controlled trial.

The main limitation of this study was the fact that the scales of the FACT-BL and LENT/SOM were not designed to be interchangeable making comparison difficult ie. grade 4 for the LENT/SOM may be perceived as being much more severe by a patient or clinician than the highest level of toxicity in the FACT-BL score. This lack of standardisation limits the comparability between PRO and CRO and may be a fundamental reason for the poor correlation between methods. The methodology used aimed to group the toxicities into none, low/medium and high in regard to the impact they would have on patients but as PRO is an interpretation by patients it is difficult to standardise this entirely. A further limitation was the fact that there were only a small number of toxicity events in certain domains and a reduced number of patients at follow up at 5 years which may limit the statistical validity of the results. Finally, although sexual dysfunction was one of the most reported side effects on both CRO and PRO, it is only investigated in one question in the FACT-BL score which focuses on interest in sex, it would be interesting to ascertain if differences continued to exist between PRO and CRO if more detailed sexual dysfunction questions were used.

Further work will be required to ascertain whether the collection of NTE data can be streamlined by using one source of data. Using PRO and CRO subscales that address the same NTE domain with equivalent severity scales will help to understand the true differences between the data collection methods in patients treated with pelvic radiotherapy. Analysis of data from current trials using other PRO scales including the PRO-CTCAE which has been designed to be used as a companion to the CTCAE scoring system may assist in reaching a conclusion to this question in the future.

Conclusion

Correlation between clinician reported outcomes and patient reported outcomes in patient's treated with radiotherapy for bladder cancer are generally poor. NTEs are reported more frequently and at higher severity using PROs than CROs. Further work is needed to evaluate whether PROs alone can be used to assess the comparative impact of trial treatments on normal tissues.

References

1. Cancer Research UK. No Title [Internet]. [cited 2021 Oct 18]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bladder-cancer/diagnosis-and-treatment#heading-Two
2. Cancer Research UK. No Title [Internet]. 2021 [cited 2021 Sep 15]. Available from: https://www.cancerresearchuk.org/about-cancer/causes-of-cancer/age-and-cancer
3. Huddart RA, Hall E, Hussain SA, Jenkins P, Rawlings C, Tremlett J, et al. Randomized noninferiority trial of reduced high-dose volume versus standard volume radiation therapy for muscle-invasive bladder cancer: Results of the BC2001 Trial (CRUK/01/004). Int J Radiat Oncol Biol Phys [Internet]. 2013;87(2):261–9. Available from: http://dx.doi.org/10.1016/j.ijrobp.2013.06.2044
4. Huddart RA, Hall E, Lewis R, Porta N, Crundwell M, Jenkins PJ, et al. Patient-reported Quality of Life Outcomes in Patients Treated for Muscle-invasive Bladder Cancer with Radiotherapy ± Chemotherapy in the BC2001 Phase III Randomised Controlled Trial[Formula presented]. Eur Urol [Internet]. 2020;77(2):260–8. Available from: https://doi.org/10.1016/j.eururo.2019.11.001
5. Henningsohn L, Wijkström H, Dickman PW, Bergmark K, Steineck G. Distressful symptoms after radical radiotherapy for urinary bladder cancer. Radiother Oncol. 2002;62(2):215–25.
6. Zietman AL, Sacco D, Skowronski U, Gomery P, Kaufman DS, Clark JA, et al. Organ conservation in invasive bladder cancer by transurethral resection, chemotherapy and radiation: Results of a urodynamic and quality of life study on long-term survivors. J Urol. 2003;170(5):1772–6.
7. Efstathiou JA, Bae K, Shipley WU, Kaufman DS, Hagan MP, Heney NM, et al. Late pelvic toxicity after bladder-sparing therapy in patients with invasive bladder cancer: RTOG 89-03, 95-06, 97-06, 99-06. J Clin Oncol. 2009;27(25):4055–61.
8. Efficace F, Fayers P, Pusic A, Cemal Y, Yanagawa J, Jacobs M, et al. Quality of patient-reported outcome reporting across cancer randomized controlled trials according to the CONSORT patient-reported outcome extension: A pooled analysis of 557 trials. Cancer. 2015;121(18):3335–42.
9. James ND, Hussain SA, Hall E, Jenkins P, Tremlett J, Rawlings C, et al. Radiotherapy with or without Chemotherapy in Muscle-Invasive Bladder Cancer. N Engl J Med. 2012;366(16):1477–88.
10. Cella BDF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. Development and Validation of the General Measure. J Clin Oncol. 1993;11(3):570–9.
11. FACIT guidelines [Internet]. FACIT administration and scoring [Internet]. Available from: www.facit.org
12. Bhattacharya IS, Haviland JS, Hopwood P, Coles CE, Yarnold JR, Bliss JM, et al. Can patient-reported outcomes be used instead of clinician-reported outcomes and photographs as primary endpoints of late normal tissue effects in breast radiotherapy trials? Results from the IMPORT LOW trial. Radiother Oncol [Internet]. 2019;134:220–30. Available from: https://doi.org/10.1016/j.radonc.2019.01.036
13. McHugh ML. Interrater reliability: the kappa statistic. Biochem medica [Internet]. 2012;22(3):276–82. Available from: https://pubmed.ncbi.nlm.nih.gov/23092060

14. Vistad I, Cvancarova M, Fosså 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(5):1335–42.
15. Mukesh MB, Qian W, Wah Hak CC, Wilkinson JS, Barnett GC, Moody AM, et al. The Cambridge Breast Intensity-modulated Radiotherapy Trial: Comparison of Clinician- versus Patient-reported Outcomes. Clin Oncol [Internet]. 2016;28(6):354–64. Available from: http://dx.doi.org/10.1016/j.clon.2016.02.011
16. Haviland JS, Hopwood P, Mills J, Sydenham M, Bliss JM, Yarnold JR. 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. Clin Oncol [Internet]. 2016;28(6):345–53. Available from: http://dx.doi.org/10.1016/j.clon.2016.01.011
17. 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(23):1624–32.
18. 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(17):3485–90.
19. Basch E, Iasonos A, McDonough T, Barz A, Culkin A, Kris MG, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. Lancet Oncol. 2006;7(11):903–9.
20. Park SB, Kwok JB, Asher R, Lee CK, Beale P, Selle F, et al. Clinical and genetic predictors of paclitaxel neurotoxicity based on patient- versus clinicianreported incidence and severity of neurotoxicity in the ICON7 trial. Ann Oncol [Internet]. 2017;28(11):2733–40. Available from: http://dx.doi.org/10.1093/annonc/mdx491

	FACT BL Question	LENT/SOM Question
Category		
Diarrhoea	" I have diarrhoea	Stool frequency
Bowel Incontinence	" I have control of my bowels"	Sphincter Control
Urinary incontinence	" I have trouble controlling my urine"	Incontinence
Urinary Frequency	"I urinate more frequently than normal"	Frequency
Dysuria	" It burns when I urinate"	Dysuria
Sexual dysfunction	" I am interested in sex"	Desire

Table 1: Paired FACT BL and LENT/SOM subjective questions.

Clinician reported outcome	Patient reported outcome				% agreement (95% CI)	Weighted Kappa (95% CI)
	None	A little bit/somewhat	Quite a bit/very much	Total		
Diarrhoea					131/169	
None	118	23	2	143	78%	0.3045
Grade 1 or 2	11	12	2	25	(70%-84%)	SE 0.069
Grade 3 or 4	0	0	1	1		
Total	129	35	5			
Bowel Incontinence					101/168	
None	97	48	19	164	60%	0.0704
Grade 1 or 2	0	4	0	4	(52%-67%)	SE 0.0248
Grade 3 or 4	0	0	0	0		
Total	97	52	19			
Urinary Frequency					70/156	
None	37	49	17	103	45%	0.187
Grade 1 or 2	2	27	15	44	(40%-53%	SE 0.0470
Grade 3 or 4	0	3	6	9		
Total	39	79	38			
Dysuria					122/159	
None	109	9	24	142	67%	0.3567
Grade 1 or 2	2	13	2	17	(70%-83%)	SE 0.0490
Grade 3 or 4	0	0	0	0		
Total	111	22	26			
Urinary Incontinence					82/159	
None	64	64	7	135	52%	0.1691
Grade 1 or 2	1	15	4	20	(44%-60%)	SE 0.0444
Grade 3 or 4	1	0	3	4		
Total	66	79	14			
Reduced Desire						
None	19	29	53	101	47/139	
Grade 1 or 2	1	3	3	7	34%	0.1195
Grade 3 or 4	0	6	25	31	(26%-42%)	SE 0.0374
Total	20	38	81			

Table 2:Frequencies, percentage agreement and Kappa statistic for CRO versus PRO at year 2

Clinician Reported Outcome	Patient reported outcome				% agreement (95% CI)	Weighted Kappa (95% CI)
	None	A little bit/somewhat	Quite a bit/very much	Total		
Diarrhoea					82/99	
None	79	15	0	94	83%	0.1974
Grade 1 or 2	2	3	0	5	(74%-90%)	SE 0.0793
Grade 3 or 4	0	0	0	0		
Total	81	18	0			
Bowel Incontinence					53/99	
None	53	40	5	98	54%	-0.019
Grade 1 or 2	1	0	0	1	(43%-64%)	SE 0.0212
Grade 3 or 4	0	0	0	0		
Total	54	40	5			
Urinary Frequency					44/93	
None	25	22	9	56	48%	0.2144
Grade 1 or 2	2	14	14	30	(37%-60%)	SE 0.0650
Grade 3 or 4	0	2	5	7		
Total	27	38	28			
Dysuria					55/92	
None	54	11	22	87	60%	0.0096
Grade 1 or 2	3	0	1	4	(49%-70%)	SE 0.0430
Grade 3 or 4	0	0	1	1		
Total	57	11	24			
Urinary Incontinence					39/90	
None	34	39	8	81	43%	0.0709
Grade 1 or 2	0	4	3	7	(33%-54%)	SE 0.0458
Grade 3 or 4	1	0	1	2		
Total	35	43	12			
Reduced desire						
None	5	17	33	55	18/75	
Grade 1 or 2	0	3	4	7	24%	0.0662
Grade 3 or 4	0	3	10	13	(14%-35%)	SE 0.0410
Total	5	23	47			

Table 3:*Frequencies, percentage agreement and Kappa statistic for CRO versus PRO at year 5*

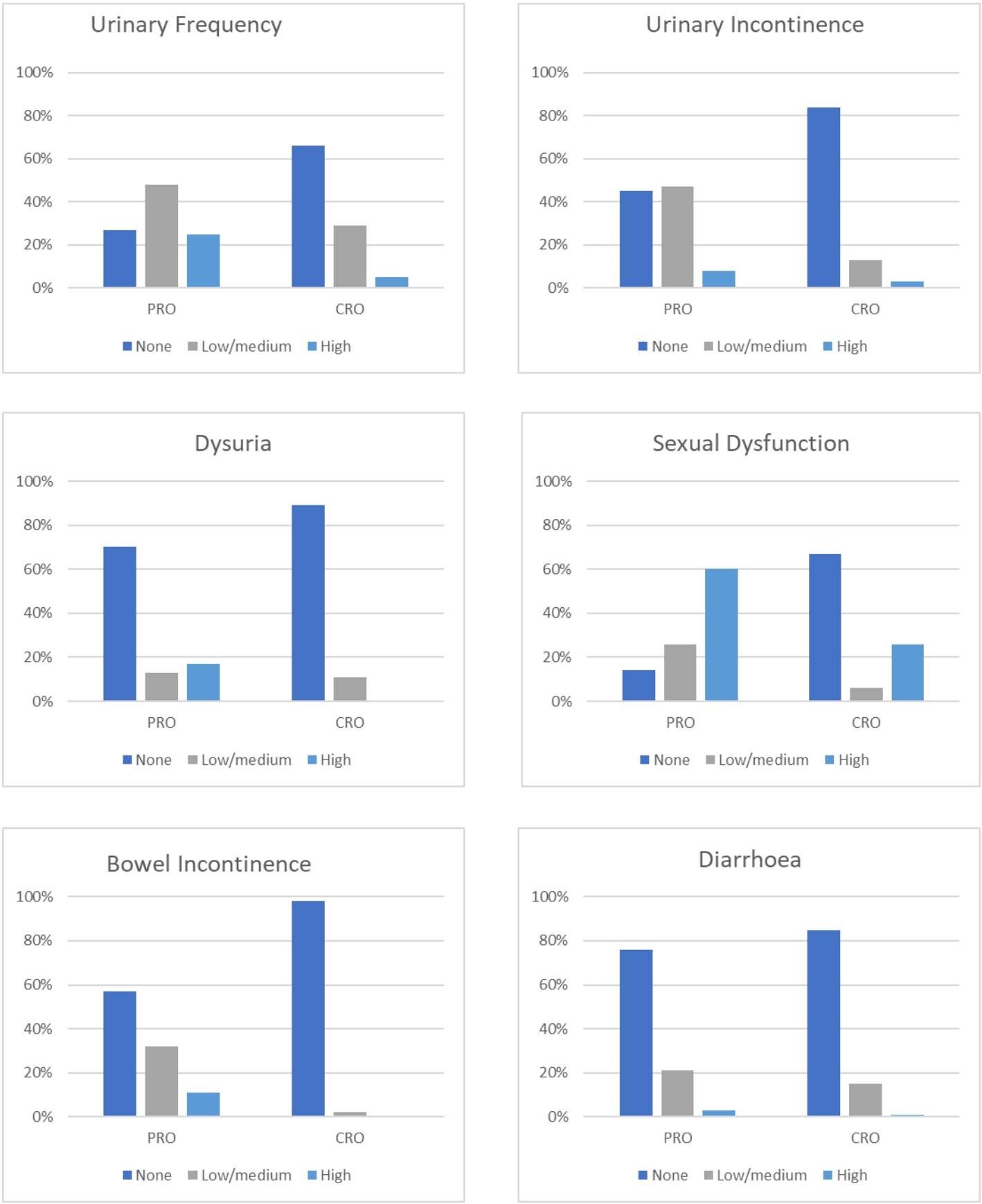

Figure 1: Percentage of none, low/medium and high-level toxicities in PRO and CROs for each question in Year 2.

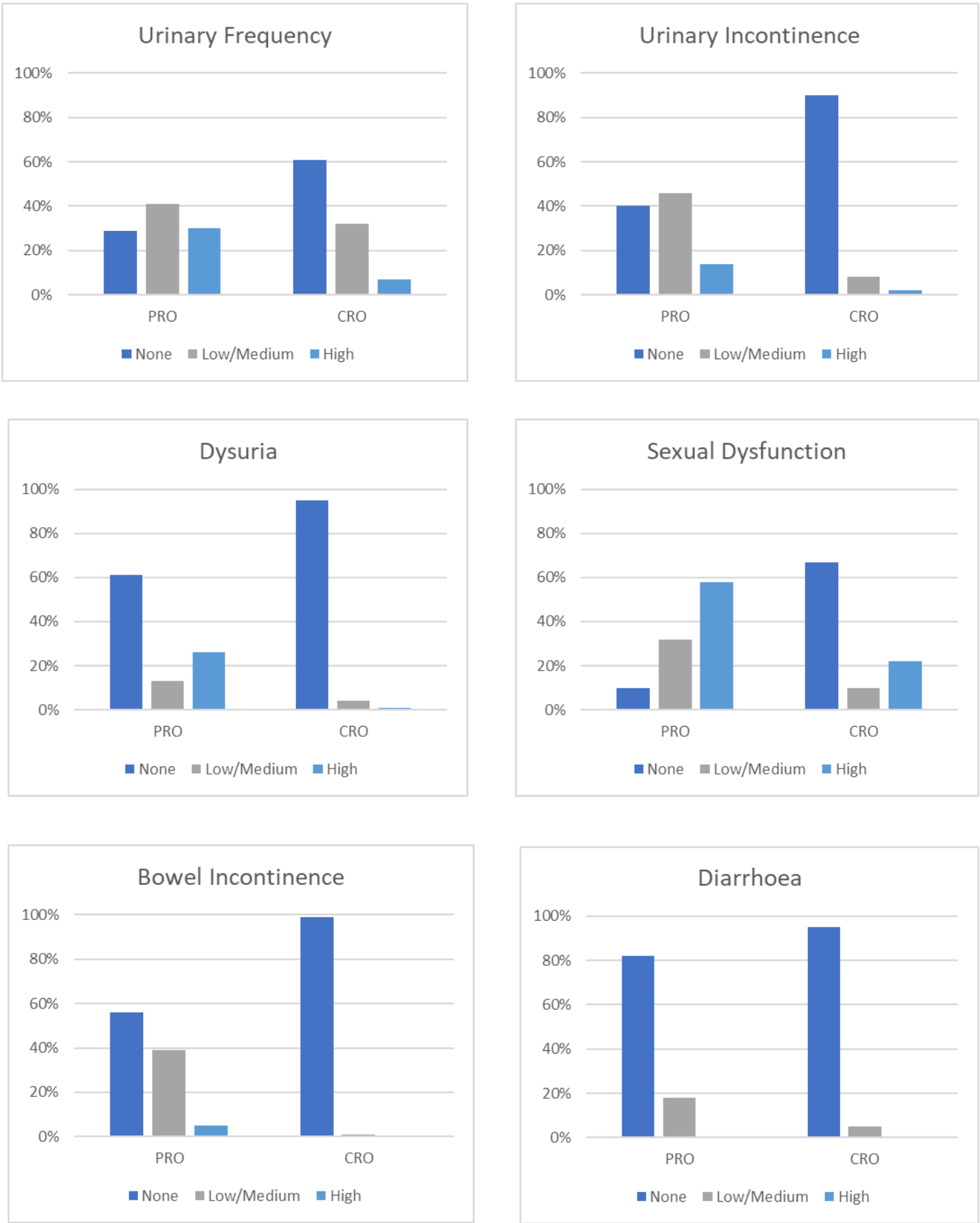

Figure 2: Percentage of none, low/medium and high-level toxicities in PRO and CROs for each question in Year 5.