Health and Quality of Life Outcomes

https://doi.org/10.1186/s12955-025-02460-1

Article in Press

Using machine learning to predict patient-reported symptom clusters in prostate cancer patients receiving radiotherapy

Received: 26 May 2025

Accepted: 25 November 2025

Published online: 29 November 2025

Cite this article as: Rammant E., Deman E., Fonteyne V. *et al.* Using machine learning to predict patient-reported symptom clusters in prostate cancer patients receiving radiotherapy. *Health Qual Life Outcomes* (2025). https://doi.org/10.1186/s12955-025-02460-1

Elke Rammant, Emile Deman, Valérie Fonteyne, Lindsay Poppe, Renée Bultijnck, Piet Dirix, Gert De Meerleer, Karin Haustermans, Ann Van Hecke, Miguel E. Aguado-Barrera, Barbara Avuzzi, David Azria, Jenny Chang-Claude, Barbara N. Chiorda, Ananya Choudhury, Patricia Calvo-Crespo, Dirk De Ruysscher, Antonio Gómez-Caamaño, Philipp Heumann, Ashley M. Hopkins, Kerstie Johnson, Maarten Lambrecht, Alan Mcwilliam, Bradley D. Menz, Filip Poelaert, Tiziana Rancati, Kato Rans, Tim Rattay, Barry S. Rosenstein, Petra Seibold, Jane Shortall, Elena Sperk, Nora Sundahl, Christopher J. Talbot, Ana Vega, Peter Vermeulen, Adam Webb, Catharine M. L. West, Liv Veldeman & Sofie Van Hoecke

Using machine learning to predict patient-reported symptom clusters in prostate cancer patients receiving radiotherapy

Elke Rammant[1], Emile Deman[2], Valérie Fonteyne[1,3], Lindsay Poppe[1], Renée Bultijnck[1,3], Piet Dirix[4], Gert De Meerleer[5], Karin Haustermans[5], Ann Van Hecke[6,7], Miguel E. Aguado-Barrera[8,9], Barbara Avuzzi[10], David Azria[11], Jenny Chang-Claude[12,13], Barbara N. Chiorda[10], Ananya Choudhury[14], Patricia Calvo-Crespo[8,15], Dirk De Ruysscher[16], Antonio Gómez-Caamaño[8, 15], Philipp Heumann[13], Ashley M. Hopkins[17], Kerstie Johnson[18], Maarten Lambrecht[5], Alan Mcwilliam[13], Bradley D. Menz[17], Filip Poelaert[19], Tiziana Rancati[20], Kato Rans[5], Tim Rattay[21], Barry S. Rosenstein[22], Petra Seibold[12], Jane Shortall[13], Elena Sperk[23], Nora Sundahl[24] Christopher J. Talbot[21], Ana Vega[8, 9, 25], Peter Vermeulen[26], Adam Webb [27], Catharine M. L. West[14], Liv Veldeman[1,3], Sofie Van Hoecke[2]

*on behalf of REQUITE consortium

Author affiliations

[1] Department of Human Structure and Repair, Ghent University, Ghent, Belgium.

[2] IDLab, Ghent University - imec, Ghent, Belgium

[3] Department of Radiation Oncology, Ghent University Hospital, Ghent, Belgium.

[4] Department of Radiation Oncology, Iridium Network, Antwerp, Belgium.

[5] Department of Radiation Oncology, Leuven University Hospitals, Leuven, Belgium.

[6] Department of Public Health and Primary Care – University Center for Nursing and Midwifery, Ghent University, Ghent, Belgium

[7] Nursing Department, Ghent University Hospital, Ghent, Belgium

[8] Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago de Compostela, Spain.

[9] Fundación Pública Galega de Medicina Xenómica (FPGMX), Santiago de Compostela, Spain.
[10] Unit of Radiation Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
[11] University Federation of Radiation Oncology of Mediterranean Occitanie, Montpellier Cancer Institute ICM, Université Montpellier, Montpellier, France.
[12] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
[13] University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany.
[14] Division of Cancer Sciences, The University of Manchester, Christie Hospital NHS Foundation Trust, Manchester, United Kingdom
[15] Servicio de Oncología Radioterápica, Hospital Clínico Universitario Santiago de Compostela, A Coruña, Santiago de Compostela, Spain
[16] Department of Radiation Oncology (Maastro), Maastricht University Medical Centre, GROW School for Oncology and Reproduction, Maastricht, The Netherlands
[17] Flinders Health and Medical Research Institute, College of Medicine and Public Health, Flinders University, Bedford Park, SA, Adelaide, Australia
[18] Department of Genetics & Cancer Sciences, University of Leicester, Leicester, UK.
[19] Department of Urology, ZAS, Antwerp, Belgium
[20] Data Science Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
[21] Leicester Cancer Research Centre, University of Leicester, Leicester, United Kingdom.
[22] Departments of Radiation Oncology & Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York City, USA
[23] Department of Radiation Oncology, Mannheim Cancer Center, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany.
[24]Department of Radiation Oncology, AZ Groeninge, Pres. Kennedylaan 4, 8500 Kortrijk, Belgium.

[25] Biomedical Network on Rare Diseases (CIBERER), Madrid, Spain
[26] Oncologisch centrum, Universiteit Antwerpen, Antwerpen
[27] Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom.

Corresponding author
Elke Rammant
Department of Human Structure and Repair, Ghent University
Corneel Heymanslaan 10, 9000 Ghent, Belgium
Elke.rammant@ugent.be

Keywords: prostate cancer; radiotherapy; symptom clusters; patient-reported outcomes; machine learning
Word count of the text: 2742
Word count of the abstract: 250

ABSTRACT

Purpose/Objective

Prostate cancer (PC) survivors frequently experience multiple co-occurring symptoms that adversely affect health-related quality of life (HRQoL). Identifying symptom clusters (SCs) may help to improve symptom management and patient care. The aim of this study is to investigate (1) SCs in prostate cancer (PC) patients, (2) associations of SCs with HRQoL, and (3) predictors of SCs.

Material/Methods

We used data from an international, multi-centre, prospective cohort study (REQUITE). SCs were identified from patient-reported outcomes collected with the EORTC Core Quality of Life questionnaire (EORTC QLQ-C30) and pelvic symptom questionnaires. Machine learning techniques identified SCs, associations with HRQoL and SCs predictors. The dataset was divided into training (80%) and validation (20%) cohorts.

Results

Data were analysed from 1538 (before radiotherapy (T0)), 1490 (end of radiotherapy (T1)), 1322 (12-months (T2)), and 1219 (24-months (T3)) patients. SCs identified at T0: SC1 (gastro-intestinal), SC2 (fatigue, urinary, emotional and cognitive functioning), and SC3 (pain, physical, role, and social functioning). SCs changed at T1: SC1 (gastro-intestinal symptoms), SC2 (fatigue, urinary problems, insomnia), SC3 (social and role functioning), and SC4 (pain, bowel problems, physical, emotional, and cognitive functioning). At T2, symptoms returned to baseline clusters. SCs including 'fatigue' or 'urinary symptoms' were most frequent across time-points. At T0, T2 and T3, HRQoL was best predicted by clusters 2 and 3 (35-45% explained variance). At T1, cluster 4 was the best predictor (52% explained variance). Planned radiotherapy target volume, prostate specific antigen (PSA) at pre-diagnostic biopsy, age and alcohol consumption were the best predictors of SC2 at T1 and SC3 and fatigue-dyspnoea at T3.

Conclusion

Although SCs including fatigue and urinary symptoms were most common, the 'pain, bowel problems, physical, emotional and cognitive functioning' SC at T1 was associated most strongly with HRQoL. The predictors can help to identify men at risk for specific SCs.

INTRODUCTION

Diagnosis and treatment of prostate cancer (PC) can adversely affect patients' health-related quality of life (HRQoL). While the diagnosis is often accompanied by psychological distress and uncertainty, treatment is associated with physical side effects such as urinary, bowel, and sexual dysfunction [1, 2]. Patients with PC are often faced with multiple co-occurring side-effects (e.g. fatigue, urinary and sexual dysfunctions), also referred to as 'symptom clusters (SC)'[3, 4]. SCs are defined as two or more concurrent symptoms that are related and tend to occur together, often due to shared underlying mechanisms. Patient-reported outcomes (PROs) capture patients' subjective experiences of symptoms and functioning and are

essential for assessing HRQoL, making them particularly useful for identifying and analyzing SCs.

Symptoms within a cluster can exacerbate one another, and their combined presence has a stronger negative impact on HRQoL than individual symptoms alone[5]. Therefore, it is important that future interventions address multiple symptoms (rather than tackling single symptoms with different interventions) to achieve greater HRQoL gains and to decrease patient burden[5, 6]. Such interventions primarily target one or two symptoms within a cluster, and alleviate the severity of other symptoms within the cluster because of common underlying mechanisms[6]. Knowledge on SCs will allow clinicians to anticipate on other related symptoms[5]. Hence, patient-provider interaction can become more efficient and overlooked symptoms can be uncovered[5].

Although emerging evidence exists in SCs research[5, 7], most studies in PC have been limited to conventional statistical techniques. Recent studies in PC have identified clusters involving fatigue, urinary symptoms, pain, and emotional distress, highlighting that symptoms often co-occur rather than occurring in isolation[8, 9]. However, it remains unclear how SCs evolve throughout PC survivorship, beyond the acute post-treatment phase. Only a few studies, mostly in mixed cancer populations, have examined changes in SCs during the first year after treatment[7, 10]. Furthermore, identifying predictors of SCs and assessing their associations with patients' overall health-related quality of life (HRQoL) is crucial. Knowledge about which patients are at risk for developing certain SCs can help healthcare providers intervene early and offer more targeted supportive care[11]. Additionally, understanding which SCs are most strongly associated with HRQoL can guide prioritization of symptom management efforts[11].

With the emergence of new data analysis techniques, machine learning can be used to advance the field of SCs research[12]. While traditional methods such as latent class analysis (LCA) have been used to identify symptom clusters in cancer populations, machine learning (ML) clustering offers additional advantages. ML can handle complex, high-dimensional data,

capture non-linear relationships among symptoms, identify clusters without pre-specifying their number, and provide intuitive visualizations (e.g., dendrograms)[13]. Papachristou et al. (2016){Papachristou, 2016 #11} showed that ML algorithms produced clinically meaningful clusters consistent with LCA results, supporting ML as a flexible and powerful tool for exploratory analyses of symptom patterns in prostate cancer patients{Papachristou, 2016 #11}. However, most prior studies using ML or other clustering approaches have been limited to cross-sectional data or small single-centre samples, often without integration of clinical and treatment-related variables. As a result, the temporal stability and clinical correlates of symptom clusters across the treatment trajectory remain unclear. The present study addresses this gap by applying hierarchical clustering and random forest models to a large, multicentre longitudinal dataset, enabling the exploration of both symptom patterns at different timepoints and their associations with HRQoL.

This study used machine learning techniques to identify (1) SCs and their changes over time in PC patients receiving radiotherapy (RT), (2) the associations of SCs with HRQoL, and (3) demographical, clinical and treatment-related predictors of SCs.

METHODS

Study design and patient population

Data were used from the REQUITE study, a multicentre European cohort study designed to identify predictors of radiotherapy-related toxicity and quality of life outcomes.[14]. Patients were recruited at 26 hospitals across eight countries (i.e. Belgium, France, Germany, Italy, Spain, The Netherlands, UK and USA). The REQUITE study was approved by the local Ethical Committees of each participating country (UK NRES Approval 14/NW/0035) and registered at http://www.controlled-trials.com, (ISRCTN98496463; Registration date: 19/03/2014).

Eligible patients were men with PC and due to receive radical RT, brachytherapy and/or postoperative radiotherapy. Further details of the REQUITE study and the eligibility criteria were previously published[14].

Data collection and outcomes

Patients were recruited between April 2014 and October 2016. Healthcare professionals completed a paper-based case report form (CRF) at baseline, including demographical, patient, clinical, radiotherapy and other treatment characteristics. PROs were collected using the EORTC QLQ-C30[15] and the pelvic symptom questionnaire at baseline (i.e. at the start of RT), at the end of RT, at month 12, and month 24 after RT. The primary method of data collection in REQUITE was a paper-based questionnaire after which data was manually entered by study centre personnel.

The EORTC QLQ-C30 is a validated questionnaire to measure cancer-specific HRQoL and includes five functional scales (physical, emotional, role, social, and cognitive functioning), nine cancer-related symptom scales (fatigue, nausea and vomiting, pain, dyspnoea, insomnia, appetite loss, constipation, diarrhoea and financial difficulties) and a global health status scale[15]. In addition, PC specific HRQoL was measured with the pelvic symptom questionnaire developed by working groups of the EORTC and the Radiation Therapy Oncology Group (RTOG) and incorporated into the National Cancer Institute's (NCI) Common Terminology Criteria for Adverse Events (CTCAE)[16]. The pelvic symptom questionnaire is a 16-item questionnaire assessing bowel (nine items: pain, pain on opening bowels, urgency bowels, unable to open bowels, diarrhoea, constipation, bowel incontinence, sticky motions and bowel pads), urinary (six items: pain on passing urine, urinary urgency, urinary frequency, urinary incontinence, nocturia, incontinence pads), and sexual (one item: erectile dysfunction) domains[16].

Statistical analyses

Descriptive statistics were used to describe the demographic, clinical and treatment sample characteristics. Missing item responses within a scale

were handled according to the EORTC scoring manual (if ≥50% of items completed, the mean of available items was used). Missing demographic or clinical data were excluded from the random forest models. Scoring procedures also adhered to the OERTC manual. The numbers (ranging from 1 to 4 on a Likert-scale) were converted into values ranging from 0 to 100 by first calculating a raw score and then performing a linear transformation according to the EORTC scoring manual[17]. A score of 100 can be interpreted as maximal function (e.g. emotional functioning) or maximal symptoms (e.g. fatigue). For further SCs analyses, all PRO scores were transformed into a score from 0 to 1 (relative to the 0 to 100 scale) with 0 being the worst outcome (i.e. minimal function or maximal symptoms) and 1 being the best outcome (i.e. maximal function or minimal symptoms[17]. No further scaling or normalization was required before fitting the random forest models, as tree-based methods are insensitive to the scale of input variables and use splits based on feature ordering rather than absolute values.

Python 3.8.5 with the scikit-learn 0.24.1 package was used for the cluster and random forest analyses. Hierarchical agglomerative clustering was used to cluster correlating symptoms (or domains)[18]. In this type of clustering, symptoms are hierarchically clustered by similarity, using Ward's linkage method with Euclidean distances. Hierarchical agglomerative clustering was chosen because it does not require specifying the number of clusters in advance and produces a dendrogram that visually shows how symptoms group together at different levels of similarity, helping to interpret complex relationships among symptoms{Pasupathi, 2021 #29}. The final clusters were identified based on the dissimilarity values (i.e. a numerical score measuring differences in data scores) in combination with expert input. Associations of SCs with HRQoL were assessed using simple linear regression models. The proportion of variance in HRQoL explained by each SC (R^2) was used to quantify the strength of these associations.

Commented [ER1]: **1. Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th ed.). Wiley.**
This foundational textbook provides comprehensive coverage of hierarchical agglomerative clustering, including its advantages such as not requiring pre-specification of the number of clusters and its ability to reveal nested relationships among data points.
DOI: 10.1002/9780470977811
2. Kim, M., Shanmuganathan, V., Madasamy, K., & Robinson Yesudhas, H. (2021). Trend analysis using agglomerative hierarchical clustering approach for time series big data. *The Journal of Supercomputing, 77*(7), 6505–6524.
This recent study demonstrates the application of agglomerative hierarchical clustering in analyzing complex datasets, highlighting its utility in identifying patterns without predefined cluster numbers.
DOI: 10.1007/s11227-020-03580-9

To uncover the predictors of the symptom clusters, random forest regression models were employed. Random forests were selected due to their robustness to multicollinearity, ability to capture non-linear relationships, and capacity to provide interpretable feature importance scores. For each cluster at each time point, a symptom cluster score was calculated as the average of the constituent symptoms for each patient. Random forest regression models were then trained to predict the cluster score from the demographic, clinical, and treatment characteristics detailed in Table 1.

Hyperparameter optimization was performed via grid search, initially considering the number of estimators, maximum depth, minimum samples split, maximum samples, and maximum features. After preliminary assessment, only the number of estimators, maximum features, and minimum samples split were retained for optimization. The dataset was split into training (80%) and test (20%) sets; no cross-validation was applied during grid search to limit computational time, but a 5-fold cross-validation was used when calculating explained variance scores to evaluate model performance.

Variable importance scores were calculated to quantify the relative contribution of each predictor. For random forest models, impurity-based feature importances were used, while for linear regression models, normalized input coefficients were used. Higher scores indicate stronger predictive influence. Patients with incomplete records (i.e., demographic, clinical, and treatment characteristics) were excluded from the random forest analyses. Although imputation of missing data could have been considered, the proportion of missing data was minimal, and imputation may introduce distortion; therefore, complete-case analysis was performed. Finally, the explained variance of the models, ranging from -1 to 1, was used to assess model accuracy, with higher values indicating better predictive performance. For each model, the ten most influential predictors of the symptom clusters were reported.

RESULTS

Data from 1538, 1490, 1322, and 1219 PC patients were analysed at baseline, the end of RT, 12 and 24 months after RT, respectively. Table 1 (patient and treatment characteristics) summarises the patient characteristics.

[insert Table 1 Patient and treatment characteristics]

SCs at the different time-points

Based on hierarchical clustering and clinical considerations three to four SCs were identified at the different time-points which are presented in the dendrograms in Figure 1.

At baseline and 12 months after RT, three SCs were identified: cluster 1a (gastro-intestinal symptoms), cluster 2a (fatigue, urinary symptoms, emotional and cognitive functioning), and cluster 3a (pain, physical, role, and social functioning). Insomnia, erectile dysfunction and dyspnoea did not cluster together with other symptoms.

At the end of RT, changes in SCs since baseline were seen and four SCs were identified: cluster 1b (gastro-intestinal symptoms without diarrhoea and bowel problems), cluster 2b (fatigue, urinary problems, insomnia), cluster 3b (social and role functioning), and a new cluster was identified (i.e. cluster 4: pain, bowel problems, physical, emotional, and cognitive functioning). Diarrhoea, erectile dysfunction, and dyspnoea did not cluster together with other symptoms.

At month 24 after RT, SCs were almost identical as at baseline and month 12. Cluster 1a (gastro-intestinal symptoms) and cluster 3a (pain, physical, role, and social functioning) remained the same but ‘fatigue’ left cluster 2c and formed another cluster together with ‘dyspnoea’ (i.e. cluster 5). Insomnia and erectile dysfunction did not cluster together with other symptoms.

Prevalence of the SCs

SCs including ‘fatigue’ or ‘urinary symptoms’ had the highest prevalence across time-points. Figure 2 presents boxplots with distribution, mean and

median scores of the different clusters over time, calculated on the complete dataset.

Associations of SCs with HRQOL

At baseline and 12 months after radiotherapy cluster 2 (41% and 39% explained variance) and cluster 3 (33% and 41% explained variance) had the strongest correlation with patients' overall HRQoL. At the end of RT, cluster 4 (50% explained variance) correlated most with overall HRQoL followed by cluster 3 (17% explained variance) and 2 (14% explained variance). At month 24, cluster 3 and 2 (both 30% explained variance) had the strongest correlation followed by cluster 1 (22% explained variance).

Figure 3 shows a visual representation of the relationship between the prevalence of SCs and their impact on HRQoL.

Predictors of SCs

The explained variances, calculated on the test dataset (20% of the total cohort), to evaluate the performance of the different random forest regression models, were low across the different SCs and time-points. However, the explained variances showed some variability between SCs and timepoints: SC 2b (i.e. fatigue, urinary symptoms, insomnia) at the end of radiotherapy and SC 3a (i.e. pain, physical, role, and social functioning) and 5 (i.e. dyspnoea, fatigue) at month 24 after radiotherapy showed the highest explained variances (i.e. 0.10828, 0.15416, and 0.08663, respectively) meaning they could be best predicted by the included predictors in this study (as detailed in table 1). Planned radiotherapy target volume, PSA at pre-diagnostic biopsy, age and current alcohol consumption were the best predictors across all SCs and time-points (higher scores predict higher frequency of the SCs). Details of the prediction models can be found in supplementary file 1.

DISCUSSION

This study adds evidence to the limited body of research concerning long-term SCs in PC patients. While minor changes in cluster composition were observed at the end of radiotherapy and at month 24, the overall structure

of symptom clusters remained largely stable from baseline throughout follow-up. SCs including fatigue and urinary symptoms were most common across time-points. The 'pain, bowel problems, physical, emotional and cognitive functioning' cluster at the end of radiotherapy had the worst impact on the patients' general HRQoL. Planned radiotherapy target volume, PSA at pre-diagnostic biopsy, age and current alcohol consumption were the best predictors of the SCs.

Although current SCs research in PC is scarce, the two most discussed SCs in the literature are the 'depression, fatigue, pain'-cluster[19] and the 'fatigue, emotional distress, PC specific symptoms (e.g. urinary symptoms)'-cluster[3]. In those studies, fatigue clusters together with symptoms such as urinary problems, emotional distress/depression and/or pain. The role of inflammation in the associations between these symptoms has been described[4]. Results of our study show that fatigue and urinary symptoms cluster together at all time-points, except at 24 months after radiotherapy. In contrast to the findings in other studies[3, 4, 19], pain did not cluster with fatigue and only clustered with emotional functioning at the end of radiotherapy. Another large study using the EORTC scales of the QLQ-C30 found that pain was clustering with fatigue, physical- and role functioning, which is partly in line with our findings[20]. Variations among studies can be attributed to distinct clustering methodologies, variations in the questionnaires employed to evaluate symptoms, or disparities in treatment modalities and population demographics.

The correlation between urinary symptoms and fatigue found in our study is not surprising as previous literature also shows that urinary dysfunction is a strong predictor of persistent fatigue in men with PC[4, 21]. The link between fatigue, urinary problems, and insomnia at the end of radiotherapy treatment may be related to nocturia, which is a common urinary symptom during and after radiotherapy. Hence, nocturia is generally known to be a common cause for fatigue and poor sleep quality in the elderly[4].

Erectile dysfunction did not cluster with other symptoms at any time point, indicating that it may represent a distinct dimension of prostate cancer survivorship. Despite this, erectile dysfunction was among the most prevalent symptoms reported, highlighting its clinical significance. The lack of clustering suggests that erectile dysfunction may not share the same underlying mechanisms as other commonly co-occurring symptoms such as fatigue, urinary, or emotional symptoms, and therefore may require targeted assessment and management strategies independent of other symptom clusters.

Emotional functioning clusters together with fatigue and urinary symptoms at baseline and 12 months after radiotherapy, which is similar to the findings of Malaski et al[3]. Men with PC who encounter urinary incontinence might be concerned about odors or urine leakage and may feel embarrassed about using incontinence pads. This, in turn, can contribute to social isolation and a decreased level of emotional functioning. Notably, the use of duloxetine, a serotonin noradrenergic reuptake inhibitor, as a treatment for depression has demonstrated an improvement in urinary symptoms[4]. In line with these results, our present study shows a correlation between emotional functioning and urinary symptoms.

In agreement with the current literature[22, 23], our study found that emotional functioning clusters together with bowel symptoms at the end of radiotherapy. Bowel symptoms are an aggravating burden for men with PC which can be reflected in the fact that this SC (i.e. emotional-, cognitive-, physical functioning, bowel symptoms, pain) had the strongest impact on HRQoL. Although it is not the most common SC among PC patients, the high impact on HRQoL calls for screening patients for these symptoms and initiating supportive treatments to alleviate these symptoms early on.

Kwekkeboom et al.[24] summarized the available evidence on guideline-recommended symptom management strategies that cross over two or more cancer symptoms. A total of 88 symptom management strategies were recommended across two or more symptoms[24]. However, only a limited number of studies have specifically designed and tested an

intervention to manage specific SC[5]. More research is warranted to test the impact of interventions (e.g. behavioural, psychological or exercise support, pain medication) on specific SCs[24].

PSA at pre-diagnostic biopsy and the planned radiotherapy target volume have been identified as common predictors of the SCs. This association is logically grounded in the fact that a more severe disease often necessitates more aggressive treatment strategies, potentially leading to increased treatment-related toxicity. However, other studies have demonstrated that the difference in toxicity between prostate (bed) only radiotherapy and prostate (bed) + whole pelvis radiotherapy is generally minimal, and the increase in toxicity associated with a larger treatment field was not observed[25]. The present study, however, suggests that the planned radiotherapy target volume predicts SCs that likely impact HRQoL.

One of the limitations of this study is that the datasets were not always complete for every patient. For machine learning, the number of complete records (i.e. patients) should be as high as possible while limiting the number of feature inputs (i.e. treatment and demographic data). Missing data were handled by case-wise deletion, and no formal assessment of missingness mechanisms or imputation methods was applied, which may have introduced attrition bias. Despite this limitation, the large and diverse sample across several countries enhances the robustness and generalizability of the study.Also, it was surprising that androgen-deprivation therapy (ADT) did not predict SCs including symptoms such as fatigue, insomnia or emotional functioning. However, this is potentially because a large proportion of our cohort (68%) received ADT. More variation was seen in our cohort for length of hormone therapy (i.e. the longer the duration of ADT, the higher the frequency of SCs) which was commonly present in the top 10 of the predictors.

In addition, the explained variance of the random forest models was relatively low (<0.2 for most clusters), indicating that the predictive models currently have limited clinical utility. Future studies could improve prediction by incorporating additional patient-level or treatment-related

variables, using longitudinal modeling approaches, or validating findings in independent cohorts. Our clustering analyses were performed cross-sectionally at each time point and therefore do not capture within-patient symptom trajectories over time. Longitudinal clustering or time-series predictive models could provide additional insights into symptom evolution. Finally, as patients contributed repeated measures, within-subject correlations were not explicitly modeled. Future research could apply mixed-effects or multilevel approaches to account for these dependencies and better understand individual symptom dynamics.

CONCLUSION

This study aimed to identify symptom clusters in prostate cancer patients during radiotherapy and survivorship, examine their associations with overall HRQoL, and explore potential predictors. We found that distinct symptom clusters can be identified at multiple time points, with some clusters most strongly associated with overall HRQoL, as indicated by the highest explained variance in regression models. While clusters including fatigue and urinary symptoms were most common, the cluster comprising pain, bowel problems, and impairments in physical, emotional, and cognitive functioning at the end of radiotherapy had the strongest association with HRQoL. Patient, clinical, and treatment characteristics, including higher planned radiotherapy target volume, higher PSA at pre-diagnostic biopsy, older age, and higher alcohol consumption, partially predicted cluster occurrence. These findings suggest that patients with these risk factors could benefit from targeted supportive care interventions, such as behavioural, psychological, or exercise support, and pain management.

List of abbreviations

Abbreviation	Full term
ADT	Androgen-deprivation therapy
C30	Core Quality of Life questionnaire (EORTC QLQ-C30)
CI	Confidence interval
CRF	Case report form
CTCAE	Common Terminology Criteria for Adverse Events
EORTC	European Organisation for Research and Treatment of Cancer
HRQoL	Health-related quality of life
LCA	Latent class analysis
ML	Machine learning
NCI	National Cancer Institute
PC	Prostate cancer
PRO	Patient-reported outcome
PSA	Prostate-specific antigen
QoL	Quality of life
REQUITE	Radiogenomics: Establishing a Quantitative Imaging and Tissue Biomarker Database for Radiotherapy Toxicity
RTOG	Radiation Therapy Oncology Group
RT	Radiotherapy
SC	Symptom cluster
SCs	Symptom clusters
SD	Standard deviation
T0	Baseline (before radiotherapy)
T1	End of radiotherapy
T2	12 months after radiotherapy
T3	24 months after radiotherapy

Declarations

Ethics approval and consent to participate

The study was carried out in compliance with the Helsinki Declaration. The REQUITE study was approved by the local Ethical Committees of each participating country (UK NRES Approval 14/NW/0035). Prior to participation, all individuals were thoroughly informed about the study's objectives and methods. Written consent was obtained from the participants before the commencement of data collection.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

Nora Sundahl reports speakers fees from Janssens-CILAG.

The salary of Ashley M. Hopkins is supported by a National Health and Medical Research Council (NHMRC) of Australia Investigator Grant. Ashley M. Hopkins reports receiving funding from Tour De Cure and The Hospital Research Foundation outside of the study. Ashley M. Hopkins reports grant funding from Boehringer Ingelheim outside of the submitted work.

Funding

This research was supported by the REQUITE project (www.requite.eu) sharing data. REQUITE received funding from the European Union's Seventh Framework Programme for research, technological development,

and demonstration under grant agreement no. 601826. In addition, this specific study was supported by a grant received from the Belgian foundation 'Kom op Tegen Kanker'.

Authors' contributions

ER, LV and RB contributed to the conception and design of the study. Data were acquired by ER, LV, VF, RB, PD, GDM, KH, MEA-B, BA, DA, JC-C, BCN, AC, PCC, DDR, AG-C, PH, AH, KJ, ML, AM, BDM, FP, TR, KR, TRa, BSR, PS, JS, ES, CT, AV, AW, and CMW. Data were analyzed and interpreted by ER, ED, SVH, VF, LP, RB, PD, and GDM. The manuscript was drafted by ER and ED. ER, ED, VF, LP, RB, PD, GDM, KH, AVH, MEA-B, BA, DA, JC-C, BCN, AC, PCC, DDR, AG-C, PH, AH, KJ, ML, AM, BDM, FP, TR, KR, TR, BSR, PS, JS, ES, NS, CT, AV, PV, AW, CMW, LV, and SVH critically revised the manuscript for important intellectual content. Statistical analyses were performed by ED and SVH. All authors read and approved the final manuscript.

Acknowledgement

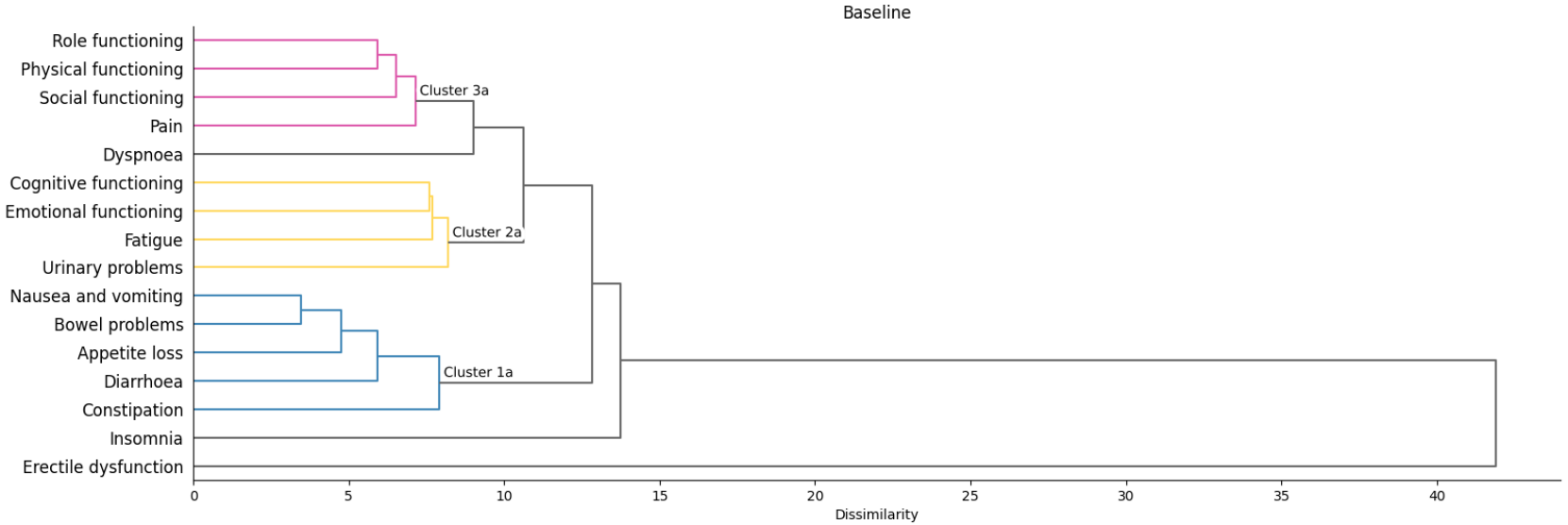
We thank all patients who participated in the study and the staff members at participating hospitals. Ana Vega: supported by Spanish Instituto de Salud Carlos III (ISCIII) funding, an initiative of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development FEDER Funds (PI22/00589, INT24/00023, DTS24/00083); the ERAPerMed JTC2018 funding (AC18/00117); the Autonomous Government of Galicia (Consolidation and structuring program: IN607B), and by the AECC (PRYES211091VEGA).

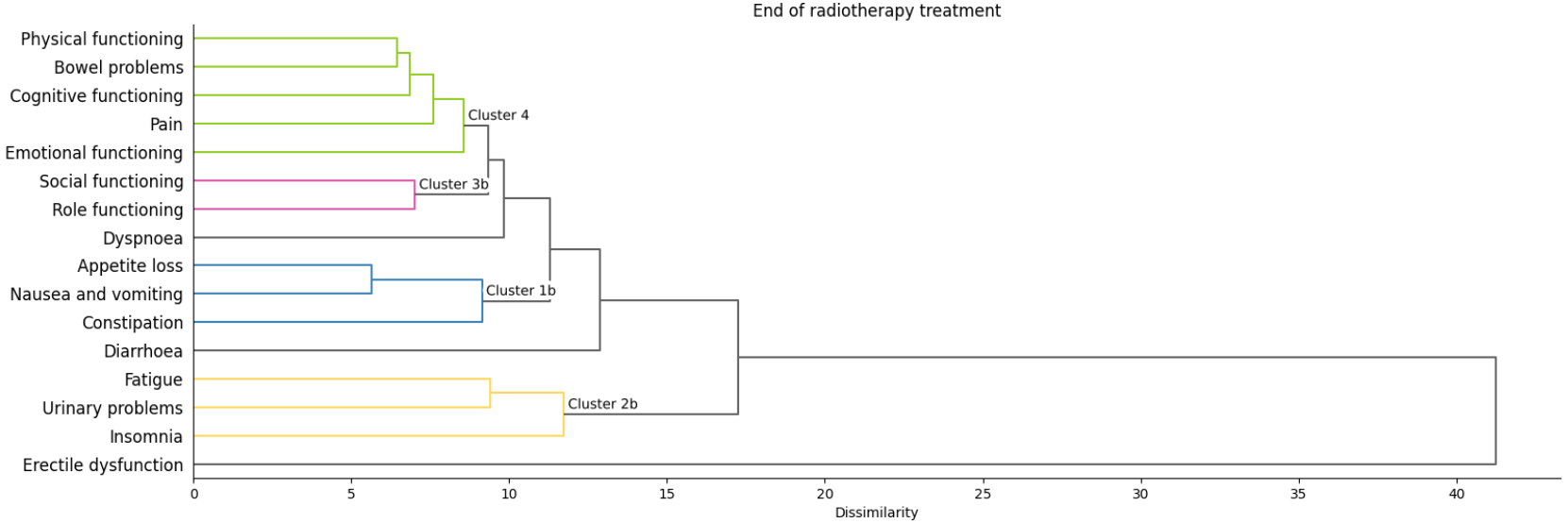
REFERENCES

1. Nakamura K, Konishi K, Komatsu T, et al. Quality of life after external beam radiotherapy for localized prostate cancer: Comparison with other modalities. International Journal of Urology. 2019;26(10):950–954.
2. Sureda A, Fumadó L, Ferrer M, et al. Health-related quality of life in men with prostate cancer undergoing active surveillance versus radical prostatectomy, external-beam radiotherapy, prostate brachytherapy and reference population: a cross-sectional study. Health and quality of life outcomes. 2019;17(1):1–9.
3. Maliski SL, Kwan L, Elashoff D, et al. Symptom clusters related to treatment for prostate cancer. Oncol Nurs Forum 2008:786–793.
4. Feng LR, Fuss T, Dickinson K, et al. Co-occurring symptoms contribute to persistent fatigue in prostate cancer. Oncology. 2019;96(4):183–191.
5. Kwekkeboom KL. Cancer symptom cluster management. Seminars in oncology nursing: Elsevier 2016:373–382.
6. So WK, Law BM, Chan DN, et al. The effect of nonpharmacological interventions on managing symptom clusters among cancer patients: a systematic review. Cancer Nursing. 2020;43(6):E304–E327.
7. Molassiotis A, Wengström Y, Kearney N. Symptom cluster patterns during the first year after diagnosis with cancer. Journal of pain and symptom management. 2010;39(5):847–858.
8. Liu Y, Yan Q, Zhou J, et al. Identification of distinct symptom profiles in prostate cancer patients with cancer-related cognitive impairment undergoing androgen deprivation therapy: A latent class analysis. Asia-Pacific Journal of Oncology Nursing. 2024;11(6):100497.
9. Adam S, Thong MS, Martin-Diener E, et al. Identifying classes of the pain, fatigue, and depression symptom cluster in long-term prostate cancer survivors—results from the multi-regional Prostate Cancer Survivorship Study in Switzerland (PROCAS). Supportive care in cancer. 2021;29(11):6259–6269.

574 10. Thomas BC, Waller A, Malhi RL, et al. A longitudinal analysis of symptom
575 clusters in cancer patients and their sociodemographic predictors. Journal of Pain
576 and Symptom Management. 2014;47(3):566–578.
577 11. Miaskowski C, Barsevick A, Berger A, et al. Advancing symptom science
578 through symptom cluster research: Expert panel proceedings and
579 recommendations. JNCI: Journal of the National Cancer Institute. 2017;109(4).
580 12. Papachristou N, Miaskowski C, Barnaghi P, et al. Comparing machine
581 learning clustering with latent class analysis on cancer symptoms' data. 2016
582 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT):
583 IEEE 2016:162–166.
584 13. Atkinson JG, Atkinson EG. Machine Learning and Health Care: Potential
585 Benefits and Issues. The Journal of Ambulatory Care Management.
586 2023;46(2):114–120.
587 14. Seibold P, Webb A, Aguado-Barrera ME, et al. REQUITE: a prospective
588 multicentre cohort study of patients undergoing radiotherapy for breast, lung or
589 prostate cancer. Radiotherapy and Oncology. 2019;138:59–67.
590 15. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization
591 for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for
592 use in international clinical trials in oncology. JNCI: Journal of the National Cancer
593 Institute. 1993;85(5):365–376.
594 16. Farnell DJ, Mandall P, Anandadas C, et al. Development of a patient-
595 reported questionnaire for collecting toxicity data following prostate
596 brachytherapy. Radiotherapy and Oncology. 2010;97(1):136–142.
597 17. Fayers P, Aaronson NK, Bjordal K, et al. EORTC QLQ-C30 scoring manual:
598 European Organisation for research and treatment of cancer 2001.
599 18. Lukasová A. Hierarchical agglomerative clustering procedure. Pattern
600 Recognition. 1979;11(5-6):365–381.
601 19. Baden M, Lu L, Drummond FJ, et al. Pain, fatigue and depression symptom
602 cluster in survivors of prostate cancer. Supportive Care in Cancer.
603 2020;28(10):4813–4824.
604 20. Machingura A, Taye M, Musoro J, et al. Clustering of EORTC QLQ-C30
605 health-related quality of life scales across several cancer types: Validation study.
606 European Journal of Cancer. 2022;170:1–9.
607 21. Jakobsson L, Lovén L, Hallberg IR. Micturition problems in relation to quality
608 of life in men with prostate cancer or benign prostatic hyperplasia: comparison
609 with men from the general population. Cancer Nurs. 2004;27(3):218–229.
610 22. Orom H, Biddle C, Underwood W, 3rd, et al. Worse Urinary, Sexual and
611 Bowel Function Cause Emotional Distress and Vice Versa in Men Treated for
612 Prostate Cancer. J Urol. 2018;199(6):1464–1469.
613 23. Sharp L, O'Leary E, Kinnear H, et al. Cancer-related symptoms predict
614 psychological wellbeing among prostate cancer survivors: results from the
615 PiCTure study. Psycho-Oncology. 2016;25(3):282–291.
616 24. Kwekkeboom K. Guideline-recommended symptom management
617 strategies that cross over two or more cancer symptoms. Number 5/September
618 2020. 2020;47(5):498–511.
619 25. Murthy V, Maitre P, Kannan S, et al. Prostate-Only Versus Whole-Pelvic
620 Radiation Therapy in High-Risk and Very High-Risk Prostate Cancer (POP-RT):
621 Outcomes From Phase III Randomized Controlled Trial. J Clin Oncol.
622 2021;39(11):1234–1242.

623


Baseline
Role functioning
Physical functioning
Social functioning
Pain
Dyspnoea
Cognitive functioning
Emotional functioning
Fatigue
Urinary problems
Nausea and vomiting
Bowel problems
Appetite loss
Diarrhoea
Constipation
Insomnia
Erectile dysfunction
Cluster 3a
Cluster 2a
Cluster 1a
0
5
10
15
20
25
30
35
40
Dissimilarity


End of radiotherapy treatment
Physical functioning
Bowel problems
Cognitive functioning
Pain
Emotional functioning
Social functioning
Role functioning
Dyspnoea
Appetite loss
Nausea and vomiting
Constipation
Diarrhoea
Fatigue
Urinary problems
Insomnia
Erectile dysfunction
Cluster 4
Cluster 3b
Cluster 1b
Cluster 2b
0
5
10
15
20
25
30
35
40
Dissimilarity

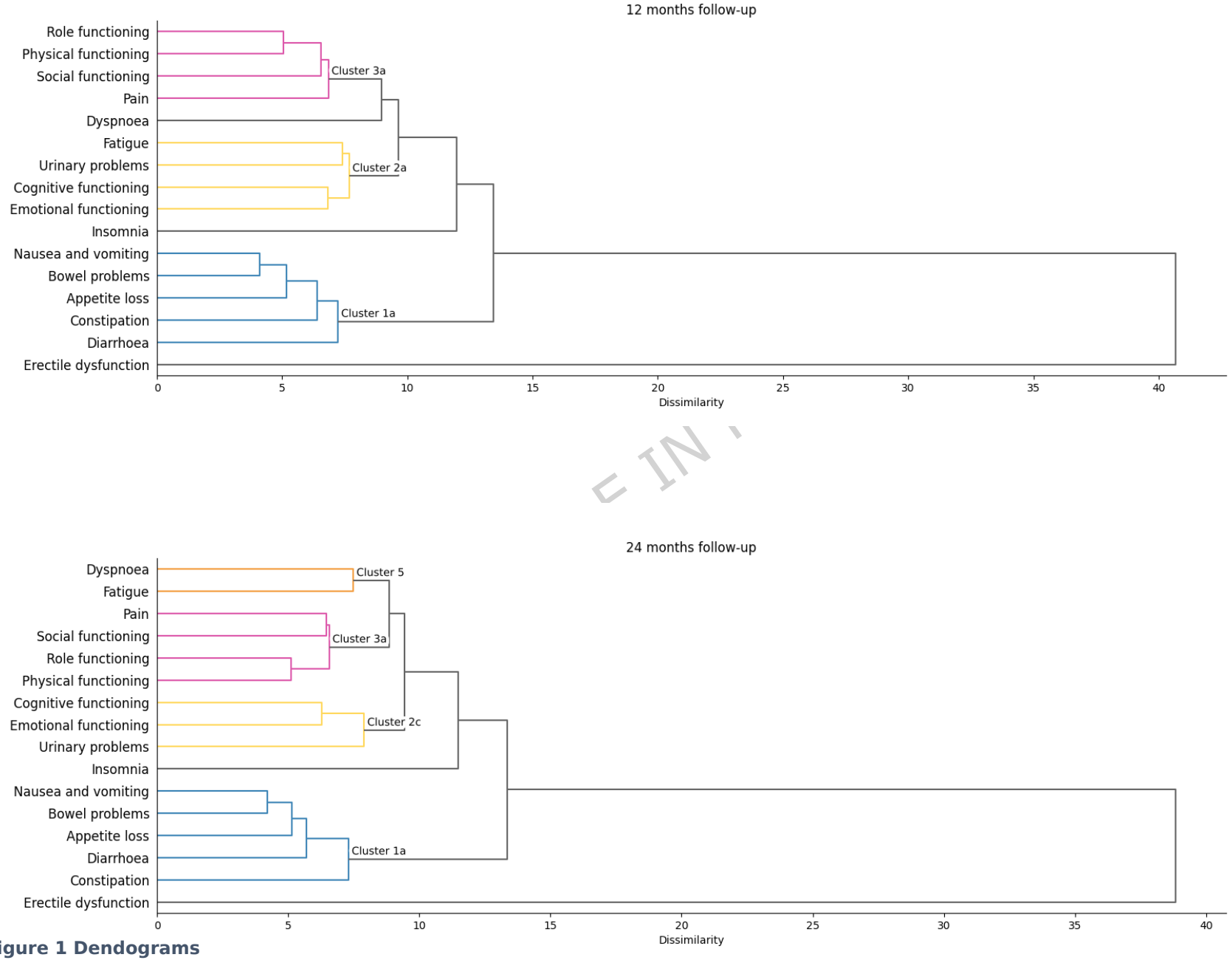


Figure 1 Dendograms

Legend: The pelvic symptom questionnaire was used to assess bowel and urinary symptoms. The EORTC QLQ-C30 was used for the other symptoms/domains of functioning. Dissimilarity is a numerical score measuring how different two data points are.

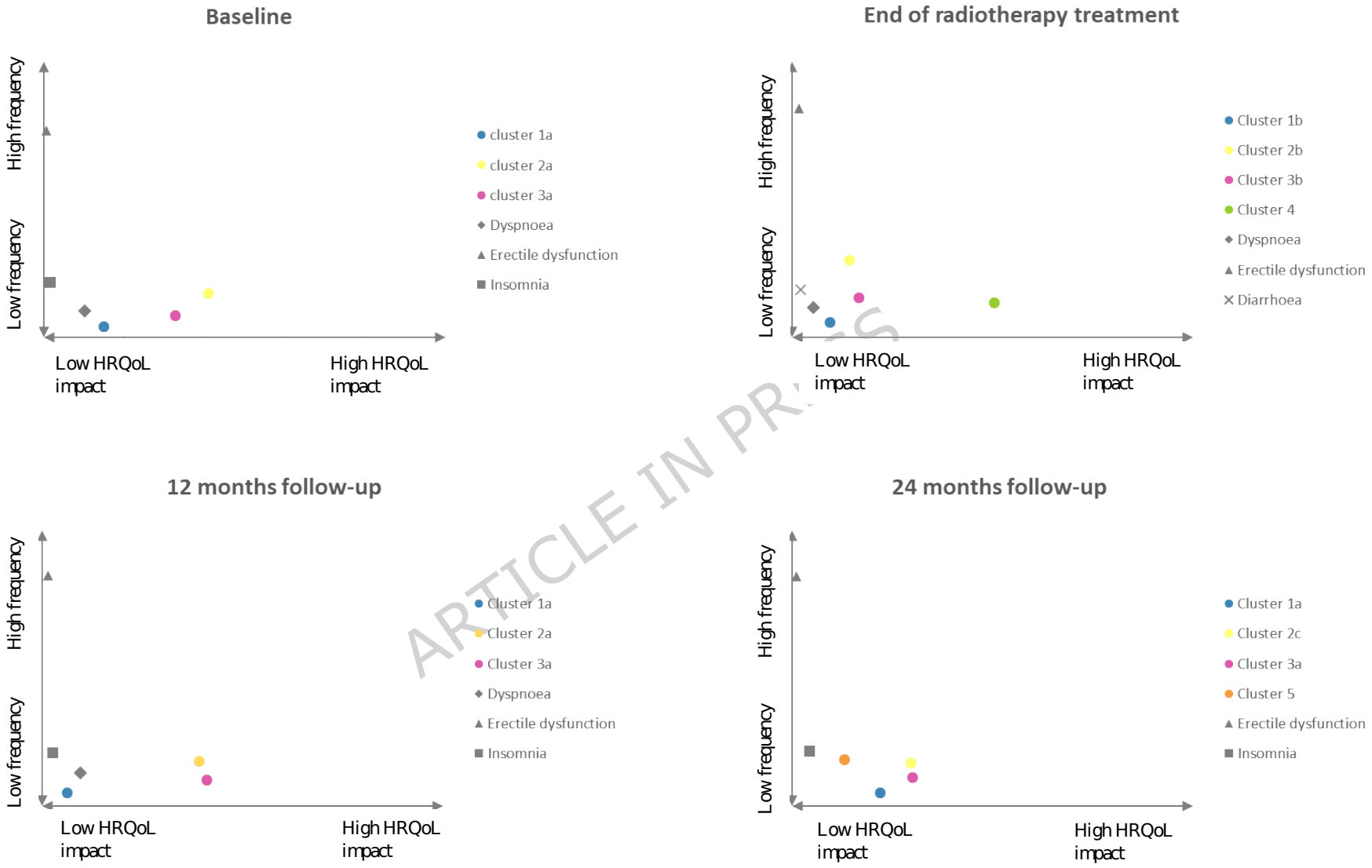


Figure 3 visual representation of prevalence and impact on HRQoL of symptom clusters plotted together

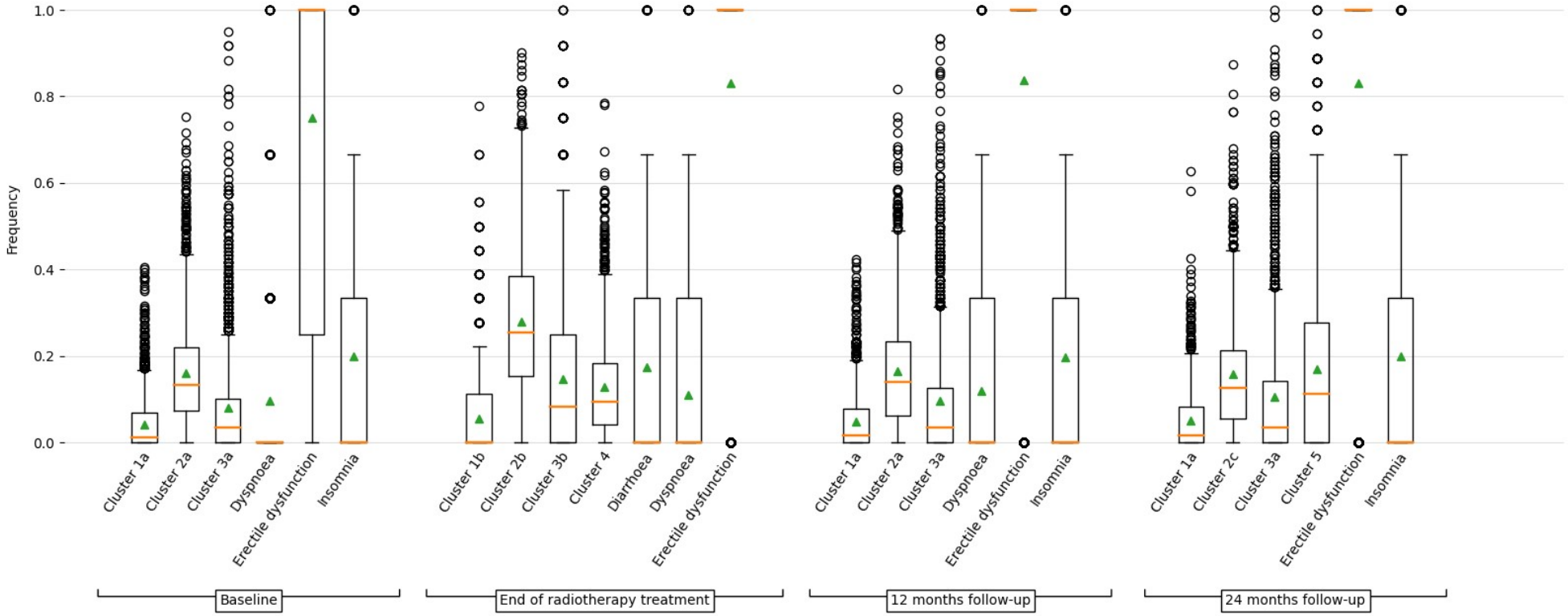


Figure 2 boxplots with distribution, mean and median scores of the different symptom clusters and single symptoms which did not cluster over time (this is a calculation based on the complete dataset).

Tabel 1 Patient and treatment characteristics

Patient characteristics (n=1538)		
		Mean (Std. Dev.)
Age (years)		69 (7)
Number of alcoholic drinks/week		9 (9)
Number of household members		2 (1)
		N(%)
BMI	<18	6 (0)
	18-25	448 (29)
	>25	1084 (70)
Ethnicity	White	1492 (97)
	Mixed	11 (0.7)
	Black or African American	15 (1)
	Asian	10 (0.7)
	Hispanic or Latino	1 (0.1)
	Other	2 (0.1)
Clinical characteristics (n=1538)		
		N(%)
M status	M0	1217 (79)
	M+	321 (21)
Clinical T stage	T1	421 (27)
	T2	582 (38)
	T3	214 (14)
	T4	17 (1)
Gleason Score	< 7	
	7	
	> 7	
		Mean (Std. Dev.)
Prostate Specific Antigen (ng/mL) pre-diagnostic biopsy		15 (24)
Comorbidities/medication use (n=1538)		
		N(%)
Cardiovascular disease		1022 (66)
Inflammatory bowel or diverticular disease		81 (5)
Haemorrhoids		322 (21)
Depression		118 (8)
Rheumatoid Arthritis		27 (2)
Diabetes		197 (13)
Use of analgesics		127 (8)
Radiotherapy characteristcs (n=1490)		
		N(%)
Radiotherapy type	3D conformal RT	265 (18)
	IMRT	196 (13)
	Rapid arc/VMAT	962 (65)
Pelvic radiotherapy		427 (29)
Brachytherapy		156 (10)
		Mean (Std. Dev.)

Radiotherapy dose per fraction (Gy/d)	2.2 (0.4)
Complete delivered dose of external beam radiation (Gy total dose)	69.2 (10.1)
Fractions per week	5 (0)
Number of fractions	33 (8)
PTV (cm3)	149 (89)
Other treatments (n=1538)	
	N(%)
Hormone therapy	1045 (68)
Lymphadenectomy	303 (20)
Radical prostatectomy	447 (29)
Bladder TUR	29 (2)
Previous abdominal surgery	524 (34)
	Mean (Std. Dev.)
Length of hormone therapy (months)	13 (13)

N(%) may not add up to 100 % due to missing data.
Abbreviations: Std. Dev., standard deviation; BMI, body mass index; IMRT, intensity modulated radiotherapy; VMAT, Volumetric modulated arc therapy; PTV, planned target volume; TUR, transurethral resection