Transformer-based AI approach to unravel long-term, time-dependent prognostic complexity in patients with advanced NSCLC and PD-L1 ≥50%: insights from the Pembrolizumab 5-year global registry

Cortellini A et al.

Supplementary Materials

Supplementary Methods.

The Pembro-real 5Y cohort was created by pooling data from an existing registry, identified as Pembro-real IT[1-5], and additional data collected through an ad-hoc international registry, forming a single registry, henceforth referred to as Pembro-real 5Y[6].

The primary eligibility criteria required patients to have received first-line pembrolizumab monotherapy outside of clinical trials, with a PD-L1 TPS of ≥50%, and to have initiated treatment by May 31, 2018. To ensure sufficient follow-up for assessing long-term outcomes, the minimum data cut-off for patients still alive was set at May 1, 2023. Patients from centers that did not participate in the updated Pembro-real IT study were excluded due to the lack of follow-up data, which prevented their inclusion in long-term outcome analyses. After harmonizing variables between the two registries, consecutive patients were screened for eligibility through medical record reviews conducted by their treating teams at each participating center. Patients lost to follow-up were excluded from the analysis. The final study population included patients treated between November 2015 and May 2018 at 61 institutions across 14 countries worldwide. Patient selection process has been already reported in details using the ESMO Guidance for Reporting Oncology real-World Evidence (ESMO-GROW) flow chart[6,7]. Patients alive beyond the 5-year mark were censored at the date of their last clinical follow-up for OS.

The large panel of baseline variables used in both the analytical approaches included clinic-pathologic oncological characteristics, demographics, concomitant medications and comorbidities: age at pembrolizumab initiation (continuous); biological sex (male vs. female); ethnicity (white vs. black/African-American vs. Asian vs. Hispanic vs. others); body mass index according to the World Health Organization (obese vs. overweight vs. normal-weight vs. underweight); Eastern Cooperative Oncology Group – Performance Status (ECOG-PS) (0–1 vs. ≥ 2); primary tumor histology (squamous cell carcinoma vs. adenocarcinoma vs. other histologies/not otherwise specified); smoking status (never smokers vs. former smokers [≥ 1 year] vs. current smokers); PD-L1 TPS value (≥90% vs. <90%); presence of central nervous system (CNS) metastases (yes vs. no); bone metastases (yes vs. no); liver metastases (yes vs. no); lung metastases (yes vs. no); pleural metastases (yes vs. no); adrenal gland metastases (yes vs. no); other metastatic sites (yes vs. no); number of metastatic sites (> 3 vs. ≤ 3); corticosteroids administration at baseline within the 30 days before treatment commencement (doses ≥10 mg/day prednisone or equivalent vs. doses <10 mg/day prednisone or equivalent vs. none); proton pump inhibitors (PPI) at baseline within the 30 days before treatment commencement (yes vs. no); systemic antibiotics at baseline within the 30 days before treatment commencement (yes vs. no); statins at baseline within the 30 days before treatment commencement (yes vs. no); metformin at baseline within the 30 days before treatment commencement (yes vs. no); other glucose lowering medications (including insulin) at baseline within the 30 days before treatment commencement (yes vs. no); epidermal growth factor receptor mutation status (positive vs. negative vs. not assessed); anaplastic lymphoma kinase translocation status (positive vs. negative vs. not assessed); ROS-1 translocation status (positive vs. negative vs. not assessed); Kirsten rat sarcoma virus receptor mutation status (positive vs. negative vs. not assessed); BRAF mutation status (positive vs. negative vs. not assessed); tumor mutational burden (TMB) was classified as high and non-high as previously described[8], following harmonization of the score provided by different platforms[9]; arterial hypertension (yes vs no); myocardial infarction (yes vs no); other cardiovascular conditions (yes vs no); type 2 diabetes (yes vs no); pulmonary disease (yes vs no); dyslipidemia (yes vs no); autoimmune disease (yes vs no); other comorbidities (yes vs no).

1. Cortellini A, Tiseo M, Banna GL, et al. Clinicopathologic correlates of first-line pembrolizumab effectiveness in patients with advanced NSCLC and a PD-L1 expression of ≥ 50%. *Cancer Immunology, Immunotherapy*. 2020/11/01 2020;69(11):2209-2221. doi:10.1007/s00262-020-02613-9

2. Cortellini A, Cannita K, Tiseo M, et al. Post-progression outcomes of NSCLC patients with PD-L1 expression ≥ 50% receiving first-line single-agent pembrolizumab in a large multicentre real-world study. *European Journal of Cancer*. 2021/05/01/ 2021;148:24-35. doi:https://doi.org/10.1016/j.ejca.2021.02.005

3. Banna GL, Tiseo M, Cortinovis DL, et al. Host immune-inflammatory markers to unravel the heterogeneous outcome and assessment of patients with PD-L1 ≥50% metastatic non-small cell lung cancer and poor performance status receiving first-line immunotherapy. *Thoracic Cancer*. 2022;13(3):483-488. doi:https://doi.org/10.1111/1759-7714.14256

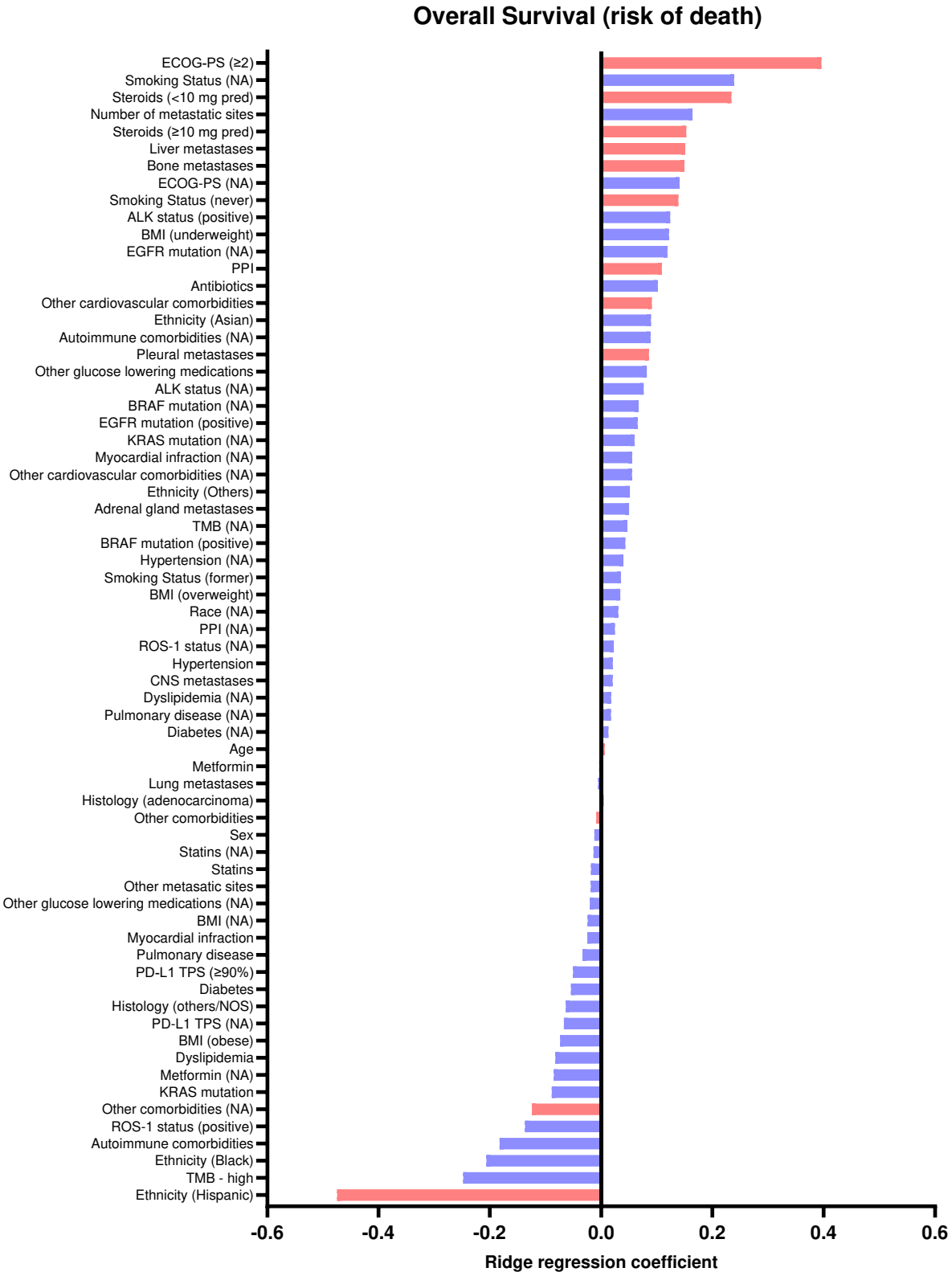
4. Cortellini A, Friedlaender A, Banna GL, et al. Immune-related Adverse Events of Pembrolizumab in a Large Real-world Cohort of Patients With NSCLC With a PD-L1 Expression ≥ 50% and Their Relationship With Clinical Outcomes. *Clinical Lung Cancer*. 2020/11/01/ 2020;21(6):498-508.e2. doi:https://doi.org/10.1016/j.cllc.2020.06.010

5. Banna GL, Cortellini A, Cortinovis DL, et al. The lung immuno-oncology prognostic score (LIPS-3): a prognostic classification of patients receiving first-line pembrolizumab for PD-L1 ≥ 50% advanced non-small-cell lung cancer. *ESMO Open*. 2021/04/01/ 2021;6(2):100078. doi:https://doi.org/10.1016/j.esmoop.2021.100078
6. Cortellini A, Garbo E, Brunetti L, et al. 534 Determinants of 5-year survival in patients with advanced NSCLC with PD-L1 ≥50% treated with first line pembrolizumab outside of clinical trials: results from the Pembro-real 5Y global registry. *Journal for ImmunoTherapy of Cancer*. 2024;12(Suppl 2):A604-A606. doi:10.1136/jitc-2024-SITC2024.0534
7. Castelo-Branco L, Pellat A, Martins-Branco D, et al. ESMO Guidance for Reporting Oncology real-World evidence (GROW). *Annals of Oncology*. 2023/12/01/ 2023;34(12):1097-1112. doi:https://doi.org/10.1016/j.annonc.2023.10.001
8. Ricciuti B, Wang X, Alessi JV, et al. Association of High Tumor Mutation Burden in Non–Small Cell Lung Cancers With Increased Immune Infiltration and Improved Clinical Outcomes of PD-L1 Blockade Across PD-L1 Expression Levels. *JAMA Oncology*. 2022;8(8):1160-1168. doi:10.1001/jamaoncol.2022.1981
9. Vokes NI, Liu D, Ricciuti B, et al. Harmonization of Tumor Mutational Burden Quantification and Association With Response to Immune Checkpoint Blockade in Non–Small-Cell Lung Cancer. *JCO Precision Oncology*. 2019;(3):1-12. doi:10.1200/po.19.00171

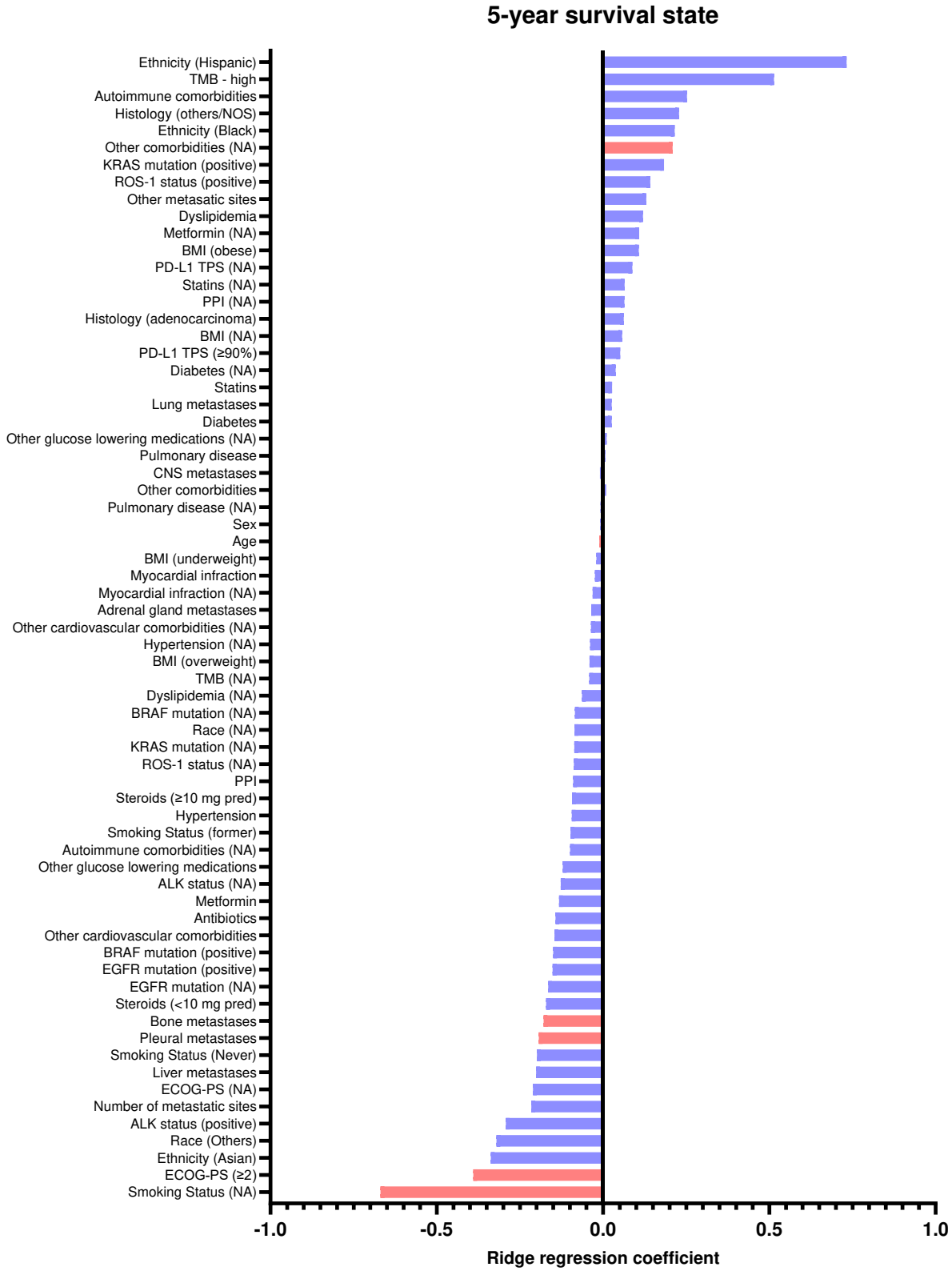
Supplementary Table 1: Participating centers list.

Participating center	N	(%)
Papardo Hospital, Messina, Italy	2	0.2%
University of Naples Federico II, Naples, Italy	4	0.4%
Henry Dunant Hospital Center, Athens, Greece	28	2.6%
IRCCS Istituto Tumori "Giovanni Paolo II", 70124 Bari, Italy	16	1.5%
Beaumont Hospital, Beaumont RCSI Cancer Centre, Dublin, Ireland	9	0.8%
University of Brescia, ASST Spedali Civili, 25123 Brescia, Italy	4	0.4%
Chelsea and Westminster Hospital, London, United Kingdom	6	0.6%
Fondazione Policlinico Universitario Campus Bio-Medico, Roma, Italy	20	1.9%
University G. D'Annunzio of Chieti-Pescara, Chieti, Italy	9	0.8%
Centre Hospitalier Intercommunal, Creteil, France	14	1.3%
Dana-Farber Cancer Institute, Boston, Massachusetts, United States	132	12.4%
Careggi University Hospital, Florence, Italy	14	1.3%
University Hospital Frankfurt, Frankfurt, Germany	20	1.9%
Cantonal Hospital Fribourg, Fribourg, Switzerland	14	1.3%
Medical University of Gdańsk, Gdańsk, Poland	11	1.0%
Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy	22	2.1%
Guy's and St Thomas' Hospitals NHS Trust, London, United Kingdom	48	4.5%
Hospital Universitario 12 De Octubre, Madrid, Spain	6	0.6%
Hospital Clinic of Barcelona, Barcelona, Spain	16	1.5%
Hospital Sírio-Libanês, São Paulo, SP, Brazil	5	0.5%
Hospital Universitario Infanta Leonor, Madrid, Spain	3	0.3%
University Hospital, Geneva, Switzerland	19	1.8%
Catalan Institute of Oncology (ICO), L'Hospitalet, Barcelona, Spain	15	1.4%
IRCCS National Cancer Institute Regina Elena, Rome, Italy	21	2.0%
Hammersmith Hospital Campus, Imperial College London, London, United Kingdom	27	2.5%
Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy	26	2.4%
IRCCS Ospedale Policlinico San Martino, Genoa, Italy	16	1.5%
King Hussein Cancer Center, Amman, Jordan	10	0.9%
Santa Maria Goretti Hospital, Latina, Italy	6	0.6%
Azienda Ospedaliero-Universitaria di Perugia, Perugia, Italy	14	1.3%
Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy	8	0.8%
AORN dei Colli Monaldi, Naples, Italy	7	0.7%
Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, New York, United States	11	1.0%
IRCCS San Gerardo dei Tintori, Monza, Italy	12	1.1%
Department of Oncology, IRCCS Sacro Cuore "Don Calabria", Negrar, Italy	3	0.3%
Niguarda Cancer Center, Grande Ospedale Metropolitano Niguarda, Milan, Italy	5	0.5%
Northumbria Healthcare NHS Foundation Trust, Cramlington, United Kingdom	14	1.3%
University of Piemonte Orientale, Novara, Italy	14	1.3%
University of Turin, San Luigi Hospital, Orbassano, Italy	31	2.9%
The Ohio State University Comprehensive Cancer Center, Columbus, OH, United States	21	2.0%
University Hospital of Parma, Parma, Italy	20	1.9%
Istituto Nazionale Tumori, IRCCS "Fondazione G. Pascale" Naples, Italy	10	0.9%
Ospedale P. Pederzoli, Peschiera del Garda (VR), Italy	5	0.5%
Fondazione IRCCS Policlinico San Matteo, Pavia, Italy	13	1.2%
AUSL della Romagna, Ravenna, Italy	9	0.8%
Erasmus MC Cancer Institute, University Medical Center, Rotterdam, the Netherlands	36	3.4%
Rush University Medical Center Chicago IL, United States	9	0.8%
Azienda Ospedaliero Universitaria Sant'Andrea, Rome, Italy	22	2.1%
Hospital de la Santa Creu I Sant Pau, Barcelona, Spain	10	0.9%
Stanford Cancer Institute, Stanford University, Palo Alto, CA, United States	16	1.5%
"G. Mazzini" Hospital of Teramo, Teramo, Italy	8	0.8%
Azienda Ospedaliera Santa Maria of Terni, Terni, Italy	6	0.6%
The UCL Cancer Institute, University College London Hospitals NHS Trust, London, United Kingdom	18	1.7%
Cliniques Universitaires St-Luc, UCLouvain, Brussels, Belgium	33	3.1%
Azienda Sanitaria Universitaria Friuli Centrale (ASUFC), Udine, Italy	17	1.6%
"Sapienza" University of Rome, Rome, Italy	4	0.4%
AUSL Latina (Aprilia) - University of Rome "Sapienza", Italy	3	0.3%
ASST Sette Laghi, Ospedale di Circolo e Fondazione Macchi, Varese, Italy	18	1.7%
Versilia Hospital, Azienda USL Toscana Nord Ovest, Lido di Camaiore, Italy	6	0.6%
Yale School of Medicine, New Haven, Connecticut, United States	84	7.9%
Roswell Park Comprehensive Cancer Center, Buffalo, NY, United States	33	3.1%
Total	1063	100.0%

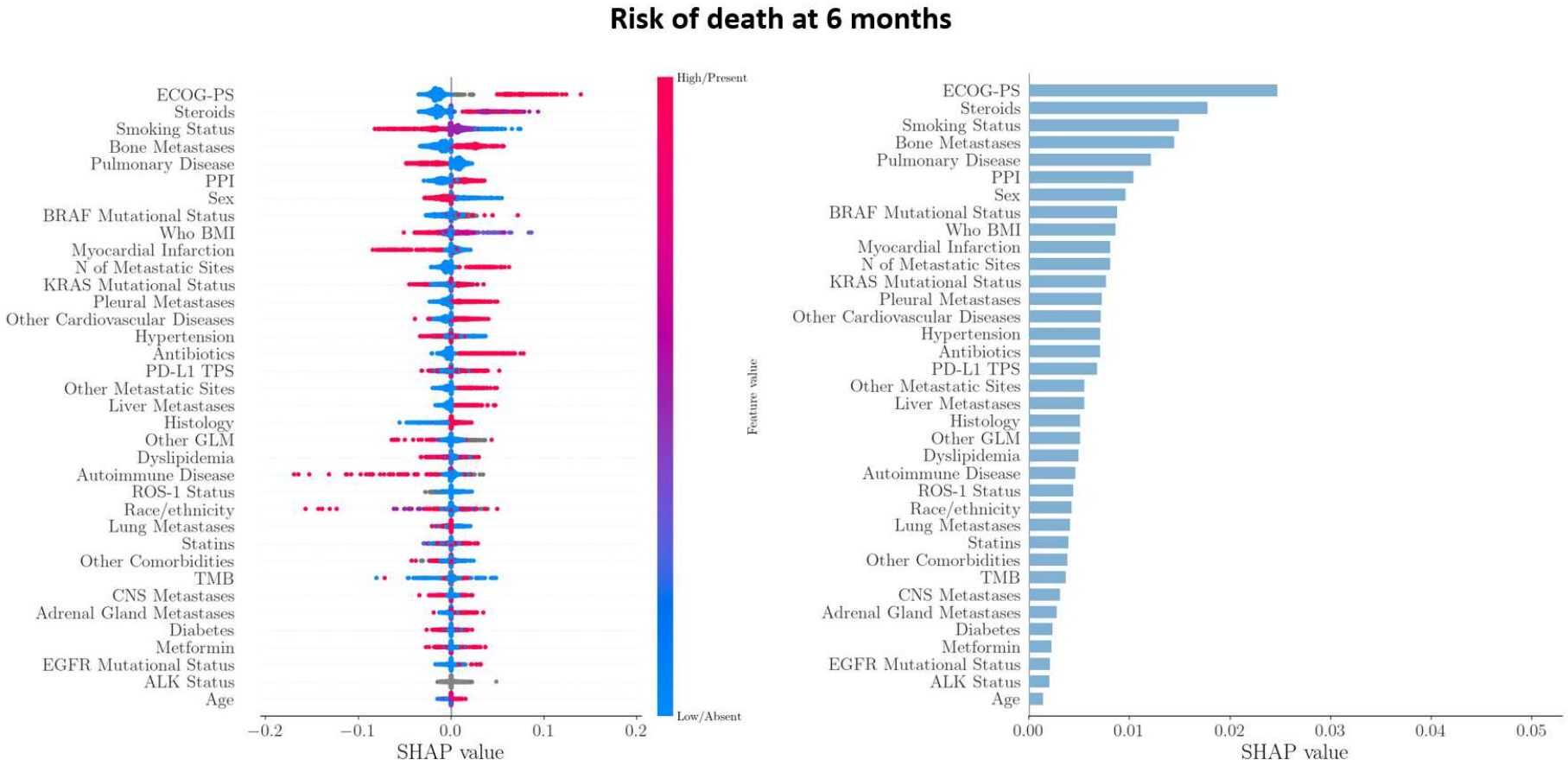
Supplementary Figure 1: Histogram plot presenting the ridge regression coefficients for each variable from the multivariable analysis of the risk of death (overall survival). Coefficients are arranged in descending order by value; positive coefficients (bars extending to the right of the 0 reference point) indicate an increased risk of death, while negative coefficients (bars extending to the left) indicate a decreased risk of death. Bars representing variables with a statistically significant impact on the outcome (hazard ratios with 95% confidence intervals not crossing unity) are highlighted in red, whereas those without significant impact are shown in blue.



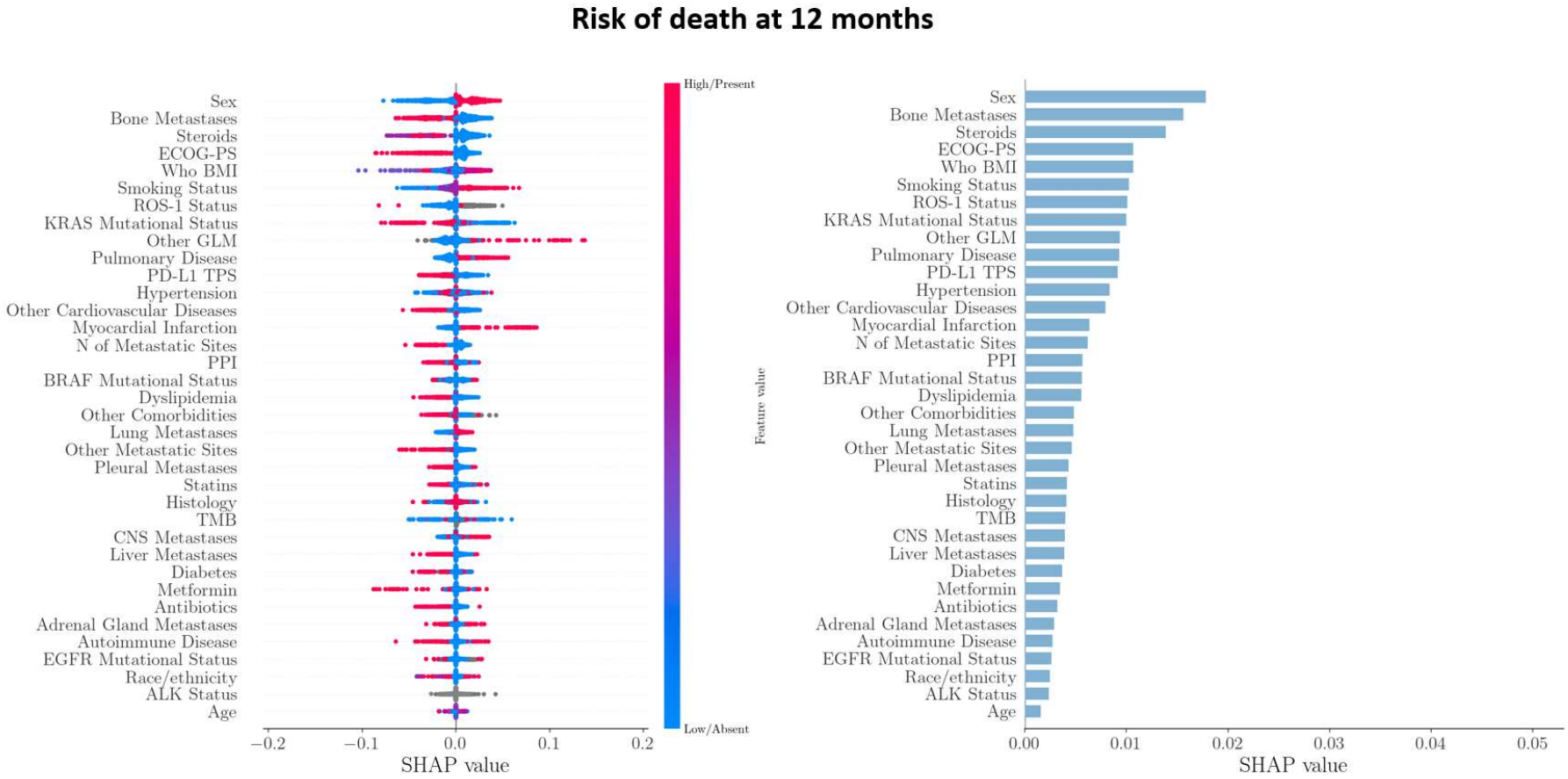
Supplementary Figure 2: Histogram plot presenting the ridge regression coefficients for each variable from the multivariable analysis of the 5 year survival state. Coefficients are arranged in descending order by value; positive coefficients (bars extending to the right of the 0 reference point) indicate an increased risk of death, while negative coefficients (bars extending to the left) indicate a decreased risk of death. Bars representing variables with a statistically significant impact on the outcome (hazard ratios with 95% confidence intervals not crossing unity) are highlighted in red, whereas those without significant impact are shown in blue.



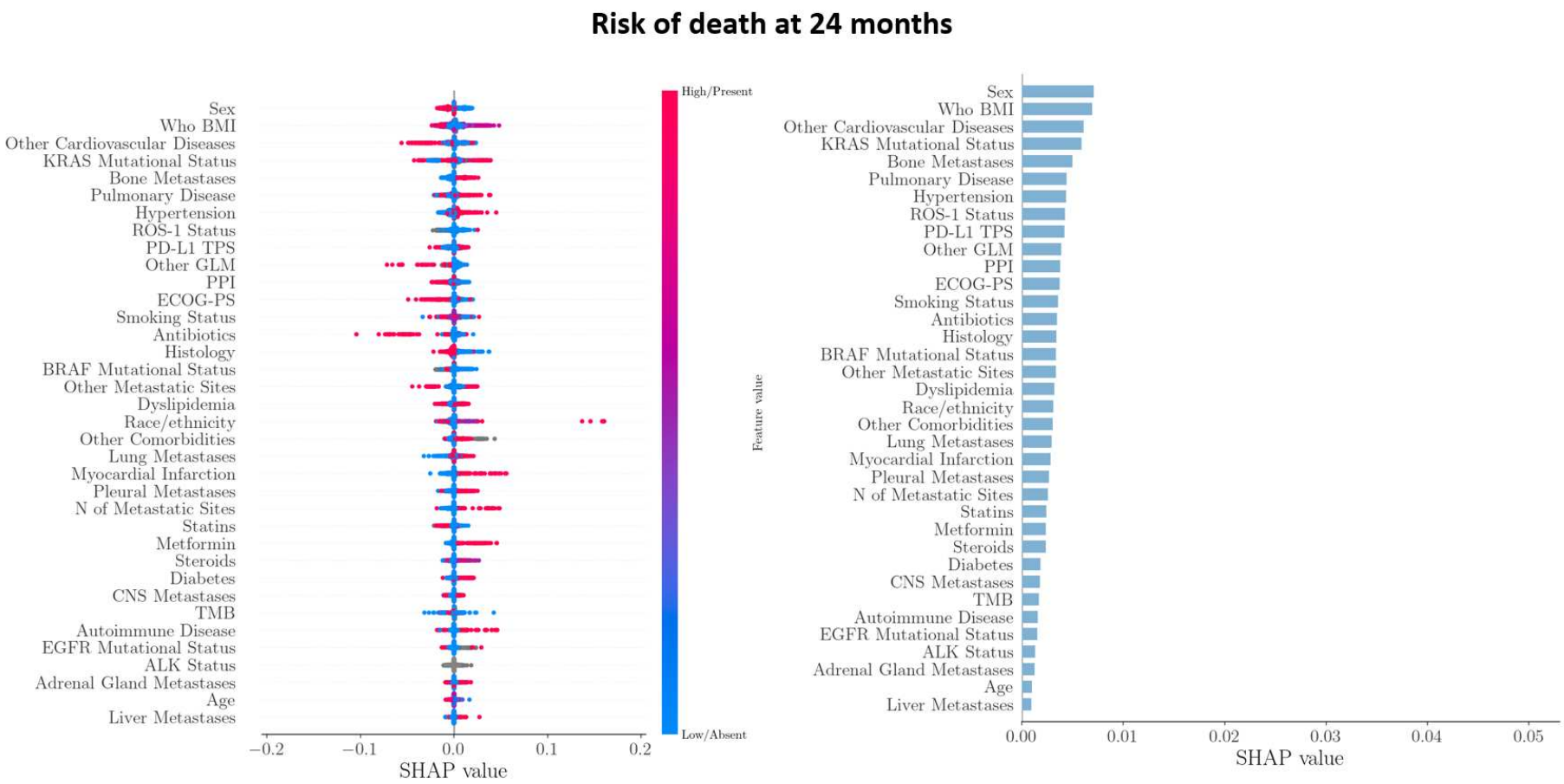
Supplementary Figure 3: Paired dot plot and histogram plot summarizing the SHAP values from the NAIM analysis for the risk of death at 6 months. Features were ordered by their absolute contribution, with high values (red) and low values (blue) positioned to indicate their influence on outcomes. For instance, red dots on the right side of the plot imply a positive association with the risk of death, while blue dots on the right side a negative association. Missing values were represented as gray dots. The length of each bar represents the SHAP value, indicating the relative importance of each variable within the model. Variable's definition and categorization details are reported in supplementary methods. Abbreviations: ECOG-PS, Eastern Cooperative Oncology Group Performance Status; WHO BMI, World Health Organization Body Mass Index levels; NOS, Not Otherwise Specified; PD-L1, Programmed Death-Ligand 1; TPS, Tumor Proportion Score; CNS, Central Nervous System; pred, Prednisone; EGFR, Epidermal Growth Factor Receptor; ALK, Anaplastic Lymphoma Kinase; KRAS, Kirsten Rat Sarcoma Virus; PPI, Proton Pump Inhibitors; GLM, Glucose Lowering Medications; TMB, Tumor Mutational Burden.



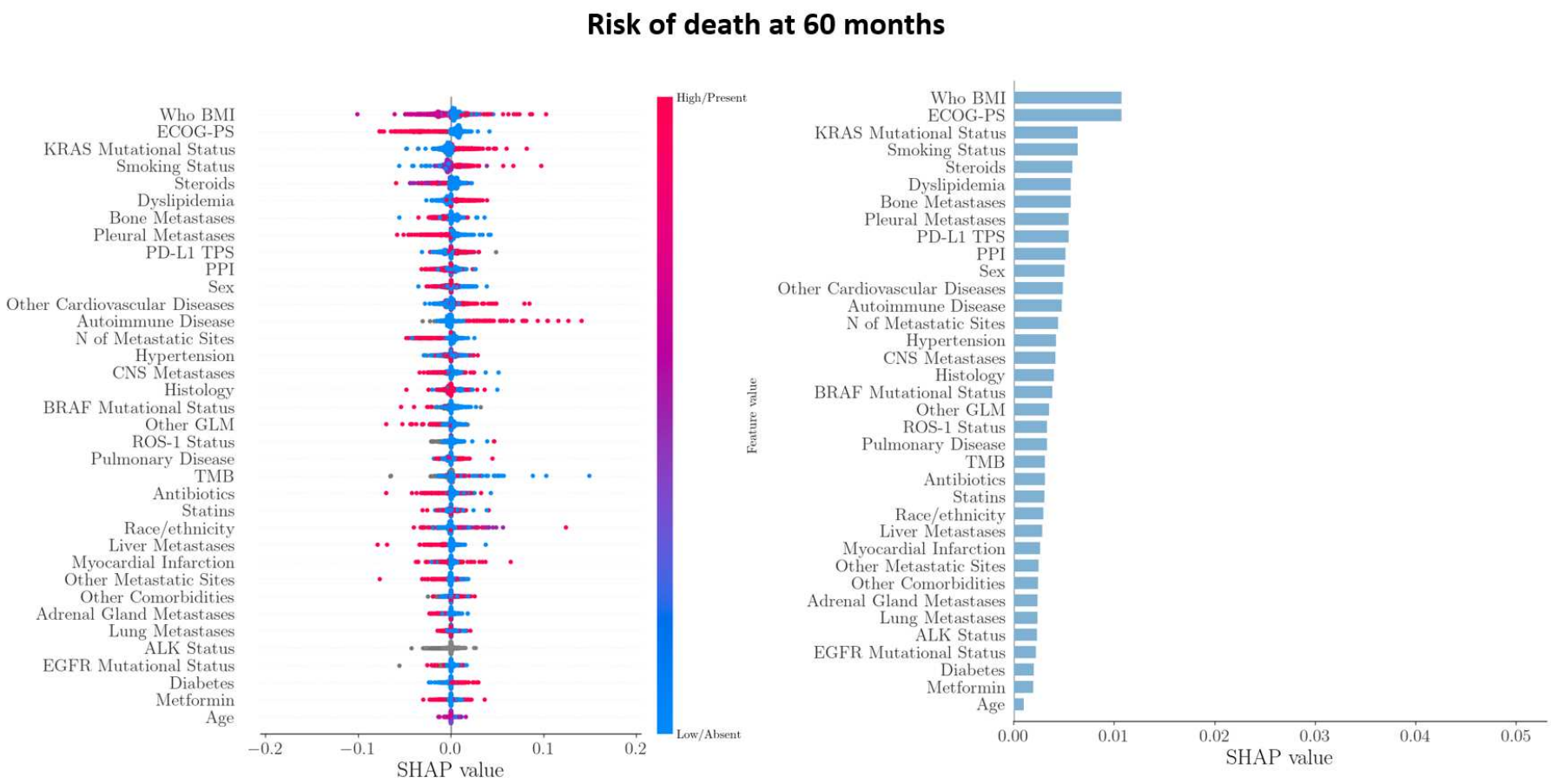
Supplementary Figure 4: Paired dot plot and histogram plot summarizing the SHAP values from the NAIM analysis for the risk of death at 12 months. Features were ordered by their absolute contribution, with high values (red) and low values (blue) positioned to indicate their influence on outcomes. For instance, red dots on the right side of the plot imply a positive association with the risk of death, while blue dots on the right side a negative association. Missing values were represented as gray dots. The length of each bar represents the SHAP value, indicating the relative importance of each variable within the model. Variable's definition and categorization details are reported in supplementary methods. Abbreviations: ECOG-PS, Eastern Cooperative Oncology Group Performance Status; WHO BMI, World Health Organization Body Mass Index levels; NOS, Not Otherwise Specified; PD-L1, Programmed Death-Ligand 1; TPS, Tumor Proportion Score; CNS, Central Nervous System; pred, Prednisone; EGFR, Epidermal Growth Factor Receptor; ALK, Anaplastic Lymphoma Kinase; KRAS, Kirsten Rat Sarcoma Virus; PPI, Proton Pump Inhibitors; GLM, Glucose Lowering Medications; TMB, Tumor Mutational Burden.



Supplementary Figure 5: Paired dot plot and histogram plot summarizing the SHAP values from the NAIM analysis for the risk of death at 24 months. Features were ordered by their absolute contribution, with high values (red) and low values (blue) positioned to indicate their influence on outcomes. For instance, red dots on the right side of the plot imply a positive association with the risk of death, while blue dots on the right side a negative association. Missing values were represented as gray dots. The length of each bar represents the SHAP value, indicating the relative importance of each variable within the model. Variable's definition and categorization details are reported in supplementary methods. Abbreviations: ECOG-PS, Eastern Cooperative Oncology Group Performance Status; WHO BMI, World Health Organization Body Mass Index levels; NOS, Not Otherwise Specified; PD-L1, Programmed Death-Ligand 1; TPS, Tumor Proportion Score; CNS, Central Nervous System; pred, Prednisone; EGFR, Epidermal Growth Factor Receptor; ALK, Anaplastic Lymphoma Kinase; KRAS, Kirsten Rat Sarcoma Virus; PPI, Proton Pump Inhibitors; GLM, Glucose Lowering Medications; TMB, Tumor Mutational Burden.



Supplementary Figure 6: Paired dot plot and histogram plot summarizing the SHAP values from the NAIM analysis for the risk of death at 60 months. Features were ordered by their absolute contribution, with high values (red) and low values (blue) positioned to indicate their influence on outcomes. For instance, red dots on the right side of the plot imply a positive association with the risk of death, while blue dots on the right side a negative association. Missing values were represented as gray dots. The length of each bar represents the SHAP value, indicating the relative importance of each variable within the model. Variable's definition and categorization details are reported in supplementary methods. Abbreviations: ECOG-PS, Eastern Cooperative Oncology Group Performance Status; WHO BMI, World Health Organization Body Mass Index levels; NOS, Not Otherwise Specified; PD-L1, Programmed Death-Ligand 1; TPS, Tumor Proportion Score; CNS, Central Nervous System; pred, Prednisone; EGFR, Epidermal Growth Factor Receptor; ALK, Anaplastic Lymphoma Kinase; KRAS, Kirsten Rat Sarcoma Virus; PPI, Proton Pump Inhibitors; GLM, Glucose Lowering Medications; TMB, Tumor Mutational Burden.


Risk of death at 60 months
Who BMI
ECOG-PS
KRAS Mutational Status
Smoking Status
Steroids
Dyslipidemia
Bone Metastases
Pleural Metastases
PD-L1 TPS
PPI
Sex
Other Cardiovascular Diseases
Autoimmune Disease
N of Metastatic Sites
Hypertension
CNS Metastases
Histology
BRAF Mutational Status
Other GLM
ROS-1 Status
Pulmonary Disease
TMB
Antibiotics
Statins
Race/ethnicity
Liver Metastases
Myocardial Infarction
Other Metastatic Sites
Other Comorbidities
Adrenal Gland Metastases
Lung Metastases
ALK Status
EGFR Mutational Status
Diabetes
Metformin
Age
−0.2
−0.1
0.0
0.1
0.2
SHAP value
High/Present
Feature value
Low/Absent
0.00
0.01
0.02
0.03
0.04
0.05
SHAP value