

An international multi-centre study to develop and validate federated learning-based prognostic models for anal cancer

Received: 27 January 2025

Accepted: 20 February 2026

Cite this article as: Theophanous, S., Lønne, P.-I., Choudhury, A. *et al.* An international multi-centre study to develop and validate federated learning-based prognostic models for anal cancer. *Nat Commun* (2026). <https://doi.org/10.1038/s41467-026-70297-3>

Stelios Theophanous, Per-Ivar Lønne, Ananya Choudhury, Maaïke Berbee, Charlotte Deijen, Andre Dekker, Matthew Field, Maria Antonietta Gambacorta, Alexandra Gilbert, Marianne Grønlie Guren, Rashmi Jadon, Rohit Kochhar, Daniel Martin, Ahmed Allam Mohamed, Rebecca Muirhead, Oriol Parés, Łukasz Raszewski, Rajarshi Roy, Andrew Scarsbrook, David Sebag-Montefiore, Emiliano Spezi, Vassilios Vassiliou, Eirik Malinen, Leonard Wee, Ane Appelt & atomCAT Consortium

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

AN INTERNATIONAL MULTI-CENTRE STUDY TO DEVELOP AND VALIDATE FEDERATED LEARNING-BASED PROGNOSTIC MODELS FOR ANAL CANCER

AUTHOR NAMES AND AFFILIATIONS

STELIOS THEOPHANOUS^{*,1,2}, PER-IVAR LØNNE³, ANANYA CHOUDHURY⁴, MAAIKE BERBEE⁴, CHARLOTTE DEIJEN^{5,6}, ANDRE DEKKER⁴, MATTHEW FIELD^{7,8,9,10}, MARIA ANTONIETTA GAMBACORTA¹¹, ALEXANDRA GILBERT^{1,2}, MARIANNE GRØNLIE GUREN^{12,13}, RASHMI JADON¹⁴, ROHIT KOCHHAR¹⁵, DANIEL MARTIN^{16,17,18}, AHMED ALLAM MOHAMED^{19,20}, REBECCA MUIRHEAD^{21,22}, ORIOL PARÉS²³, ŁUKASZ RASZEWSKI²⁴, RAJARSHI ROY²⁵, ANDREW SCARSBROOK^{1,2}, DAVID SEBAG-MONTEFIORE², EMILIANO SPEZI²⁶, VASSILIOS VASSILIOU²⁷, EIRIK MALINEN^{+,28}, LEONARD WEE^{+,4}, ANE APPELT^{+,1,2}, ATOMCAT CONSORTIUM[‡]

¹ Leeds Teaching Hospitals NHS Trust, Leeds, UK

² Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, UK

³ Department of Medical Physics, Oslo University Hospital, Oslo, Norway

⁴ Department of Radiation Oncology (Maastr), GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, The Netherlands

⁵ Department of Radiation Oncology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek (NKI-AVL), Amsterdam, The Netherlands

⁶ Department of Radiation Oncology, Amsterdam University Medical Center, Amsterdam, The Netherlands

⁷ South Western Sydney Clinical Campus, School of Clinical Medicine, UNSW, Sydney, New South Wales, Australia

⁸ Ingham Institute for Applied Medical Research, Liverpool, New South Wales, Australia

⁹ Department of Radiation Oncology, Liverpool and Macarthur Cancer Therapy Centres, Sydney, New South Wales, Australia

¹⁰ School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, New South Wales, Australia

¹¹ Fondazione Policlinico Universitario A.Gemelli IRCCS, Università Cattolica S.Cuore, Rome, Italy

¹² Department of Oncology, Oslo University Hospital, Oslo, Norway

¹³ Institute of Clinical Medicine, University of Oslo, Oslo, Norway

¹⁴ Cambridge University Hospital NHS Foundation Trust, Cambridge, UK

¹⁵ The Christie NHS Foundation Trust, Manchester, UK

¹⁶ Goethe University Frankfurt, University Hospital, Department of Radiotherapy and Oncology, Frankfurt, Germany

¹⁷ German Cancer Research Center (DKFZ), Heidelberg, Germany, German Cancer Consortium (DKTK), Partner Site, Frankfurt, Germany

¹⁸ Frankfurt Cancer Institute (FCI), Frankfurt, Germany

¹⁹ RWTH Aachen University Medical Centre, Aachen, Germany

²⁰ Center for Integrated Oncology Aachen Bonn Köln Düsseldorf (CIO ABCD), Aachen, Germany

²¹ Department of Oncology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

²² Department of Oncology, University of Oxford, Oxford, UK

²³ Champalimaud Foundation, Lisbon, Portugal

²⁴ Greater Poland Cancer Centre, Poznań, Poland

²⁵ Hull University Teaching Hospitals NHS Trust, Hull, UK

²⁶ School of Engineering, Cardiff University, Cardiff, UK

²⁷ Bank of Cyprus Oncology Centre, Nicosia, Cyprus

²⁸ Department of Radiation Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway

* Corresponding author: stelios.theophanous@nhs.net

+ These authors jointly supervised this work.

‡ A list of authors and their affiliations appears at the end of the paper.

ARTICLE IN PRESS

KEYWORDS

Anal cancer; squamous cell carcinoma; chemoradiotherapy; federated learning; distributed learning; outcome modelling; overall survival; locoregional control; freedom from distant metastases

ABSTRACT

Precision oncology relies on access to high-quality data for increasingly smaller patient subgroups. The international atomCAT consortium investigates the potential of federated learning to support this, using anal cancer as a rare cancer exemplar. Here, we show that federated multivariable Cox models trained across 14 centres (1,428 patients) and externally validated in two additional centres (277 patients) achieve consistent calibration and discrimination during leave-one-centre-out and external validation (c-indices 0.68-0.79). Lower T stage, absence of nodal involvement, smaller tumour volume, female sex, younger age, and mitomycin- or cisplatin-based chemotherapy are associated with improved overall survival. Lower T stage, smaller tumour volume, and female sex are associated with improved locoregional control, while absence of nodal involvement and smaller tumour volume are associated with better freedom from distant metastases. These findings demonstrate that federated learning enables robust, privacy-preserving prognostic modelling for rare cancers using real-world data, supporting international collaboration without data sharing.

INTRODUCTION

Precision oncology has brought increasing use of molecular profiling and sub-stratification, with more patients being treated with targeted agents guided by tumour genetic profiles. Consequently, many previously deemed 'common cancers' (e.g. lung cancer) can now be considered as a collection of rarer tumours. Therefore, rare cancers are diagnosed in an increasingly large proportion of patients¹, and ultimately, all cancers are becoming rare cancers. This poses a key challenge for clinical trials and research: small patient numbers and wide geographic spread make it difficult to generate evidence by standard methods². Recruitment for large clinical trials is complicated and often impossible³, but traditional registry data do not have the depth needed for detailed subgrouping and analyses. Consequently, this constitutes a major barrier to research and development in an era of stratified medicine.

There has been growing recognition of the significant role of real-world data (RWD) in addressing the challenges related to rare cancer research. RWD can be broadly defined as 'data relating to patient health or experience or care delivery which is collected outside of clinical trials'^{4,5}. Given that rare cancers are often dispersed across different healthcare institutions and geographic regions, collating large, high-quality datasets with sufficient depth is complex when reliant on centralised analysis. Patient privacy and data protection in particular are often the key obstacles in this setting⁶. Federated learning (FL), however, enables collaborative analysis without the need to pool data centrally^{7,8}. Through this approach, individual institutions retain control over their data, ensuring patient privacy and compliance with data governance regulations. FL with RWD offers a solution to data scarcity for rare cancers, allowing for better understanding of prognostic factors as well as outcomes in patient sub-groups^{9,10}.

The international atomCAT consortium (Anal cancer Treatment Outcome Modelling with Computer Aided Theragnostics) was established in 2021 with the aim of demonstrating the value of FL in rare cancers, using anal cancer as the exemplar. Specifically, the consortium set out to identify prognostic factors for anal cancer using FL¹¹, building on previous proof-of-concept work¹². Anal carcinoma is rare, comprising approximately 0.3% of all cancer cases^{13,14}. However, its incidence is rising substantially, with studies showing a 35% increase in men and a 75% increase in women between 1988 and 2012 across Europe¹⁵. This trend is largely attributed to increased exposure to high-risk HPV, following demographic and behavioural shifts since the introduction of the oral contraceptive pill in the 1960s. Despite the rising incidence, most centres treat relatively few cases per year and may struggle to identify appropriate evidence-based treatment for patient subgroups, many of whom are under-represented in clinical trials. As an example, if an oncologist were to see an elderly patient with anal cancer in their clinic today, the treatment offered might differ substantially across cancer centres and countries¹⁶.

Current standard treatment for localised anal cancer consists of concurrent chemotherapy and radiotherapy¹⁷⁻¹⁹ delivered with intensity-modulated radiation therapy (IMRT) or volumetric modulated arc therapy (VMAT)¹⁹. This approach confers favourable outcomes: modern UK and US multi-centre cohorts have reported three-year overall survival and disease-free survival rates of 80-85% and 75%, respectively^{20,21}. Historically, most disease relapses have

been reported to occur locoregionally, i.e. at the site of the primary disease or in involved lymph nodes, although this pattern may be changing with modern treatment strategies²¹. Previous research has conveyed a clear need for improved treatment stratification with respect to locoregional treatment intensification or deintensification, as well as the use of novel agents for systemic disease management. The design of new treatment stratification strategies depends on robust assessment of risk groups with respect to both locoregional and distant metastatic control^{22,23}.

In this report, we demonstrate the value offered by international real-world curated data research utilising federated learning through a prospectively planned and pre-registered study of a large international anal cancer patient cohort. We report on the development and validation of federated prognostic models for overall survival (OS), locoregional tumour control (LRC), and freedom from distant metastases (FFDM), based on data from anal cancer patients treated with modern chemoradiotherapy at 16 cancer centres across Europe and Australia.

RESULTS

A total of 1428 patients, treated from 2004 to 2022 across 14 participating centres, were identified for inclusion in the primary cohort analysis, with an additional 277 patients available for external validation. Table 1 summarises the patient characteristics of the primary and validation cohorts; with data stratified on individual centres available in Supplementary Material: Supplementary Note 5. More than 80% of patients were treated between 2010 and 2022.

The overall cohort had a weighted mean age of 62 at the start of radiotherapy, with 28% of patients older than 70 years. The majority of patients (70%) were female. Primary tumour volume varied considerably across centres, from a mean of 23.9 cm³ in Centre 6 to a mean of 120.5 cm³ in Centre 1. Mean prescribed primary tumour dose (in equivalent dose in 2Gy per fraction, $\alpha/\beta=10\text{Gy}$, EQD2 $_{\alpha/\beta=10\text{Gy}}$) was relatively consistent across centres, ranging from 50.6Gy to 60.1Gy. The most common chemotherapy regimen was Mitomycin C (MMC) and 5-fluorouracil (5FU) / capecitabine (Cap), which 91% (n=1292) of patients received. Only 6% (n=83) of patients received no chemotherapy.

Summary survival statistics for the three outcomes are provided in Table 2, and presented visually for all centres in Figure 1, with full per centre data in the Supplementary Material: Supplementary Note 5. Weighted mean potential follow-up times across the entire cohort were 48.7 months for OS, 42.0 months for LRC, and 41.8 months for FFDM. At 3 years, the weighted means (range) for OS, LRC, and FFDM were 83% (75%-97%), 83% (73%-90%), and 87% (57%-96%), respectively. OS rates varied more than LRC and FFDM rates across centres.

The results from the OS, LRC, and FFDM global models, trained using all available data from all primary cohort centres, are presented in Table 3. Lower T stage, lack of nodal involvement, female sex, younger age, smaller primary tumour, and use of doublet chemotherapy with MMC or Cisplatin and 5FU or Cap (relative to no chemotherapy) were associated with better OS. Moreover, lower T stage, smaller primary tumour size, and female sex were

associated with better LRC. Lastly, lack of nodal involvement and smaller primary tumour size were associated with better FFDM.

The multivariable outcome models allow for individual patient risk prediction using the reference outcome rates summarised in Supplementary Material: Supplementary Note 5. Observed weighted mean three-year outcome rates across centres were markedly different between predicted high- and low-risk groups: 73% vs 90% for OS; 76% vs 91% for LRC, and 83% vs 94% for FFDM. Plots for high- and low-risk groups, indicating the observed weighted mean outcome rates across centres at two, three, and five years are presented in Supplementary Material: Supplementary Note 5. Figure 2 demonstrates the failure risk distribution for locoregional control and distant metastases for individual patients in a single centre cohort (Centre 2, n=210); with the impact of tumour volume on failure risk highlighted.

The global model performance on the primary study cohort was acceptable, with weighted mean c-indices of 0.68, 0.71, 0.69 for OS, LRC, and FFDM, respectively. The performance was stable for the leave-one-centre-out internal validation (0.67, 0.69, 0.66), indicating very little overfitting. This was confirmed by consistent model performance in the external validation cohorts (0.72, 0.75, 0.79). Full details of model performance per centre can be found in Supplementary Material: Supplementary Note 5.

To illustrate the added value of the federated approach, comparator models for all three outcomes were developed using the same specifications, but only data from a single centre (Centre 2, n=210). When evaluated on the external validation cohorts, this single-centre models yielded substantially lower C-indices for all outcomes (OS: 0.60, LRC: 0.70, FFDM: 0.62) compared to the federated models (0.72, 0.75, and 0.79, respectively), indicating reduced generalisability.

None of the secondary models (Supplementary Material: Supplementary Note 5) performed considerably better than the primary cohort models. For OS and LRC, the secondary models including performance status and treatment compliance performed minimally better, but at the cost of having additional parameters. This took the total model parameters beyond the number supported by the available sample size. Using a simple split on staging (high risk: T4 N_{any} or T_{any} N+, vs low risk: T1-3 N0), instead of including T stage and N stage individually, was less useful for LRC (where it became non-significant) but still provided relevant information for OS and FFDM. For distant metastases, removing GTV size from the model resulted in a significant effect of T stage on outcome, which was not observed with GTV included. None of the alternative parametrisations of continuous factors (age, GTV size) performed better than the ones used in the primary models.

Figure 3 presents calibration plots for three-year OS, LRC, and FFDM. Points closer to the reference line indicate stronger calibration and signify that the predicted outcome rates correspond more closely to actual (observed) outcome rates. The results indicate that calibration was better for LRC prediction compared to OS and (to a lesser

extent) FFDM. Calibration was generally good in the external validation cohorts (indicated with black circles in Figure 3).

DISCUSSION

This international consortium study aimed to demonstrate the use of federated learning on real-world data to develop and validate prediction models for anal cancer outcomes after chemoradiotherapy. For three clinically important outcomes, consistent model performance was demonstrated in independent external validation datasets. Importantly, we identified a robust set of prognostic factors with differential effects on survival, locoregional control and risk of distant metastatic disease. Although some of the prognostic variables identified (e.g., T stage, N stage, GTV) are well supported in previous studies, our findings show that federated learning can reproduce and externally validate these variables. This highlights the capability of federated learning for developing robust, privacy-preserving prognostic models for rare cancers using real-world data.

To our knowledge, the multi-national atomCAT cohort of 1705 patients is one of the largest contemporary international anal cancer cohorts analysed to date. We found that lower T stage was associated with better OS and LRC, while nodal involvement was associated with poorer OS and higher risk of distant metastases. Sex was associated with OS and LRC (but not distant metastases), with females doing considerably better than males for both outcomes. These results confirm findings from previous smaller studies, summarised in our systematic review²⁴. We demonstrated the importance of primary tumour size as a prognostic factor for all three outcomes examined. This emphasises the importance of imaging-specific factors on cancer outcomes; data which are not usually available e.g. in centralised cancer registries, although noting the overlap of this factor with T stage. Interestingly, our findings suggest that prescribed radiotherapy dose is not prognostic for any of the three outcomes, which contradicts previous reports of radiotherapy dose-response for tumour control based on literature data^{25,26}. This lack of dose-effect relationship could reflect cohort heterogeneity and/or confounding by indication, as higher doses tend to be prescribed to patients with more advanced disease. In addition, variations in dose reporting across institutions may dilute this relationship. Prospective, randomised studies with risk-adapted doses such as PLATO²⁷ are expected to clarify this issue.

While the discrimination metrics from the outcome models were moderate (c-indices 0.68 - 0.79), these values are comparable with other externally validated real-world prognostic models in oncology in general and anal cancer specifically²⁸⁻³⁰. They indicate clinically meaningful performance in a rare cancer setting, where predictive model performance is often limited by cohort size and heterogeneity. Moreover, slightly poorer calibration for OS and FFDM relative to LRC was observed, which likely reflects greater heterogeneity in non-cancer mortality across

centres. Despite this, model performance remained within acceptable limits for clinical interpretation. Formal statistical comparisons between the training and validation cohorts were not performed, as these were independent datasets used to test model generalisability across different populations. The observed differences in demographics and treatment reflect real-world variability, which the federated model is designed to accommodate.

Our achievement was galvanised by use of RWD connected through FL, avoiding complicated data protection issues and data sharing agreements. Although comparison to a fully centralised model was not feasible due to data governance constraints, the federated Cox regression algorithm used in this study has been shown to be mathematically equivalent to a centralised model under the Breslow approximation when applied to the same data³¹. As such, this study demonstrates the potential of the FL approach for generation of robust evidence for rare cancers, where high-quality data for large patient cohorts are otherwise challenging to access. Given the increasing focus on molecular sub-stratification of many cancers, this will become important for a growing number of patients in the coming years. However, using RWD to explore comparative effectiveness (such as effects of different dose regimens) has numerous potential pitfalls⁵, including confounding by indication, and should preferably be considered in a formal causal analysis framework³². As an example, the limitations for analysis of radiotherapy dose and its impact on LRC are discussed above. This question has been explored in the PLATO trial²⁷, which is expected to provide a definitive answer to this question on treatment stratification in the near future. In other words, RWD should be seen as complementary to and not a replacement of clinical trials. It may offer crucial guidance for future trial design, such as identification of particular risk groups, and may augment knowledge from clinical trials³³⁻³⁵. Importantly, older patients comprised nearly one-third of our cohort, yet remain poorly studied in prospective trials¹⁶. Models trained on RWD may therefore offer a tool to support individualised treatment discussions and improve prognostication for such under-represented groups.

In addition to the FL methodology discussed above, the atomCAT consortium study has several other strengths. Firstly, feature selection was based on expert consensus and literature review, rather than automated methods, avoiding known limitations with techniques like LASSO and stepwise selection in smaller, heterogeneous datasets³⁶⁻³⁸. This approach ensured clinical relevance, reproducibility, and supported prospective sample size calculation. The prospective nature of various aspects of this study, including the sample size calculation and study pre-registration, is a particular strength. The former aimed to minimise model overfitting and ensure that the overall risk of each outcome was estimated precisely³⁹, while the latter guarded against data dredging and post-hoc model tuning^{40,41}. Consequently, model performance as well as individual factor effect sizes are likely robust and not over-optimistic. This is supported by consistent model performance in the internal-external (“leave-one-centre-out”) and external cohort validation⁴². Additionally, use of a large and diverse set of patients, covering a range of geographical regions and healthcare systems with a variety of standard treatment protocols, increases the likelihood of the findings being generalisable beyond the study population.

Data availability, quality, and curation are well-known challenges for studies of RWD generated in routine care setting, and also underlies some of the main limitations of our study⁵. Some potential prognostic factors were not selected for inclusion in the primary model analysis due to their unavailability for most patients at many consortium centres. HIV and HPV status were excluded due to inconsistent data availability across centres, and performance status was omitted because of high missingness and potential informative bias. Frailty may ultimately represent a more clinically meaningful predictor of treatment tolerance and outcome than performance status or age; however, its incorporation into federated analyses would require standardised definitions to ensure comparability across centres. The study protocol and data analysis plan adopted a pragmatic approach, exploring factors routinely recorded in daily practice, rather than e.g. emerging or novel biomarkers. This minimised any issues with missing data but may have introduced unknown confounding by indication. Data curation and quality assurance was facilitated by keeping all data within individual centres, where review and validation for individual patient cases were straightforward. This proved to be an additional benefit of the federated learning approach compared to centralised data analysis.

Some of the study limitations are related to technical aspects of the federated learning architecture; especially with respect to standard prognostic model development steps which were not available in our federated implementation at the time of analysis. This includes advanced methods for missing data imputation^{43,44} and correlation analysis. Instead, simple imputation methods were used, and correlated variables were handled during the study design phase. For example, secondary models explored the overlap between T stage and GTV size and indicated that some (but not all) of the information from GTV size can be captured by T stage. Due to privacy restrictions within the federated learning framework, individual-level Kaplan-Meier curves stratified by clinical subgroups could not be generated, although future Vantage6 algorithms may allow secure aggregation of subgroup survival estimates. Additionally, although the federated Cox model is mathematically equivalent to the centralised version under the Breslow approximation, small numerical deviations may occur due to asynchronous updates or local preprocessing. These differences are likely to be negligible and do not affect overall model conclusions. Moreover, two centres were excluded from the summary 5-year outcome rates due to incomplete follow-up; however, this represents a small fraction of the total cohort (14 centres), and the two- and three-year endpoints remain unaffected. Subgroup analyses were not prespecified and were underpowered in the available cohorts. Future federated learning studies with larger datasets will allow more reliable evaluation of subgroups, including older, frailer, and comorbid patients.

Future enhancements to federated learning are likely to come from advances in data curation, interoperability, and infrastructure. Large language models and multimodal generative pre-trained transformer (GPT) systems could assist with local data cleaning, mapping, and harmonisation, reducing the burden of manual curation. Adoption of common data models (CDMs) and interoperability frameworks, such as the OMOP CDM⁴⁵ and the operational ontology for oncology (O3)⁴⁶, can enable collaboration and analysis across heterogeneous clinical datasets. Generative AI could also be deployed locally to create synthetic patient data to support imputation or mitigate inclusion bias without

sharing identifiable information. The Vantage6 platform supports containerised implementation of a large range of model types, including machine learning and deep learning algorithms. Recent examples include federated survival forests and convolutional neural networks for radiotherapy and imaging research⁴⁷, illustrating the potential for broader applications as the infrastructure matures further.

In conclusion, the atomCAT consortium analysed a large, geographically distributed multi-centre anal cancer cohort. The study provides unique insights into the distinct prognostic effect of different patient and disease characteristics on OS, LRC, and FFDM, and may inform the design of future clinical trials and risk stratification for treatment personalisation. It demonstrates the value offered by high-quality international real-world data research utilising federated learning. This approach may now also be used for more complex research questions and data, such as tissue, blood and imaging biomarkers or additional radiotherapy-specific data. Our robust and validated prognostic models, developed using large cohorts, lay the foundation for decision support tools that can inform shared decision-making, enable more personalised treatment strategies, and may be considered in prospective clinical trials for patient inclusion and stratification. In the future, this federated approach could also guide biomarker-driven subgroup analyses and support policy development, such as reimbursement decisions for new therapies and informing of international guidelines, through privacy-preserving analyses of real-world data at scale.

METHODS

This study was conducted in accordance with all relevant ethical regulations, with approval and data access permissions obtained from the HRA and the institutional review boards or equivalent ethics committees at each participating centre.

INSTITUTIONAL DATA ACCESS & DATA PROTECTION APPROVALS

The atomCAT2 study used FL methodology to develop prognostic models using local datasets from 16 participating centres (see Supplementary Material: Supplementary Note 2), without exchange of any sensitive individual patient-level data. Only non-identifiable aggregated information in the form of mathematical parameters, such as model coefficients, was shared between centres to train and validate the federated models. Each institution obtained approval from the relevant institutional review board or equivalent ethics committee for accessing and using patient data for research, and provided a copy of their local data access approval, including approving body and reference number, to the central study coordinator. For UK centres, central approval was provided by the Health Research Authority (HRA) (IRAS project ID: 303103, REC reference: 22/WA/0081). Full details of the approving body, approval reference, and informed consent status for each centre are provided in Supplementary Note 2.

The atomCAT2 study involved no protocol-specific intervention beyond standard-of-care radiotherapy. Given the retrospective, non-interventional nature of the study and the use of routinely collected, de-identified data that

remained at the originating institutions, the requirement for individual patient informed consent was waived or deemed not required by the approving bodies, as detailed in Supplementary Note 2.

PROSPECTIVE STUDY PROTOCOL & STATISTICAL ANALYSIS PLAN

A prospective study protocol and comprehensive statistical analysis plan were developed collaboratively, published¹¹, and registered in Open Science Foundation⁴⁸. Details of data item selection and definition, outcome definition, model specification, sample size calculation, and model evaluation are available in the protocol¹¹. Model development, validation procedure and results were reported in accordance with the TRIPOD statement and checklist⁴⁹ (see Supplementary Material: Supplementary Note 1). Figure 4 illustrates the workflow for the entire study, including previously published preliminary and preparatory work¹².

STUDY DESIGN & PATIENT POPULATION

Patients treated with radical intent external beam radiotherapy for primary anal squamous cell carcinoma, with or without concomitant chemotherapy, were included. Inclusion was restricted to patients treated with conformal radiotherapy techniques (3D conformal radiotherapy (3D-CRT) or IMRT/VMAT). Individuals treated with palliative intent, and patients who had received prior pelvic radiotherapy or brachytherapy (either primary or as boost treatment) were excluded. Patients were treated according to each participating centre's protocols, which consisted of concurrent radiotherapy and chemotherapy with varying regimens, or radiotherapy only.

OUTCOME DEFINITIONS

Three outcomes were employed: OS, LRC, and FFDM. The Core Outcome Research Measures in Anal Cancer (CORMAC) initiative has identified these as key outcome measures for anal cancer research⁵⁰. Complete definitions for all three outcomes can be found in the study protocol¹¹, in brief: OS included death from any event, with censoring at last clinical follow-up; LRC events included lack of complete response at evaluation following chemoradiotherapy and locoregional disease progression during follow-up. Patients were censored at death, at last clinical follow-up, if undergoing abdominoperineal resection for non-disease related reasons, or in case of distant metastases; distant metastasis was defined as disease recurrence outside of the primarily treated volumes, including any non-irradiated nodes outside the pelvis or other organs e.g. lung or liver. Patients were censored at local recurrence, at death, or at last clinical follow-up. All times were calculated from first radiotherapy treatment date.

IDENTIFICATION OF RELEVANT PROGNOSTIC FACTORS

To identify potential prognostic factors for the three outcomes, a systematic review of the literature was conducted²⁴; this included studies published after 2000, with medium-to-large cohorts (>100 patients), reporting

on anal cancer outcomes after treatment with conformal radiotherapy. Factors identified as prognostic through multivariable analysis in several studies were selected and prioritised (based on the number of studies reporting on them), which formed an initial list of relevant data to be collected. This list was reviewed by three senior clinical oncologists (AG, MGG, MB), who added additional relevant factors. A data scoping exercise across participating centres limited the list to factors consistently available in routine clinical data. HIV and HPV status, though clinically relevant, were excluded due to inconsistent availability and high risk of informative missingness across participating centres. Finally, a prioritised list of factors for each prognostic model (see Sample Size section below) was created by the senior oncologist team. A full description of the variable selection and correlation assessment process, including clinical rationale and literature sources, is provided in the published protocol¹¹. A data dictionary was created and shared between all centres for standardised data collection, formatting, and reporting. Additional detail on missing data handling and standardisation procedures across centres is provided in Supplementary Material: Supplementary Note 3.

PATIENT DATA COLLECTION & MISSING DATA

Patient data were identified and extracted from existing research and clinical databases at each participating site. To ensure good data quality, each institution spot checked all extracted data, ensuring adherence to the coding system specified in the data dictionary and identifying any outliers. The study protocol¹¹ pre-specified a framework on how to deal with missing data at individual centres, in multiple different scenarios, including use of data imputation. Only gross tumour volume (GTV, cm³) required use of imputation in two centre datasets; this was implemented based on a global median of means from all other centres.

SAMPLE SIZE

A prospective sample size calculation was performed using the framework devised by Riley et al.³⁹, and implemented using the “pmsampsize” package in R, to determine the minimum sample size required to fit a Cox proportional hazards model for each of the three outcomes. The number of prognostic factors included in the final models was based on the total number of patients available in the primary consortium cohort; with the specific set of factors based on the prioritised list developed from a systematic literature review and expert oncologist input. The detailed methodology used to calculate sample size and prioritised list of factors is provided in the study protocol¹¹.

DESCRIPTIVE DATA ANALYSIS

Descriptive data analysis included calculation of summary and survival statistics from each centre. Overall summary statistics for continuous variables were calculated using weighted means. Two-year, three-year, and five-year OS,

LRC, and FFDM were estimated by Kaplan-Meier methods. Potential follow-up times were based on the inverse Kaplan-Meier estimator⁵¹.

MODEL DEVELOPMENT & SPECIFICATION

Primary analysis consisted of development and internal validation (TRIPOD Type 2b⁵²) of federated multivariable Cox Proportional Hazards models³¹ across 14 centres; separately for OS, LRC, and FFDM. Secondly, external validation (TRIPOD Type 3) was performed using previously unseen data from two additional institutions (see below). Based on available patient numbers in the 14 primary cohort centres (n=1428), eight parameters could be included in the models. Predictors used were: age, biological sex, T stage, nodal involvement, primary GTV (using a \log_{10} transformation), prescribed primary tumour dose (in $\text{EQD2}_{\alpha/\beta=10\text{Gy}}$), and histology (as per the prioritisation lists in the study protocol). Biological sex was included as a predictor in all multivariable models and was recorded based on clinical documentation at each participating centre. Gender identity was not separately collected or verified. Sex-specific analyses beyond inclusion as a covariate were not pre-specified and were not performed.

A set of secondary models were additionally developed to explore the impact of different factor parametrisation (e.g. for age and GTV), different definitions (e.g. for tumour staging), and including secondary factors only available for a subset of patients (e.g. performance status). All analyses were run as multivariable models, with no data driven feature selection or model reduction. Patients with metastatic disease at baseline were excluded from all models for FFDM. An additional exploratory comparison between the federated models and a single-centre model (Centre 2) was performed to illustrate the added value of the federated approach.

FEDERATED MODEL LEARNING

Description of the FL architecture, as well as further technical details of the model training algorithm's aggregation process, convergence thresholds, and iteration scheme are provided in Supplementary Material: Supplementary Note 4. In brief, models were trained iteratively through exchange of local model coefficients and fit errors between each participating centre and a central server. At each iteration, the central server aggregated model coefficients computed locally in each centre and a single globally-convergent model was determined by minimising the total error⁵³⁻⁵⁵. The process was repeated until pre-specified convergence criteria were fulfilled. The source code for infrastructure implementation and the federated Cox regression algorithm utilised for all prognostic models are freely accessible on GitHub (see Supplementary Material: Supplementary Note 4).

MODEL REPORTING

For each model developed, the estimated factor effects were reported as Hazard Ratios (HR), along with 95% confidence intervals (CI). Factors were deemed prognostic if their 95% CI did not overlap with 1. For each of the three outcomes, reference outcome rates at 2 years, 3 years and 5 years were calculated. The reference outcome rate can be defined as the outcome rate when all model factors are set to their baseline value. To calculate reference

outcome rates, all categorical factors were set to 0, age at the start of radiotherapy was set to 35 years, prescribed dose to the primary tumour was set to 40 EQD2 $_{\alpha/\beta=10\text{Gy}}$, and \log_{10} of GTV was set to 0.02572 (GTV of 1.1 cm³). The combination of reference outcome rates and factor effect estimates (HR) enables prediction of outcomes for future patients, rendering the models useable for individual patient outcome prediction.

EVALUATION AND VISUALISATION OF MODEL PERFORMANCE

Model performance was assessed using Harrell's concordance index (c-index, providing a measure of model discrimination)⁵⁶ on a per-centre basis, and a global weighted average calculated. A closed-loop internal-external "leave-one-centre-out" cross-validation method⁵⁷ was applied to obtain the out-of-sample performance for the primary study cohort, providing an initial estimate of the over-optimism of the global model. The weighted mean of c-index values generated in all centres was reported. For model calibration, predicted three-year OS, LRC, and FFDM rates for each patient were calculated based on the primary models, and patients were allocated to 'high risk' or 'low risk' groups according to the global predicted mean for each outcome. For each centre, the actual (observed) three-year outcome rates for 'high risk' and 'low risk' patients were evaluated using Kaplan-Meier estimates and subsequently plotted against the mean predicted outcome rates for each group to indicate model calibration. Finally, summary plots were generated to illustrate the observed weighted mean outcome rates for predicted 'high risk' and 'low risk' groups across centres at two, three, and five years.

EXTERNAL VALIDATION

The models developed in the primary consortium cohort were further evaluated for their calibration and discrimination ability in two external validation cohorts. The external validation centres were selected based on cohort size, completeness of routinely collected variables, and geographical representation (one European and one Australian centre). Both were independent from the 14 training centres to ensure external validation across distinct populations. This process aimed to assess the reproducibility and generalisability of the models. Patient data collection and formatting followed the same study protocol as for the primary consortium cohort, and analyses retained the principles of no individual patient-level data sharing.

ROLE OF THE FUNDING SOURCE

None of the study sponsors 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.

DATA AVAILABILITY

Due to the nature of this research, the individual-level patient data analysed in this study cannot be made publicly available. These data include demographic variables, tumour characteristics, treatment details, and clinical outcomes collected at the participating centres. Even after de-identification, such data are considered potentially

re-identifiable and fall outside the scope of the ethical approvals obtained for this study, which do not permit redistribution of patient-level data.

The study protocol and statistical analysis plan, including a detailed data dictionary defining all variables, have been previously published (Theophanous et al., *Diagn Progn Res* 2022; <https://doi.org/10.1186/s41512-022-00128-8>). Aggregated results supporting the findings of this study are provided in the figures and tables of the main manuscript and Supplementary Materials. Source data underlying the reported figures are provided with this paper.

CODE AVAILABILITY

The custom code developed for this study is publicly available at <https://github.com/MaastrichtU-CDS/atomcat2> and is released under the GNU General Public License v3.0. The repository contains all scripts used for data processing, federated model execution, and aggregation of results. The code has been archived on Zenodo and assigned a persistent DOI⁵⁸.

The open-source Vantage6 (v2.3.4) federated learning infrastructure was used to carry out all federated learning analyses: <https://github.com/vantage6/vantage6>. Two Vantage6 federated learning algorithms were employed to run the federated analysis in this study. Federated Cox regression algorithm: <https://github.com/IKNL/vtg.coxph>, and federated validation algorithm: https://github.com/MaastrichtU-CDS/vtg.coxph_val. Descriptive data analysis was conducted in each participating centre individually using R/RStudio, and aggregation of results (including federated results) was carried out by the coordinating centre using R/RStudio.

REFERENCES

- 1 Botta L, Gatta G, Trama A, *et al.* Incidence and survival of rare cancers in the US and Europe. *Cancer Med* 2020; **9**: 5632–42.
- 2 Riley RD, Pate A, Dhiman P, Archer L, Martin GP, Collins GS. Clinical prediction models and the multiverse of madness. *BMC Med* 2023; **21**: 502.
- 3 Komatsubara KM, Carvajal RD. The promise and challenges of rare cancer research. *Lancet Oncol* 2016; **17**: 136–8.
- 4 Kent S, Kincaid L, Sharma M, *et al.* NICE real-world evidence framework. 2022; published online June. <https://www.nice.org.uk/corporate/ecd9/chapter/overview> (accessed Sept 21, 2023).
- 5 Booth CM, Karim S, Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. *Nat Rev Clin Oncol* 2019; **16**: 312–25.
- 6 Lawlor RT. The impact of GDPR on data sharing for European cancer research. *Lancet Oncol* 2023; **24**: 6–8.
- 7 Skripcak T, Belka C, Bosch W, *et al.* Creating a data exchange strategy for radiotherapy research: Towards federated databases and anonymised public datasets. *Radiother Oncol* 2014; **113**: 303–9.
- 8 Jochems A, Deist TM, van Soest J, *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.
- 9 Zerka F, Barakat S, Walsh S, *et al.* Systematic Review of Privacy-Preserving Distributed Machine Learning From Federated Databases in Health Care. *JCO Clin Cancer Inform* 2020; 184–200.
- 10 Kirienko M, Sollini M, Ninatti G, *et al.* Distributed learning: a reliable privacy-preserving strategy to change multicenter collaborations using AI. *Eur J Nucl Med Mol Imaging* 2021; **48**: 3791–804.
- 11 Theophanous S, Lønne P-I, Choudhury A, *et al.* Development and validation of prognostic models for anal cancer outcomes using distributed learning: protocol for the international multi-centre atomCAT2 study. *Diagn Progn Res* 2022; **6**: 14.
- 12 Theophanous S, Choudhury A, Lønne P-I, *et al.* Predicting outcomes in anal cancer patients using multi-centre data and distributed learning – A proof-of-concept study. *Radiother Oncol* 2021; **159**: 183–9.
- 13 Islami F, Ferlay J, Lortet-Tieulent J, Bray F, Jemal A. International trends in anal cancer incidence rates. *Int J Epidemiol* 2016; dyw276.
- 14 Salati SA. Anal Cancer : A Review. *Int J Health Sci* 2012; **6**: 206–30.
- 15 Kang Y-J, Smith M, Canfell K. Anal cancer in high-income countries: Increasing burden of disease. *PLoS One* 2018; **13**: e0205105.
- 16 Gilbert A, Rostoft S, Vendrely V, *et al.* Treatment of anal cancer in older patients and patients with frailty. *ESMO Gastrointest Oncol* 2025; 100155.
- 17 Nigro ND, Vaitkevicius VK, Considine B. Combined therapy for cancer of the anal canal: A preliminary report. *Dis Colon Rectum* 1974; **17**: 354–6.
- 18 Rao S, Guren MG, Khan K, *et al.* Anal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2021; **32**: 1087–100.
- 19 Glynn-Jones R, Nilsson PJ, Aschele C, *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.

- 20 Shakir R, Adams R, Cooper R, *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.
- 21 Kachnic LA, Winter KA, Myerson RJ, *et al.* Long-Term Outcomes of NRG Oncology/RTOG 0529: A Phase 2 Evaluation of Dose-Painted Intensity Modulated Radiation Therapy in Combination With 5-Fluorouracil and Mitomycin-C for the Reduction of Acute Morbidity in Anal Canal Cancer. *Int J Radiat Oncol* 2022; **112**: 146–57.
- 22 Hemingway H, Croft P, Perel P, *et al.* Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ* 2013; **346**: e5595–e5595.
- 23 Kent P, Cancelliere C, Boyle E, Cassidy JD, Kongsted A. A conceptual framework for prognostic research. *BMC Med Res Methodol* 2020; **20**: 172.
- 24 Theophanous S, Samuel R, Lilley J, *et al.* Prognostic factors for patients with anal cancer treated with conformal radiotherapy—a systematic review. *BMC Cancer* 2022; **22**: 607.
- 25 Muirhead R, Partridge M, Hawkins MA. A tumor control probability model for anal squamous cell carcinoma. *Radiother Oncol* 2015; **116**: 192–6.
- 26 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.
- 27 ISRCTN registry [Internet]. London: BMC. ISRCTN88455282, PLATO - Personalising anal cancer radiotherapy dose. 2016; published online Aug 3. DOI:10.1186/ISRCTN88455282.
- 28 Phung MT, Tin Tin S, Elwood JM. Prognostic models for breast cancer: a systematic review. *BMC Cancer* 2019; **19**: 230.
- 29 Beumer BR, Buettner S, Galjart B, *et al.* Systematic review and meta-analysis of validated prognostic models for resected hepatocellular carcinoma patients. *Eur J Surg Oncol* 2022; **48**: 492–9.
- 30 Frood R, Mercer J, Brown P, *et al.* Training and external validation of pre-treatment FDG PET-CT-based models for outcome prediction in anal squamous cell carcinoma. *Eur Radiol* 2023; **34**: 3194–204.
- 31 Lu C-L, Wang S, Ji Z, *et al.* WebDISCO: A web service for distributed cox model learning without patient-level data sharing. *J Am Med Inform Assoc* 2015; ocv083.
- 32 Cnossen M, Van Essen TA, Ceyisakar IE, *et al.* Adjusting for confounding by indication in observational studies: a case study in traumatic brain injury. *Clin Epidemiol* 2018; **Volume 10**: 841–52.
- 33 Mahajan R. Real world data: Additional source for making clinical decisions. *Int J Appl Basic Med Res* 2015; **5**: 82.
- 34 Azoulay L. Rationale, Strengths, and Limitations of Real-World Evidence in Oncology: A Canadian Review and Perspective. *The Oncologist* 2022; **27**: e731–8.
- 35 Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Med Res Methodol* 2022; **22**: 287.
- 36 Sun G-W, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol* 1996; **49**: 907–16.
- 37 Van Calster B, Van Smeden M, De Cock B, Steyerberg EW. Regression shrinkage methods for clinical prediction models do not guarantee improved performance: Simulation study. *Stat Methods Med Res* 2020; **29**: 3166–78.
- 38 Leeuwenberg AM, Van Smeden M, Langendijk JA, *et al.* Performance of binary prediction models in high-correlation low-dimensional settings: a comparison of methods. *Diagn Progn Res* 2022; **6**: 1.
- 39 Riley RD, Snell KI, Ensor J, *et al.* Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 2019; **38**: 1276–96.
- 40 Davey Smith G. Data dredging, bias, or confounding. *BMJ* 2002; **325**: 1437–8.

- 41 Steyerberg EW, Moons KGM, van der Windt DA, *et al.* Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med* 2013; **10**: e1001381.
- 42 Collins GS, Dhiman P, Ma J, *et al.* Evaluation of clinical prediction models (part 1): from development to external validation. *BMJ* 2024; e074819.
- 43 Chang C, Deng Y, Jiang X, Long Q. Multiple imputation for analysis of incomplete data in distributed health data networks. *Nat Commun* 2020; **11**: 5467.
- 44 Brink C, Hansen CR, Field M, *et al.* Distributed learning optimisation of Cox models can leak patient data: Risks and solutions. 2022; published online April 12. <http://arxiv.org/abs/2204.05856> (accessed May 27, 2022).
- 45 Reich C, Ostropolets A, Ryan P, *et al.* OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization. *J Am Med Inform Assoc* 2024; **31**: 583–90.
- 46 Mayo CS, Feng MU, Brock KK, *et al.* Operational Ontology for Oncology (O3): A Professional Society-Based, Multistakeholder, Consensus-Driven Informatics Standard Supporting Clinical and Research Use of Real-World Data From Patients Treated for Cancer. *Int J Radiat Oncol* 2023; **117**: 533–50.
- 47 Choudhury A, Volmer L, Martin F, *et al.* Advancing Privacy-Preserving Health Care Analytics and Implementation of the Personal Health Train: Federated Deep Learning Study. *JMIR AI* 2025; **4**: e60847.
- 48 Theophanous S. atomCAT2 - A multicentre study of overall survival, locoregional control and distant metastasis in anal cancer utilising distributed learning. Open Science Framework. <https://doi.org/10.17605/OSF.IO/J7AUH> (2021).
- 49 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015; **162**: 55.
- 50 Fish R, Sanders C, Adams R, *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.
- 51 Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. *Control Clin Trials* 1996; **17**: 343–6.
- 52 Moons KGM, Altman DG, Reitsma JB, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015; **162**: W1.
- 53 Deist TM, Jochems A, van Soest J, *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.
- 54 Moncada-Torres A, Martin F, Sieswerda M, Van Soest J, Geleijnse G. VANTAGE6: an open source priVAcY preserviNg federaTed leArninG infrastruCTurE for Secure Insight eXchange. *AMIA Annu Symp Proc AMIA Symp* 2020; **2020**: 870–7.
- 55 Deist TM, Dankers FJWM, Ojha P, *et al.* Distributed learning on 20 000+ lung cancer patients – The Personal Health Train. *Radiother Oncol* 2020; **144**: 189–200.
- 56 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; **30**: 1105–17.
- 57 Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. *J Clin Epidemiol* 2016; **69**: 245–7.
- 58 Theophanous S, Lønne P-I, Choudhury A, *et al.* An international multi-centre study to develop and validate federated learning-based prognostic models for anal cancer: atomCAT2 source code. Zenodo. <https://doi.org/10.5281/zenodo.18431456> (2026).

ACKNOWLEDGEMENTS

The authors would like to acknowledge Cancer Research UK funding (Grant Ref. C19942/A28832 and RRCOER-Jun24/100004) for the Leeds Radiotherapy Research Centre of Excellence (RadNet), which facilitated this work and provided academic salary support for Andrew Scarsbrook. Ane Appelt is supported by Yorkshire Cancer Research Academic Fellowship funding (grant L389AA). Leonard Wee is in receipt of funding from the Dutch Research Council (NWO) – BIONIC (grant no. 629.002.204), TRAIN (grant no. 629.002.212), CARRIER (grant no. 628.011.212), and STRaTegy (grant no. 14930), and from the Hanarth Foundation. Andrew Scarsbrook also acknowledges academic salary support from Leeds Hospitals Charity (Ref: 9R01/1403). 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.

AUTHOR CONTRIBUTIONS

AA and LW conceived the study idea. AA, LW, EM, ST and PIL established the atomCAT consortium and formulated the overarching research goals and aims. ST and AA coordinated the planning and execution of all research activity. AA and ST contributed to obtaining the required legal and ethics (HRA and REC) study approvals in the UK. AA, LW, EM, ST, PIL, MGG, MB, and AG designed the methodology. AC and LW contributed to the development of the federated learning infrastructure. AC, LW, ST and PIL contributed to the technical implementation of federated learning. ST, AA, AG, MGG, PIL, MB, CD, RR, RJ, LR, MAG, ES, VV, AAM, OP, RM, DM, RK, and MF were responsible for the local data collection at each participating centre. Members of the atomCAT consortium contributed to the local aspects of atomCAT2 in individual centres, including data collection, local dataset preparation, infrastructure setup, and local coordination. Please see the atomCAT Consortium section for the full list of atomCAT consortium members who contributed to this work. AG, MGG, MB, AS and DSM contributed to the discussion of the clinical relevance of the study findings. ST and AA prepared the initial draft manuscript. LW, EM, PIL, AG, MB, MGG, AD, AS and DSM critically reviewed the draft manuscript and provided feedback. ST and AA revised the draft manuscript. All co-authors critically reviewed the updated version of the manuscript and provided feedback. ST and AA prepared the final version. All authors read and approved the final version of the manuscript.

COMPETING INTERESTS

Leonard Wee receives consultancy fees when providing continuing professional development courses for radiotherapy physicists via Elekta AB (Stockholm, Sweden). Andre Dekker is a founder and employee of Medical Data Works B.V, which provides commercial support for Vantage6-based federated learning infrastructures. The remaining authors declare no competing interests.

TABLES

Table 1. Patient and treatment characteristics of the primary and validation cohorts.

Summary statistics for patient and treatment characteristics across the entire training cohort (14 centres) and validation cohort (2 centres). Means and medians are all weighted based on centre cohort size. SD: Standard deviation; GTV: Gross tumour volume; CI: Confidence interval; SCC: Squamous cell carcinoma; MMC: Mitomycin C; 5FU: 5-fluorouracil; Cap: Capecitabine; Cispl: Cisplatin. N/A: Information not available.

	Primary cohort (14 centres)	Validation cohort (2 centres)
Number of patients	1428	277
Treatment period	2004-2022	2008-2023
Biological sex		
Male	435 (30%)	114 (41%)
Female	993 (70%)	163 (59%)
Age at the start of radiotherapy (years)		
Mean	62.4	61.0
(SD, range)	(11.3, 29-94)	12.5 (21-92)
Age <50	197 (14%)	52 (19%)
Age 50-69	835 (58%)	149 (54%)
Age >= 70	396 (28%)	76 (27%)
T stage at diagnosis		
T1-2	815 (57%)	187 (68%)
T3-4	613 (43%)	90 (32%)
N stage at diagnosis		
N0	697 (49%)	155 (56%)
N+	731 (51%)	122 (44%)
M stage at diagnosis		
M0	1390 (97%)	272 (98%)
M1	38 (3%)	5 (2%)
Primary tumour GTV (cm ³)		
Mean	70.6	39.7
(SD, range)	(157.3, 0.6-974.4)	(46.4, 1.4-314.6)
Median	41.5	29.6
Histology		
SCC	1233 (86%)	250 (90%)
Basaloid SCC	195 (14%)	27 (10%)
Primary tumour dose (EQD2 $\alpha/\beta=10$)		
Mean	53.2	55.3
(SD, range)	(4.2, 40.7-66.0)	(3.1, 33.6-63.7)
Radiotherapy technique		
3D-CRT	117 (8%)	38 (14%)
IMRT / VMAT	1311 (92%)	239 (86%)
Chemotherapy regimen		
No chemotherapy	83 (6%)	52 (19%)
MMC and 5FU	927 (65%)	194 (70%)
MMC and Cap	365 (26%)	22 (8%)
Cispl and 5FU	12 (<1%)	3 (1%)
Cispl and Cap	3 (<1%)	0 (0%)
Other	38 (3%)	6 (2%)

ARTICLE IN PRESS

Table 2. Outcome summary statistics for the primary and validation cohorts.

Summary of survival statistics, including overall number of events, as well as estimated overall survival (OS), locoregional control (LRC), and freedom from distant metastases (FFDM) at two, three, and five years, for the training cohort (14 centres) and validation cohort (2 centres).

	Primary cohort (14 centres)	Validation cohort (2 centres)
Number of patients	1428	277
Number of events		
Deaths	286 (20%)	44 (16%)
Locoregional failures	214 (15%)	32 (12%)
Distant metastases	163 (11%)	30 (11%)
Overall survival		
2-year OS	88%	89%
3-year OS	83%	87%
5-year OS	76%	79%
Locoregional control		
2-year LRC	86%	90%
3-year LRC	83%	89%
5-year LRC	81%	84%
Freedom from distant metastases		
2-year FFDM	89%	89%
3-year FFDM	87%	89%
5-year FFDM	86%	88%

Table 3. Multivariable Cox proportional hazards model results.

Results from multivariable Cox proportional hazards models for overall survival (OS), locoregional control (LRC), and freedom from distant metastases (FFDM), trained on the full primary cohort (14 centres). Hazard ratios (HRs) and 95% confidence intervals (CI) are shown for each included variable. All models were adjusted for the full set of covariates listed (apart from radiotherapy technique, which was not included in the FFDM model).

	Overall survival	Locoregional control	Freedom from distant metastases
Mean global model c-index	0.68	0.71	0.69
Mean leave-one-centre-out validation c-index	0.67	0.69	0.66
	Hazard ratio (95% CI)		
Nodal involvement (N+ relative to N0)	1.45 (1.11-1.89)	1.24 (0.92-1.68)	2.09 (1.42-3.08)
T stage (T3-4 relative to T1-2)	1.42 (1.07-1.89)	1.46 (1.05-2.03)	1.18 (0.80-1.74)
Biological sex (Female relative to male)	0.65 (0.51-0.83)	0.56 (0.43-0.73)	0.82 (0.58-1.16)
Age at start of radiotherapy (per 10 years)	1.20 (1.07-1.34)	1.08 (0.96-1.22)	1.00 (0.86-1.16)
Gross tumour volume (cm ³)	2.02 (1.47-2.76)	2.47 (1.73-3.53)	2.14 (1.40-3.27)
Prescribed dose to primary tumour (log ₁₀ EQD2, per 10 Gy)	0.96 (0.71-1.29)	1.17 (0.82-1.67)	1.21 (0.79-1.86)
Histology (Basaloid SCC relative to SCC)	0.88 (0.61-1.28)	0.64 (0.39-1.06)	1.04 (0.64-1.69)
Radiotherapy technique (IMRT/VMAT relative to 3D-CRT)	0.96 (0.67-1.39)	1.55 (0.91-2.64)	N/A
Chemotherapy regimen (all relative to no chemotherapy)			
Mitomycin-based	0.35 (0.23-0.53)	0.67 (0.35-1.25)	0.59 (0.28-1.23)
Cisplatin-based	0.32 (0.11-0.92)	0.72 (0.22-2.30)	0.80 (0.21-3.09)
Other chemotherapy	0.81 (0.42-1.56)	0.83 (0.30-2.27)	0.94 (0.31-2.92)

FIGURE LEGENDS

Figure 1. Outcome rates across participating centres.

(a) Overall survival, (b) locoregional control and (c) freedom from distant metastases rates, calculated using Kaplan-Meier methods. Patients at risk at two years, three years, and five years are specified below. Shaded areas represent the 95% confidence interval bands from 12 participating centres in the primary cohort, while the solid line represents the weighted mean across all 12 centres. Primary cohort centres 6 and 14 were not included in this figure due to lack of 5-year follow-up data. Source data are provided as a Source Data file.

Figure 2. Predicted failure risk distribution and impact of tumour volume.

Failure risk distribution for locoregional control and distant metastases for individual patients in a single centre cohort (Centre 2, n=210). In (a), individual patients are stratified into locoregional failure and distant metastasis risk groups according to their predicted risk score; (b) highlights the impact of tumour volume on failure risk.

Figure 3. Calibration of federated prognostic models for 3-year outcomes.

Calibration curves for 3-year (a) overall survival, (b) locoregional control and (c) freedom from distant metastases rates. Size of circles corresponds to individual cohort size. The black circles represent the calibration results from the two external validation cohorts. Source data are provided as a Source Data file.

Figure 4. atomCAT2 study design workflow.

Workflow of the atomCAT2 study, showing progression from a pilot network of three centres (atomCAT1) to the establishment of the 16 centre atomCAT network for federated model development and external validation.

ATOMCAT CONSORTIUM

The following individuals are members of the atomCAT consortium and contributed to this work:

Richard Adams²⁶, Krystyna Adamska²⁴, Muhammad Amin²⁹, Ane Appelt^{1,2}, Maike Berbee⁴, Nikola Dino Capocchiano¹¹, Philip Chlap^{7,8,9}, Ananya Choudhury⁴, Peter Colley²⁵, Andrea Damiani¹¹, Viola De Luca¹¹, Charlotte Deijen^{5,6}, Andre Dekker⁴, Antri Demetriou²⁷, Shrikant Deshpande^{7,9}, Michael J. Eble¹⁹, Anthony Espinoza^{7,9}, Emmanouil Fokas^{16,17,18,30}, Matthew Field^{7,8,9,10}, Maria Antonietta Gambacorta¹¹, Loukia Georgiou²⁷, Alexandra Gilbert^{1,2}, Marianne Grønlie Guren^{12,13}, Ann Henry², Andrew Hoole¹⁴, Lois C. Holloway^{8,9,10,31,32}, Rashmi Jadon¹⁴, Thomas Jansen¹⁹, Tomas Janssen⁵, Rohit Kochhar¹⁵, Alexandros Kritikopoulos²⁷, Joanna Y.G. Lau¹⁴, Mark T. Lee^{7,9}, John Lilley¹, Gisela Lima²⁵, Per-Ivar Lønne³, Patricia Lopes²³, Adam Loveday¹⁴, Eirik Malinen²⁸, Stefania Manfrida¹¹, Jenny Marsden³³, Daniel Martin^{16,17,18}, Carlotta Masciocchi¹¹, Joseph Mercer¹⁵, Ahmed Allam Mohamed^{19,20}, Rebecca Muirhead^{21,22}, Elisavet Papageorgiou²⁷, Oriol Parés²³, Gareth Price¹⁵, Thomas Rackley²⁹, Łukasz Raszewski²⁴, Claus Michael Rödel^{16,17,18}, Rajarshi Roy²⁵, Mariachiara Savino¹¹, Andrew Scarsbrook^{1,2}, Athina Sdrolia²⁵, David Sebag-Montefiore², Emiliano Spezi²⁶, Joep Stroom²³, Ioannis Stylianou²⁷, Stelios Theophanous^{1,2}, David Thwaites^{2,32}, Maciej Trojanowski²⁴, Vincenzo Valentini¹¹, Rens van Haveren⁵, Baukelien van Triest⁵, Vassilios Vassiliou²⁷, Amy Walker^{7,8,9,31}, Leonard Wee⁴

²⁹ Velindre University NHS Trust, Cardiff, UK

³⁰ University Hospital Cologne, Cologne, Germany

³¹ Centre for Medical Radiation Physics, University of Wollongong, New South Wales, Australia

³² Institute of Medical Physics, The University of Sydney, New South Wales, Australia

³³ University Hospitals of Leicester, Leicester, UK

Editorial Summary

Precision oncology requires access to high quality data, as well as the application of current technologies. Here, the authors use federated learning to predict clinical outcomes in anal cancer patients across 16 centres as part of the atomCAT consortium.

Peer review information: *Nature Communications* thanks Momina Shaheen, Du Cai, Soledad Iseas and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

ARTICLE IN PRESS







