frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 10 June 2024
DOI 10.3389/fimmu.2024.1383644

Check for updates

OPEN ACCESS

EDITED BY
Mojtaba Amani,
Ardabil University of Medical Sciences, Iran

REVIEWED BY
Patrick C. Gedeon,
Brigham and Women's Hospital and Harvard Medical School, United States
Harrison Brown,
Expression Therapeutics, United States
Soheila Shabani-mashcool,
University of Tehran, Iran

*CORRESPONDENCE
Kedar A. Patwardhan
kedar.patwardhan@astrazeneca.com

†PRESENT ADDRESSES
Nikolaos Nikolaou,
Department of Physics and Astronomy, University College London, London, United Kingdom
Ignacio Gonzalez-García,
Cognigen Corporation, a Division of Simulations Plus, Inc., Buffalo, NY, United States
Joachim Reischl,
Diagnostics and Precision Medicine, Danaher Corp, Gothenburg, Sweden

RECEIVED 07 February 2024
ACCEPTED 21 May 2024
PUBLISHED 10 June 2024

CITATION
Patwardhan KA, RaviPrakash H, Nikolaou N, Gonzalez-García I, Salazar JD, Metcalfe P and Reischl J (2024) Towards a survival risk prediction model for metastatic NSCLC patients on durvalumab using whole-lung CT radiomics.
*Front. Immunol.* 15:1383644.
doi: 10.3389/fimmu.2024.1383644

COPYRIGHT
© 2024 Patwardhan, RaviPrakash, Nikolaou, Gonzalez-García, Salazar, Metcalfe and Reischl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Towards a survival risk prediction model for metastatic NSCLC patients on durvalumab using whole-lung CT radiomics

Kedar A. Patwardhan[1*], Harish RaviPrakash[1], Nikolaos Nikolaou[2†], Ignacio Gonzalez-García[3†], José Domingo Salazar[2], Paul Metcalfe[2] and Joachim Reischl[4†]

[1]Oncology Data Science, AstraZeneca, Waltham, MA, United States, [2]Oncology Data Science, AstraZeneca, Cambridge, United Kingdom, [3]Clinical Pharmacology and Quantitative Pharmacology, Clinical Pharmacology and Safety Sciences, AstraZeneca, Cambridge, United Kingdom, [4]Diagnostic Science, Precision Medicine, AstraZeneca, Gothenburg, Sweden

**Background:** Existing criteria for predicting patient survival from immunotherapy are primarily centered on the PD-L1 status of patients. We tested the hypothesis that noninvasively captured baseline whole-lung radiomics features from CT images, baseline clinical parameters, combined with advanced machine learning approaches, can help to build models of patient survival that compare favorably with PD-L1 status for predicting 'less-than-median-survival risk' in the metastatic NSCLC setting for patients on durvalumab. With a total of 1062 patients, inclusive of model training and validation, this is the largest such study yet.

**Methods:** To ensure a sufficient sample size, we combined data from treatment arms of three metastatic NSCLC studies. About 80% of this data was used for model training, and the remainder was held-out for validation. We first trained two independent models; Model-C trained to predict survival using clinical data; and Model-R trained to predict survival using whole-lung radiomics features. Finally, we created Model-C+R which leveraged both clinical and radiomics features.

**Results:** The classification accuracy (for median survival) of Model-C, Model-R, and Model-C+R was 63%, 55%, and 68% respectively. Sensitivity analysis of survival prediction across different training and validation cohorts showed concordance indices ([95 percentile]) of 0.64 ([0.63, 0.65]), 0.60 ([0.59, 0.60]), and 0.66 ([0.65,0.67]), respectively. We additionally evaluated generalization of these models on a comparable cohort of 144 patients from an independent study, demonstrating classification accuracies of 65%, 62%, and 72% respectively.

**Conclusion:** Machine Learning models combining baseline whole-lung CT radiomic and clinical features may be a useful tool for patient selection in immunotherapy. Further validation through prospective studies is needed.

KEYWORDS

NSCLC, radiomics, multi-modal, immunotherapy, survival risk

# Introduction

Lung cancer is the leading cause of cancer death and makes up around 25% of cancer deaths worldwide (1). The 5-year survival for patients with non-small cell lung cancer (NSCLC) is approximately 25% but drops to almost 7% in the case of metastatic NSCLC (mNSCLC). Recent advances have led to development of targeted immune checkpoint inhibitors such as durvalumab, targeting programmed death ligand 1 (PD-L1), which has shown durable clinical benefit to patients with locally advanced NSCLC (2). In this case levels of PD-L1 expression, derived from evaluation of biopsy tissue, have been shown to impact overall survival (OS) following durvalumab treatment (3). Lung biopsies are invasive, often challenging due to anatomical access, do not account for inter/intra lesion heterogeneity, and are unable to provide evaluations for all lesions. Radiological imaging e.g., computed tomography (CT), provide non-invasive characterization of all lesions as well as the surrounding anatomical region. This information is available at baseline as well as longitudinal time-points, whereas longitudinal biopsies are impractical. This has been a key motivation for our exploration of non-invasive imaging based patient stratification approaches in the mNSCLC setting.

## Literature review

Given that OS is typically a primary endpoint for clinical trials to demonstrate efficacy, we searched the PubMed database to identify relevant prior work in the field of mNSCLC where baseline CT images (radiomics) were used for OS prediction in patients on immunotherapy (IO) treatment (4). The most relevant publications as of December 1, 2022, are highlighted in Table 1[1].

While such prior work has demonstrated the benefit of CT radiomics and clinical models, unlike our work, all the models reported in literature have been trained on relatively small datasets of <200 patients, which can limit the variability encountered and thereby model generalization in unseen datasets. More importantly, all prior work requires accurate delineation of the target tumor. This typically requires a slice-by-slice annotation by an experienced radiologist which is a time-intensive task. Such a primary-lesion (and/or tumor macro-environment) focused approach does not take into consideration the contribution of other lesions, anatomical structures (e.g., lymph nodes), as well as the characteristics of healthy lung tissue that may be relevant to the OS of the patients (e.g., overall health status, tolerance to IO, etc).

## Summary of contributions

In this study we aim to establish the value of non-invasive CT radiomic features in patient selection for durvalumab monotherapy or in combination in an mNSCLC setting. This is the largest study so far looking at baseline CT radiomic features for IO survival prediction in mNSCLC (N > 1000 including training, testing, and independent cohort evaluation). Apart from the sample size, since our data covers multiple sites and clinical trials, there is reasonable diversity in the imaging data, as shown in Table 2, to ensure that models developed are robust to such variations. Additionally, compared to existing literature, we do not use any lesion annotation for extraction of radiomic features while demonstrating favorable performance compared to prior art for survival-based patient risk prediction. The following sections will describe in detail the data, methods, and results, followed by a brief discussion towards next steps in our research.

# Materials and methods

## Data preparation

We performed a retrospective analysis using data from three phase III clinical studies (ARCTIC [NCT02352948] (10, 11), MYSTIC [NCT02125461] (3, 12), and NEPTUNE [NCT04571658] (13, 14)) investigating durvalumab alone or in combination (durvalumab with tremelimumab) in patients with a diagnosis of mNSCLC. The analysis presented in the manuscript uses data from patients that provided informed consent including for secondary reuse. All studies were conducted in compliance with the Declaration of Helsinki and the United States (US) Food and Drug Administration (FDA) Guidelines for Good Clinical Practice.

The dataset of 1320 patients was assessed for eligibility (see Figure 1A). We manually reviewed the images as a quality check for image resolution and motion artifacts and excluded patients with in-plane resolution of over 1 mm or slice thickness (out-of-plane resolution) over 10 mm (n = 107). Finally, patients with missing clinical parameters were also excluded (n = 295) resulting in a total of 918 patients as shown in Figure 1B and Supplementary Figure 1. A complete list of the included clinical parameters can be found in the Supplementary Material (see Supplementary Table 3).

Prior to splitting the analysis data into training and test cohorts, Kaplan-Meier (KM) (15) survival curve of the treatment arms of the three studies was generated (see Supplementary Figure 2A) and no significant difference was observed in the survival times when analyzed as a combined cohort (though it must be noted that there are patient subpopulations where there are differences in survival) (3, 16). Thus, data from the three clinical trials were combined to create a large analysis cohort (referred to as the primary study cohort) that comprised of 918 patients. PD-L1 status for each patient was determined at 25% cut-off[2]. Patients with over 25% PD-L1 expression were labeled as PD-L1-positive and PD-L1-negative otherwise. KM curve for PD-L1 status showed a significant difference between PD-L1- positive and negative patients (Supplementary Figure 2B). The patient cohorts were

1 For additional details regarding the selection of these specific references, please refer to the exact search link[4] and Supplementary Material - Literature Review.

2 The data for 25% cutoff is consistently available across all the studies used for our analysis. KM curves for different sub-cohorts where we may have 50% cutoff information available are provided in Supplementary Figure 3.

10.3389/fimmu.2024.1383644

TABLE 1 Overview of Relevant Prior Work Towards OS Prediction Using Baseline CT Images in Patients on IO, in a mNSCLC Setting.

| Reference | Total no. of Patients | Accuracy (%) in Test Cohort (mutually exclusive training and test sets sampled from same cohort) | AUC in Test Cohort (mutually exclusive training and test sets sampled from same cohort) | Concordance Index (robustness to training/test splits) | Generalization Accuracy (%) (independent cohort) |
|---|---|---|---|---|---|
| Jazieh et al. (5) | 133 | N/A ($N_{test}$ = 69) | N/A | 0.69 (N/A) | N/A |
| Gong et al. (6) | 224 | 50.79 ($N_{test}$ = 63) | 0.52 (0.39–0.65) | 0.51 (0.43–0.6) | 52.94 ($N_{test}$ = 68) |
| He et al. (7) | 236 | 83.3 ($N_{test}$ = 48) | N/A | 0.75 (0.59–0.90) | N/A |
| Liu et al. (8) | 46 | N/A ($N_{test}$ = 1) | 0.64 (0.48–0.79) | 0.86 (0.74–0.97) | N/A |
| Trebeschi et al. (9) | 123 | N/A ($N_{test}$ = 42) | 0.76 (N/A) | N/A | N/A |
| Work Presented* | 917 | 66.8 ($N_{test}$ = 199) | 0.71 (0.65–0.78) | 0.66 (0.65–0.67) | 73.5 ($N_{test}$ = 144) |

Values not reported are indicated with N/A. $N_{test}$ indicates the number of "held-out" samples used to test a model (these samples are not seen by the model during the training phase).
*Results reported are for Model-C+R. Evaluation metrics can be found in Supplementary Table 5.
AUC, area under the curve; CT, computed tomography; IO, immunotherapy; OS, overall survival; mNSCLC, metastatic non-small cell lung cancer; NA, not applicable.

TABLE 2 Patient Cohorts and Demographics.

| Patient Cohorts | | | | | | |
|---|---|---|---|---|---|---|
| Study Name | Training Set | | Test Set | | Independent Study | |
| | PD-L1 Status | | PD-L1 Status | | PD-L1 Status | |
| | Neg. | Pos. | Neg. | Pos. | Neg. | Pos. |
| ARCTIC | 97 | 26 | 29 | 7 | – | – |
| MYSTIC | 191 | 158 | 51 | 43 | – | – |
| NEPTUNE | 147 | 99 | 42 | 28 | – | – |
| CP1108 | – | – | – | – | 66 | 78 |
| All data by PD-L1 status | 435 | 283 | 122 | 78 | 66 | 78 |
| All data combined | 718 | | 200 | | 144 | |

| Patient Demographics and Imaging Characteristics | | | |
|---|---|---|---|
| Covariate | Training Cohort | Test Cohort | Independent Study |
| Sex = Female | 31% | 27% | 43% |
| Age (years): Mean (± SD) | 62.99 (± 9.71) | 62.9 (± 9.99) | 64.03 (± 11.57) |
| Overall survival: Median (IQR) | 317 (126–699) | 385 (149–836) | 282 (107–856) |
| Baseline tumor size: Median (IQR) | 82 (49–117) | 83 (50–121) | 77 (54–114) |
| In-plane resolution (mm): Mean (± SD) | 0.78 (± 0.10) | 0.78 (± 0.10) | 0.80 (± 0.10) |
| Slice thickness (mm): Mean (± SD) | 4.3 (± 1.27) | 4.4 (± 1.24) | 4.1 (± 1.13) |
| CT scanner manufacturers (Top 4) | SIEMENS (33%) GE (32%) TOSHIBA (16%) PHILIPS (16%) | GE (33%) SIEMENS (31%) PHILIPS (20%) TOSHIBA (15%) | SIEMENS (46%) GE (43%) PHILIPS (4%) TOSHIBA (4%) |

(Continued)

TABLE 2 Continued

Based on data available from all the studies, we used a cutoff of 25% to indicate PD-L1 positive or negative to ensure a consistent cutoff across studies.
IQR, interquartile range; Neg., negative; PD-L1, programmed death ligand 1; Pos., positive; SD, standard deviation.

thus generated via stratifying based on PD-L1 and study, and were split 80:20 into training and test datasets.

In order to use a completely blinded cohort (previously unseen by the model) for validation, a matching cohort (with respect to drug regimen and indication) was selected from a drug-escalation study CP-1108 (NCT01693562) (17) (n = 224). Patients with disease stage III and below at the time of entry to the clinical trial were excluded from the analysis. Patients were further excluded based on the criteria used in the primary cohort. The final independent validation cohort consisted of 144 patients. Number of patients in each study and the PD-L1 distribution is shown in Figure 1A.

Table 2 shows the stratification of patients in the different cohorts. The patient demographics along with the baseline tumor size are included in Table 2 for each cohort. With a median survival of 332 days (see Figure 1C) in the primary cohort, 1-year survival was selected as the outcome for modeling.

## Feature extraction

Pretreatment clinical-radiological information collected include patient demographics, routine blood work, and high-resolution CT imaging. The process of extracting features from both these modalities has been described below.

### Clinical features

The clinical features available included demographic and blood work values. Categorical demographic information such as gender, smoking history, Eastern Cooperative Oncology Group (ECOG) performance status and treatment combination were encoded into binary variables. Other numerical features collected were body mass

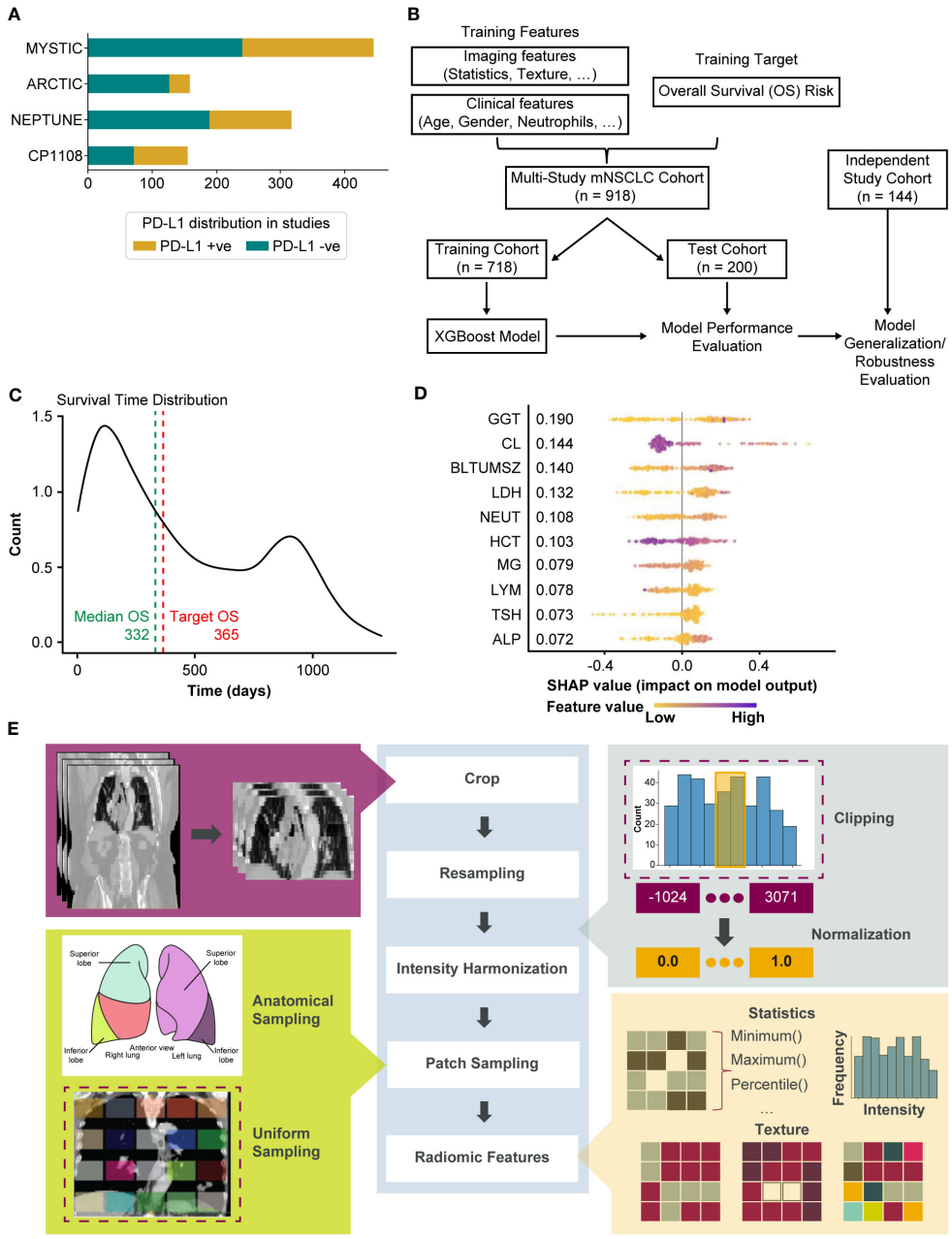


FIGURE 1
Overview of the modeling for the generation of whole-lung radiomics for survival risk prediction. **(A)** Bar chart showing the number of patients in each of the clinical trials. PD-L1 status was determined at 25% cutoff. **(B)** Overview of the training and testing procedure using XGBoost model. The MYSTIC, ARCTIC, and NEPTUNE datasets were stratified into training (80%) and test (20%). The model was tuned by further dividing the training set into training (80%) and validation (20%). The resulting trained model was evaluated on the test set. Robustness is evaluated by testing on an independent study cohort. **(C)** Distribution of survival times in the study cohort. **(D)** Top-10 features from Model-C as computed by SHAP. **(E)** Annotation-free radiomics feature extraction pipeline. Dashed boxes indicate currently implemented methods in the pipeline where other options exist (e.g., sampling from pre-specified sites from an anatomical atlas as shown). ALP, alkaline phosphatase; BLTUMSZ, baseline tumor size; CL, chlorine; GGT, gamma glutamyl transferase; HCT, hematocrit; LDH, lactate dehydrogenase; MG, magnesium; NEUT, neutrophils; PD-L1, programmed death ligand 1; SHAP, SHapley Additive exPlanations; TSH, thyroid stimulating hormone.

index (BMI), blood work values such count of neutrophils, count of lymphocytes, amount of albumin, gamma-glutamyl transferase (GGT), alkaline phosphatase (ALP), and other available baseline models. Features that were highly correlated with each other were dropped from the analysis. The complete list of included clinical parameters and their units can be found in Supplementary Material (see Supplementary Table 3). Figure 1D shows the top features selected in the clinical model (Model-C).

### Whole-lung radiomic features

CT images of the lung region were captured for each patient at baseline using whole-body, thoracic, or chest and abdominal CT imaging. Our in-house annotation tool was used to delineate the lung extents along the coronal and axial views in the abdominal and whole-body CT images, and the images were cropped to include only this lung region. Since the images were captured from multiple sites using several different scanners, intensity harmonization was performed by clipping CT gray values in the range of [-120–300] Hounsfield Units to focus on lung tissues (18). This was then followed by sampling patches from the image. Different sampling strategies can be employed such as anatomically informed sampling (based on lobes, nodes, etc.), uniform sampling, etc. We uniformly sampled the entire lung region and extracted n=80 3D patches of size 5cm$^3$, as illustrated in Figure 1E. Finally, radiomic features such as statistical radiomic features and texture features including gray-level cooccurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level difference matrix (GLDM), gray-level size zone (GLSZM), and neighborhood gray tone difference matrix (NGTDM) were extracted using PyRadiomics (19) with the 3D patches serving as the region of interest. The above proposed pipeline is novel and is shown in Figure 1E.

## Machine learning modeling

The proposed whole-lung radiomics approach yielded radiomic features from a fixed number of patches in a given CT image. These features were then aggregated using simple summary statistics across the patches, namely mean, median, minimum, maximum, sum, mean absolute deviation, skew, and kurtosis, to provide a holistic view of the lung as well as to reduce the dimensionality problem that would otherwise arise. Feature selection was performed on the aggregated features on the training set using correlation and mutual information approaches with bootstrapping to further reduce the feature dimension. Specifically, from each pair of features with absolute Spearman's correlation >0.9 a single feature is retained. Features with the highest mutual information with the target (OS) were selected. The process was repeated five times on separate bootstraps of the training set and only features selected at least twice made the final feature set. This was performed independently for the clinical and radiomic features. The selected features were then used as input to train an XGBoost (20) model. The parameters of the model were tuned using the training data and the threshold for prediction was determined using the receiver operating characteristics curve on the validation data. The final model was then generated by finetuning on the validation data[3]. This final model was used to evaluate the performance on the test set as well as the generalizability to an independent study cohort. SHapley Additive exPlanations (SHAP) (21) were used to identify important features contributing to the model performance as shown in Figure 1D. The overall analysis pipeline is shown in Supplementary Figure 4. Python (22) programming language was used for the feature extraction and selection pipelines. PyRadiomics (19) was used for feature extraction. The modeling and evaluation were implemented in R programming language (23). The final model parameters are shown in Supplementary Table 4. Cox Proportional Hazards (24) (Cox PH) was used to compute the hazard ratios in the evaluation and the KM method was used for survival analysis.

# Results

We evaluated individual models using clinical features (Model-C), radiomic features (Model-R), and the combined clinical-radiomic model (Model-C+R). The following sections describe the survival risk stratification experiments and robustness experiments.

## Survival-based risk stratification

Following the approach detailed in the Machine Learning modeling, tuned XGBoost models were evaluated on the test set. First Model-C, trained using the selected clinical parameters was evaluated on the held-out test set and an accuracy of 63% was achieved. Survival curves were generated using KM analysis and the patient stratification was quantified via the $p$-value. The clinical model achieved a significant $p$-value (0.00011) suggesting that it is capable of stratifying high- and low-risk patients on IO.

Next, the radiomics features only model (Model-R) was evaluated on the held-out test set and an accuracy of 54% was achieved which was much lower than that of the clinical model. The KM analysis similarly shows a non-significant separation with a $p$-value of 0.091. However, the separation in median survival times between low-risk and high-risk patients is 243 days, similar to that found in the PD-L1 status KM curves as shown in Figure 2B.

We then evaluated the potential complementary nature of the clinical and radiomic features using Model-C+R. The combined model yielded an accuracy of 68% on the held-out test set, a 5% gain over Model-C. The SHAP feature importance plot for the model was generated and the importance and contribution of radiomic features alongside clinical ones was noted in Supplementary Figure 6 (top panel). The KM analysis yielded a $p$-value <0.0001 suggesting a significant separation of high- and low-risk patients. Furthermore, separation in median survival times was largest using Model-C+R, at 486 days.

3 Further details on the training of the XGBoost model can be found in Supplementary Material – *Analysis Pipeline*.

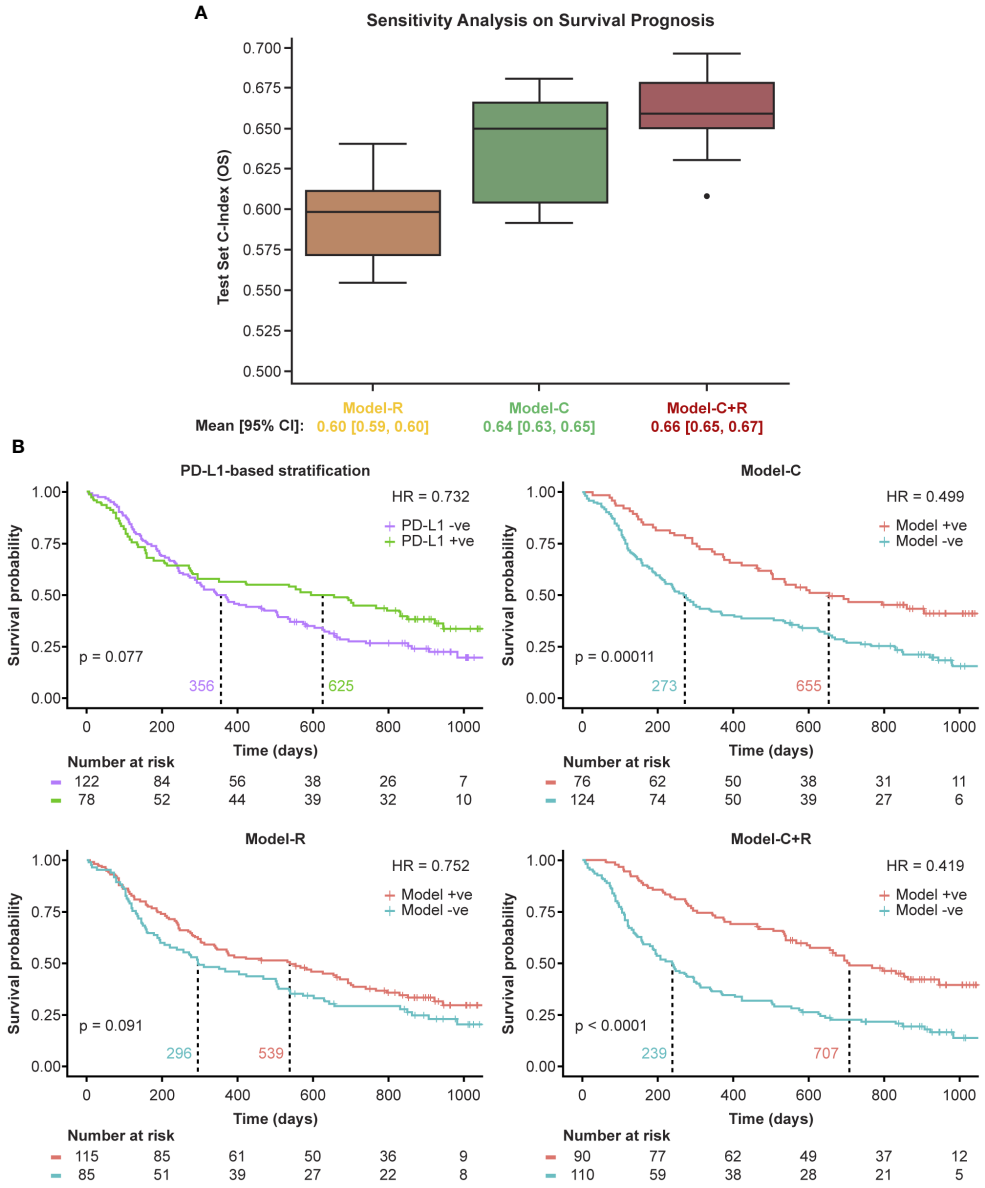


FIGURE 2
Model performance on the primary study cohort. Model+ve indicates model prediction of OS ≥365 and Model-ve indicates model prediction of OS <365. **(A)** Sensitivity analysis using 10 random train/test splits and **(B)** Kaplan-Meier curves of the model performance on the held-out test set. +ve, positive; -ve, negative; HR, hazard ratio; PD-L1, programmed death ligand 1.

The large separation in median survival times suggests the benefit of combining radiomic and clinical models. The KM curves for the models are shown in Figure 2B.

## Model robustness

### Sensitivity to data splits

Machine learning algorithms are often affected by potential unseen biases in the selection of the training/test splits. To test the sensitivity of the selected features to the selection of training/test splits, 10 random training/test splits were tested on the OS prediction task. The results, shown in Figure 2A, validate the stability of the selected features to training/test splits with consistent concordance indices and a tight 95% confidence interval across all models.

### Generalization to independent study cohort

The independent study cohort was used to validate the models trained on the training set. The survival distribution of the independent study cohort, shown in Figure 3A, had a median survival of 282 days, 35 days lower than that of the training set (primary study cohort). Previously selected features were extracted from the independent study cohort and tested using the three models: Model-C, Model-R, and Model C+R trained on the primary study cohort training data. Similar to the performance observed on the test set (primary study cohort),

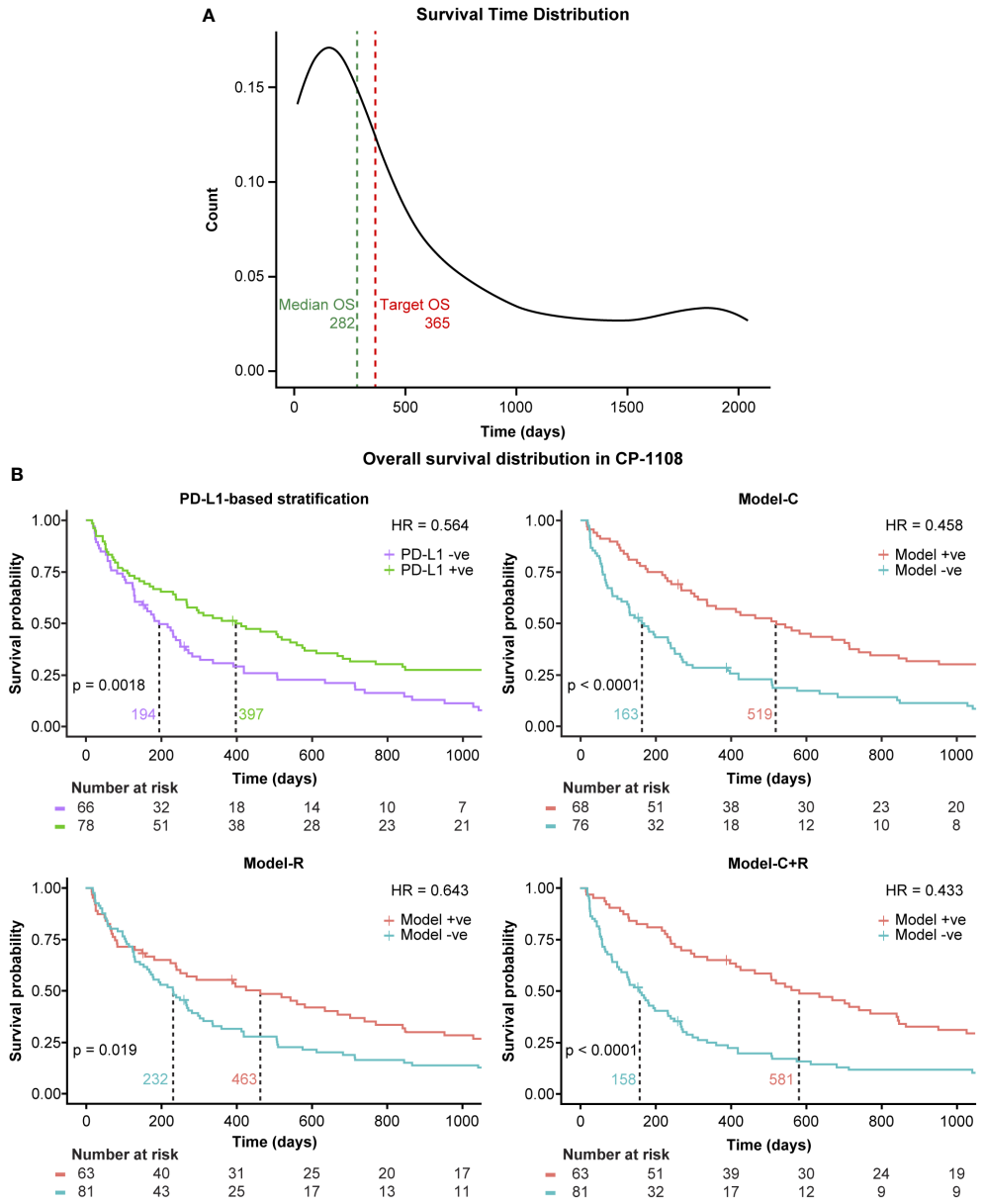


FIGURE 3
Generalization to an independent study cohort. Model+ve indicates model prediction of OS ≥365 and Model-ve indicates model prediction of OS <365. **(A)** Survival distribution in the independent study cohort and **(B)** Kaplan-Meier curves of the model performance on the independent study cohort. +ve, positive; -ve, negative; HR, hazard ratio; PD-L1, programmed death ligand 1.

Model-C outperformed Model-R, and Model-C+R outperformed both the clinical and radiomic models, further confirming the complementary nature of the clinical and radiomic features. The model achieved an accuracy of ~72% and stratified the patients into high- and low-risk groups with a difference in median survival of 423 days as shown in Figure 3B. The comparison of the performance of the different models across both test cohort and independent study cohorts in terms of sensitivity, specificity, precision and accuracy are reported in Supplementary Table 6. SHAP feature importance plot of Model-C +R on the independent study cohort revealed similar top features with the same radiomic features contributing to the prediction (Supplementary Figure 6, bottom panel). Further details on the SHAP analyses as well as the building of more traditional Cox models using the top predictors of overall survival are included in the Supplementary Material – SHAP Analysis.

## Discussion

The emergence of PD-L1 inhibitors as first line treatment for non-resectable advanced NSCLC[2] has provided improved quality of life and an increased survival time. In this paper we have presented a novel approach to the prediction of patients' survival that integrates clinical and radiomic features and requires no previous tumor annotation. Our approach is competitive in terms of concordance index (when predicting overall survival) and

accuracy (when predicting risk of less than median survival). Finally, our approach compares very favorably with predictions based on PD-L1 status and has been validated against a multi-trial test set as well as an independent additional trial.

The potential for such an approach as a diagnostic tool is explored by including the standard of care (SoC) arms from the clinical trials of MYSTIC and NEPTUNE[4]. The KM curves from the model predictions are illustrated in Figure 4. Model-C+R offers the potential to select patients who would benefit from IO therapy. As shown in Figure 4A, Model-C+R identifies patients who would not benefit from IO and have similar survival curves to the SoC arms. It also identifies patients who would benefit from IO therapy and when combined with PD-L1 expression helps identify patients with the highest odds of survival (Figure 4B). The Model-C+R identified patients that also satisfy a higher (50%) cutoff for PD-L1 show the best improvement in median OS as shown in Figure 4D. Thus, the proposed non-invasive approach opens new avenues for patient selection that are complementary to PD-L1 in the mNSCLC setting.

4 Patient Demographics from the SoC arms can be found in Supplementary Material – Data Preparation.

## Challenges and limitations

In order to build clinician confidence around using such diagnostic tools, it is important to build an intuitive understanding of the interpretability of these features. The radiomics features identified by our models to be most informative are shown in the Supplementary Material – SHAP Analysis. Our models suggest that low texture heterogeneity corresponds to better OS. This is similar to results reported in literature where texture features representing tumor heterogeneity correlated with OS and PFS (5). However, with the exploration of the whole lung, these features need to be further investigated in terms of interpretability and consistency with the biology of cancer.

When the clinical trials were originally designed, the optimal cutoff for PD-L1 expression in first line mNSCLC was not well known; consequently, while valid based on data at the time, we have only 25% cutoff available consistently for PD-L1 in the studies explored herein. It is now more broadly known that 50% cutoff is more clinically relevant for IO monotherapy. The performance of Model-C+R on the patient subset with 50% PD-L1 cutoff is shown in Supplementary Figure 5 and further analysis with additional datasets will be performed to establish clinical relevance and clinician confidence.

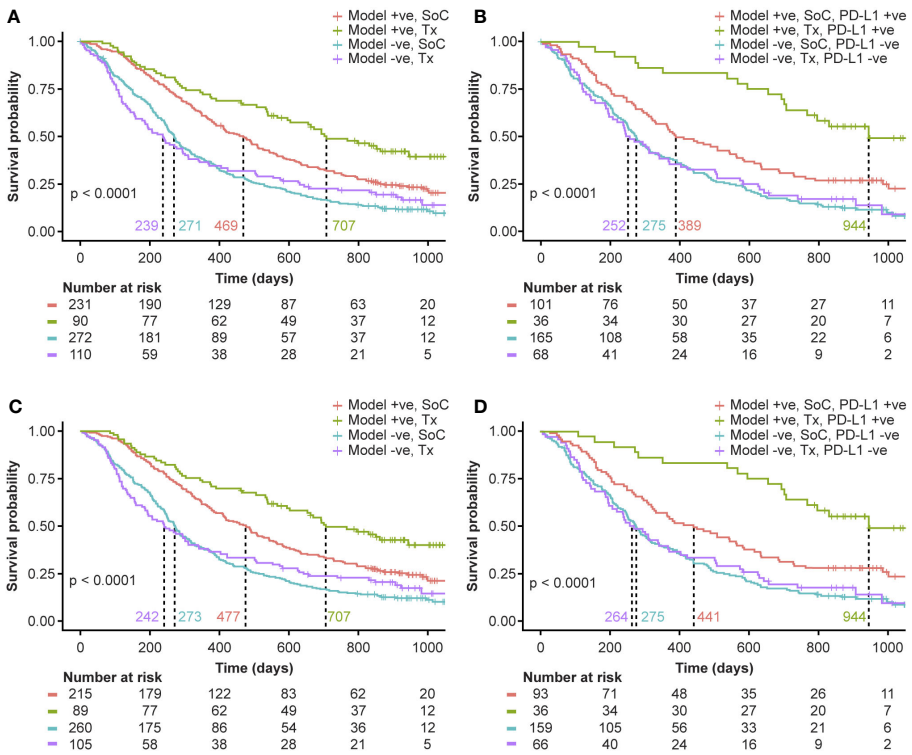


FIGURE 4
Evaluating the potential of Model-C+R as a diagnostic tool for survival-based patient risk stratification. Model+ve indicates model prediction of OS ≥365 and Model-ve indicates model prediction of OS <365. **(A)** Comparing model risk stratification on Tx arm against SoC. **(B)** Survival risk stratification on PD-L1 status and model prediction intersection in the test set (primary study cohort) and SoC. Patients identified as low-risk on Tx have median survival 944 days, 692 days more than high-risk patients on Tx. **(C)** Comparing model risk stratification on Tx arm against SoC for sub-cohort with PD-L1 expression 50% cutoff available. **(D)** Survival risk stratification on PD-L1 status at 50% cutoff and model prediction intersection in the test set (primary study cohort) and SoC. Patients identified as low-risk on Tx have median survival of 944 days. +ve, positive; -ve, negative; PD-L1, programmed death ligand 1; SoC, standard of care; Tx, treatment.

The ML models presented in this work were developed using only data from IO treatment arms. This makes it challenging to tease out the predictive/prognostic nature of the features used by the model. We plan to address this limitation as a key next step in our future research.

Though we have shown generalizability of our model to a completely independent study cohort, prospective validation of such models will be the key towards the development of diagnostic tools that can be used in the clinic.

### Future work

An important next step is to identify the predictive and prognostic nature of the model features to assist in development of further diagnostic tools. Additionally, we also plan to explore the possibility of using imaging data from follow-ups (*delta*-radiomics) to further improve our models. Results in these directions will be reported elsewhere.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Data underlying the findings described in this manuscript may be obtained in accordance with AstraZeneca's data sharing policy described at https://astrazenecagrouptrials.pharmacm.com/ST/Submission/Disclosure. Data for studies directly listed on Vivli can be requested through Vivli at www.vivli.org. Data for studies not listed on Vivli could be requested through Vivli at https://vivli.org/members/enquiries-about-studies-not-listed-on-the-vivli-platform/. The AstraZeneca Vivli member page is also available outlining further details: https://vivli.org/ourmember/astrazeneca/. Requests to access these datasets should be directed to www.vivli.org.

## Ethics statement

Ethical approval for the use of data from the studies involving humans that were utilized in this modelling study was not required, in accordance with the local legislation and institutional requirements. Written informed consent for inclusion in this modelling study from the participants in the studies involving humans that were utilized in this modelling study (or the participants' legal guardians/next of kin) was not required, in accordance with the national legislation and the institutional requirements.

## Author contributions

KP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. HR: Writing – original draft, Writing – review & editing, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization. NN: Writing – original draft, Writing – review & editing, Formal analysis, Investigation, Methodology, Software, Validation, Visualization. IG-G: Writing – original draft, Writing – review & editing, Data curation, Investigation, Methodology. JS: Writing – original draft, Writing – review & editing, Formal analysis, Investigation, Methodology, Software, Validation. PM: Resources, Writing – review & editing. JR: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The work done was supported by AstraZeneca.

## Acknowledgments

The authors would like to thank Paul Agapow and Evan Wu for useful inputs, technical discussions and annotations.

## Conflict of interest

All authors were employees of AstraZeneca, the funder, at the time of contribution to the manuscript.

The authors declare that this study received funding from AstraZeneca. The funder had the following involvement in the study: employer of the authors and controller for the data that was used for this retrospective analysis.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1383644/full#supplementary-material

# References


1. Cancer.Net. *Lung cancer - non-small cell: statistics*. ASCO. (2022). Available at: https://www.cancer.net/cancer-types/lung-cancer-non-small-cell/statistics.

2. Paz-Ares L, Spira A, Raben D, Planchard D, Cho B, Özgüroğlu M, et al. Outcomes with durvalumab by tumor PD-L1 expression in unresectable, stage III non-small-cell lung cancer in the PACIFIC trial. *Ann Oncol*. (2020) 31(6):798–806. doi: 10.1016/j.annonc.2020.03.287

3. Rizvi NA, Cho BC, Reinmuth N, Lee KH, Luft A, Ahn M-J, et al. Durvalumab with or without tremelimumab vs standard chemotherapy in first-line treatment of metastatic non–small cell lung cancer: the MYSTIC phase 3 randomized clinical trial. *JAMA Oncol*. (2020) 6:661–74. doi: 10.1001/jamaoncol.2020.0237

4. PUBMED. Available online at: https://pubmed.ncbi.nlm.nih.gov/?term=%22nsclc%22%20and%20%22immunotherapy%22%20and%20%22radiomics%22%20and%20%22CT%22%20and%20%22advanced%22%20not%20%22PET%22&filter=simsearch1.fha&sort=date.

5. Jazieh K, Khorrami M, Saad A, Gad M, Gupta A, Patil P, et al. Novel imaging biomarkers predict outcomes in stage III unresectable non-small cell lung cancer treated with chemoradiation and durvalumab. *J Immunother Cancer*. (2022) 10(3):e003778. doi: 10.1136/jitc-2021-003778

6. Gong J, Bao X, Wang T, Liu J, Peng W, Shi J, et al. A short-term follow-up CT based radiomics approach to predict response to immunotherapy in advanced non-small-cell lung cancer. *Oncoimmunology*. (2022) 11(1):2028962. doi: 10.1080/2162402X.2022.2028962

7. He BX, Zhong YF, Zhu YB, Deng JJ, Fang MJ, She YL, et al. Deep learning for predicting immunotherapeutic efficacy in advanced non-small cell lung cancer patients: a retrospective study combining progression-free survival risk and overall survival risk. *Transl Lung Cancer Res*. (2022) 11(4):670–85. doi: 10.21037/tlcr-22-244

8. Liu C, Gong J, Yu H, Liu Q, Wang S, Wang J. A CT-based radiomics approach to predict nivolumab response in advanced non-small-cell lung cancer. *Front Oncol*. (2021) 11:544339. doi: 10.3389/fonc.2021.544339

9. Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Călin AM, Delli Pizz A, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol*. (2019) 30(6):998–1004. doi: 10.1093/annonc/mdz108

10. Planchard D, Reinmuth N, Orlov S, Fischer JR, Sugawara S, Mandziuk S, et al. ARCTIC: durvalumab with or without tremelimumab as third-line or later treatment of metastatic non-small-cell lung cancer. *Ann Oncol*. (2020) 31(5):609–18. doi: 10.1016/j.annonc.2020.02.006

11. Clinicaltrials.gov. *A Global Study to Assess the Effects of MEDI4736 (Durvalumab), Given as Monotherapy or in Combination With Tremelimumab Determined by PD-L1 Expression Versus Standard of Care in Patients With Locally Advanced or Metastatic Non-Small Cell Lung Cancer (ARCTIC)*.

12. Clinicaltrials.gov. *Phase III Open Label First Line Therapy Study of MEDI 4736 (Durvalumab) With or Without Tremelimumab Versus SOC in Non-Small-Cell Lung Cancer (NSCLC)*.

13. de Castro G Jr., Rizvi NA, Schmid P, Syrigos K, Martin C, Yamamoto N, et al. NEPTUNE: phase 3 study of first-line Durvalumab plus tremelimumab in patients with metastatic NSCLC. *J Thorac Oncol*. (2023) 18(1):106–19. doi: 10.1016/j.jtho.2022.09.223

14. Clinicaltrials.gov. *Study of Durvalumab With Tremelimumab Versus SoC as 1st Line Therapy in Metastatic Non Small-Cell Lung Cancer (NSCLC) (NEPTUNE)*.

15. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. (1958) 53(282):457–81. doi: 10.1080/01621459.1958.10501452

16. Johnson ML, Cho BC, Luft A, Alatorre-Alexander J, Geater SL, Laktionov K, et al. Durvalumab with or without tremelimumab in combination with chemotherapy as first-line therapy for metastatic non–small-cell lung cancer: the phase III POSEIDON study. *J Clin Oncol*. (2023) 41(6):1213–27. doi: 10.1200/JCO.22.00975

17. Clinicaltrials.gov. *A Phase 1/2 Study to Evaluate MEDI4736*.

18. Trebeschi S, Bodalal Z, Boellaard TN, Tareco Bucho TM, Drago SG, Kurilova I, et al. Prognostic value of deep learning-mediated treatment monitoring in lung cancer patients receiving immunotherapy. *Front Oncol*. (2021) 11:609054. doi: 10.3389/fonc.2021.609054

19. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339

20. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2 1*, no. 4 (2015) 1:1–4.

21. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems (NIPS'17)*. Red Hook, NY, USA: Curran Associates Inc. (2017) 30:4768–77.

22. Van Rossum G, Drake FL. *Python 3 reference manual*. Scotts Valley, CA: CreateSpace. (2009).

23. Team, R. C. R: A language and environment for statistical computing. *Open J Stat*. (2013) 13(2).

24. Cox DR. Regression models and life-tables. *J R Stat Society: Ser B*. (1972) 34:187–202. doi: 10.1111/j.2517-6161.1972.tb00899.x