Predictive Modelling of Patient Reported Radiotherapy-Related Toxicity by the Application of Symptom Clustering and Autoregression

A. Lemanska[1,*], A. Cox[1], N.F. Kirkby[2], T. Chen[2] and S. Faithfull[1]

[1]*School of Health Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford GU2 7XH, UK*

[2]*Department of Chemical and Process Engineering, University of Surrey, Guildford GU2 7XH, UK*

Abstract: Patient reported outcome measures (PROMs) are increasingly being used in research to explore experiences of cancer survivors. Techniques to predict symptoms, with the aim of providing triage care, rely on the ability to analyse trends in symptoms or quality of life and at present are limited. The secondary analysis in this study uses a statistical method involving the application of autoregression (AR) to PROMs in order to predict symptom intensity following radiotherapy, and to explore its feasibility as an analytical tool. The technique is demonstrated using an existing dataset of 94 prostate cancer patients who completed a validated battery of PROMs over time. In addition the relationship between symptoms was investigated and symptom clusters were identified to determine their value in assisting predictive modeling. Three symptom clusters, namely urinary, gastrointestinal and emotional were identified. The study indicates that incorporating symptom clustering into predictive modeling helps to identify the most informative predictor variables. The analysis also showed that the degree of rise of symptom intensity during radiotherapy has the ability to predict later radiotherapy-related symptoms. The method was most successful for the prediction of urinary and gastrointestinal symptoms. Quantitative or qualitative prediction was possible on different symptoms. The application of this technique to predict radiotherapy outcomes could lead to increased use of PROMs within clinical practice. This in turn would contribute to improvements in both patient care after radiotherapy and also strategies to prevent side effects. In order to further evaluate the predictive ability of the approach, the analysis of a larger dataset with a longer follow up was identified as the next step.

Keywords: Predictive Modeling, Patient Reported Outcome Measures, Autoregression, Radiotherapy-Related Side Effects, Longitudinal Study.

INTRODUCTION

Cancer is increasingly being viewed as a chronic disease with the number of patients surviving for 10 years having doubled over the last 40 years [1]. Traditional cancer services have attempted to adapt to this change with innovative models of surveillance, shared care and remote support being provided to manage the increased volume of patients [2], but to date little has been done to predict or pre-empt post-radiotherapy symptoms. Furthermore, combined therapies and longer survival have an impact on the prevalence of late effects. This is evidenced in increased healthcare needs reported in the UK [3, 4]. However, to detect and meet these needs, tailored healthcare approaches are required to match service models to individuals' predicted risk [5, 6]. Looking at symptoms from the patients' perspective is critical to understanding the consequences of cancer treatment and the impact of late effects on people's lives [7, 8]. To this end, patient reported outcome measures (PROMs) have become an important tool to reflect the quality of health care from the patient perspective.

Although PROMs are clearly an area of healthcare policy interest [9], there is no evidence based guidance on the inclusion of PROMs in clinical practice or in research. PROMs have been proposed as a predictor of survival for cancer patients [10] and identified as an essential requirement of clinical trials and research [11]. We propose, that through careful analysis, PROMs can be used to predict radiotherapy side effects, which could lead to improvements in symptom management in cancer survivors. One obstacle to the extensive use of PROMs in clinical practice is the lack of consistency in current methods and published guidelines [12]. There is a need for evaluation of PROMs based methods to improve their reliability and validity [13], as well as a need for standardization [14]. One area of PROMs that has not received much attention is that of the change in symptoms over time and by investigating these trends we intend to improve the potential of PROMs for clinicians. Two analytical approaches that could be particularly useful to evaluate PROMs are predictive modelling and symptom clustering.

*Address correspondence to this author at the School of Health Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford GU2 7XH, UK; Tel: +44(0)1483 689384; Fax: +44(0)1483 300803; E-mail: a.lemanska@surrey.ac.uk

A feasibility study involving the prediction of post-radiotherapy symptoms using an autoregression (AR) technique was the primary objective in this paper. The focus is methodological and the main purpose was to investigate and present the application of AR for the prediction of PROMs. In addition the effect of symptom clustering was also explored to assist predictive modelling. Symptom clusters have been well defined in cancer populations and methods of how these are measured have been critiqued [15, 16]. In fact patients rarely exhibit symptoms in isolation [17], and although research has traditionally focused on individual symptoms, more recent studies have shown that complex relationships occur between multiple patient health-related outcomes [18]. It has been observed that two or more symptoms can have a catalytic effect on one another [19, 20]. The occurrence of specific symptom clusters can have a profound and direct effect on patient outcomes, resulting in reduced health-related quality of life (HRQOL) [21]. Conversely, treating symptoms in isolation may not necessarily improve HRQOL or enable the prediction of patient outcomes [22]. Multiple symptoms are common in over 40% of patients [23] and those patients who report a higher number of symptoms often have poorer quality of life and poorer long term survival [24]. At present the identification of clusters does not necessarily aid intervention and change practice [25].

In contrast to the significant interest in symptom clusters, little has been done to predict patient health-related outcomes into the future. The majority of literature concerned with the prediction of patient outcomes focuses on the prediction of overall survival, in particular with regard to cancer survival [26, 27]. This may be partly due to the lack of understanding of the factors that can lead to late complications [28] or those that affect HRQOL over time [29]. Therefore, the primary objective of this study was to explore the feasibility of predicting patient reported symptoms after radiotherapy. We present here a new application of AR for this purpose. We propose here that the method may enable the prediction of later health-related outcomes on the basis of early outcomes. We explored the feasibility of the method to predict post-radiotherapy symptoms from PROMs recorded during treatment. The intensity of symptoms experienced by patients changes over time during radiotherapy, and the aim was to use this change when analyzing the data from radiotherapy in order to predict post-radiotherapy health-related outcomes. The ability to accurately predict patient outcomes is important because it offers

great potential to evaluate PROMs in clinical practice and can lead to improved development of PROMs. Few previous studies have been able to identify factors that affect HRQOL over time [29]. This paper explores how the analysis of time profiles from PROMs and the use of symptom clusters can be used to assist and target the predictive modeling of later health-related outcomes.

MATERIALS AND METHODS

Dataset

This study was based around a secondary analysis of an already existing and previously published dataset [30]. Ethical approval for the secondary data analysis was gained from the local NHS Caldecott Guardian. The dataset for the original study was collected from 115 men who were diagnosed with prostate cancer (n=94) and bladder cancer (n=21), and subsequently underwent radical radiotherapy. However, only prostate cancer patients were included in this study in order to preserve the homogeneity of the dataset. The participants were aged from 51 to 80 years old. The radiotherapy treatment was computerised tomography planned and delivered to the prostate gland on a daily basis, using 2 Gy fractions to give a total dose of 64-74 Gy. The dataset and treatment techniques were described in detail in a previous paper [30]. This is a small dataset, however it was selected for this study to serve as an exemplar for the purpose of presenting this new approach, because it had the particular advantage of frequent collection of PROMs during radiotherapy, making it ideal for the development of AR predictive models. The success of this technique requires serial measurements in PROMS for prediction, and in many other datasets it is difficult to correlate acute and late toxicity in any detail, as toxicity is often measured pre and post treatment and looking only at change over large time intervals.

Measures

Two well validated tools were utilised for the collection of PROMs. These were the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Core 30 (EORTC QLQ-C30) [31] and Memorial Symptom Assessment Scale (MSAS) [32, 33]. EORTC QLQ-C30 is a general tool used to assess global components of health and overall HRQOL and it was used here to explore patients' psychological reaction to the illness and treatment. The EORTC QLQ-C30 emotional functioning

domain that consist of four symptoms was included in the analysis. MSAS that had been previously redesigned and validated [34] was used in the dataset to measure 10 symptoms specific to pelvic radiotherapy. MSAS provides multidimensional information about symptoms, where each symptom is assessed in terms of frequency, severity and distress of symptom occurrence. Although these three dimensions are usually highly correlated, some studies have demonstrated that the multidimensional assessment broadens the acquired information about the impact of symptoms [32], giving rise to a more detailed understanding of the symptoms. Prior to analysis, the three dimensions were averaged for each symptom to obtain one composite score (referred to in this paper as symptom intensity), which would reflect the level of symptom burden.

Both MSAS and EORTC QLQ-C30 capture a detailed response when given several categories to a question by using the Likert scale [35]. However, because the aim of the dataset was to capture changes in symptoms over time, a continuous visual analogue scale (VAS) from 0 to 100 [36] was applied during the data collection with MSAS. Participants were asked to mark their responses to the questions about the symptoms on a 100mm line, with a higher score representing a higher intensity. This method gave a more sensitive monitoring approach in comparison to the Likert scale [37, 38]. Symptom assessment was performed on a weekly basis during the six weeks of radiotherapy, and then every six weeks after radiotherapy until the 18th week, when the study

ended. EORTC QLQ-C30 assessment was performed three times, in weeks 1 (the beginning of radiotherapy), 6 (the end of radiotherapy) and 18 (the end of the study) and it was aligned with the MSAS data using linear interpolation to create one comprehensive dataset.

Table **1** summarises the symptoms included in the study according to the represented disorder and the corresponding PROMs tools. In total, 10 symptoms were monitored with MSAS. They were complemented by four symptoms representing the emotional functioning scale from EORTC QLQ-C30. Incorporating both MSAS and EORTC QLQ-C30 facilitated the analysis and understanding of the interactions between the physical symptoms and psychological measures.

Statistical Analysis and Prediction

Data Pre-Processing

Data handling was performed using Matlab version 7.14.0.739 (R2012a) with in-house developed codes. The original data were pre-processed to address the irregular collection times (when data were not collected exactly at the specified weeks) and missing values (when some symptoms were not recorded). The time irregularity was corrected using linear interpolation, and the missing data, apart from dropouts [39], were imputed using the expectation-maximization algorithm [39]. Dropouts are common problem in patient studies. There were 88, 76, 34 and 16 patients, at week 1, 6, 12 and 18 respectively, who completed both symptoms and HRQOL questionnaires. Dropouts were not

Table 1:   Summary of the Symptoms Analysed in the Study According to the Disorder Group and the Corresponding PROMs Tool as a Symptom Source

Disorder group	Symptom name	Symptom source
Urinary	Urinary frequency Urinary leak Pain when passing urine Nocturia	MSAS
Gastrointestinal	Sore anus Abdominal cramp Constipation Diarrhoea Feeling sick	
Emotional	Feeling tired Feeling tense Feeling worry Feeling irritable Feeling depressed	EORTC QLQ-C30

imputed and only available patients were used for symptom clustering each week. In addition only the 16 patients that completed the 18 week study were included in predictive modelling.

Clustering

The identification of symptom clusters *per se* was not the ultimate goal of this research, but a tool to support predictive modelling. We propose that the information contained within clusters may help to identify the predictor variables without modelling the entire dataset and so making modelling more efficient. Therefore for the brevity of this paper, we do not explore or go into too much detail about the various ways of exploring symptom clusters. Rather we are interested in the contribution of how specific clusters can facilitate prediction and help to build AR models.

The Pearson correlation coefficient [40], in the form of the Pearson correlation distance, was used to capture numerically the similarity between symptoms. It was calculated between each pair of symptom intensity variables and collected in a matrix known as the distance matrix. Principal coordinates analysis (PCO) [40] was used to analyze the relationship. Similar to principal component analysis (PCA), PCO is an exploratory technique that summarises the variance in the dataset in the form of principal components. PCO, however, uses information about the similarity or dissimilarity between the variables, instead of raw data. It can be applied directly to the distance matrix, which makes it particularly suitable for cluster analysis. To investigate clustering between the symptoms, the first two PCO components were used in order to visualise the symptom variance in the form of a two dimensional scores plots.

Autoregression

Autoregression (AR) [41, 42] is a simple and practical procedure for predictive modelling of time series data. It is used to predict the future outputs of a time series from the past values of those outputs. The linear AR model relates the predicted variable, x_t, to its past values according to equation:

$$x_t = \mu + \sum_{i=1}^{I} \alpha_i x_{t-i} + e_t \qquad (1)$$

where t is the time index, $\alpha_i \ldots \alpha_I$ are the weights (autoregression coefficients), $x_{t-1}, \ldots, x_{t-I}$ are past values of the variable at time $t-1, \cdots, t-I$, and I is the number of past time points considered in the model; e_t is the term of the residual associated with the noise

and μ is a constant that represents the offset. In the terminology of regression modelling, $x_{t-1}, \ldots, x_{t-I}$ (the variables at the right-hand side of eq. (1)) are also termed predictor variables, and the variable to be predicted, x_t (the variable at the left-hand size of eq. (1)), is called response.

Apart from using time-lagged endogenous terms ($x_{t-1}, \ldots, x_{t-I}$), it is also possible to include some exogenous variables that are considered to be good predictors. For example, the autoregressive model with one exogenous variable (ARX) is given by

$$x_t = \mu + \sum_{i=1}^{I} \alpha_i x_{t-i} + \sum_{i=1}^{I} \beta_i u_{t-i} + e_t \qquad (2)$$

where u_{t-i} represents a time-lagged exogenous variable and β_i its coefficients. To determine the ARX coefficients, the partial least squares (PLS) [43] method was used because of high collinearity between the predictor variables ($x_{t-i}'s$ and $u_{t-i}'s$).

The quality of model prediction can be evaluated by calculating the differences between observed and predicted values, i.e. the residuals. Root mean square error (RMSE) is a common measure that is based on the sum of squares of residuals. Another measure of prediction capability is the coefficient of determination (R^2), which indicates the proportion of variance within the data that have been explained by the model [44]. Both RMSE and R^2 were used in this analysis.

RESULTS

Clustering

Figure **1** presents the PCO results for four points in time (weeks 1, 6, 12 and 18), where the first two PCO components explain the majority (78%, 82%, 77% and 78%, respectively) of variance in the distance matrices. At these four time points, the three symptom disorder groups (urinary, gastrointestinal and emotional), as listed in Table **1**, can be broadly identified as clusters on the scatter plots with a few exceptions. The urinary cluster is the most distinguishable, especially in week 1 and week 12. The emotional cluster is the most distinct at the end of the treatment in week 6. The gastrointestinal cluster, which is clearly distinguished from the other symptoms especially in week 1, usually shows similar trends to the urinary symptoms. It appears in the same half of all graphs as the urinary cluster, except in week 12, according to the first principal component (comp 1).
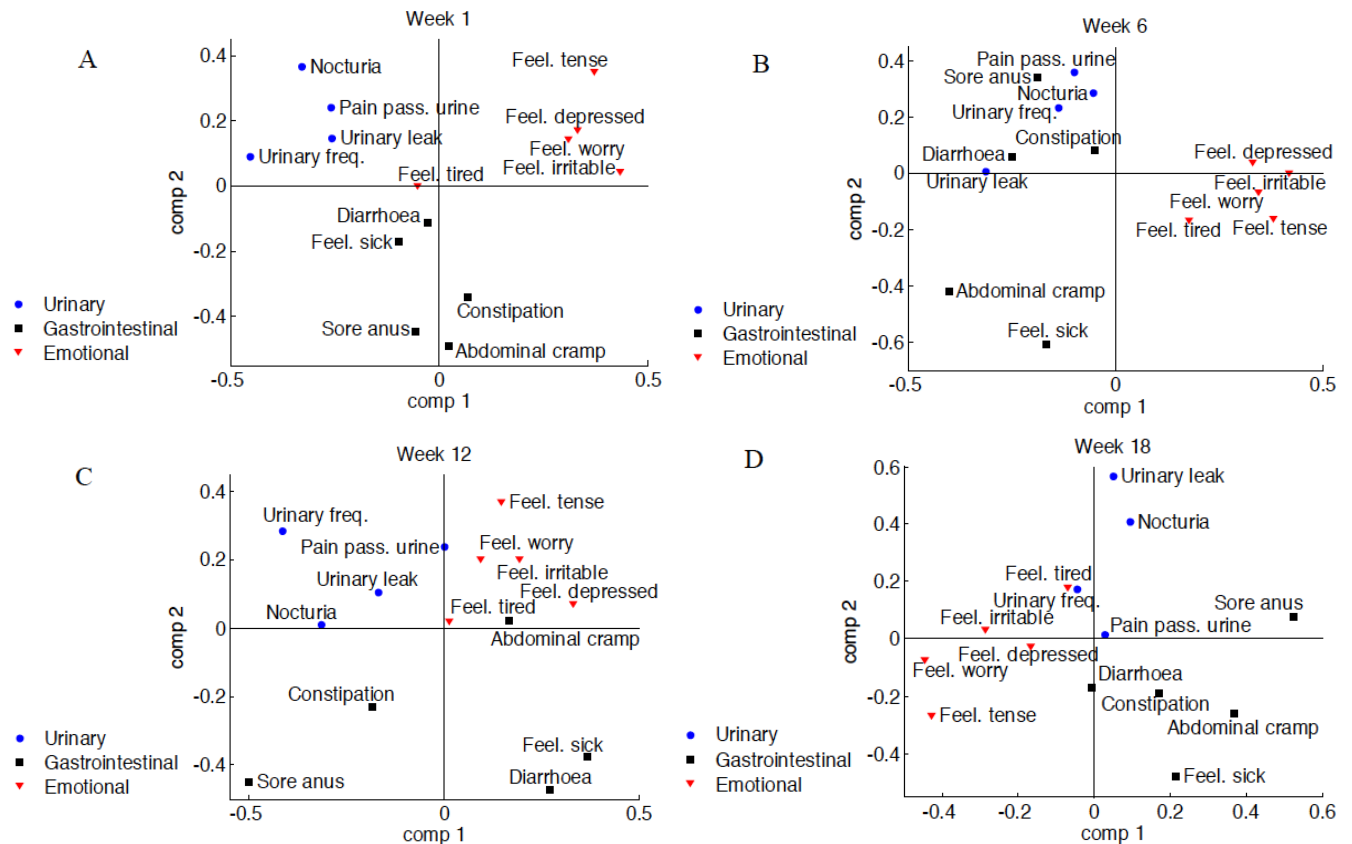
Figure 1: Clustering results for four time points: (**A**) week 1, (**B**) week 6, (**C**) week 12, and (**D**) week 18. Three symptom clusters were identified, namely urinary, gastrointestinal and emotional clusters.

Prediction

Data from the initial six weeks (1 to 6) were used to predict the symptom intensity in week 18. As it was explained in the *Data pre-processing* section, only the 16 patients that completed the 18 week study were included in predictive modelling. The prediction accuracy was evaluated by using the leave-one-out cross-validation (LOOCV) method [45]. Without clustering, all the symptoms from these six weeks were used as predictor variables (denoted by "Pred-All"), and each symptom in week 18 was modelled individually as the response variable (denoted by "Resp-One"). The clustering results can be used in two ways: (i) the predictor variables include only the symptoms that are in the same cluster as the response variables (denoted by "Pred-Cluster"), and (ii) the response variables to be predicted include all symptoms from a cluster (denoted by "Resp-Cluster"). Therefore, four types of models were investigated: "Pred-All & Resp-One", "Pred-All & Resp-Cluster", "Pred-Cluster & Resp-One", and "Pred-Cluster & Resp-Cluster". Simultaneous prediction of multiple symptoms in a cluster, i.e. in "Pred-All & Resp-Cluster" and "Pred-Cluster & Resp-Cluster", was

implemented by the PLS2 algorithm, which builds models for multiple response variables [43].

The results of the four types of ARX models are presented in Table **2** in terms of R^2 and RMSE (in parentheses). A general observation is that some, but not all symptoms, can be well predicted with large R^2 and small RMSE values. In particular, Table **2** shows that without clustering ("Pred-All & Resp-One"), the value of R^2 for nocturia is greater than 0.6, indicating satisfactory prediction, (R^2 values greater than 0.6 are in bold in the table). In addition, it appears that symptom clustering is helpful to improve prediction accuracy in some cases. When clustering is used for either the predictor or the response variables, or both, four more symptoms are well predicted (urinary leak, sore anus, diarrhoea, feeling tired). In particular, using clustering for both predictor and response variables ("Pred-Cluster & Resp-Cluster") gives the most satisfactory results in terms of the capability of predicting four symptoms. However, no matter what model type is used, the prediction results for nine out of 14 symptoms are not satisfactory, and for the successful cases, it is not clear *a priori* what type of

Table 2: **Prediction Results Using the Raw PROMs Data in Terms of R^2 and RMSE in Parenthesis: (a) Predicting One Symptom at a Time (Resp-One); (b) Predicting Symptoms in Clusters (Resp-Cluster). R^2 Values Greater than 0.6 are in Bold, Indicating Satisfactory Prediction**

Response variable	(a) Resp-One				(b) Resp-Cluster			
	Pred-Cluster		Pred-All		Pred-Cluster		Pred-All	
Urinary freq.	0.41	(7.61)	0.33	(8.11)	0.32	(8.15)	0.34	(8.05)
Urinary leak	0.12	(12.95	-0.02	(13.97)	**0.80**	(6.17)	0.23	(12.11)
Pain pass. urine	0.23	(9.22)	0.00	(10.47)	0.28	(8.89)	0.05	(10.24)
Nocturia	**0.88**	(6.21)	**0.70**	(9.65)	**0.86**	(6.67)	**0.72**	(9.26)
Sore anus	**0.68**	(9.32)	0.44	(12.33)	0.45	(12.25)	**0.64**	(9.87)
Abdominal cramp	0.03	(8.64)	-0.12	(9.27)	0.08	(8.40)	-0.13	(9.30)
Constipation	0.32	(7.61)	-0.11	(9.68)	0.04	(8.99)	-0.15	(9.88)
Diarrhoea	-0.14	(14.18)	0.16	(12.15)	**0.63**	(8.13)	0.18	(12.02)
Feel. sick	0.52	(9.68)	-0.15	(14.96)	0.22	(12.29)	-0.14	(14.91)
Feel. tired	**0.70**	(10.03)	0.39	(14.35)	**0.87**	(6.62)	0.57	(12.00)
Feel. tense	0.30	(17.23)	-0.36	(24.14)	0.02	(20.42)	-0.22	(22.79)
Feel. worry	-0.02	(20.75)	-0.30	(23.42)	0.05	(20.02)	-0.08	(21.41)
Feel. irritable	0.31	(19.81)	0.08	(22.81)	0.30	(19.89)	0.08	(22.77)
Feel. depressed	0.35	(15.89)	0.46	(14.50)	0.58	(12.66)	0.34	(15.95)

model (whether using clustering on predictor and/or response variables) gives the best accuracy.

Besides the raw PROMs data, it is also possible to use the slopes of the symptoms' time profiles during the first six weeks as predictor variables. The slope represents the change of symptom intensity during the course of radiotherapy and thus the patient's acute sensitivity to treatment, and it is believed to be an important factor that determines later risks for the patient [46]. Conceptually, a steep slope would indicate a higher risk of side effects [28, 47]. In addition, using slope removes the effect of baseline symptom recorded in week 1 and focuses on the change from the baseline to week 6 when acute side effects are expected to be the most pronounced. The results of predicting symptoms intensity in week 18, by using the slope as predictor variables, are given in Table **3**. Similar to the previous method where the raw PROMs data were used, four types of models were built depending on whether symptom clustering was used: "Pred-All & Resp-One", "Pred-All & Resp-Cluster", "Pred-Cluster & Resp-One", and "Pred-Cluster & Resp-Cluster"

Table **3** shows that only one symptom (nocturia) can be well predicted if no clustering is used, and additional four (urinary leak, sore anus, constipation, diarrhoea) when clustering is applied at either the predictor or the response variables, or both. Symptom clustering appears to be a useful method to improve prediction accuracy; in particular using clustering at both predictor and response variables ("Pred-Cluster & Resp-Cluster"), as was observed using the raw PROMs data.

Combining the results in Tables **2** and **3** reveals that by using symptom clustering in one of the three types of models, and using either raw PROMs data or slope, it is possible to obtain acceptable prediction results (R^2>0.6 with relatively small RMSE) for six of the 14 symptoms, including two (urinary leak and nocturia) from the urinary cluster, three (sore anus, constipation and diarrhoea) from the gastrointestinal cluster, and one (feeling tired) from the emotional cluster. However, the remaining eight symptoms appeared to be unpredictable in the dataset reported. The other challenge is that no single model type is universally superior to the others for different symptoms. This situation calls for the careful evaluation of predictive models before using them in practice (such as the use of cross-validation in this study).

Figure **2** illustrates the predicted versus observed values for the two best predicted symptoms (nocturia and feeling tired) and the four quantitatively unpredictable symptoms (urinary frequency, pain passing urine, constipation, and diarrhoea), when using the "Pred-Cluster & Resp-One" model with raw PROMs data.

Table 3: **Prediction Results Using the Slope of Symptom Time-Profile in Terms of R^2 and RMSE in Parenthesis: (a) Predicting One Symptom at a Time (Resp-One); (b) Predicting Symptoms in Clusters (Resp-Cluster). R^2 Values Greater than 0.6 are in Bold, Indicating Satisfactory Prediction**

Response variable	(a) Resp-One				(b) Resp-Cluster			
	Pred-Cluster		Pred-All		Pred-Cluster		Pred-All	
Urinary freq.	0.45	(7.32)	0.46	(7.29)	0.45	(7.32)	0.44	(7.39)
Urinary leak	0.23	(12.15)	0.57	(9.11)	**0.66**	(8.03)	0.55	(9.25)
Pain pass. urine	0.29	(8.84)	0.11	(9.90)	0.22	(9.29)	0.53	(7.21)
Nocturia	**0.72**	(9.42)	**0.80**	(7.85)	**0.64**	(10.58)	**0.73**	(9.11)
Sore anus	**0.85**	(6.32)	0.50	(11.63)	**0.75**	(8.26)	0.44	(12.29)
Abdominal cramp	-0.13	(9.32)	-0.03	(8.88)	-0.19	(9.56)	-0.13	(9.32)
Constipation	**0.76**	(4.48)	0.24	(8.03)	0.22	(8.11)	0.31	(7.63)
Diarrhoea	-0.36	(15.50)	0.46	(9.79)	**0.60**	(8.38)	0.46	(9.72)
Feel. sick	0.42	(10.64)	0.14	(12.94)	-0.10	(14.63)	-0.14	(14.89)
Feel. tired	0.45	(13.60)	0.44	(13.80)	0.53	(12.60)	0.51	(12.88)
Feel. tense	0.36	(16.56)	0.11	(19.48)	0.40	(16.07)	0.24	(18.05)
Feel. worry	0.21	(18.30)	-0.05	(21.02)	0.21	(18.28)	-0.02	(20.72)
Feel. irritable	0.32	(19.61)	0.51	(16.61)	0.38	(18.77)	0.47	(17.23)
Feel. depressed	0.08	(18.86)	0.31	(16.27)	0.13	(18.27)	0.46	(14.38)

For symptoms such as nocturia and feeling tired, which have the best results of prediction (based on the highest R^2 value achieved), points on the graph fall close to the diagonal line. This indicates that quantitative prediction of these symptoms was possible. For symptoms with a value of R^2 lower than 0.6, where reliable quantitative prediction was not fully possible, prediction can be evaluated by qualitative assessment in terms of symptom intensity being either high or low. For example, by choosing a value of 10 as the threshold between high and low symptom intensity, we can observe from the plot of diarrhoea, where R^2 is as low as -0.14 and RMSE=14.18, that six patients experienced a high level of intensity for diarrhoea. Out of those six patients, three were predicted true positive, in a group of four patients who were predicted to experience acute diarrhoea, and three were false negative. For one patient the predicted symptom intensity was false positive, while for the remaining nine the results were true negative, and correctly predicted to experience no serious diarrhoea symptoms. In the case of urinary frequency there were only four patients with a high observed intensity for that symptom. Two of them were predicted true positive and two false negative. The remaining patients were correctly predicted as true negatives. The qualitative assessment method was characterized with an average sensitivity of 50% and specificity of over 93%.

With high specificity this method can be a valuable screening technique to identify patients that are not at risk of serious side effects. However, we can observe that with its low sensitivity, the qualitative method was not very effective in identifying patients that will develop high intensity of a symptom in the future.

DISCUSSION

In the light of current results we believe that qualitative prediction of symptom intensity could be used to inform clinical decision making. The poor performance of the model in predicting symptoms that have high intensity can be accounted for by the fact that in the current dataset only a small number of patients experienced a high observed intensity of symptoms. The performance of the method should be further investigated on other datasets containing patients that experience high intensity of a symptom, as well as patients that do not. In addition, the choice of the decision threshold between high and low symptom intensity in this model was set arbitrarily. Further research, working together with the expertise of clinicians, is required to establish this arbitrary cut-off value, and this should be done empirically, and also independently for each symptom and condition.

In this study, we endeavoured to explore the clusters present in the dataset for the purpose of
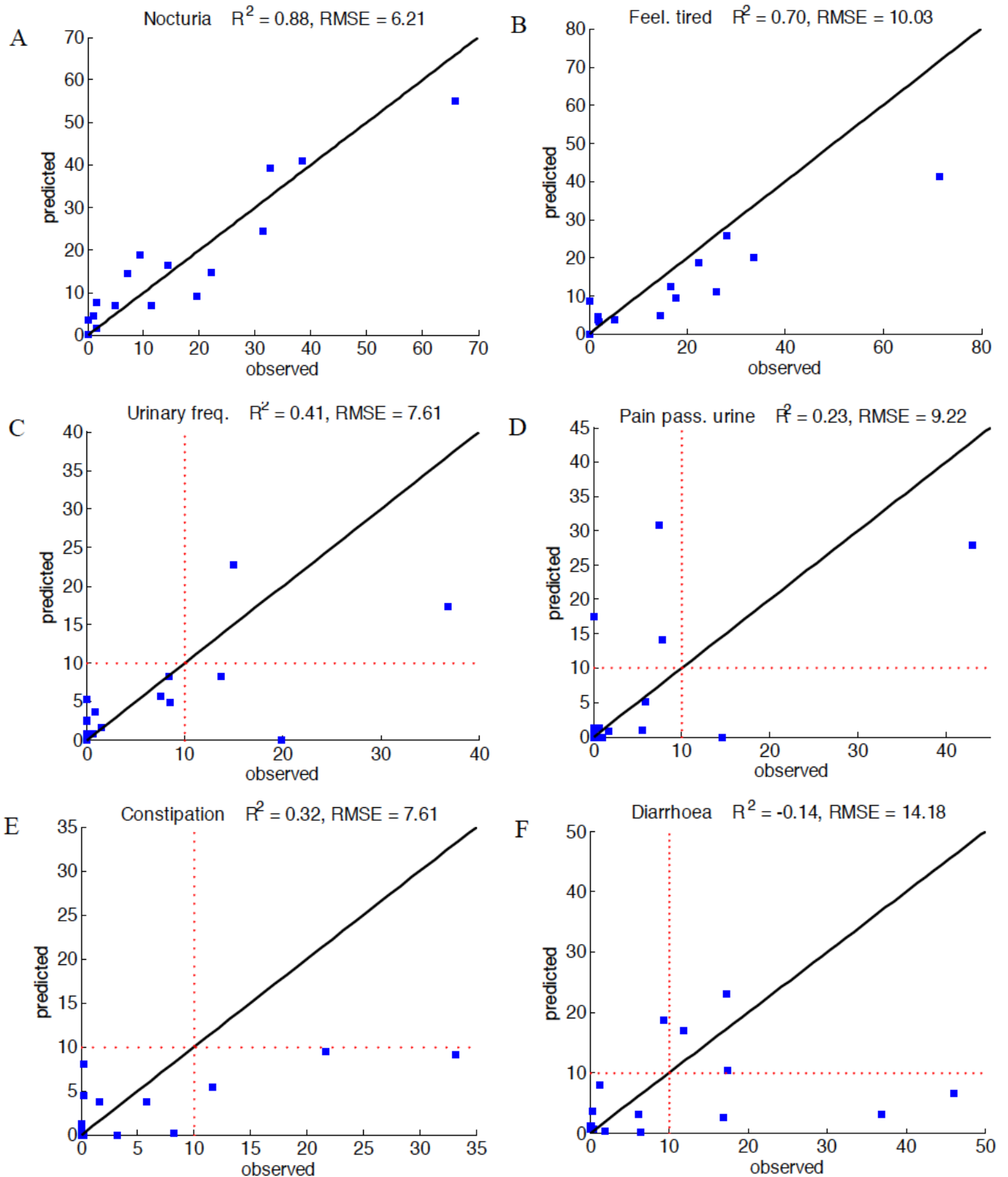
Figure 2: The predicted versus observed symptom intensity for patients in week 18, of 6 symptoms (**A**) nocturia, (**B**) feeling tired (**C**) urinary frequency, (**D**) pain when passing urine, (**E**) constipation, and (**F**) diarrhoea.

prediction of post-radiotherapy health-related outcomes. As would be expected, the clusters identified follow the trends that have been defined in pelvic toxicity literature [48], confirming the reliability of

the clustering analysis approach. We investigated how the information that was gained during cluster analysis can be utilised for predictive modelling, and four types of ARX models were explored. The results suggest that

clustering analysis is potentially helpful in identifying the informative and appropriate predictors and response variables. The effectiveness of symptom clustering may be because the symptoms within a cluster are highly correlated and have mechanistic relationships and therefore inform prediction. In many cases, the inclusion of clustering data gave better results than using all symptoms indiscriminately. This finding can be beneficial in studies where the dataset is large, with many symptoms to consider, as prediction can be performed with specific symptoms rather than using all data, which is a more economical approach, and has the advantage of building simpler ARX models that are easier to interpret. The clinical relevance of symptom clusters and their relationship within PROMs remains a key question for nurse researchers and merits further research in this area.

In addition to modelling the raw PROMs data, the use of slope calculated from time profiles was investigated, and both methods gave similar results. Slope is an important and direct indication of the patient's sensitivity to radiotherapy and may be preferred by clinicians because of its clear physical meaning. Finally, it should be noted that not all symptoms are well predicted as judged by the quantitative measures R^2 and RMSE. Whilst a better modelling approach and data collection procedure are likely to improve prediction results, the other possibility is to deliver qualitative prediction in terms of whether the symptom intensity is high or low.

For the purpose of this feasibility study a small, historical dataset was used, which made the new application of statistical techniques manageable. The dataset in question was particularly suitable on a number of levels for examining the concept of predictive modelling. As mentioned previously, it is the frequent collection of PROMs that makes the dataset suitable for the development of AR models. Traditional methods for evaluating PROMS look at change over time, using the baseline compared to the endpoint with ANOVA or multiple regression. This does not utilise multiple data points or capture the breadth of data or trajectory over time, and could potentially miss predictors. In addition the dataset provided the opportunity to test how effective the method was working outside optimal conditions, for example with irregular or missing data. There were however, some limitations in using this existing dataset, such as the fact that the dataset followed patients for only 18 weeks. There was also a large number of dropouts. The next step towards further validation of this

approach to prediction would be to work with a more complete dataset. A PROMs dataset with a longer time series would be of benefit in predicting late side effects. The research team has secured larger datasets over a longer time frame, and there are plans in place to further test the performance of the methods presented here. However, as an initial study, this research was a valuable first step toward modelling PROMs data for use in radiotherapy. This analysis technique could have potential for wider use outside radiotherapy because symptom clusters were previously identified and linked to health related outcomes in chronic diseases such as multiple sclerosis (Motl & McAuley, 2010).

ACKNOWLEDGEMENTS

REFERENCES

[1]   Richards M, Corner J, Maher J. The National Cancer Survivorship Initiative: new and emerging evidence on the ongoing needs of cancer survivors. Br J Cancer 2011; 105(S1): S1-S4.
http://dx.doi.org/10.1038/bjc.2011.416

[2]   Maher J, McConnell H. New pathways of care for cancer survivors: adding the numbers. Br J Cancer 2011; 105(S1): S5-S10.
http://dx.doi.org/10.1038/bjc.2011.417

[3]   Khan NF, Evans J, Rose PW. A qualitative study of unmet needs and interactions with primary care among cancer survivors. Br J Cancer 2011; 105(S1): S46-S51.
http://dx.doi.org/10.1038/bjc.2011.422

[4]   Elliott J, Fallows A, Staetsky L, *et al.* The health and well-being of cancer survivors in the UK: findings from a population-based survey. Br J Cancer 2011; 105(S1): S11-S20.
http://dx.doi.org/10.1038/bjc.2011.418

[5]   Armes J, Crowe M, Colbourne L, *et al.* Patients' supportive care needs beyond the end of cancer treatment: a prospective, longitudinal survey. J Clin Oncol 2009; 27(36): 6172-9.
http://dx.doi.org/10.1200/jco.2009.22.5151

[6]   Ream E, Quennell A, Fincham L, *et al.* Supportive care needs of men living with prostate cancer in England: a survey. Br J Cancer 2008; 98(12): 1903-9.
http://dx.doi.org/10.1038/sj.bjc.6604406

[7]   Davidson SE, Faithfull S. Late radiotherapy effects: is bowel morbidity adequately documented or patients' needs managed appropriately? Clin Oncol 2006; 18(5): 419-20.
http://dx.doi.org/10.1016/j.clon.2006.03.005

[8]   Basch E, Abernethy AP, Mullins CD, *et al.* Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. J Clin Oncol 2012; 30(34): 4249-55.
http://dx.doi.org/10.1200/JCO.2012.42.5967

[9]   Department of Health, Macmillan Cancer Support and NHS Improvement. National Cancer Survivorship Initiative (NCSI) vision. 2010. Accessed: 10 October 2014, http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.

uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/@ps/documents/digitalasset/dh_111477.pdf

[10] Osoba D. Health-related quality of life and cancer clinical trials. Ther Adv Med Oncol 2011; 3(2): 57-71.
http://dx.doi.org/10.1177/1758834010395342

[11] FDA. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. Accessed: 10 October 2014, http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf

[12] Wild D, Grove A, Martin M, *et al.* Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health 2005; 8(2): 94-104.

[13] Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health 2007; 10(2): S94-S105.
http://dx.doi.org/10.1111/j.1524-4733.2007.00272.x

[14] Two R, Verjee-Lorenz A, Clayson D, *et al.* A methodology for successfully producing global translations of patient reported outcome measures for use in multiple countries. Value Health 2010; 13(1): 128-31.
http://dx.doi.org/10.1111/j.1524-4733.2009.00585.x

[15] Skerman HM, Yates PM, Battistutta D. Identification of cancer-related symptom clusters: an empirical comparison of exploratory factor analysis methods. J Pain Symptom Manage 2012; 44(1): 10-22.
http://dx.doi.org/10.1016/j.jpainsymman.2011.07.009

[16] Skerman HM, Yates PM, Battistutta D. Multivariate methods to identify cancer-related symptom clusters. Res Nurs Health 2009; 32(3):345-60.
http://dx.doi.org/10.1002/nur.20323

[17] Miaskowski C, Dodd M, Lee K. Symptom clusters: the new frontier in symptom management research. J Natl Cancer Inst Monogr 2004; 32: 17-21.
http://dx.doi.org/10.1093/jncimonographs/lgh023

[18] Kirkova J, Walsh D. Cancer symptom clusters - a dynamic construct. Support Care Cancer 2007; 15(9): 1011-3.
http://dx.doi.org/10.1007/s00520-007-0259-2

[19] Barsevick AM, Whitmer K, Nail LM, Beck SL, Dudley WN. Symptom cluster research: conceptual, design, measurement, and analysis issues. J Pain Symptom Manage 2006; 31(1): 85-95.
http://dx.doi.org/10.1016/j.jpainsymman.2005.05.015

[20] Parker KP, Kimble LP, Dunbar SB, Clark PC. Symptom Interactions as Mechanisms Underlying Symptom Pairs and Clusters. J Nurs Scholarsh 2005; 37(3): 209-15.
http://dx.doi.org/10.1111/j.1547-5069.2005.00037.x

[21] Motl RW, McAuley E. Symptom cluster and quality of life: preliminary evidence in multiple sclerosis. J Neurosci Nurs 2010; 42(4): 212-6.
http://dx.doi.org/10.1097/JNN.0b013e3181e26c5f

[22] Fan G, Filipczak L, Chow E. Symptom clusters in cancer patients: a review of the literature. Curr Oncol 2007; 14(5): 173-9.

[23] Esther Kim J-E, Dodd MJ, Aouizerat BE, Jahan T, Miaskowski C. A review of the prevalence and impact of multiple symptoms in oncology patients. J Pain Symptom Manage 2009; 37(4): 715-36.
http://dx.doi.org/10.1016/j.jpainsymman.2008.04.018

[24] Degner LF, Sloan JA. Symptom distress in newly diagnosed ambulatory cancer patients and as a predictor of survival in lung cancer. J Pain Symptom Manage 1995; 10(6): 423-31.
http://dx.doi.org/10.1016/0885-3924(95)00056-5

[25] Xiao C. The state of science in the study of cancer symptom clusters. Eur J Oncol Nurs 2010; 14(5): 417-34.
http://dx.doi.org/10.1016/j.ejon.2010.05.011

[26] Quinten C, Maringwa J, Gotay CC, *et al.* Patient self-reports of symptoms and clinician ratings as predictors of overall cancer survival. J Natl Cancer Inst 2011; 103(24): 1851-8.
http://dx.doi.org/10.1093/jnci/djr485

[27] Osoba D. Health-related quality of life and predicting survival in cancer: not a simple matter. Support Care Cancer 2007; 15(4): 353-5.
http://dx.doi.org/10.1007/s00520-006-0187-6

[28] Wedlake LJ, Thomas K, Lalji A, *et al.* Predicting Late Effects of Pelvic Radiotherapy: Is There a Better Approach? Int J Radiat Oncol Biol Phys 2010; 78(4): 1163-70.
http://dx.doi.org/10.1016/j.ijrobp.2009.09.011

[29] Sarenmalm K, Elisabeth, Öhlén J, Odén A, Gaston-Johansson F. Experience and predictors of symptoms, distress and health-related quality of life over time in postmenopausal women with recurrent breast cancer. Psychooncology 2008; 17(5): 497-505.
http://dx.doi.org/10.1002/pon.1258

[30] Faithfull S, Corner J, Meyer L, Huddart R, Dearnaley D. Evaluation of nurse-led follow up for patients undergoing pelvic radiotherapy. Br J Cancer 2001; 85(12): 1853-64.
http://dx.doi.org/10.1054/bjoc.2001.2173

[31] Aaronson NK, Ahmedzai S, Bergman B, *et al.* The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993; 85(5): 365-76.
http://dx.doi.org/10.1093/jnci/85.5.365

[32] Portenoy RK, Thaler HT, Kornblith AB, *et al.* The Memorial Symptom Assessment Scale: an instrument for the evaluation of symptom prevalence, characteristics and distress. Eur J Cancer 1994; 30(9): 1326-36.
http://dx.doi.org/10.1016/0959-8049(94)90182-1

[33] Portenoy RK, Thaler HT, Kornblith AB, *et al.* Symptom prevalence, characteristics and distress in a cancer population. Qual Life Res 1994; 3(3): 183-9.
http://dx.doi.org/10.1007/BF00435383

[34] Faithfull S. 'Just grin and bear it and hope that it will go away': coping with urinary symptoms from pelvic radiotherapy. Eur J Cancer Care 1995; 4(4): 158-65.
http://dx.doi.org/10.1111/j.1365-2354.1995.tb00087.x

[35] Bowling A. Measuring health: a review of quality of life measurement scales. 2nd ed. Milton Keynes: Open University Press 1997.

[36] Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. J Clin Nurs 2005; 14(7): 798-804.
http://dx.doi.org/10.1111/j.1365-2702.2005.01121.x

[37] Jensen MP. The validity and reliability of pain measures in adults with cancer. J Pain 2003; 4(1): 2-21.
http://dx.doi.org/10.1054/jpai.2003.1

[38] Hjermstad MJ, Fayers PM, Haugen DF, *et al.* Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage 2011; 41(6): 1073-93.
http://dx.doi.org/10.1016/j.jpainsymman.2010.08.016

[39] Molenberghs G. Missing data in clinical studies. 1st ed. Chichester: John Wiley & Sons 2007.

[40] Brereton RG. Chemometrics for pattern recognition. 1st ed. Chichester: John Wiley & Sons 2009.

[41] Chatfield C. The analysis of time series. An introduction. 4th ed. New York, NY: Chapman and Hall 1989.

[42] Peng H, Ozaki T, Toyoda Y, *et al.* RBF-ARX model-based nonlinear system modeling and predictive control with application to a NOx decomposition process. Control Eng Pract 2004; 12(2): 191-203.
http://dx.doi.org/10.1016/s0967-0661(03)00050-9

[43]   Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. Anal Chim Acta 1986; 185(0): 1-17.
http://dx.doi.org/10.1016/0003-2670(86)80028-9

[44]   Campbell MJ, Machin D. Medical statistics: a commonsense approach. 3rd ed. Chichester: John Wiley & Sons 1999.

[45]   Martens HA, Dardenne P. Validation and verification of regression in small data sets. Chemometr Intell Lab Syst 1998; 44(1–2): 99-121.
http://dx.doi.org/10.1016/s0169-7439(98)00167-1

[46]   Larsen A, Bjørge B, Klementsen B, *et al*. Time patterns of changes in biomarkers, symptoms and histopathology during pelvic radiotherapy. Acta Oncol 2007; 46(5): 639-50.
http://dx.doi.org/10.1080/02841860601099241

[47]   Denham JW, Hauer-Jensen M. The radiotherapeutic injury – a complex 'wound'. Radiother and Oncol 2002; 63(2): 129-45.
http://dx.doi.org/10.1016/s0167-8140(02)00060-9

[48]   Andreyev HJN, Wotherspoon A, Denham JW, Hauer-Jensen M. Defining pelvic-radiation disease for the survivorship era. Lancet Oncol 2010; 11(4): 310-2.
http://dx.doi.org/10.1016/s1470-2045(10)70026-7