Machine Learning–Based Prediction of Distant Recurrence Risk and Ribociclib Treatment Effect in HR+/HER2– Early Breast Cancer Using Real-World and NATALEE Data

Frederick M. Howard[1], Peter A. Fasching[2], Cesar A. Santa-Maria[3], Elgene Lim[4,5], Joseph A. Sparano[6], Maryam B. Lustberg[7], Thomas Bachelot[8], Oleg Blyuss[9], Christine Brezden-Masley[10], Yeon Hee Park[11], Murat Akdere[12], Fen Ye[13], Kristyn Pantoja[14], Christoph Kurz[15], Patricia Dominguez Castro[16], and Pedram Razavi[17]

[1]Department of Medicine, University of Chicago, Chicago, IL, USA; [2]University Hospital Erlangen Comprehensive Cancer Center Erlangen-EMN, Friedrich Alexander University Erlangen-Nuremberg, Erlangen, Germany; [3]Breast and Gynecological Malignancies Group, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA; [4]Garvan Institute of Medical Research, Sydney, NSW, Australia; [5]Faculty of Medicine, St Vincent's Clinical School, University of New South Wales, Sydney, NSW, Australia; [6]Icahn School of Medicine at Mount Sinai, Tisch Cancer Institute, New York, NY, USA; [7]Department of Medicine, Yale University School of Medicine, New Haven, CT, USA; [8]Department of Medical Oncology, Centre Léon-Bérard, Lyon, France; [9]Centre for Cancer Screening, Prevention, and Early Diagnosis, Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom; [10]Division of Medical Oncology and Hematology, Faculty of Medicine, Mount Sinai Hospital, University of Toronto, Toronto, ON, Canada; [11]Division of Hematology-Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea; [12]Novartis Pharma AG, Basel, Switzerland; [13]Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA; [14]Novartis

Pharmaceuticals Corporation, Cambridge, MA, USA; [15]Novartis Pharma GmbH, Munich, Germany; [16]Novartis Ireland, Dublin, Ireland; [17]Memorial Sloan Kettering Cancer Center, New York, NY, USA

Running Title

Machine learning–based breast cancer recurrence prediction

Corresponding Author

Frederick M. Howard

5841 S. Maryland Ave. | MC 2115

Chicago, Illinois 60637-1470

Frederick.Howard@bsd.uchicago.edu

ORCID: 0000-0002-6039-1141

Declaration of Competing Interest

F.M. Howard reports personal fees from Novartis, AG, Leica Biosystems. P.A. Fasching reports personal fees from Novartis, Daiichi-Sankyo, AstraZeneca, Eisai, and Merck Sharp & Dohme; grants from Lilly, SeaGen, Roche, Agendia, Gilead, Mylan, Menarini, Veracyte, and GuardantHealth; grants from Biontech and Cepheid; grants and personal fees from Pfizer; and other from Translational Research in Oncology (TRIO). C.A. Santa-Maria, E. Lim, and O. Blyuss have nothing to disclose. J.A. Sparano reports personal fees from AstraZeneca, Genentech, Lyell, Novartis, and Pfizer. M.B. Lustberg reports personal fees from Novartis, Gilead, AstraZeneca, Daiichi Lilly, and Menarini. T. Bachelot reports personal fees and nonfinancial support from Roche; grants, personal fees and nonfinancial support from Novartis, AstraZeneca/Daiichi, and Pfizer; and grants and personal fees from SeaGen. C. Brezden-Masley reports grants and other from Astellas, AstraZeneca, Gilead, Eli Lilly, Merck, Novartis, Pfizer, Knight; other from Amgen, Beigene, and Daiichi Sankyo; nonfinancial support and other from BMS; and other from Sanofi. Y.H. Park reports grants and personal fees from MSD and AstraZeneca; grants, personal fees, and nonfinancial support from Pfizer and Roche; personal fees from Novartis; grants from Gencurix and Genome Insight; and personal fees from Daiichi-Sankyo, Gilead, and Lilly. M. Akdere, F. Ye, K. Pantoja, C. Kurz, and P. Dominguez Castro report employment and stock ownership from Novartis. P. Razavi reports grants from Grail/Illumina; grants and other from Novartis, AstraZeneca, Epic Sciences, Tempus, and Inivata; grants from Invitae/ArcherDx and Biotheranostics; and other from Foundation Medicine, Pfizer, Daiichi, Natera, Odyssey Biosciences, Paige.AI, and SAGA Diagnostics.

Notes

Word count: 4,907

Figures and tables: 6

Translational Relevance

Despite standard-of-care endocrine therapy, distant recurrence remains a major concern for patients with hormone receptor–positive/HER2-negative early breast cancer. Recently, CDK4/6 inhibitors (ribociclib, abemaciclib) have been shown to reduce the risk of recurrence in their respective clinical trial populations; better understanding individual patient risk will improve clinical decision making. In this study, a machine learning model was developed that can accurately predict distant recurrence risk over 10 years. The model was extensively validated with data from real-world patients and clinical trial participants eligible for treatment with ribociclib. Moreover, the model was used to predict the benefit of treatment with ribociclib on distant recurrence risk in a real-world population. This model supports the use of machine learning models in optimizing treatment decisions in this patient population.

Abstract

Purpose: Despite current standard-of-care endocrine therapy, distant recurrence remains a concern for patients with HR+/HER2− early breast cancer (EBC). Understanding individual recurrence risk would aid in clinical decision making. We used machine learning to identify risk factors and develop recurrence risk prediction models.

Experimental Design: Predictor variables were identified by gradient boosting and used to train models on a large, diverse real-world dataset of patients with stage I–III HR+/HER2− EBC obtained from the US-based, electronic health record-derived deidentified Flatiron Health Research Database. An elastic net–penalized Cox proportional hazards model was validated internally with real-world data and externally with data from the NATALEE trial of ribociclib in patients with HR+/HER2− EBC. Prediction and outcome concordance for distant recurrence and treatment effect were analyzed with Harrell's concordance index (C-index) and integrated Brier score (IBS); model performance over time was determined by dynamic AUC analysis.

Results: The model accurately predicted distant recurrence in the real-world cohort (n=7842; C-index: 0.85 [95% CI: 0.8461–0.8598]; IBS: 0.05 [95% CI: 0.0443–0.0495]) over time (AUC >0.7 through 10 years); internal validation and sensitivity analyses confirmed model performance. External validation with the NATALEE NSAI alone arm yielded a lower but still discriminative performance (C-index: 0.66). Training on NATALEE data improved concordance (C-index: 0.70); the NATALEE-trained model predicted a 3.2% reduction in distant recurrence at 48 months with ribociclib treatment in the real-world cohort.

Conclusion: A machine learning model was developed that accurately predicted distant recurrence in HR+/HER2− EBC. The identified predictor variables and developed models may aid risk-based personalized treatment decision making.

Introduction

Breast cancer remains the most common cancer in women, causing 2.3 million new cases in 2022, and continues to increase in incidence and mortality globally, with predicted increases in new cases and deaths of 38% and 68%, respectively, by 2050 (1, 2). Hormone receptor–positive (HR+)/human epidermal growth factor receptor 2–negative (HER2–) breast cancer is the most common subtype, accounting for approximately 70–75% of all cases (3, 4), and most patients with breast cancer are diagnosed with early breast cancer (EBC) (5).

Current standard of care for HR+/HER2− EBC consists of a combination of surgery, radiotherapy, and systemic adjuvant treatment, such as chemotherapy and/or endocrine therapy (ET) (2). However, disease recurrence remains a substantial concern (6-9). Moreover, most recurrence presents as distant recurrence (6-10), which is generally uncurable (11), dramatically affecting survival and quality of life. Recently, adjuvant treatment with the cyclin-dependent kinase 4/6 inhibitors (CDK4/6i) ribociclib and abemaciclib have shown promise in reducing recurrence risk, significantly prolonging invasive disease–free survival (iDFS) for patients with HR+/HER2– EBC (6, 12) in their respective trial populations (NATALEE, monarchE), and have received regulatory approval based on these trials (13, 14). Accurately predicting recurrence risk is critical for guiding treatment decisions and improving outcomes.

Traditional factors used to predict recurrence risk include tumor size, nodal involvement, histological grade, and lymphovascular invasion (15); the application of these anatomical factors in clinical practice has improved risk prediction. Subsequently, tumor biology has been characterized to augment these traditional risk factors, including biomarker tests such as Ki-67 biomarker score and genomic tests examining gene expression. Tools including Oncotype DX, EndoPredict, PREDICT, and MammaPrint have been applied to breast cancer management with some success; for example, Oncotype DX application impacted approximately one-third of treatment decisions across studies (16). Methodologies to combine genomic test results with other risk factors have been developed, such as RSClin, a tool that combines Oncotype DX Recurrence Score with several patient characteristics and tumor anatomical features (17), and EPclin, which combines EndoPredict results with tumor anatomical features (18). However, careful selection of additional predictive variables has been shown to improve the predictive performance of models (19, 20), suggesting that algorithms accounting for a wide breadth of predictive variables will result in improved model accuracy.

Machine learning models can integrate and analyze complex data, with the promise of more precise risk estimation through the integration of clinical, pathologic, radiologic, demographic, and genomic data (21). Thus, these models can be leveraged to improve precision medicine approaches in the management of patients with HR+/HER2– EBC. Nevertheless, challenges for the clinical implementation of these tools remain, including model bias owing to insufficient training data in terms of size and breadth of the datasets (22). Due to its breadth and applicability, real-world data (RWD) provides the opportunity to train robust machine learning models. Moreover, clinical trial data can be used to validate RWD-based machine learning models to improve estimation accuracy and clinical applicability (23). In this study, we used machine learning to identify predictive factors for distant recurrence via model training on RWD, develop and validate distant recurrence risk models by training on RWD and clinical trial data separately, and assess the ability of the validated clinical trial data–trained model to predict the treatment effect on distant recurrence risk in a real-world setting. The aim of this study was to develop a personalized prognostic assessment for a broad range of patients with HR+/HER2– EBC to ultimately improve patient management and treatment decision making.

Methods

RWD-based machine learning model study design

This study used the US-based, electronic health record-derived deidentified Flatiron Health Research Database (24). The identification period was from 1 January 2011 to 30 April 2024; the index date for patients was the date of initiation of adjuvant therapy (i.e., first treatment, either drug therapy or radiotherapy, following surgery). The post-period included all available data following and including the initial diagnosis date.

Patients

From the Flatiron Health electronic health record (EHR)–derived deidentified database, de-identified patients aged ≥18 years with stage I–III HR+/HER2− EBC at diagnosis who had undergone surgery and initiated adjuvant ET were included. Exclusion criteria included a metastatic diagnosis before the index date, treatment with clinical study drugs in any setting (i.e., neoadjuvant, adjuvant, locoregional, or metastatic) during the study period, and treatment with a CDK4/6i (**Supplementary Fig. 1**). This study, which used deidentified data per expert determination, does not constitute human subjects research under the Common Rule and thus did not require IRB oversight or patient informed consent.

Data from the NATALEE trial were used for external validation studies (Supplementary Methods). The NATALEE study design has been described previously (25). The NATALEE trial was conducted in accordance with the provisions of the Declaration of Helsinki and Good Clinical Practice guidelines. The trial protocol and all amendments were approved by an institutional review board or independent ethics committee at each site. The conduct of the trial was overseen by a steering committee that included participating international investigators, representatives of the sponsor, and a patient advocate. An independent data monitoring committee assessed efficacy and safety data in accordance with the trial protocol. All patients provided written informed consent.

Study endpoints

Distant recurrence, defined as whether an individual experienced an event of distant recurrence, was the primary endpoint of this prediction exercise. Distant recurrence–free survival (DRFS), defined as whether an individual experienced an event of distant recurrence or death (any cause), and overall survival (OS; i.e., death from any cause) were also assessed (26). Survival endpoints such as DRFS and OS can be used to account for possible bias arising from death as a competing risk event. Additional details are provided in the **Supplementary Methods**.

RWD-trained machine learning model development and predictor variable identification

To identify variables to include in machine learning model training, 19 baseline characteristics closely associated with distant recurrence or death were examined using a gradient-boosting embedded method. The characteristics tested were chosen based on consultation with clinical experts, published literature, and data availability in the Flatiron Health database. The variables evaluated were age at diagnosis, race, menopausal status, Eastern Cooperative Oncology Group (ECOG) performance status, Charlson comorbidity index (CCI) score, breast cancer laterality, body mass index (BMI), percentage of ER+ cells, percentage of PR+ cells, HER2 immunohistochemistry status, Ki-67 biomarker score, Oncotype DX Recurrence Score (cutoff values were adapted from commercial cutoff points defining low risk [score <18], intermediate risk [18–30], and high risk [≥31] categories) [27]; however, a single cutoff of 18 was used as few patients fell into the high-risk category), tumor size, nodal status, tumor grade, tumor histology, time from diagnosis to surgery, re-excision after primary surgery, and primary surgery margin. All numerical variables with missingness above 60% were excluded from the analyses. For categorical variables with missingness above 30%, a category designated "unknown/not documented" was created; this approach retains all observations in the dataset while maintaining information about the extent of missingness. Missing values were handled via k-nearest neighbor imputation.

A decision tree was used to identify cutoff values for some variables (**Supplementary Methods**); these variables were subsequently treated as binary. For the prediction of distant recurrence, these variables included percentage of ER+ cells (≤88.0% vs >88.0%), percentage of PR+ cells (≤66.7% vs >66.7%), BMI (<18.5, 18.5–24.99, 25–29.99, 30–34.99, ≥35), Oncotype DX Recurrence score (<18, 18–30, >30), and Ki-67 biomarker score (≤20% vs >20%). For the OS model, the binary cutoff for percentage of PR+ cells was ≤85.1% vs >85.1%; all other variable binary cutoffs were the same across the models.

The 10 predictor variables most closely associated with distant recurrence were used to develop a “base model” using an elastic net–penalized Cox proportional hazards approach (**Supplementary Methods**), and this approach was compared to multiple machine learning models such as gradient boosting, survival support vector machine, random survival forest, and extra survival tree. Model selection was performed via cross-validation in the training set consisting of 80% of the Flatiron Health dataset.

Machine learning risk prediction model development

A RWD-trained model with parameters and model architecture selected as described above was trained to predict distant recurrence using data from 80% of patients in the Flatiron Health database that met the selection criteria described above. For sensitivity analysis, separate models with subgroups defined by follow-up time and adjuvant treatment to assess distant recurrence risk as well as models to assess DRFS and OS endpoints were also developed. Additional details are provided in the **Supplementary Methods**.

The performance of the machine learning models was evaluated on the remaining 20% of patient data that made up the held-out test set, using Harrell’s concordance index (C-index, 0.5–1.0; 1.0 = perfect concordance) and, in the case of ties, the integrated Brier score (IBS; 0–0.25; 0 = perfect concordance). All computations were performed using 1000 bootstrap iterations. We report the mean across these iterations and the 95% confidence intervals (CIs). The contribution of each predictor variable to the risk of distant recurrence, DRFS, or OS events was assessed with Shapley values and proportional hazards hazard ratios. To evaluate model performance over time, the standard area under the curve (AUC) metric was extended using a time-dependent approach. Specifically, a dynamic AUC was applied, which accounted for changes in event status (e.g., recurrence) by distinguishing between individuals who experienced the event by a given time point t (cumulative cases) and those who remained event-free at t (dynamic controls). This method enabled assessment of how accurately the model predicted the occurrence of an event within a specified time horizon rather than at a fixed point. Dynamic AUC analysis was conducted to determine the accuracy of each model in predicting distant recurrence, DRFS, or OS and to evaluate the influence of each predictor variable on model performance over time. *P* values were determined by Wald test and are reported at the $\alpha = 0.05$ level.

Validation of the RWD-trained machine learning distant recurrence risk prediction model with NATALEE trial data

The selection of predictor variables used in the RWD-trained machine learning model was validated using clinical trial data by training a separate model on the NATALEE nonsteroidal aromatase inhibitor (NSAI) alone arm (i.e., the control arm of the NATALEE study without ribociclib treatment), using variables in the NATALEE dataset that matched as closely as possible to the 10 selected predictor variables in the Flatiron Health dataset.

The generalizability of the RWD-trained model was assessed by predicting on the NATALEE NSAI alone arm data, using the C-index to determine model accuracy. Sensitivity analysis was carried out using both weighted and unweighted models trained on RWD for patients with stage I–III disease and a model trained on RWD for patients with stage II/III disease, which were applied to the NATALEE NSAI alone arm data for risk prediction. Additional details are provided in the **Supplementary Methods**.

Estimation of the ribociclib treatment effect on distant recurrence risk using a weighted Cox model

Similar to the RWD-trained machine learning model, the clinical trial data–based model was trained using the same 10 predictor variables with 80% of the NATALEE population, and the model's performance was confirmed using the C-index and IBS on the 20% held-out test set. To assess the ribociclib treatment effect on distant recurrence risk in a real-world setting, the optimal machine learning architecture from discovery was retrained to best estimate ribociclib treatment benefit on distant recurrence risk. Thus, a Cox proportional hazards model weighted based on identified variable differences between the Flatiron Health patients with stage II/III disease and NATALEE NSAI alone arm patients (**Supplementary Fig. 2**) was trained with 80% of the entire NATALEE cohort and fitted using the 10 predictor variables along with treatment (ribociclib plus NSAI vs NSAI alone) as a covariate. The model was then tested in the remaining 20% of the NATALEE cohort.

After model-fitting and evaluation, the weighted Cox proportional hazards model trained on the full NATALEE cohort was used to predict the survival outcomes of the Flatiron Health cohort (stage I–III, without CDK4/6i treatment) to estimate the potential ribociclib treatment effect on distant recurrence risk. In a sensitivity analysis, an unweighted Cox proportional hazards model was also fitted to predict distant recurrence risk.

Data availability

Novartis is committed to sharing patient-level data and supporting clinical documents from eligible studies with qualified external researchers. Requests are reviewed by an independent panel based on scientific merit. All shared data are anonymized in accordance with applicable laws and regulations to protect patient privacy. Details on data availability and the request process are provided at www.clinicalstudydatarequest.com. The data supporting findings from this study also originated from Flatiron Health, Inc. These deidentified data may be made available upon request and are subject to a license agreement with Flatiron Health, Inc. Interested researchers should contact cgdb-fmi@flatiron.com to inquire about licensing terms. Sample code for reproducibility can be found at https://osf.io/rnsua/.

Results

RWD-trained machine learning models and predictor variables

From the Flatiron Health database, 7842 patients with stage I–III HR+/HER2− EBC who received adjuvant ET, but not a CDK4/6i, between January 2011 and April 2024 were included (**Table 1**), 382 (4.9%) of whom experienced distant recurrence, 994 (12.7%) a DRFS event, and 829 (10.6%) an OS event.

According to gradient-boosting modeling and determination of Shapely values and Cox proportional hazards hazard ratios for each predictor variable, the variables most closely associated with increased distant recurrence risk in the “base model” were greater nodal involvement, larger tumor size, greater tumor grade, Ki-67 biomarker score >20%, longer time from diagnosis to surgery, and older age at diagnosis, whereas Oncotype DX Recurrence Score <18, percentage of PR+ cells >66.7%, percentage of ER+ cells >88.0%, and postmenopausal status were the variables most closely associated with decreased distant recurrence risk (**Fig. 1**). Several variables in the dataset contained missing values, which were addressed through imputation. The proportion of missing data varied by variable. Oncotype DX Recurrence Scores and Ki-67 biomarker scores were missing in 58.5% and 55.5% of cases, respectively. Given that these tests are typically ordered based on specific clinicopathologic features, the assumption of missing completely at random was considered to be unlikely. Instead, missing at random, in which the probability of missingness is thought to be to depend on observed covariates, was assumed. Additional missing data were observed for race (11.8%), menopausal status (6.5%), and tumor grade (0.6%).

Multiple types of machine learning models, developed using the 10 predictor variables most closely associated with distant recurrence identified above, were applied to determine the best-performing model regarding concordance of predicted and actual distant recurrence in the Flatiron Health cohort outcome data. While random survival forest, survival support vector machine, elastic net–penalized Cox proportional hazards model, gradient boosting, and extra survival trees produced similar concordance between predicted and actual distant recurrence events, the elastic net–penalized Cox proportional hazards model (**Fig. 2, Supplementary Table 1**), with a C-index of 0.85 (95% CI: 0.8461–0.8598) and an IBS of 0.05 (95% CI: 0.0443–0.0495), was chosen for the distant recurrence "base model" based on its interpretability, clinical plausibility, and ease of deployment compared with the other approaches; thus, this model was used for all downstream analyses. Results from the elastic net–penalized Cox proportional hazards model without imputation, as well as a version excluding nodal status, are also reported (**Supplementary Tables 2 and 3**).

Given the inherent potential biases in RWD, multiple sensitivity analyses were performed, and distant recurrence prediction models were developed to account for potential confounders due to differences in follow-up time (**Supplementary Table 4**, **Supplementary Fig. 3A**) or adjuvant treatment with chemotherapy (**Supplementary Table 5, Supplementary Fig. 3B**), which displayed similar importance for predictor variables, except a notable addition of BMI and race for the "adjuvant chemotherapy model." DRFS (**Supplementary Fig. 3C**) and OS (**Supplementary Fig. 3D**) prediction models showed similar variable importance, with a slightly greater influence of age at diagnosis and ECOG performance status. All sensitivity analysis models demonstrated similar concordance between predicted and actual outcomes (5-year follow-up model: C-index: 0.83; IBS: 0.04; adjuvant chemotherapy model: C-index: 0.75; IBS: 0.09; DRFS model: C-index: 0.74; IBS: 0.09; and OS model: C-index: 0.77; IBS: 0.07).

Accuracy of the RWD-trained machine learning models and contributions of the predictor variables to distant recurrence risk over time

For the elastic net-penalized Cox model trained on data from patients with stage I–III HR+/HER2− EBC who received adjuvant ET, but not a CDK4/6i ("base model"), distant recurrence risk prediction accuracy was high, particularly for the first 48 months after diagnosis, with a dynamic AUC >0.85 (**Fig. 3**). The prediction accuracy decreased slightly over time, but the dynamic AUC remained >0.70 through 120 months. Bootstrapping (1000 iterations) indicated high model stability, as reflected by narrow CIs for the overall time-dependent AUC curve. Regarding the contribution of the individual predictor variables to the accuracy of the machine learning model prediction, Oncotype DX Recurrence Score <18, postmenopausal status, percentage of PR+ cells >66.7%, percentage of ER+ cells >88.0%, and older age at diagnosis increased in importance over time, whereas the importance of tumor grade and time from diagnosis to surgery remained stable.

Validation of the RWD-trained machine learning model in Flatiron Health RWD and NATALEE trial data

To externally validate the developed machine learning model, NATALEE trial data were employed. The characteristics of the Flatiron Health and NATALEE populations were first compared. Patients with stage II/III disease from the Flatiron Health cohort and those from the NATALEE trial NSAI alone arm had generally similar characteristics; differences between the populations included age (median: 52 vs 62 y; p<.001) (**Supplementary Table 6**) (8). Moreover, the survival without distant recurrence of the Flatiron Health population was also similar to that of the NATALEE NSAI alone and ribociclib plus NSAI arm populations (**Supplementary Fig. 4**). The “base model” trained on data from patients with stage I–III HR+/HER2– EBC was applied to the NATALEE NSAI alone arm and remained predictive, with slightly lower performance than in the Flatiron Health cohort (C-index: 0.66). For additional sensitivity analyses, we determined whether retraining the model on only patients with stage II-III disease in the Flatiron Health cohort (to better match the inclusion criteria of NATALEE, which excludes stage I disease) or reweighting parameters according to differences in variable distribution between the Flatiron Health and NATALEE cohorts altered model performance. Both of these approaches resulted in performance similar to that of the original “base model, without a significant increase in predictive accuracy ” (C-index: 0.65 and 0.66, respectively) (**Supplementary Fig. 4**).

NATALEE data–trained machine learning model and prediction of ribociclib treatment effect on distant recurrence risk

Training the machine learning model on data from 80% of the NATALEE NSAI alone arm, using the same 10 predictor variables, and evaluating it on the remaining 20% yielded slightly improved concordance between distant recurrence risk prediction and actual recurrence (C-index: 0.70, IBS: 0.06).

Training a weighted Cox proportional hazards model on NATALEE data, now including ribociclib treatment status as an 11th predictor variable, was performed to evaluate the generalizability of model training with the 10 selected variables to the NATALEE cohort. This model, applied to the NATALEE cohort, yielded a C-index of 0.68; tumor size, nodal status, percentage of PR+ cells, and Ki-67 biomarker score were confirmed as significantly contributing variables (**Fig. 4**). Ribociclib plus NSAI treatment, as a predictor variable, had a hazard ratio of 0.66 (95% CI: 0.55–0.80), representing a 34% relative reduction in risk of distant recurrence across the follow-up period.

This machine learning model was then applied to the full Flatiron Health cohort (patients with stage I–III disease, without CDK4/6i treatment) to predict the risk of distant recurrence in a real-world population over time. At 36 months, there was a 2.4% reduction in predicted distant recurrence risk with ribociclib plus NSAI vs NSAI alone (95.2% vs 92.9%), which increased to a 3.2% reduction in predicted distant recurrence risk at 48 months (93.4% vs 90.2%) (**Fig. 5**). For the unweighted model, reductions in predicted distant recurrence risk with ribociclib plus NSAI vs NSAI alone were identical to the weighted model (36 months: 95.2% vs 92.9%; 48 months: 93.4% vs 90.2%).

Discussion

To improve patient outcomes in this era of novel therapeutics for patients with EBC, better prediction of distant recurrence and survival risk is needed to optimize treatment decisions. In this study, a machine learning model was developed on the basis of clinical, pathologic, and genomic factors, which predicted distant recurrence in HR+/HER2− EBC with high discriminatory performance.

Nineteen variables were assessed for their association with distant recurrence. The Shapely analysis (**Fig. 1**) in the distant recurrence model revealed that tumor status, nodal status, and tumor grade were among the most influential predictors. Their importance in our model aligns with traditional clinicopathologic risk stratification approaches such as the Nottingham Prognostic Index and clinical tools like NHS PREDICT (27). Additionally, the strong contribution of the Oncotype DX score reinforces the recognized value of genomic risk profiling in HR+/HER2− disease, as supported by the TAILORx trial and related work (28). The proportions of ER- and PR-positive cells and the Ki-67 index also emerged as key features, reflecting tumor biology and proliferative capacity—attributes that have well-documented associations with recurrence and endocrine resistance (29-32). Other variables, such as menopausal status, time from diagnosis to surgery, and age at diagnosis, have similarly been implicated in previous studies as prognostically relevant in EBC (33-38). Notably, we identified time from diagnosis to surgery as an important variable for estimating distant recurrence risk, which has recently been associated with breast cancer–free survival (33, 39), but to our knowledge has not been previously examined in prognostic algorithms (17, 18, 40-42). In addition, while other prognostic tools (e.g., RSClin, EPclin, breast cancer index [BCI], PREDICT) are derived from a similar range of clinical, pathologic, and genomic factors, they currently include a lower number of variables (ranging from 3 to 8, depending on the model) (17, 18, 40-42). Importantly, PREDICT has undergone multiple updates to include additional variables, which has led to corresponding improvements in model performance (20, 41, 43, 44). However, unlike the machine learning model presented here, PREDICT does not include genomic testing results, time from diagnosis to surgery, or menopausal status. Additionally, while RSClin did not report C-index or IBS in their validation cohort (17), an independent analysis using the National Cancer Data Base reported a C-index of 0.574 for OS (45). Similarly, EPclin has demonstrated a C-index of 0.69 (46), which is lower than the performance observed with the model presented in this study when applied to the Flatiron dataset and comparable to its performance in the NATALEE dataset. Together, these studies demonstrate that inclusion of additional variables in model training for this patient population can potentially improve prediction performance.

We also trained models with DRFS and OS as endpoints to determine whether this approach could support prediction of outcomes with a survival component with similar accuracy and to account for death as a competing risk event. These models had slightly lower predictive accuracy than the “base model” for distant recurrence. Systemic health indicators such as ECOG performance status and age at diagnosis were more predictive in the models incorporating survival events than in the distant recurrence models, whereas biomarkers such as percentages of HR+ cells and Ki-67 biomarker score were less important. This finding suggests that factors directly linked to health measures were more important to the models predicting survival endpoints and that recurrence endpoints were more closely related to disease biology characteristics.

Taken together, these data suggest that models intended to predict distant recurrence risk should focus on recurrence-specific outcomes, rather than on composite outcomes (i.e., DRFS) or mortality (i.e., OS), to achieve optimal performance.

The external validation of RWD-trained machine learning models on the NATALEE randomized clinical trial dataset with good predictive accuracy supports the generalizability of the model. This validation step is a notable advantage of our study because many contemporary studies do not include formal validation with independent clinical trial data (47). External validation with a dataset representative of the target population is recommended for the development of models with the goal of clinical application (23). Each of the tested models performed similarly, supporting the validity of the machine learning model and its applicability in the clinical setting. Notably, the elastic net–penalized Cox proportional hazards model demonstrated strong predictive performance even without nodal status, supporting its potential utility in clinical settings in which this information may be missing or unavailable—as in trials such as Choosing Wisely, SOUND, and INSEMA, in which nodal status was not available.

Analysis of the predictive ability of the RWD-trained machine learning model over time revealed a high level of accuracy through 10 years, with the highest accuracy in the first 48 months. The contributions of the individual predictor variables to model accuracy shifted over time, with most variables increasing in their contribution, such as percentage of ER+/PR+ cells, Oncotype DX Recurrence Score, menopausal status, and age at diagnosis. Notably, markers of ET sensitivity (i.e., percentage of ER+/PR+ cells) may therefore be more predictive of later distant recurrence than of early distant recurrence. This result may indicate that a higher percentage of ER+/PR+ cells is protective against distant recurrence via prolonged ET efficacy, whereas a lower percentage may result in ET insensitivity, diminishing the effectiveness of standard-of-care ET. Previous studies of RSClin, which includes age and tumor grade and size along with Oncotype DX Recurrence Score, have shown that Oncotype DX Recurrence Score alone is not as accurate as the combination of Oncotype DX Recurrence Score and clinical variables; additionally, the Oncotype DX Recurrence Score has a more distant association with distant recurrence risk over time (48). In contrast, the impact of Oncotype DX Recurrence Score on model performance increased over time in our model, indicating that when considered in combination with the predictor variables used in this model, Oncotype DX Recurrence Score remains an important prognostic factor.

The machine learning model was also trained on data from the NATALEE NSAI alone arm to tailor the model to the approved patient population eligible for ribociclib treatment, which comprises patients with a high risk of recurrence (25); this model performed similarly for predicting distant recurrence risk. Estimation of the ribociclib treatment effect on distant recurrence risk in the full NATALEE cohort by the NATALEE data–trained model yielded a predicted treatment effect on distant recurrence (HR: 0.66, 95% CI: 0.55–0.80) in a range similar to the actual measured iDFS (HR: 0.75, 95% CI: 0.62–0.91) and DRFS (HR: 0.72, 95% CI: 0.58–0.89) benefits of ribociclib plus NSAI in the NATALEE trial (8). Furthermore, applying this model to the entire real-world population in this study yielded a predicted reduction in distant recurrence risk, which increased over time to 3.2% at 48 months. To our knowledge, this is the first application of a machine learning model to predict the ribociclib treatment benefit for distant recurrence risk in a real-world population.

This study has some limitations. Systemic differences in the recording of clinical variables across clinics from which data was retrieved by Flatiron Health may exist, which could lead to unexpected biases in model performance. Moreover, real-world studies relying on electronic health records can have residual confounding signals from unmeasured, mis-measured, or mis-specified confounders (49). High missingness in important variables (e.g., Oncotype DX Recurrence Score in the NATALEE cohort) can also impact the reliability of the study results. Additionally, selection bias may exist if patients die prior to the end of the follow-up period, which is complicated by the fact that survival and/or follow-up is likely to differ between patients with and without distant recurrence, leading to potential under- or overestimation of distant recurrence risk. Another key limitation was the use of data from a US-based real-world population; future studies may be needed to verify the generalizability of the results to a global population. Finally, the index in the real-world and NATALEE data differed (i.e., start of treatment vs day of randomization, respectively), allowing for up to a 12-month delay to treatment initiation in the NATALEE cohort compared with the real-world cohort, which could influence estimation of the ribociclib treatment effect on distant recurrence risk.

Nevertheless, the extensive validation of the model results with multiple machine learning models, training datasets, and validation datasets and similarity of predictions to actual outcomes observed in the NATALEE trial suggest that model performance was not substantially affected. Additionally, the use of real-world demographically diverse data from the Flatiron Health database of US patient data enhances the generalizability of the machine learning model. External validation conducted with the global NATALEE trial data also provides robust support for the generalizability of the distant recurrence risk prediction model for application to patients with HR+/HER2− EBC.

This study supports the possibility of accurately predicting personalized risk of distant recurrence and survival, and the identified predictor variables and developed machine learning models may aid clinicians and patients with HR+/HER2− EBC in risk stratification. Moreover, the model can be applied to treatment decision making in the adjuvant setting, informing personalized surveillance or adjuvant therapy escalation/de-escalation strategies. Future studies with larger, more diverse real-world datasets representing a more diverse EBC population (including Asian patients) with additional clinical and genomic variables may improve model performance, and additional treatments could be included to expand treatment effect prediction. Finally, incorporation of additional multimodal data (e.g., digital pathology, ctDNA testing) could improve the predictive accuracy and personalization of the model output. Together, future studies can lead to the development of a decision support tool for clinical application.

Acknowledgments

We thank Venugopala Rao Manneni and Huilin Hu for their expert statistical support and contributions to the data analysis, Rima Izem for helpful discussion, and Jeffrey Melvin and Colton Smith of Nucleus Global for medical editorial assistance with this manuscript. Ribociclib was discovered by Novartis Institutes for BioMedical Research in collaboration with Astex Pharmaceuticals. This work was supported by Novartis, which also provided financial support for medical editorial assistance.

References

1. Kim J, Harper A, McCormack V, Sung H, Houssami N, Morgan E, et al. Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nat Med. 2025;31(4):1154-62.
2. Society AC. Cancer Facts & Figures 2025. Atlanta: American Cancer Society; 2025.
3. Network TCGA. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61-70.
4. Howlader N, Altekruse SF, Li CI, Chen VW, Clarke CA, Ries LA, et al. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. J Natl Cancer Inst. 2014;106(5).
5. Iqbal J, Ginsburg O, Rochon PA, Sun P, Narod SA. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. JAMA. 2015;313(2):165-73.
6. Johnston SRD, Toi M, O'Shaughnessy J, Rastogi P, Campone M, Neven P, et al. Abemaciclib plus endocrine therapy for hormone receptor-positive, HER2-negative, node-positive, high-risk early breast cancer (monarchE): results from a preplanned interim analysis of a randomised, open-label, phase 3 trial. Lancet Oncol. 2023;24(1):77-90.
7. Mayer EL, Dueck AC, Martin M, Rubovszky G, Burstein HJ, Bellet-Ezquerra M, et al. Palbociclib with adjuvant endocrine therapy in early breast cancer (PALLAS): interim analysis of a multicentre, open-label, randomised, phase 3 study. Lancet Oncol. 2021;22(2):212-22.
8. Slamon D, Lipatov O, Nowecki Z, McAndrew N, Kukielka-Budny B, Stroyakovskiy D, et al. Ribociclib plus Endocrine Therapy in Early Breast Cancer. N Engl J Med. 2024;390(12):1080-91.
9. Group EBCTC. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. Lancet. 2015;386(10001):1341-52.
10. Gnant M, Dueck AC, Frantal S, Martin M, Burstein HJ, Greil R, et al. Adjuvant Palbociclib for Early Breast Cancer: The PALLAS Trial Results (ABCSG-42/AFT-05/BIG-14-03). J Clin Oncol. 2022;40(3):282-93.
11. Caswell-Jin JL, Plevritis SK, Tian L, Cadham CJ, Xu C, Stout NK, et al. Change in Survival in Metastatic Breast Cancer with Treatment Advances: Meta-Analysis and Systematic Review. JNCI Cancer Spectr. 2018;2(4):pky062.
12. Hortobagyi GN, Lacko A, Sohn J, Cruz F, Ruiz Borrego M, Manikhas A, et al. A phase III trial of adjuvant ribociclib plus endocrine therapy versus endocrine therapy alone in patients with HR-positive/HER2-negative early breast cancer: final invasive disease-free survival results from the NATALEE trial. Ann Oncol. 2025;36(2):149-57.
13. FDA expands early breast cancer indication for abemaciclib with endocrine therapy: US FDA; 2023 [updated 3/3/2023. Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-expands-early-breast-cancer-indication-abemaciclib-endocrine-therapy.
14. FDA approves Kisqali with an aromatase inhibitor and Kisqali Femara co-pack for early high-risk breast cancer: FDA; 2024 [updated 9/17/2024. Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-ribociclib-aromatase-inhibitor-and-ribociclib-and-letrozole-co-pack-early-high-risk.
15. Voogd AC, Nielsen M, Peterse JL, Blichert-Toft M, Bartelink H, Overgaard M, et al. Differences in risk factors for local and distant recurrence after breast-conserving therapy or mastectomy for stage I and II breast cancer: pooled results of two large European randomized trials. J Clin Oncol. 2001;19(6):1688-97.

16. Ward S, Scope A, Rafia R, Pandor A, Harnan S, Evans P, et al. Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2013;17(44):1-302.

17. Sparano JA, Crager MR, Tang G, Gray RJ, Stemmer SM, Shak S. Development and Validation of a Tool Integrating the 21-Gene Recurrence Score and Clinical-Pathological Features to Individualize Prognosis and Prediction of Chemotherapy Benefit in Early Breast Cancer. J Clin Oncol. 2021;39(6):557-64.

18. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res. 2011;17(18):6012-20.

19. Brinkman N, Shah R, Doornberg J, Ring D, Gwilym S, Jayakumar P. Artificial neural networks outperform linear regression in estimating 9-month patient-reported outcomes after upper extremity fractures with increasing number of variables. OTA Int. 2023;6(5 Suppl):e284.

20. Wishart GC, Rakha E, Green A, Ellis I, Ali HR, Provenzano E, et al. Inclusion of KI67 significantly improves performance of the PREDICT prognostication and prediction model for early breast cancer. BMC Cancer. 2014;14:908.

21. Silveira JA, da Silva AR, de Lima MZT. Harnessing artifcial intelligence for predicting breast cancer recurrence: a systematic review of clinical and imaging data. Discov Oncol. 2025;16(1):135.

22. Hunter B, Hindocha S, Lee RW. The Role of Artificial Intelligence in Early Cancer Diagnosis. Cancers (Basel). 2022;14(6).

23. Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ. 2024;384:e074819.

24. Database Characterization Guide: Flatiron Health; 2025 [updated March 18, 2025. Available from: https://flatiron.com/database-characterization.

25. Slamon DJ, Fasching PA, Hurvitz S, Chia S, Crown J, Martin M, et al. Rationale and trial design of NATALEE: a Phase III trial of adjuvant ribociclib + endocrine therapy versus endocrine therapy alone in patients with HR+/HER2- early breast cancer. Ther Adv Med Oncol. 2023;15:17588359231178125.

26. Tolaney SM, Garrett-Mayer E, White J, Blinder VS, Foster JC, Amiri-Kordestani L, et al. Updated Standardized Definitions for Efficacy End Points (STEEP) in Adjuvant Breast Cancer Clinical Trials: STEEP Version 2.0. J Clin Oncol. 2021;39(24):2720-31.

27. Copson ER, Abraham JE, Braybrooke JP, Cameron D, McIntosh SA, Michie CO, et al. Expert UK consensus on the definition of high risk of recurrence in HER2-negative early breast cancer: A modified Delphi panel. Breast. 2023;72:103582.

28. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. N Engl J Med. 2018;379(2):111-21.

29. Sleightholm R, Neilsen BK, Elkhatib S, Flores L, Dukkipati S, Zhao R, et al. Percentage of Hormone Receptor Positivity in Breast Cancer Provides Prognostic Value: A Single-Institute Study. J Clin Med Res. 2021;13(1):9-19.

30. Kwak Y, Jang SY, Choi JY, Lee H, Shin DS, Park YH, et al. Progesterone Receptor Expression Level Predicts Prognosis of Estrogen Receptor-Positive/HER2-Negative Young Breast Cancer: A Single-Center Prospective Cohort Study. Cancers (Basel). 2023;15(13).

31. Lee J, Lee YJ, Bae SJ, Baek SH, Kook Y, Cha YJ, et al. Ki-67, 21-Gene Recurrence Score, Endocrine Resistance, and Survival in Patients With Breast Cancer. JAMA Netw Open. 2023;6(8):e2330961.

32. Ma Q, Liu YB, She T, Liu XL. The Role of Ki-67 in HR+/HER2- Breast Cancer: A Real-World Study of 956 Patients. Breast Cancer (Dove Med Press). 2024;16:117-26.

33. Zhu S, Li S, Huang J, Fei X, Shen K, Chen X. Time interval between breast cancer diagnosis and surgery is associated with disease outcome. Sci Rep. 2023;13(1):12091.

34. Andre F, Ismaila N, Allison KH, Barlow WE, Collyar DE, Damodaran S, et al. Biomarkers for Adjuvant Endocrine and Chemotherapy in Early-Stage Breast Cancer: ASCO Guideline Update. J Clin Oncol. 2022;40(16):1816-37.

35. Sheffield KM, Peachey JR, Method M, Grimes BR, Brown J, Saverno K, et al. A real-world US study of recurrence risks using combined clinicopathological features in HR-positive, HER2-negative early breast cancer. Future Oncol. 2022;18(21):2667-82.

36. Lao C, Elwood M, Kuper-Hommel M, Campbell I, Lawrenson R. Impact of menopausal status on risk of metastatic recurrence of breast cancer. Menopause. 2021;28(10):1085-92.

37. Arvold ND, Taghian AG, Niemierko A, Abi Raad RF, Sreedhara M, Nguyen PL, et al. Age, breast cancer subtype approximation, and local recurrence after breast-conserving therapy. J Clin Oncol. 2011;29(29):3885-91.

38. Pedersen RN, Esen B, Mellemkjær L, Christiansen P, Ejlertsen B, Lash TL, et al. The Incidence of Breast Cancer Recurrence 10-32 Years After Primary Diagnosis. J Natl Cancer Inst. 2022;114(3):391-9.

39. Holzel D, Schlesinger-Raab A, Schubert-Fritschle G, Halfter K. Prolonged time to breast cancer surgery and the risk of metastasis: an explorative simulation analysis using epidemiological data from Germany and the USA. Breast Cancer Res Treat. 2025;211(1):151-60.

40. Boyages J, Taylor R, Chua B, Ung O, Bilous M, Salisbury E, et al. A risk index for early node-negative breast cancer. Br J Surg. 2006;93(5):564-71.

41. Grootes I, Wishart GC, Pharoah PDP. An updated PREDICT breast cancer prognostic model including the benefits and harms of radiotherapy. NPJ Breast Cancer. 2024;10(1):6.

42. van der Plas-Krijgsman WG, Giardiello D, Putter H, Steyerberg EW, Bastiaannet E, Stiggelbout AM, et al. Development and validation of the PORTRET tool to predict recurrence, overall survival, and other-cause mortality in older patients with breast cancer in the Netherlands: a population-based study. Lancet Healthy Longev. 2021;2(11):e704-e11.

43. Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res. 2010;12(1):R1.

44. Wishart GC, Bajdik CD, Dicks E, Provenzano E, Schmidt MK, Sherman M, et al. PREDICT Plus: development and validation of a prognostic model for early breast cancer that includes HER2. Br J Cancer. 2012;107(5):800-7.

45. Vannier AGL, Dhungana A, Zhao F, Chen N, Shubeck S, Hahn OM, et al. Validation of the RSClin risk calculator in the National Cancer Data Base. Cancer. 2024;130(8):1210-20.

46. Martin M, Brase JC, Ruiz A, Prat A, Kronenwett R, Calvo L, et al. Prognostic ability of EndoPredict compared to research-based versions of the PAM50 risk of recurrence (ROR) scores in node-positive, estrogen receptor-positive, and HER2-negative breast cancer. A GEICAM/9906 sub-study. Breast Cancer Res Treat. 2016;156(1):81-9.

47. Gonzalez R, Nejat P, Saha A, Campbell CJV, Norgan AP, Lokker C. Performance of externally validated machine learning models based on histopathology images for the diagnosis,

classification, prognosis, or treatment outcome prediction in female breast cancer: A systematic review. J Pathol Inform. 2023;15:100348.

48. Sparano JA, Crager M, Gray RJ, Tang G, Hoag J, Baehner FL, et al. Clinical and Genomic Risk for Late Breast Cancer Recurrence and Survival. NEJM Evid. 2024;3(8):EVIDoa2300267.

49. Vafaei Sadr A, Li J, Hwang W, Yeasin M, Wang M, Lehmann H, et al. Flexible imputation toolkit for electronic health records. Sci Rep. 2025;15(1):17176.

Figures and Tables

Table 1. Flatiron Health patient baseline characteristics.

Characteristic	*N* = 7842
Median age at dx (range), years	64 (22–85)
Race, n (%)	
White	5409 (69.0)
Black or African American	622 (7.9)
Other than White or Black	889 (11.3)
Unknown/not reported	922 (11.8)
Menopausal status, n (%)	
Pre-/perimenopausal	1522 (19.4)
Postmenopausal	5814 (74.1)
Unknown/not documented	506 (6.5)
T status, n (%)[a]	
T1	5167 (65.9)
T2	2173 (27.7)
T3	393 (5.0)
T4	109 (1.4)
N status, n (%)[b]	
N0	5745 (73.3)
N1	1668 (21.3)
N2	282 (3.6)
N3	147 (1.9)
Tumor grade, n (%)	
Grade 1	2275 (29.0)
Grade 2	4176 (53.3)
Grade 3	1341 (17.1)
Unknown/not documented	50 (0.6)
Oncotype DX Recurrence Score, n (%)	
<18	2017 (25.7)
18-30	990 (12.6)
>30	244 (3.1)
Unknown/not documented	4591 (58.5)
Ki-67 biomarker score, n (%)	
≤20%	2053 (26.2)
>20%	1439 (18.3)
Unknown/not documented	4350 (55.5)
Received chemotherapy, n (%)	
Neoadjuvant	388 (4.9)
Adjuvant	1503 (19.2)
No	5997 (76.5)
Received adj. radiotherapy, n (%)	4418 (56.3)
Received adj. targeted treatment, n (%)	10 (0.1)
Median follow-up, years (IQR)	4.8 (4.75)

[a] T1, ≤2 cm; T2, >2 cm but ≤5 cm; T3, >5 cm; T4, any size with direct extension to chest wall and/or skin (ulceration or macroscopic nodules) (AJCC 8th edition). [b] N0, node negative; N1, 1–3 axillary lymph nodes; N2, 4–9 axillary lymph nodes; N3, 10+ axillary lymph nodes (AJCC 8th edition).

adj., adjuvant; AJCC, American Joint Committee on Cancer; dx, diagnosis; IQR, interquartile range.

Figure 1. Mean absolute Shapley values to predict distant recurrence (A), impacts of variables on model output (B), and Cox proportional hazards hazard ratios (C) for a machine learning model trained on patients with stage I–III HR+/HER2− EBC who received adjuvant endocrine therapy, but not a CDK4/6i, from the Flatiron Health database.

[a] Ordinal variable treated as continuous. Linear assumption supported by data. Hazard ratio in **C** indicates change with one step of variable (e.g., N0 to N1). [b] Multivariate imputation by chained equations used to impute missing values. [c] Unknown/not documented category created when summarizing baseline variables with >30% missing data was excluded from model analyses. [d] Binary variable indicating condition that is met. Hazard ratio in **C** indicates change associated with meeting stated condition. [e] Cutoff associated with recurrence according to machine learning decision tree method. [f] Red indicates a high value (continuous variables) or that stated condition is met (binary variables). [g] Negative Shapley value indicates association with reduced DR risk and vice versa. Individual data points with same value (e.g., pts with N0 disease) are stacked.

CDK4/6i, cyclin-dependent kinase 4/6 inhibitor; dx, diagnosis; ER, estrogen receptor; N, nodal; PR, progesterone receptor; T, tumor.

Figure 2. Machine learning model performance scores with 95% confidence intervals indicated for each model.

C-index, concordance index; IBS, integrated Brier score; SVM, support vector machine.

Figure 3. Accuracy of 10-variable machine learning model trained on patients with stage I–III HR+/HER2− EBC who received adjuvant ET, but not a CDK4/6i, from the Flatiron Health database to predict distant recurrence and variables' contributions to model accuracy over time.

Relative importance was calculated as the ratio of the AUC of the independent variable to the AUC for the machine learning model. Gray shading indicates 95% confidence interval calculated for the overall population.

AUC, area under the curve; CDK4/6i, cyclin-dependent kinase 4/6 inhibitor; EBC early breast cancer; ER, estrogen receptor; ET, endocrine therapy; HER2−, human epidermal growth factor receptor 2–negative; HR+, hormone receptor–positive; N, nodal; PR, progesterone receptor; T, tumor.

Figure 4. Treatment effect of ribociclib in NATALEE population on the risk of distant recurrence.

Adapted weighted NATALEE-trained Cox proportional hazards model using treatment as a predictor and the NATALEE full population.

[a] Ordinal variable treated as continuous. Linear assumption supported by data. Hazard ratio indicates change with one step of variable (e.g., N0 to N1). [b] Multivariate imputation by chained equations used to impute missing values. [c] Unknown/not documented category created when summarizing baseline variables with >30% missing data was excluded from model analyses. [d] Binary variable indicating condition that is met. Hazard ratio indicates change associated with meeting stated condition. [e] Cutoff associated with recurrence according to machine learning decision tree method.

dx, diagnosis; ER, estrogen receptor; N, nodal; PR, progesterone receptor; T, tumor.

Figure 5. Treatment effect of ribociclib in real-world Flatiron Health population on the risk of distant recurrence.

Adapted weighted NATALEE-trained Cox proportional hazards model using the NATALEE full population to predict risk of distant recurrence in a real-world population, assuming all patients were treated with ribociclib plus NSAI or with NSAI.

NSAI, nonsteroidal aromatase inhibitor; RIB, ribociclib.

Figure 1

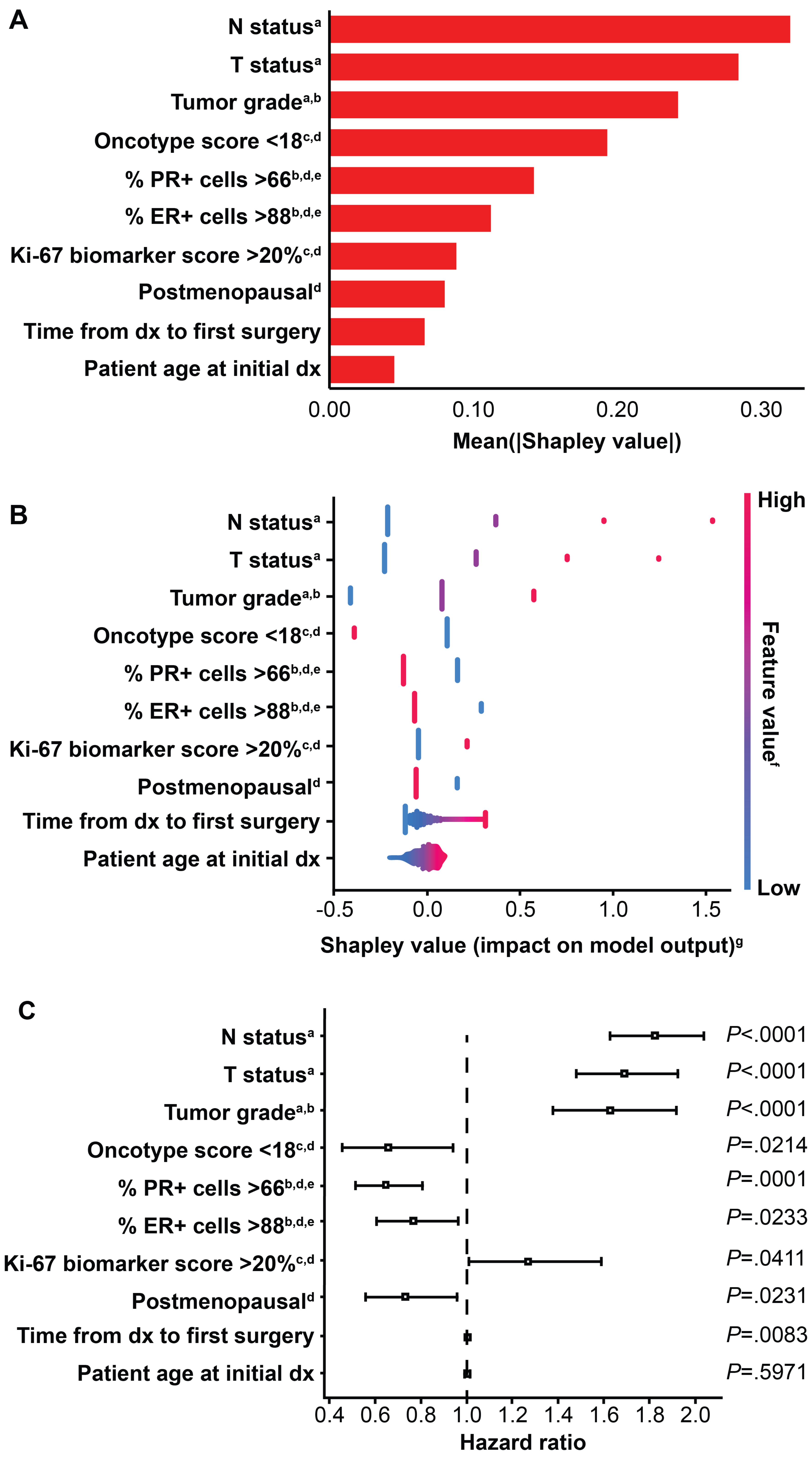
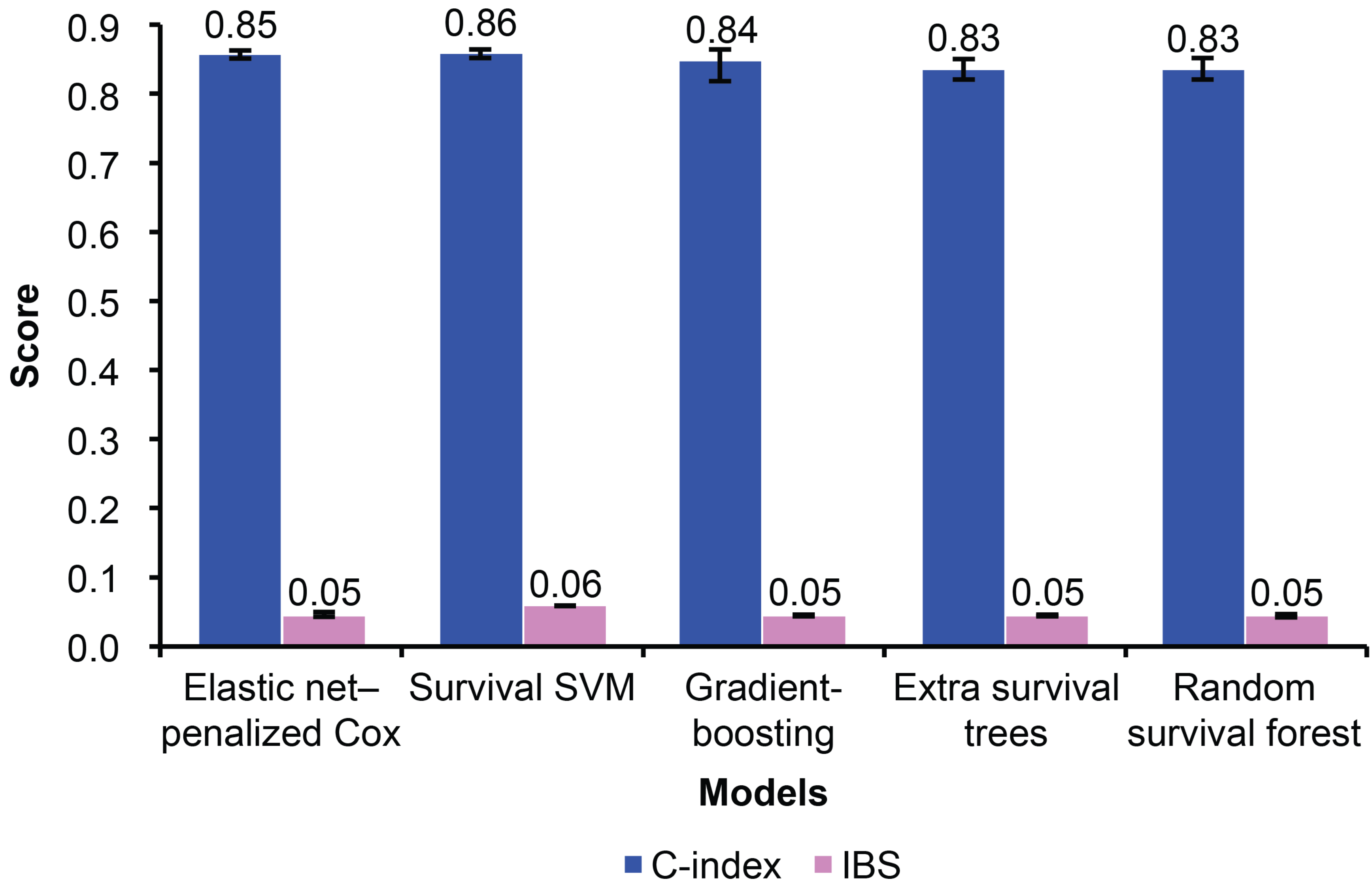
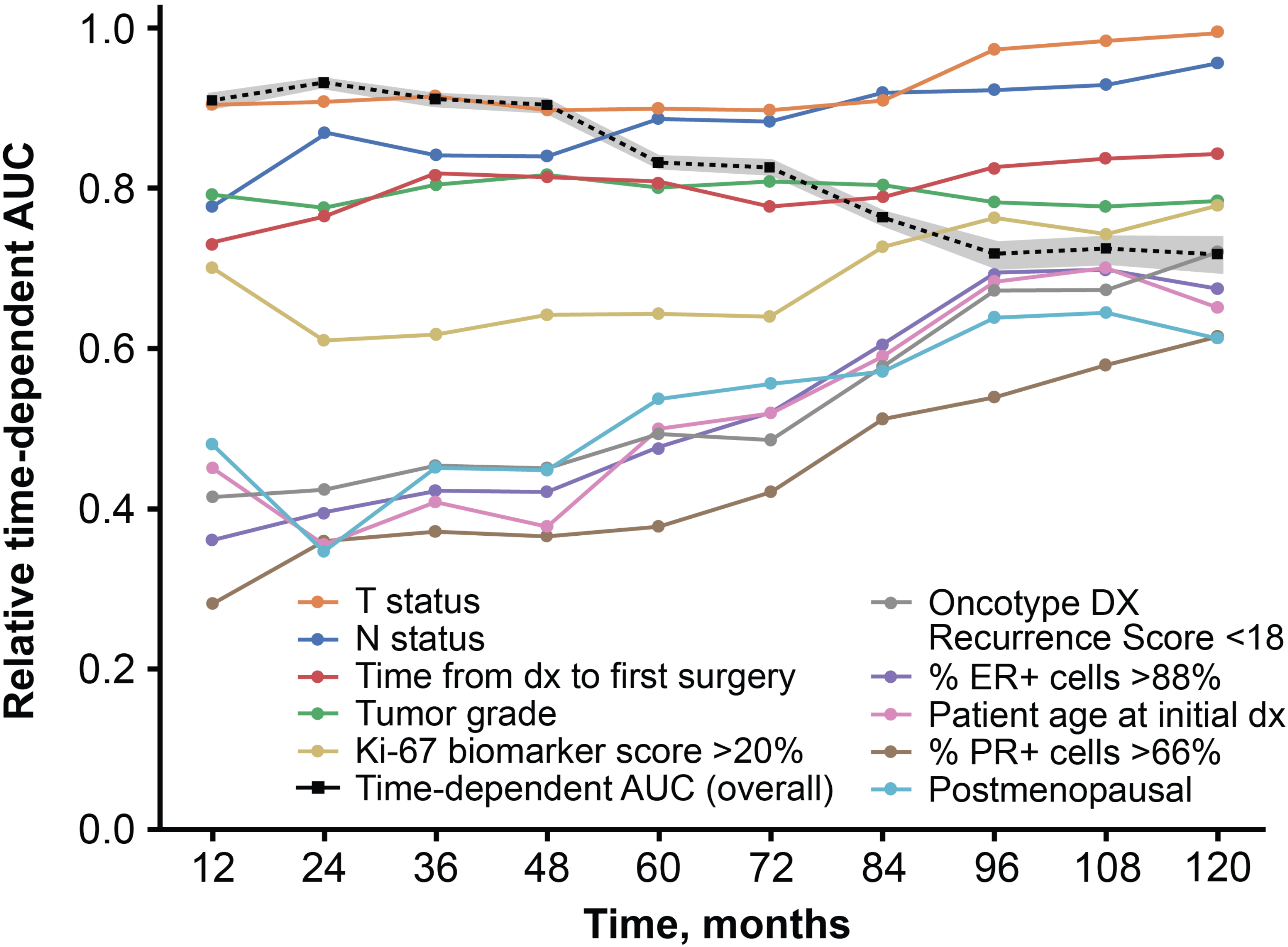

A
N status[a]
T status[a]
Tumor grade[a,b]
Oncotype score <18[c,d]
% PR+ cells >66[b,d,e]
% ER+ cells >88[b,d,e]
Ki-67 biomarker score >20%[c,d]
Postmenopausal[d]
Time from dx to first surgery
Patient age at initial dx
0.00
0.10
0.20
0.30
Mean(|Shapley value|)
B
N status[a]
T status[a]
Tumor grade[a,b]
Oncotype score <18[c,d]
% PR+ cells >66[b,d,e]
% ER+ cells >88[b,d,e]
Ki-67 biomarker score >20%[c,d]
Postmenopausal[d]
Time from dx to first surgery
Patient age at initial dx
-0.5
0.0
0.5
1.0
1.5
Shapley value (impact on model output)[g]
High
Feature value[f]
Low
C
N status[a]
T status[a]
Tumor grade[a,b]
Oncotype score <18[c,d]
% PR+ cells >66[b,d,e]
% ER+ cells >88[b,d,e]
Ki-67 biomarker score >20%[c,d]
Postmenopausal[d]
Time from dx to first surgery
Patient age at initial dx
P<.0001
P<.0001
P<.0001
P=.0214
P=.0001
P=.0233
P=.0411
P=.0231
P=.0083
P=.5971
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
Hazard ratio

Figure 2



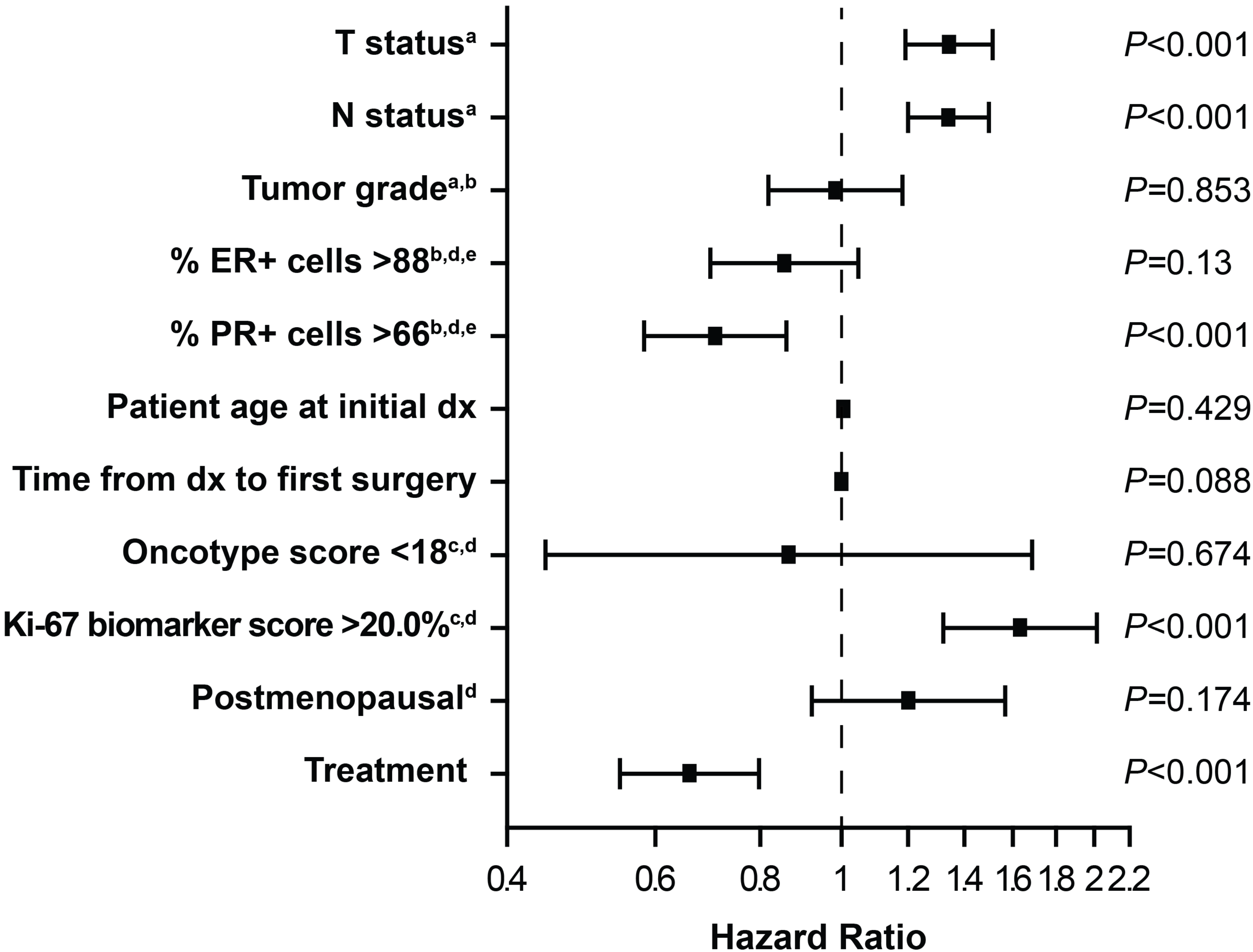
0.85
0.86
0.84
0.83
0.83
0.05
0.06
0.05
0.05
0.05
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Score
Elastic net–penalized Cox
Survival SVM
Gradient-boosting
Extra survival trees
Random survival forest
Models
C-index
IBS

Figure 3



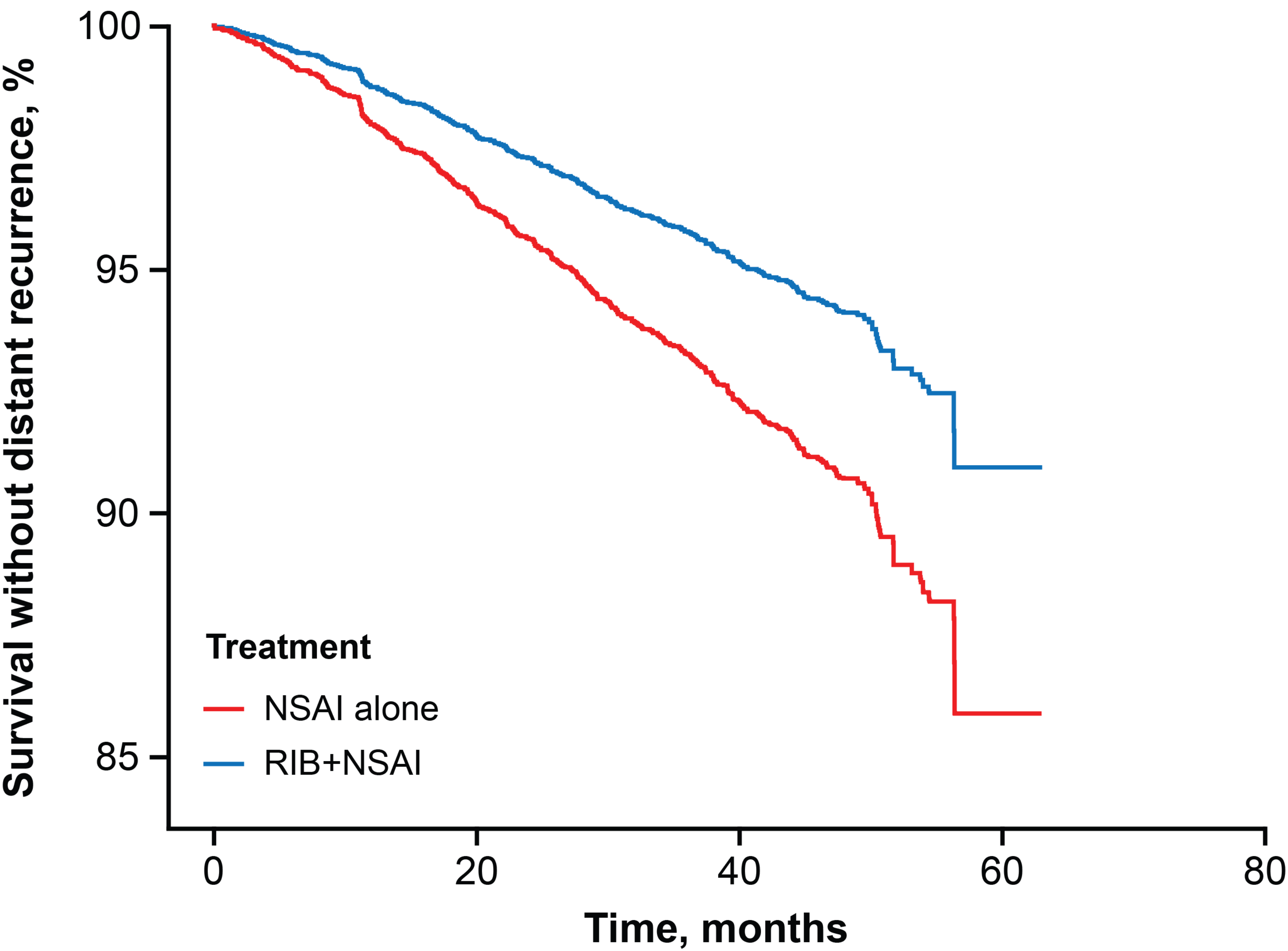
Relative time-dependent AUC
1.0
0.8
0.6
0.4
0.2
0.0
12
24
36
48
60
72
84
96
108
120
Time, months
T status
N status
Time from dx to first surgery
Tumor grade
Ki-67 biomarker score >20%
Time-dependent AUC (overall)
Oncotype DX Recurrence Score <18
% ER+ cells >88%
Patient age at initial dx
% PR+ cells >66%
Postmenopausal

Figure 4



T status[a]
N status[a]
Tumor grade[a,b]
% ER+ cells >88[b,d,e]
% PR+ cells >66[b,d,e]
Patient age at initial dx
Time from dx to first surgery
Oncotype score <18[c,d]
Ki-67 biomarker score >20.0%[c,d]
Postmenopausal[d]
Treatment
P<0.001
P<0.001
P=0.853
P=0.13
P<0.001
P=0.429
P=0.088
P=0.674
P<0.001
P=0.174
P<0.001
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
2.2
Hazard Ratio

Figure 5



Survival without distant recurrence, %
100
95
90
85
0
20
40
60
80
Time, months
Treatment
NSAI alone
RIB+NSAI