Article:
Choudhury, A, Theophanous, S orcid.org/0000-0002-4148-3905, Lønne, P-I et al. (11 more authors) (2021) Predicting outcomes in anal cancer patients using multi-centre data and distributed learning - a proof-of-concept study. Radiotherapy and Oncology, 159. pp. 183-189. ISSN 0167-8140

https://doi.org/10.1016/j.radonc.2021.03.013

Predicting outcomes in anal cancer patients using multi-centre data and distributed learning - a proof-of-concept study

Author names and affiliations

1 Ananya Choudhury[+]
2 Stelios Theophanous[+]
3 Per-Ivar Lønne[+]
2 Robert Samuel
4 Marianne Grønlie Guren
1 Maaike Berbee
5 Peter Brown
6 John Lilley
1,7 Johan van Soest
1 Andre Dekker
2 Alexandra Gilbert
3,8 Eirik Malinen
1 Leonard Wee[*]
2 Ane L Appelt[*]

1 MAASTRO (Dept of Radiotherapy), GROW School of Oncology and Developmental Biology, Maastricht University & Maastricht University Medical Centre+, Dr Tanslaan 12, Limburg 6229 ET, The Netherlands.

2 Leeds Institute of Medical Research at St James's, University of Leeds, Beckett Street, Leeds, LS9 7TF, UK

3 Department of Medical Physics, Oslo University Hospital, PO Box 4953 Nydalen, 0424 Oslo, Norway

4 Department of Oncology, Oslo University Hospital, PO Box 4953 Nydalen, 0424 Oslo, Norway

5 Department of Clinical Radiology, Leeds Teaching Hospitals NHS Trust, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, UK

6 Department of Medical Physics, Leeds Cancer Centre, St James's University Hospitals, Beckett Street, Leeds, LS9 7TF, UK

7 Brightlands Institute for Smart Society (BISS, Faculty of Science & Engineering, Maastricht University, Smedestraat 2, 6411 CR Heerlen, The Netherlands

8 Department of Physics, University of Oslo, P.O. Box 1048 Blindern, 0316 Oslo, Norway

+ Joint first authors
* Joint senior authors and corresponding authors:
leonard.wee@maastro.nl
a.l.appelt@leeds.ac.uk

Keywords
Anal cancer; squamous cell carcinoma; chemoradiotherapy; distributed learning; federated learning; outcome modelling; overall survival

Running head
Distributed learning for anal cancer

Highlights

- This proof-of-concept study demonstrates the feasibility of privacy preserving distributed learning for anal cancer outcome modelling.
- A Cox proportional hazards regression model was trained and validated using data from three international institutions, with clinical and treatment-related risk factors, and exhibited robust performance.
- The study data represent one of the largest available series of anal cancer patients treated with modern radiotherapy techniques.
- Distributed learning may be an attractive approach for outcome modelling in rare cancers.

Abstract

Background and purpose: Predicting outcomes is challenging in rare cancers. Single-institutional datasets are often small and multi-institutional data sharing is complex. Distributed learning allows machine learning models to use data from multiple institutions without exchanging individual patient-level data. We demonstrate this technique in a proof-of-concept study of anal cancer patients treated with chemoradiotherapy across multiple European countries.

Materials and methods: atomCAT is a three-centre collaboration between Leeds Cancer Centre (UK), MAASTRO Clinic (The Netherlands) and Oslo University Hospital (Norway). We trained and validated a Cox proportional hazards regression model in a distributed fashion using data from 281 patients treated with radical, conformal chemoradiotherapy for anal cancer in three institutions. Our primary endpoint was overall survival. We selected disease stage, sex, age, primary tumour size, and planned radiotherapy dose (in EQD2) *a priori* as predictor variables.

Results: The Cox regression model trained across all three centres found worse overall survival for high risk disease stage (HR=2.02), male sex (HR=3.06), older age (HR=1.33 per 10 years), larger primary tumour volume (HR=1.05 per 10cm^3) and lower radiotherapy dose (HR=1.20 per 5 Gy). A mean concordance index of 0.72 was achieved during validation, with limited variation between centres (Leeds=0.72, MAASTRO=0.74, Oslo=0.70). The global model performed well for risk stratification for two out of three centres.

Conclusions: Using distributed learning, we accessed and analysed one of the largest available multi-institutional cohorts of anal cancer patients treated with modern radiotherapy techniques. This demonstrates the value of distributed learning in outcome modelling for rare cancers.

Introduction

Prediction models for cancer outcomes can support clinical decision making, and hold the promise for individualisation of cancer treatment and radiotherapy plan optimisation. Development of robust and validated models is often hampered by lack of access to data, however, especially across countries and institutions. This is particularly the case for rare cancers.

"Distributed learning" facilitates the development and validation of statistical models using data across multiple institutions without transferring individual patient data outside the originating institution. This is one of several novel methodologies developed to preserve patient data privacy [1,2], such as differential privacy and encryption [3]. Our distributed learning approach is an open-source solution (Vantage6) which prevents insider attacks by blocking any direct connection between data hosts [4,5]. Only locally aggregated statistics (model coefficients and fit errors) are exchanged between the data centres and the central server. Model development in the distributed learning framework is an iterative mathematical optimization problem where the coefficients of a single globally-convergent model will be determined by minimizing the total error [6]. The general methodology has been shown to be scalable up to vast numbers of patients [7].

The distributed learning approach may be ideally suited to rare diseases, where single-institutional datasets are limited in size and sharing data between institutions is restricted by data protection regulations and related ethical considerations [8]. One such example is anal cancer; a rare disease with an incidence rate around 2.1 per 100,000 person-years in Northern Europe and twice the incidence in women relative to men [9]. Currently, the standard treatment for localised disease involves concomitant radiotherapy and chemotherapy [10], which leads to a complete response in approximately 3 out of 4 patients. 5-year overall survival rates of 75% have previously been reported [11–13]. Further improvements in disease control and survival have proven challenging, and questions remain around optimal tumour dose [14–16]. Additionally, patients that undergo standard treatment commonly suffer from various early and late side effects, such as gastrointestinal symptoms that range from mild to severe [17]. This highlights the need for a personalised approach to anal cancer chemoradiotherapy. Such individualisation will be dependent on the development of outcome prediction models [18], which again require sufficient data for model training and validation. A distributed learning approach may help obtaining sufficient patient data from different institutions in order to develop robust and generalisable models, while circumventing many of the barriers associated with individual-level patient data sharing.

In this proof-of-concept study, we aimed to show the feasibility of our distributed learning approach for patients with anal cancer receiving radical chemoradiotherapy. A prediction model for overall survival (OS), employing established baseline clinical factors and radiotherapy dose as predictors, was applied on data across institutions in three European countries. OS was chosen as our outcome of interest as this is an important outcome measure in anal cancer research [19] and a robust endpoint across institutions.

We hypothesise that a global Cox proportional hazards model developed without exchange of any individual-level patient data is highly reproducible in a multi-centre setting, when evaluated through an “internal-external” validation cycle [20], despite the small sample sizes within each participating centre. Furthermore, we hypothesize that we can define risk groups across institutions.

Materials & Methods

The study protocol was developed collaboratively by the three participating institutions prior to study initiation: Leeds Cancer Centre (UK), MAASTRO Clinic (The Netherlands), Oslo University Hospital (Norway). Patients were treated with chemoradiotherapy with radical intent for anal squamous cell carcinoma (ASCC), with conformal radiotherapy (forward-planned 3D conformal (3D-CRT) or intensity-modulated radiation therapy/volumetric modulated arc therapy (IMRT/VMAT)). Baseline, treatment and outcome data were available. The main outcome of interest for this proof-of-concept study was overall survival (OS). Death from any cause was counted as an event, with patients censored at the time of local data collection. Survival interval was calculated from the date of the first fraction of radiotherapy, to either date of death or the last follow-up date if alive.

For candidate outcome predictors, the literature on anal cancer chemoradiotherapy was reviewed, and expert input sought from three consultant clinical oncologists specialising in anal cancer. Importantly, we considered only predictors available at start of treatment (thus not radiotherapy compliance or treatment gaps). The following predictor variables were chosen, based on published data, clinical experience, and data availability in participating institutions: disease stage - low risk (Stage I-II, T1N0 or T2N0 or T3N0) versus high risk (Stage III, T4N(any) or T(any)N+) according to TNM v8 [21]; sex; age; primary tumour size (gross tumour volume, GTV, on planning CT); and primary tumour prescribed dose (converted from physical dose to equivalent dose in 2 Gy per fraction, $EQD2_{\alpha/\beta=10Gy}$). For disease stage, there is ongoing debate as to whether T3N0 tumours should be regarded as low or high risk [16]. The model was thus also fitted with T3N0 tumours assigned to the high rather than the low risk group. Additionally, histology (basaloid SCC: yes/no) was identified as a potential predictor, but was not included in the final analysis due to a large proportion of missing SCC subtype data in one institution. A data code book was shared between all institutions, for standardised data collection and reporting.

Patient data collection

For Leeds Cancer Centre, a subset of patients treated for anal cancer between 2015 and 2018 with baseline and outcome data available were included. All patients were treated with VMAT and simultaneous integrated boost (SIB). Patients were identified through existing research databases, and additional data was sourced as

necessary from clinical databases. Tumour volumes were extracted manually from radiotherapy plans. Survival data were based on patient electronic records, which are automatically linked to the NHS England death registry.

At MAASTRO Clinic, patients treated by radiotherapy for primary anal cancer with radical intent between 2008 and 2017 were retrieved from electronic treatment records. All radiotherapy was in the form of either 3D-CRT (n=26; prior to 2013) or VMAT (n=55; after 2013), with dose to the primary tumour escalated by either sequential boost or SIB. Tumour volumes were extracted manually from radiotherapy planning delineations. Dates of death were obtained from the electronic patient records, which were automatically updated from a Dutch citizens registry.

For Oslo University Hospital, anal cancer patients enrolled in the prospective ANCARAD trial (ClinicalTrials registration NCT01937780) receiving treatment between 2013 and 2017 were included. All patients received chemoradiotherapy using 3D-CRT (40 patients), IMRT (11 patients) or VMAT (69 patients), with boosts delivered either sequentially (109 patients) or as SIB (11 patients). Baseline and outcome data were prospectively collected as part of the ANCARAD trial. Additional baseline data were retrieved as necessary from clinical databases. Tumour volumes were extracted from radiotherapy structure sets in the treatment planning system using an in-house script.

Details on the radiotherapy and concomitant chemotherapy schedules used at each centre are shown in Table 1.

Table 1. Radiotherapy and concomitant chemotherapy treatment schedules used at each of the three centres.

	Leeds	MAASTRO	Oslo
Radiotherapy regimen	Most patients were prescribed 50.4-53.2 Gy to the primary tumour, 50.4Gy to involved nodes and 40 Gy to elective nodal volumes in 28 fractions. 5 patients were treated with doses above 53.2 Gy.	All patients were prescribed 54-66 Gy to the primary tumour and 39-49.5 Gy to elective lymph nodes in 30-33 fractions.	All patients were prescribed 54-58 Gy to the primary tumour and pathological lymph nodes and 46 Gy to elective nodal volumes in 27-29 fractions.
Chemotherapy regimen	Mitomycin-C (12 mg/m2 bolus day 1, capped at 20 mg) and 5-FU (1000 mg/m2 in 1 L normal saline over 24 hours, days 1-4 and days 29-32, capped at 2 m2).	Mitomycin-C (10 mg/m2 bolus day 1) plus either capecitabine (2 x 825 mg/m2 per radiotherapy treatment day) or continuous 5-FU (750 mg/m2 days 1-5 and 29-33); 11 patients who were elderly/frail or had T1N0M0 disease were treated with 66 Gy radiotherapy only.	Mitomycin-C (10 mg/m2 bolus day 1, capped at 20 mg) and 5-FU (1000 mg/m2 in 1 L normal saline over 24 hours, days 1-4), according to national guidelines. Patients with T1-T2 and N0 tumours received a single cycle (5-FU: days 1-4, MMC: day 1); patients with T3-4 tumours or N+ received two cycles (additional cycle in the fifth treatment week; 5-FU days 29-32, MMC day 29).

Institutional data access & data protection approvals

Each institution acquired separate local approvals for accessing and collecting patient data for research. As no individual patient data were exchanged between institutions, no data sharing agreements or additional patient consent were needed. Local information governance and data protection review of the distributed learning infrastructure were obtained wherever appropriate. In Leeds, the study was approved by LeedsCAT; a radiotherapy-specific institutional research governance board. In MAASTRO, IRB approval was obtained to extract patient data from electronic records. In Oslo, Regional Ethics Committee approval was obtained for re-use of data from the ANCARAD trial (via an amendment), and the local data protection officer reviewed and approved the distributed learning infrastructure.

Distributed learning architecture

We used the Vantage6 v0.2.4 software to set up three components; (1) "nodes" where patient-level data is accessed and where local model coefficients are computed, (2) a trusted coordinating "server" that performs aggregation of coefficients, and (3) a "researcher" that provides the model to be trained. The purpose was to fit a distributed Cox model for overall survival for anal cancer (see Figure 1). For additional security, all patient data were pseudonymized and stripped of protected health information (e.g. dates of treatments, dates of birth/death, generic medical record numbers, etc).

Nodes were set up on common personal computers (either physical or virtual) running any one of well-supported operating systems (Windows/MacOS/Ubuntu) with an installation of Python (v3.6 or later), Docker Desktop community edition, and Vantage6 v0.2.4. The complete source code for the infrastructure implementation is available [https://github.com/IKNL/vantage6 - Version 0.2.4]. Network connectivity was fully compliant with local institutional policies, and only one secured network port through the institution firewall was enabled for Vantage6 traffic.

The Leeds node was set up as a Windows 10 Pro virtual machine (Intel(R) Xeon(R) Gold 5118 CPU, 16GB RAM), and only accessible by NHS Trust users granted the appropriate permissions. Patient data were extracted from a clinical database, de-identified, and forwarded to the virtual machine. The MAASTRO node was set up as a physical Surface Book 2 laptop (Intel(R) Core i7-8650 CPU, 16GB RAM) running Windows 10 Pro, and pseudonymized patient data accessed via a mapped folder directing to an internal storage server. The Oslo node was set up on a Lenovo ThinkPad laptop (Intel(R) i7-4600M CPU, 16GB RAM), running Ubuntu Linux 18.04 as a virtual machine, which can easily be cloned when setting up nodes for new projects. The Oslo node was physically decoupled from the hospital network, and pseudonymized data was transferred to the machine via an encrypted external hard disk drive.

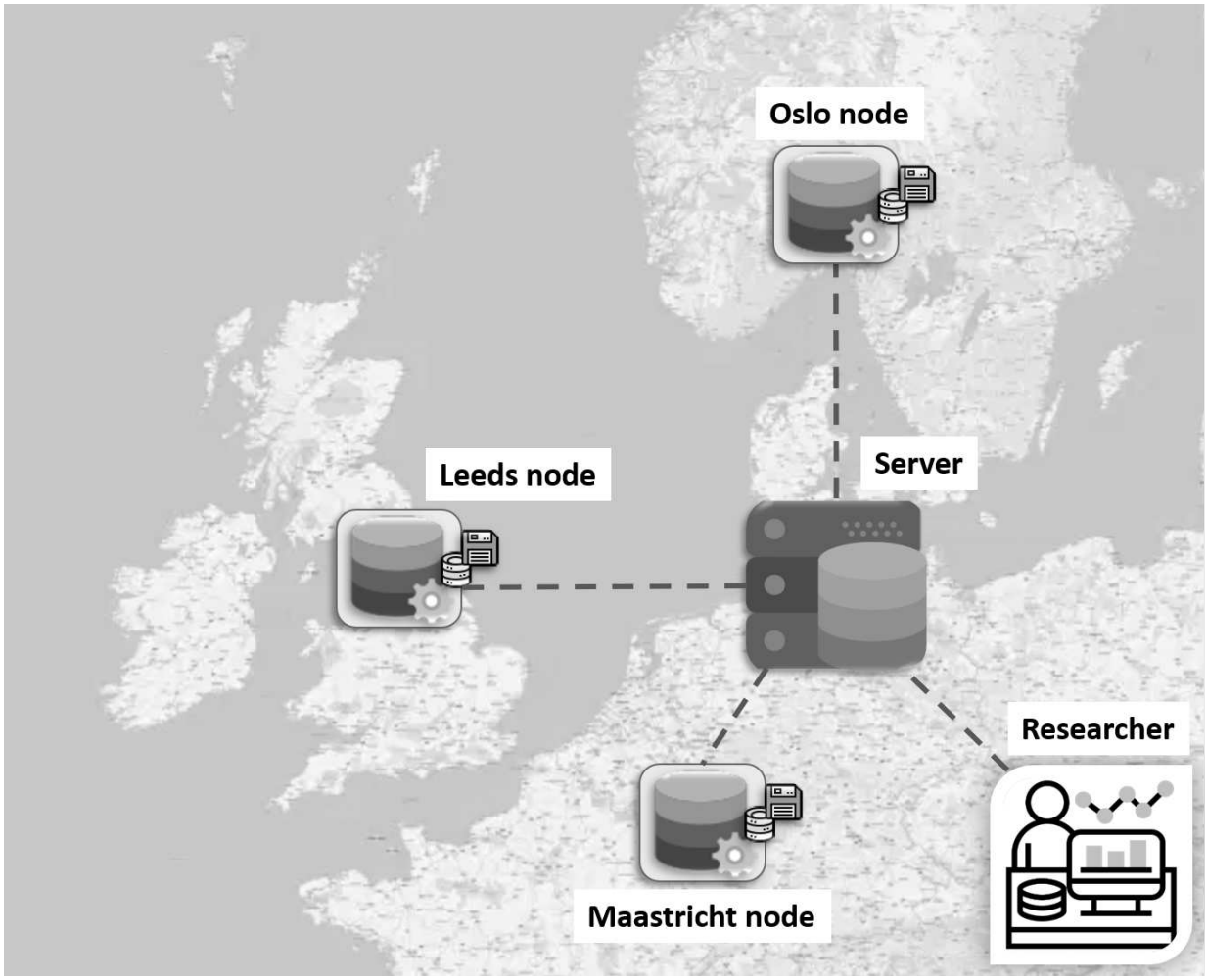

Figure 1. Distributed learning as a multinational collaboration to train a Distributed Cox model for overall survival in anal cancer across three data nodes with a trusted coordination server in the Microsoft Azure cloud.

The central coordination server takes the role of trusted messaging "broker" for the collaboration network. Only key-authenticated messages were allowed to pass between researcher and server, and between node and server. The server administrator maintains a registry of collaborations, researchers, institutions and institution administrators, as well as unique encryption keys for each role. For this proof-of-concept run, the server was set up by MAASTRO as an Ubuntu Linux 18.04 virtual instance (30GB storage, 4 GB memory) on the Microsoft Azure cloud computing service based in Europe.

Descriptive data analysis

Summary statistics were exchanged between centres in order to explore cohort differences prior to modelling. Categorical variables were tested using a chi-squared test, and numerical variables were tested using a one-way ANOVA test. All tests were carried out using summary statistics (number of patients, mean and standard deviation values) rather than individual patient data. Estimated 3-year survival rates and potential follow-up times were calculated by each centre individually using the 'survival' package

in R [22], employing the Kaplan-Meier estimator. Median follow-up time was based on the inverse Kaplan Meier estimator [23].

Distributed Cox algorithm

The Distributed Cox algorithm developed by Lu et al. [2] was adapted to the Vantage6 v0.2.4 infrastructure as R scripts (v.3.6.2). The source code has been made openly accessible on GitHub (https://github.com/AnanyaCN/d_coxph). Scripts for computing model coefficients, median risk score, and leave-one-centre-out model validation were packaged as application "containers" (via Docker) that were locally executed in each node.

Cox model development and validation

The primary analysis involved the development and validation of a Cox proportional hazards model across all centres. The performance of the model was initially assessed using Harrell's concordance index (c-index) [24] on a per-centre basis. The global model's performance was assessed on all data from all three institutions, which has been recommended by Steyerberg and Altman and TRIPOD [20,25] since small datasets should not be split during the model training phase. A more robust estimate for out-of-sample performance was obtained using a closed-loop "leave-one-centre-out" method [20], where new models were trained using data from two sites and then validated on the third site. This was repeated three times to cover the possible combinations, thus resulting in different c-indices which provide an estimate of the over-optimism of the global model. Additionally, the Schoenfeld residuals for each model variable were examined on a per centre level for the global model, and were tested for association with time, in order to examine whether the proportional hazard assumptions were fulfilled [26].

Visualisation of model performance

We evaluated the performance of the global model for risk stratification on a centre level. The individual patient risk score was defined as the overall risk for a patient relative to the baseline and was calculated as the exponent of the patient's linear predictor (LP) value ($risk=e^{[LP]}$). A global median risk score from the global Cox regression model was estimated in an iterative procedure, with the median of the medians as a starting value. The global median risk score was used as cut-off for defining risk categories (high vs low risk), based on individual patient risk scores. Each centre subsequently produced a Kaplan-Meier data object independently in R, with their local survival curves stratified by risk categories, and then shared these objects. These only contained the coordinate points required to plot events and censored patients in a figure.

Results

A total of 281 patients were included in the analysis - 80 patients from Leeds, 81 patients from MAASTRO, and 120 patients from Oslo; see Table 2 for patient characteristics.

Table 2. Overview of patient and treatment characteristics categorised by centre. P-values represent cohort comparisons using either chi-squared or one-way ANOVA tests. GTV: Gross tumour volume. EQD2: Equivalent dose in 2 Gy fractions (α/β=10Gy). IQR: Interquartile range. CI: Confidence interval.

	Leeds	MAASTRO	Oslo	p-value
Disease stage				
Low risk (T1-3N0)	28 (35%)	33 (41%)	58 (48%)	0.16
High risk (T4N(any) or T(any)N+)	52 (65%)	48 (59%)	62 (52%)	
Sex				
Female	53 (66%)	46 (57%)	88 (73%)	0.05
Male	27 (34%)	35 (43%)	32 (27%)	
Age at the start of radiotherapy (years)				
Mean (sd, range)	60 (12, 29-86)	61 (11, 28-84)	62 (10, 40-89)	0.44
Primary tumour GTV (cm3)				
Mean (sd, range)	64.8 (58.7, 2.1-284.9)	57.5 (72.4, 0.8-433.0)	78.1 (69.4, 4.1-459.4)	0.09
Primary tumour dose (EQD2)				
Mean (sd, range)	52.8 (2.7, 49.1-62.6)	60.2 (2.7, 59.4-66.2)	56.3 (2.0, 54.0-58.1)	<0.0001
Potential follow-up time (months)				
Median (IQR)	46 (38-51)	42 (32-63)	49 (39-61)	N/A
Estimated 3-year survival				
Survival (std error, 95%CI)	83% (4%, 76-92%)	78% (5%, 70-88%)	93% (2%, 89-98%)	N/A
Outcome				
Alive	66 (83%)	63 (78%)	107 (89%)	N/A
Dead	14 (17%)	18 (22%)	13 (11%)	

There were no significant differences in disease stage, age at the start of radiotherapy, or primary tumour GTV between the three cohorts. The Oslo cohort had a significantly higher proportion of female to male patients, as expected from the Norwegian anal cancer epidemiology [27]. EQD2 had the highest variance between cohorts, with a difference of 7.4Gy between the highest (MAASTRO) and lowest (Leeds) in mean dose. Moreover, all three cohorts had comparable outcomes and follow-up times. The 3-year survival estimates of Leeds were comparable to both other centres, while the 95% confidence intervals of MAASTRO and Oslo did not overlap.

The results of the global Cox regression model, trained on all three nodes, are summarised in Table 3 in the form of hazard ratio (HR) estimates.

Table 3. Results of the global distributed multivariate Cox regression analysis across all three centres. Age, primary tumour GTV and primary tumour dose were treated as continuous variables. The HRs represent a change of 10 years in age; $10cm^3$ in primary tumour GTV; and 5 Gy in primary tumour dose (EQD2). GTV: Gross tumour volume. EQD2: Equivalent dose in 2 Gy fractions (α/β=10Gy). CI: Confidence interval

	Hazard ratio (95% CI)
High risk disease (compared to low risk disease)	2.02 (0.90-4.54)
Male sex (compared to female sex)	3.06 (1.54-6.11)
Age at the start of RT	1.33 (0.98-1.82)
Primary tumour GTV	1.05 (1.02-1.09)
Primary tumour dose (EQD2)	0.83 (0.48-1.43)

The results of the global model suggest that higher risk disease, older age at the start of radiotherapy, male sex, lower radiotherapy dose, and a greater volume primary tumour (GTV) are associated with worse overall survival. The global model's performance was assessed on each node, yielding a c-index of 0.72 for Leeds, 0.74 for MAASTRO, and 0.70 for Oslo. The c-indices from all three nodes are similar, suggesting that the model performs consistently well across centres.

In addition, the c-indices from the leave-one-centre-out validation runs (Table 4) suggest that the model performance remains stable when model training is carried out using data from only two centres and validated on a third, completely independent dataset. Moreover, the effects of factors are similar across centres, as all three runs produced similar hazard ratios for all variables. The only exception is prescription dose, where one model showed somewhat discordant effects. Notably, the effect of the primary tumour GTV is most consistent across the three validation runs. The overall results of the global model as well as the leave-one-centre-out validation runs were not considerably impacted when including T3N0 tumours in the high risk group (Appendix

A). The Schoenfeld test results convey that the proportional hazard assumptions were fulfilled for all variables in all three centres (Appendix B).

Table 4. Results from the three leave-one-centre-out validation runs. Each column represents one run, consisting of model training (and associated hazard ratios, HR) on two nodes and validation on the third, independent node. Factor effects are presented in terms of hazard ratios with 95% confidence intervals; HR (95% CI). The HRs represent a change of 10 years in age; 10cm^3 in primary tumour GTV; and 5 Gy in primary tumour dose (EQD2). The resulting c-index from each validation run is also reported. GTV: Gross tumour volume. EQD2: Equivalent dose in 2 Gy fractions (α/β=10Gy).

Training nodes	MAASTRO Oslo	Leeds Oslo	Leeds MAASTRO
Validation node	Leeds	MAASTRO	Oslo
High risk disease (compared to low risk disease)	2.52 (0.93-6.78)	1.96 (0.68-5.67)	1.85 (0.71-4.86)
Male sex (compared to female sex)	3.59 (1.55-8.33)	3.83 (1.57-9.37)	2.12 (0.92-4.90)
Age at the start of RT	1.10 (0.74-1.64)	1.47 (0.99-2.17)	1.48 (1.05-2.10)
Primary tumour GTV	1.04 (1.00-1.08)	1.08 (1.03-1.13)	1.07 (1.03-1.11)
Primary tumour dose (EQD2)	0.97 (0.46-2.04)	0.35 (0.14-0.87)	0.97 (0.59-1.59)
Validation c-index	0.70	0.73	0.68

Risk scores were calculated using the global model. A global median risk score of 0.98 was used as the cut-off to define risk categories. Patients with individual risk scores lower than 0.98 were assigned in the low risk category, whereas patients with risk scores greater than 0.98 were assigned in the high risk category. The low risk category consisted of 141 patients (Leeds: 41, MAASTRO: 40, Oslo: 60); the high risk category included 140 patients (Leeds: 39, MAASTRO: 41, Oslo: 60). The Kaplan-Meier curves (Figure 2) convey that there is a good separation in overall survival between the low and high risk categories for two of the centres. For the third centre, the separation is small compared to the other centres.

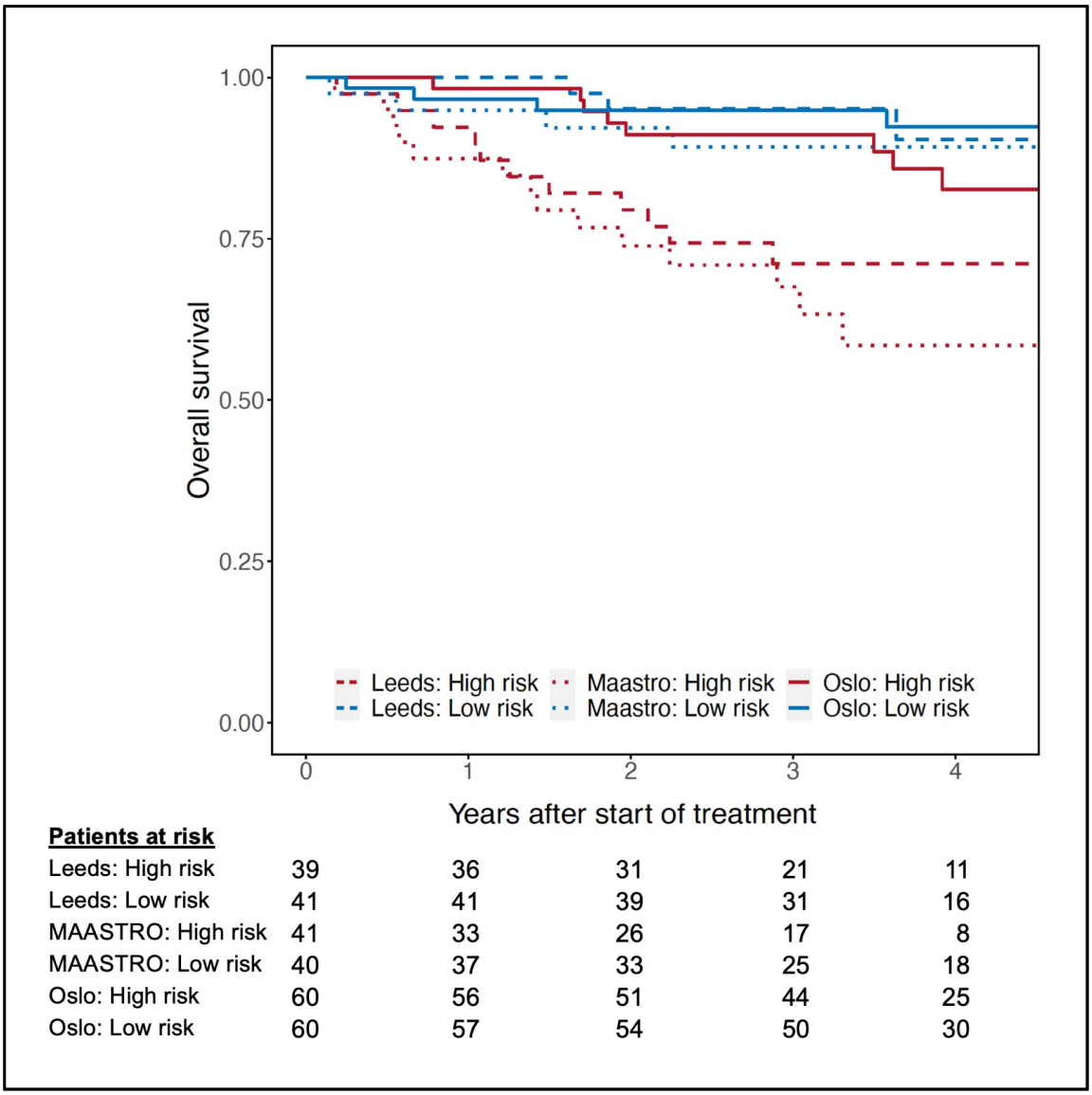

Figure 2. Kaplan-Meier overall survival curves for each centre's cohort, stratified into low and high risk categories. The curves were constructed using the global model, which was trained on data from all three centres. The HR of the high risk category relative to the low risk category is 4.39 [95% CI = 1.22-15.73] for Leeds, 4.02 [1.32-12.23] for MAASTRO and 1.73 [0.56-5.31] for Oslo.

Discussion

This proof-of-concept study demonstrates the feasibility of privacy preserving distributed learning for anal cancer. We trained and validated a Cox proportional hazards regression model [28] in a distributed fashion, using patient data from three European institutions, with clinical and treatment-related factors, and demonstrated robust model performance. Our approach is unique compared to previously published studies employing distributed learning, since we developed and applied a Cox proportional hazards regression model with a time-to-event outcome for a rare cancer.

In contrast, other studies have explored binary outcomes using support-vector machines [6] or logistic regression [7]. In addition, the distributed learning architecture employed in our study is public, open-source and uses Docker containers for enhanced security.

Our analysis involved data for 281 patients treated with modern conformal radiotherapy techniques, including radiotherapy-specific data (GTV volume and prescription dose). This makes for one of the largest available cohorts of anal cancer patients treated with modern radiotherapy, and the only such study with robust multi-centre validation of outcome predictors. Shakir et al [29] reported outcome data from 385 patients treated with IMRT in five UK centres, with median follow-up of 24 months. de Meric de Bellefon et al [30] recently published long-term outcomes, including late toxicity data, for 193 patients treated with IMRT in a single French centre. No other studies have reported on cohorts of this size, and none with multi-national data. Our study could only be realised using the distributed learning methodology, which averted any need for data sharing agreements and data protection reviews.

We found, as expected, worse outcomes for patients with more advanced disease. This, and worse survival for males, mirrors previous results in the literature, including long-term data from RTOG 98-11 [31] and data from a large, prospective Nordic database [32]. Uniquely, by utilising data from 3D planned radiotherapy, we were able to include a volumetric measure of primary tumour size (GTV volume); with an increased risk observed for larger tumours even in multivariate analysis taking staging into account. Tumour size appears to be the most stable factor across all model runs. The relatively weak predictive power of radiation dose was expected as overall survival, and not tumour control, was used as endpoint. Still, the observed effect size was equivalent to that reported for local control in the study by Johnsson et al [14].

Our analysis was limited to data available in routine clinical records for two of the participating centres, and as such potential predictors for outcome were restricted. We selected up front the three clinical factors which we expected to have the largest impact on survival (stage, age, sex), in addition to two radiotherapy-related factors (GTV volume, dose). This process necessarily required some prioritisation, and other factors could equally well have been included such as HPV status, chemotherapy prescription, anatomical site (anal canal versus anal margin), and performance status. We did not examine non-linear effects of age, dose or GTV volume, nor interactions between factors; all of which might be of interest in a more definitive study. Other limitations include variation in staging and GTV definition between centres, as one would expect from a non-prospective multi-centre analysis.

Importantly, the current study was designed to test the feasibility of distributed learning in a rare cancer, with the prospect of accessing combined patient cohorts rivalling the largest reported in the literature. It was not designed as a quality improvement exercise, and as such did not attempt to compare outcomes between centres for specific tumour stages or other patient subgroups. Neither did we set out to produce a definitive model to guide treatment or to test novel predictors for outcome. In its current state, this model is not ready to be used for individual patient predictions. In addition to the inherent limitations related to the medium-size data set, a global baseline

survival curve cannot be provided, which prevents individual patient survival risk estimates. This is a deficiency in the current implementation of Vantage6, which will be addressed in future versions. We examined the use of our global model for risk stratification on an individual centre level, and found good results for two centres. The inability of the model to properly stratify patients in the third centre (Figure 2) may possibly be caused by the high overall survival in that data subset. This emphasises that more centres, with more diverse data, will be needed to develop definitive models.

For optimisation and individualisation of anal cancer radiotherapy, models for locoregional tumour control and late toxicity are needed. For this, more complex radiotherapy data, such as dose volume histogram metrics for both tumour and normal tissue and detailed toxicity and recurrence data are required. Studies also suggest a role for imaging biomarkers for outcome prediction [33–35]. We plan to extend our distributed learning analysis to include both, in a larger network of centres.

We note further that distributed learning per se is not unique and is not perfect when the number of patients per centre is low. We used containerised applications, which provide an isolated execution space to the software and are easily shareable. Containerisation technologies also make it difficult for external parties to tamper with the software. This makes the model algorithms re-usable and agnostic to the specifics of each node installation. We have shown that this implementation works with a diverse collection of hardware and operating systems.

In conclusion, we have demonstrated the utility of privacy preserving distributed learning for analysing multi-national cohorts of patients with rare cancers. We aim to expand the network with more institutions, and also the complexity of our outcome prediction models.

Conflict of interest

The authors report no conflicts of interest.

Acknowledgements

The authors would like to thank Bob Wheller at Leeds Cancer Centre for his assistance with the local implementation of the distributed learning infrastructure.

Role of the funding source

The authors would like to acknowledge Cancer Research UK funding (grant C19942/A28832) for the Leeds Radiotherapy Research Centre of Excellence (RadNet). Per-Ivar Lønne and Eirik Malinen are supported by a grant from the Norwegian Cancer Society (grant no. 182672-2016). Marianne G. Guren is also supported by a grant from the Norwegian Cancer Society (grant no. 190188-2017). Ane Appelt is supported by Yorkshire Cancer Research Academic Fellowship funding (grant L389AA). Alexandra Gilbert is supported by CRUK Clinical trial Fellowship funding (grant C52878/A28301). Peter Brown is supported by Leeds Cares charitable trust funding. Robert Samuel is supported by National Institute for Health Research funding. None of the funders have had any involvement in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

References

[1] Jochems A, Deist TM, van Soest J, Eble M, Bulens P, Coucke P, et al. Distributed learning: Developing a predictive model based on data from multiple hospitals without data leaving the hospital – A real life proof of concept. Radiother Oncol 2016;121:459–67. https://doi.org/10.1016/j.radonc.2016.10.002.

[2] Lu C-L, Wang S, Ji Z, Wu Y, Xiong L, Jiang X, et al. WebDISCO: A web service for distributed cox model learning without patient-level data sharing. J Am Med Inform Assoc 2015:ocv083. https://doi.org/10.1093/jamia/ocv083.

[3] Dwork C. Differential Privacy: A Survey of Results. In: Agrawal M, Du D, Duan Z, Li A, editors. Theory Appl. Models Comput., vol. 4978, Berlin, Heidelberg: Springer Berlin Heidelberg; 2008, p. 1–19. https://doi.org/10.1007/978-3-540-79228-4_1.

[4] Martin F, Sieswerda M, Soest JV, Moncada-Torres A, IntGRen, Codacy Badger. IKNL/vantage6: 1.0.0a1. Zenodo; 2020. https://doi.org/10.5281/ZENODO.3686944.

[5] Moncada-Torres A, Martin F, Sieswerda M, van Soest J, Gelijnse G. VANTAGE6: an open source priVAcy preserviNg federaTed leArninG infrastructurE for Secure Insight eXchange. AMIA Annu Symp Proc 2020.

[6] Deist TM, Jochems A, van Soest J, Nalbantov G, Oberije C, Walsh S, et al. Infrastructure and distributed learning methodology for privacy-preserving multi-centric rapid learning health care: euroCAT. Clin Transl Radiat Oncol 2017;4:24–31. https://doi.org/10.1016/j.ctro.2016.12.004.

[7] Deist TM, Dankers FJWM, Ojha P, Scott Marshall M, Janssen T, Faivre-Finn C, et al. Distributed learning on 20 000+ lung cancer patients – The Personal Health Train. Radiother Oncol 2020;144:189–200. https://doi.org/10.1016/j.radonc.2019.11.019.

[8] Skripcak T, Belka C, Bosch W, Brink C, Brunner T, Budach V, et al. Creating a data exchange strategy for radiotherapy research: Towards federated databases and anonymised public datasets. Radiother Oncol 2014;113:303–9. https://doi.org/10.1016/j.radonc.2014.10.001.

[9] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global Cancer Observatory: Cancer Today 2018. https://gco.iarc.fr/today (accessed September 14, 2020).

[10] Glynne-Jones R, Nilsson PJ, Aschele C, Goh V, Peiffert D, Cervantes A, et al. Anal cancer: ESMO-ESSO-ESTRO clinical practice guidelines for diagnosis, treatment and follow-up. Eur J Surg Oncol J Eur Soc Surg Oncol Br Assoc Surg Oncol 2014;40:1165–76. https://doi.org/10.1016/j.ejso.2014.07.030.

[11] Ajani JA. Fluorouracil, Mitomycin, and Radiotherapy vs Fluorouracil, Cisplatin, and Radiotherapy for Carcinoma of the Anal Canal: A Randomized Controlled Trial. JAMA 2008;299:1914. https://doi.org/10.1001/jama.299.16.1914.

[12] James RD, Glynne-Jones R, Meadows HM, Cunningham D, Myint AS, Saunders MP, et al. Mitomycin or cisplatin chemoradiation with or without maintenance chemotherapy for treatment of squamous-cell carcinoma of the anus (ACT II): a randomised, phase 3, open-label, 2×2 factorial trial. Lancet Oncol 2013;14:516–24. https://doi.org/10.1016/S1470-2045(13)70086-X.

[13] Peiffert D, Tournier-Rangeard L, Gérard J-P, Lemanski C, François E, Giovannini M, et al. Induction Chemotherapy and Dose Intensification of the Radiation Boost in Locally Advanced Anal Canal Carcinoma: Final Analysis of the Randomized UNICANCER ACCORD 03 Trial. J Clin Oncol 2012;30:1941–8. https://doi.org/10.1200/JCO.2011.35.4837.

[14] Johnsson A, Leon O, Gunnlaugsson A, Nilsson P, Höglund P. Determinants for local tumour control probability after radiotherapy of anal cancer. Radiother Oncol 2018;128:380–6. https://doi.org/10.1016/j.radonc.2018.06.007.

[15] Muirhead R, Partridge M, Hawkins MA. A tumor control probability model for anal squamous cell carcinoma. Radiother Oncol 2015;116:192–6. https://doi.org/10.1016/j.radonc.2015.07.014.

[16] ISRCTN registry [Internet]. London: BMC. ISRCTN88455282, PLATO - Personalising anal cancer radiotherapy dose 2016. https://doi.org/10.1186/ISRCTN88455282 (accessed December 21, 2020).
[17] Gilbert A, Drinkwater K, McParland L, Adams R, Glynne-Jones R, Harrison M, et al. UK national cohort of anal cancer treated with intensity-modulated radiotherapy: One-year oncological and patient-reported outcomes. Eur J Cancer 2020;128:7–16. https://doi.org/10.1016/j.ejca.2019.12.022.
[18] El Naqa I, editor. A guide to outcome modeling in radiotherapy and oncology: listening to the data. Boca Raton, FL: CRC Press, Taylor & Francis Group; 2018.
[19] Fish R, Sanders C, Adams R, Brewer J, Brookes ST, DeNardo J, et al. A core outcome set for clinical trials of chemoradiotherapy interventions for anal cancer (CORMAC): a patient and health-care professional consensus. Lancet Gastroenterol Hepatol 2018;3:865–73. https://doi.org/10.1016/S2468-1253(18)30264-4.
[20] Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. J Clin Epidemiol 2016;69:245–7. https://doi.org/10.1016/j.jclinepi.2015.04.005.
[21] Amin MB, American Joint Committee on Cancer, American Cancer Society, editors. AJCC cancer staging manual. Eight edition / editor-in-chief, Mahul B. Amin, MD, FCAP ; editors, Stephen B. Edge, MD, FACS [and 16 others] ; Donna M. Gress, RHIT, CTR-Technical editor ; Laura R. Meyer, CAPM-Managing editor. Chicago IL: American Joint Committee on Cancer, Springer; 2017.
[22] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: 2014.
[23] Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. Control Clin Trials 1996;17:343–6. https://doi.org/10.1016/0197-2456(96)00075-X.
[24] Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 2011:n/a-n/a. https://doi.org/10.1002/sim.4154.
[25] Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med 2015;162:W1. https://doi.org/10.7326/M14-0698.
[26] Schoenfeld D. Partial residuals for the proportional hazards regression model. Biometrika 1982;69:239–41. https://doi.org/10.1093/biomet/69.1.239.
[27] Guren MG, Aagnes B, Nygård M, Dahl O, Møller B. Rising Incidence and Improved Survival of Anal Squamous Cell Carcinoma in Norway, 1987-2016. Clin Colorectal Cancer 2019;18:e96–103. https://doi.org/10.1016/j.clcc.2018.10.001.
[28] Singh R, Mukhopadhyay K. Survival analysis in clinical trials: Basics and must know areas. Perspect Clin Res 2011;2:145. https://doi.org/10.4103/2229-3485.86872.
[29] Shakir R, Adams R, Cooper R, Downing A, Geh I, Gilbert D, et al. Patterns and Predictors of Relapse Following Radical Chemoradiation Therapy Delivered Using Intensity Modulated Radiation Therapy With a Simultaneous Integrated Boost in Anal Squamous Cell Carcinoma. Int J Radiat Oncol 2020;106:329–39. https://doi.org/10.1016/j.ijrobp.2019.10.016.
[30] de Meric de Bellefon M, Lemanski C, Castan F, Samalin E, Mazard T, Lenglet A, et al. Long-term follow-up experience in anal canal cancer treated with Intensity-Modulated Radiation Therapy: Clinical outcomes, patterns of relapse and predictors of failure. Radiother Oncol J Eur Soc Ther Radiol Oncol 2020;144:141–7. https://doi.org/10.1016/j.radonc.2019.11.016.
[31] Gunderson LL, Moughan J, Ajani JA, Pedersen JE, Winter KA, Benson AB 3rd, et al. Anal carcinoma: impact of TN category of disease on survival, disease relapse, and colostomy failure in US Gastrointestinal Intergroup RTOG 98-11 phase 3 trial. Int J Radiat Oncol Biol Phys 2013;87:638–45. https://doi.org/10.1016/j.ijrobp.2013.07.035.
[32] Leon O, Guren M, Hagberg O, Glimelius B, Dahl O, Havsteen H, et al. Anal carcinoma - Survival and recurrence in a large cohort of patients treated according to Nordic

guidelines. Radiother Oncol J Eur Soc Ther Radiol Oncol 2014;113:352–8. https://doi.org/10.1016/j.radonc.2014.10.002.
[33] Brown PJ, Zhong J, Frood R, Currie S, Gilbert A, Appelt AL, et al. Prediction of outcome in anal squamous cell carcinoma using radiomic feature analysis of pre-treatment FDG PET-CT. Eur J Nucl Med Mol Imaging 2019;46:2790–9. https://doi.org/10.1007/s00259-019-04495-1.
[34] Jones M, Hruby G, Coolens C, Driscoll B, Stanwell P, Kumar M, et al. A prospective, multi-centre trial of multi-parametric MRI as a biomarker in anal carcinoma. Radiother Oncol 2020;144:7–12. https://doi.org/10.1016/j.radonc.2019.10.001.
[35] Rusten E, Rekstad BL, Undseth C, Klotz D, Hernes E, Guren MG, et al. Anal cancer chemoradiotherapy outcome prediction using 18F-fluorodeoxyglucose positron emission tomography and clinicopathological factors. Br J Radiol 2019;92:20181006. https://doi.org/10.1259/bjr.20181006.

Appendices

Appendix A

When including T3N0 tumours in the high risk group, 2 patients changed category in the Leeds cohort, 3 in the MAASTRO cohort and 12 in the Oslo cohort. Table A1 presents the results of the global model with this alternative grouping for disease stage. The validation c-indices derived from the global model are 0.74 for Leeds, 0.73 for MAASTRO, and 0.71 for Oslo. The results from the leave-one-centre-out validation runs using the alternative grouping for disease stage are shown in Table A2.

Table A1. Results of the global distributed multivariate Cox regression analysis across all three centres, carried out with the alternative grouping for disease stage, for which T3N0 tumours were assigned to the "high risk" group. Age, primary tumour GTV and primary tumour dose were treated as continuous variables. The HRs represent a change of 10 years in age; $10cm^3$ in primary tumour GTV; and 5 Gy in primary tumour dose (EQD2). GTV: Gross tumour volume. EQD2: Equivalent dose in 2 Gy fractions (α/β=10Gy). CI: Confidence interval

	Hazard ratio (95% CI)
High risk disease (compared to low risk disease)	1.46 (0.62-3.45)
Male sex (compared to female sex)	2.40 (1.26-4.59)
Age at the start of RT	1.30 (0.95-1.78)
Primary tumour GTV	1.06 (1.02-1.09)
Primary tumour dose (EQD2)	0.87 (0.54-1.41)

Table A2. Results from the three leave-one-centre-out validation runs, carried out with the alternative grouping for disease stage, for which T3N0 tumours were assigned to the "high risk" group. Each column represents one run, consisting of model training on two nodes and validation on the third, independent node. Factor effects are presented in terms of hazard ratios with 95% confidence intervals; HR (95% CI). The HRs represent a change of 10 years in age; 10cm3 in primary tumour GTV; and 5 Gy in primary tumour dose (EQD2). The resulting c-index from each validation run is also reported. GTV: Gross tumour volume. EQD2: Equivalent dose in 2 Gy fractions (α/β=10Gy).

Training nodes	MAASTRO Oslo	Leeds Oslo	Leeds MAASTRO
Validation node	Leeds	MAASTRO	Oslo

High risk disease (compared to low risk disease)	1.58 (0.53-4.70)	1.89 (0.62-5.70)	1.85 (0.71-4.86)
Male sex (compared to female sex)	3.02 (1.32-6.87)	2.70 (1.19-6.16)	2.12 (0.92-4.90)
Age at the start of RT	1.06 (0.71-1.58)	1.46 (0.98-2.17)	1.48 (1.05-2.10)
Primary tumour GTV	1.05 (1.00-1.09)	1.07 (1.02-1.12)	1.07 (1.03-1.11)
Primary tumour dose (EQD2)	0.98 (0.47-2.06)	0.45 (0.22-0.94)	0.97 (0.59-1.59)
Validation c-index	0.69	0.73	0.67

Appendix B

Table B1. Schoenfeld test results for all variables included in our model, categorised by centre.

Variable	Schoenfeld test p-value		
	Leeds	MAASTRO	Oslo
Disease stage	0.74	0.62	0.22
Sex	0.10	0.38	0.14
Age at the start of RT	0.11	0.98	0.99
Primary tumour GTV	0.22	0.20	0.38
Primary tumour dose (EQD2)	0.76	0.45	0.77