

Independent burden and functional status trajectories define prognostic phenotypes in stage IV NSCLC

Alon Pomerantz

Sheba Medical Center

Asaf Zlotnik

Sheba Medical Center

Roni Gillis

Shaare Zedek Medical Center

Reham Basheer

Shaare Zedek Medical Center

Ido David Dechtman

Sheba Medical Center

Noam Asna

Shaare Zedek Medical Center

Liran Levy

Sheba Medical Center

Nir Peled

Shaare Zedek Medical Center

Adar Yaacov

`adar.yaacov@mail.huji.ac.il`

Shaare Zedek Medical Center

Research Article

Keywords: Non-small cell lung cancer (NSCLC), Metastatic burden, ECOG performance status, Natural language processing (NLP), Trajectory analysis, Real-world data, Host fitness, Prognostic modeling

Posted Date: May 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9804101/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Prognostic assessment in advanced non-small cell lung cancer (NSCLC) relies on static measures of metastatic burden and ECOG performance status. Functional decline is assumed to be correlated with disease progression, implying these trajectories may be partly redundant. Whether they provide independent prognostic information has not been tested at scale.

Methods

Using the MSK-CHORD real-world dataset, which integrates longitudinal tumor burden and ECOG status, we identified 1,486 stage IV NSCLC patients with both trajectories at a pre-specified 6-month landmark. Trajectory worsening was tested for statistical independence (chi-square, equivalence testing). Additive prognostic value was assessed using cross-validated Cox models. Four concordance-based risk groups were compared by Kaplan-Meier and adjusted Cox analyses.

Results

Burden and ECOG trajectory worsening were statistically independent ($\phi = -0.033$; TOST $P = 0.005$; $BF_{01} = 17.3$), with 47.2% discordant. Each trajectory contributed independent prognostic information ($\Delta C \approx 0.02$; combined $\Delta C = +0.054$). Concordance defined four groups with median post-landmark survival from 16.2 months (both worsened; adjusted HR 2.16, $P < 0.001$) to 32.0 months (both stable). Independence held across landmarks and clinical subgroups.

Conclusions

Burden and functional trajectories capture distinct disease dimensions; their discordance might distinguish reversible functional decline from concordant progression, informing treatment versus supportive-care decisions.

Introduction

Stage IV non-small cell lung cancer (NSCLC) accounts for the majority of lung cancer diagnoses and carries a five-year survival below 10% despite meaningful advances in targeted therapy and immune checkpoint inhibition.^{1,2} Prognostic assessment in this setting continues to rely on variables measured at a single timepoint, applying a static framework to a disease defined by continuous biological and clinical evolution.^{3,4}

Two measures anchor prognostic thinking in advanced NSCLC: ECOG performance status and metastatic burden. An ECOG score of two or above is a widely recognized threshold that restricts therapeutic options and independently predicts inferior survival across treatment modalities, including in the era of immune checkpoint inhibition.⁴⁻⁷ In parallel, three decades of work on the oligometastatic paradigm have established disease burden and distribution as an independent prognostic axis, with patients carrying limited metastatic involvement benefiting from aggressive local consolidative strategies in randomized trials.⁸⁻¹⁰ Both constructs, however, are operationalized as static snapshots and provide no information about how the tumor or the patient is changing over time.

Embedded in current practice is an untested causal assumption: that functional decline in advanced cancer is principally a downstream consequence of tumor progression, such that ECOG trajectory might be redundant with burden trajectory.¹¹ Under this redundancy, the chain burden $\uparrow \rightarrow$ ECOG $\downarrow \rightarrow$ death implies that monitoring either dimension alone would suffice. If false, the two trajectories might reflect distinct axes, tumor kinetics on one hand and host fitness on the other. The latter integrates treatment toxicity, comorbidity burden, and physiological reserve alongside any tumor-related decline. Their concordance or discordance would then define phenotypically distinct patient groups requiring different clinical responses. This question has not been tested at scale, and its resolution has direct consequences for how serial assessments are collected, how prognostic models are built, and how treatment versus supportive care decisions are made.^{12,13}

We used real-world dataset to test directly whether metastatic burden trajectory and ECOG performance status trajectory are independent or redundant prognostic dimensions in stage IV NSCLC. We hypothesized that these trajectories capture largely non-overlapping dimensions of disease evolution and that their concordance or discordance defines clinically distinct phenotypes with differential implications for treatment adaptation and supportive care.

Methods

Data Source and Study Population

This study used data from the MSK-CHORD dataset, a comprehensive real-world oncology data resource from Memorial Sloan Kettering Cancer Center^{14,15}. MSK-CHORD integrates genomic profiling (MSK-IMPACT), NLP-derived clinical phenotypes from electronic health records, and structured clinical data across more than 25,000 patients with cancer. We identified NSCLC patients using ONCOTREE codes (LUAD, LUSC, NSCLC, LUNE, LUNG, LUAS, LUMEC, LUACC, LUPC, LECLC). The metastatic burden analysis was restricted to Stage IV patients (N = 3,133 with ≥ 2 tumor site assessments and complete survival data) (Fig. 1a). The ECOG analysis included all NSCLC patients with ≥ 2 ECOG assessments (N = 5,438). The combined analysis included the 1,486 patients present in 6-month landmark cohorts.

NLP-Derived Clinical Phenotypes

Metastatic burden: NLP algorithms extracted tumor site information from radiology reports at each imaging timepoint, identifying involvement across anatomical sites (lung, liver, bone, CNS/brain, lymph nodes, adrenal, pleura, and others). Metastatic burden was defined as the count of distinct tumor sites. Oligometastatic was defined as ≤ 3 sites and polymetastatic as ≥ 4 sites^{8,9,16}. ECOG performance status: ECOG scores were extracted from clinical notes.^{14,17}

Landmark Analysis Framework and Trajectory Feature Definitions

To prevent immortal time bias, we implemented a landmark analysis framework at 3, 6, and 12 months from the first clinical assessment (Fig. 1b). The 6-month landmark was pre-specified as the primary analysis. At each landmark: (1) patients must have ≥ 2 assessments before the landmark and be alive at the landmark; (2) trajectory features were computed using only pre-landmark data; (3) overall survival was measured from the landmark forward.

Within this framework, we defined two categories of trajectory features computed up to the landmark:

Burden trajectory features: baseline burden (first assessment), burden change (most recent – baseline), burden slope (linear regression of site count over time), any increase (binary), oligo-to-poly transition (binary), trajectory classification (Decreasing, Stable, Variable, Increasing).

ECOG trajectory features: baseline ECOG, ECOG slope (linear regression, capped at $[-1, 1]$ points/month), maximum ECOG, had deterioration (any increase above baseline), reached ECOG ≥ 2 (among baseline 0–1), trajectory classification (Stable, Declining, Improving, Variable).

Statistical Analysis

Independence testing. The primary independence test was a chi-square test of the 2×2 contingency table. We report the phi coefficient with Fisher z-transform 95% confidence intervals. Because a non-significant chi-square test does not establish equivalence, we additionally performed two one-sided tests (TOST) equivalence procedure¹⁸ with an equivalence margin of $|\phi| < 0.10$, which represents the conventional threshold for a negligible effect size. The standard error of phi was estimated as $1/\sqrt{(N - 3)}$. We also computed a Bayes factor (BF_{01}) using the BIC approximation ($BF_{01} \approx \exp [0.5 \times (\ln N - \chi^2)]$), where $BF_{01} > 1$ favors independence. We additionally computed Spearman and Kendall rank correlations between continuous trajectory slopes. Independence was assessed at each landmark and within clinical subgroups.

Survival analysis. Kaplan-Meier estimates with log-rank tests were used for survival comparisons. Cox proportional hazards models with L2 regularization (penalizer = 0.01) were used for multivariable analysis. A modest L2 penalty ($\lambda = 0.01$) was applied to stabilize coefficient estimates in models with up to 14 covariates relative to the effective sample size; results were insensitive to the penalty value across $\lambda \in [0.001, 0.1]$. Multivariable models adjusted for age, sex, smoking status, and key driver mutations

(EGFR, KRAS, TP53, STK11, KEAP1). Pairwise log-rank tests and Cox HRs with 95% CIs were computed for all six group pairs.

Model discrimination. As a sensitivity analysis, models were additionally adjusted for first-line treatment class (chemo-immunotherapy, targeted therapy, chemotherapy, immunotherapy, other). Concordance indices were computed via 10-fold stratified cross-validation, repeated 3 times (30 total folds). Progressive models were built on the combined cohort: (1) demographics only, (2) + baseline ECOG and burden, (3) + trajectory features only (no baselines), (4) + baseline + trajectory. The symmetric ΔC increments (ECOG added over burden, and vice versa) tested additive independence. The statistical significance of ΔC between nested models was assessed using 1,000 bootstrap resamples (percentile 95% CI; one-sided $P = \text{proportion of resamples with } \Delta C \leq 0$) and likelihood ratio tests (LRT) on unpenalized Cox models.

Risk group characterization. Patients were classified into four groups based on concordant vs. discordant trajectory worsening. Hazard ratios for each group versus both-stable (reference) were computed both unadjusted and adjusted for baseline ECOG, baseline burden, age, sex, and smoking status, to control for confounding by baseline clinical characteristics. Pairwise log-rank P -values were corrected for multiple testing using the Holm-Bonferroni method (6 comparisons). Group differences in demographics, mutations, and treatment patterns were tested using chi-square tests (categorical) and ANOVA (continuous). ECOG recovery was assessed by comparing post-landmark ECOG scores to pre-landmark maximum.

All analyses were performed in Python (v3.12) using lifelines (v0.30) for survival analysis, scikit-learn (v1.7) for cross-validation, and SciPy (v1.15) for statistical tests.

Results

Study Population

The study population was drawn from MSK-CHORD NSCLC patients (Fig. 1a). The metastatic burden cohort comprised 3,133 Stage IV NSCLC patients with longitudinal NLP-derived tumor site data (median 9 imaging timepoints per patient, IQR 5–15). The ECOG trajectory cohort comprised 5,438 NSCLC patients with longitudinal ECOG assessments (mean 7.9 assessments per patient). At the 6-month landmark, 2,387 patients were evaluable in the burden cohort and 3,026 in the ECOG cohort. The combined analysis included 1,486 patients with both burden and ECOG trajectory data at the 6-month landmark (Fig. 1a-b). The median age of the population was 68.0 years (IQR, 60.0–75.0), with a predominance of female patients (60.0%) and adenocarcinoma histology (83.5%) (Table 1). The cohort included a high prevalence of *EGFR* (33.5%) and *KRAS* (23.3%) mutations. At baseline, most patients presented with favorable functional status (94.0% ECOG 0–1), oligometastatic disease burden (78.5%; median 2 sites) and first-line treatment patterns reflected contemporary practice, consisting primarily of chemo-immunotherapy (36.4%) or targeted therapy (33.2%).

**Table 1. Baseline Patient Characteristics by Trajectory Risk Group
(6-Month Landmark, N = 1,486)**

Characteristic	Overall (N=1,486)	P	Both Stable (N=151)	Burden Only (N=457)	ECOG Only (N=244)	Both Worsened (N=634)
Demographics						
Age, median (IQR)	68.0 (60.0–75.0)	0.090	68.0 (60.0–74.0)	68.0 (60.0–75.0)	69.0 (62.0–76.0)	67.0 (60.0–74.0)
Female sex, n (%)	892 (60.0)	0.533	91 (60.3)	277 (60.6)	155 (63.5)	369 (58.2)
Never smoker, n (%)	532 (35.8)	0.028	65 (43.0)	176 (38.5)	73 (29.9)	218 (34.4)
Histology						
Adenocarcinoma, n (%)	1241 (83.5)	0.604	128 (84.8)	389 (85.1)	201 (82.4)	523 (82.5)
Squamous cell, n (%)	106 (7.1)	—	11 (7.3)	29 (6.3)	20 (8.2)	46 (7.3)
NSCLC NOS, n (%)	102 (6.9)	—	9 (6.0)	28 (6.1)	13 (5.3)	52 (8.2)
Other, n (%)	37 (2.5)	—	3 (2.0)	11 (2.4)	10 (4.1)	13 (2.1)
Baseline Clinical Features						
Baseline ECOG, median (IQR)	1 (0–1)	<0.001	1 (1–1)	1 (1–1)	0 (0–0)	0 (0–0)
ECOG 0, n (%)	686 (46.2)	—	0 (0.0)	0 (0.0)	206 (84.4)	480 (75.7)
ECOG 1, n (%)	710 (47.8)	—	135 (89.4)	395 (86.4)	36 (14.8)	144 (22.7)
ECOG ≥2, n (%)	90 (6.1)	—	16 (10.6)	62 (13.6)	2 (0.8)	10 (1.6)
Baseline metastatic sites, median (IQR)	2 (1–3)	<0.001	3 (2–4)	2 (1–3)	2 (2–3)	2 (1–3)
Oligometastatic at baseline, n (%)	1167 (78.5)	<0.001	100 (66.2)	365 (79.9)	207 (84.8)	495 (78.1)
Imaging assessments to landmark, median (IQR)	5 (4–6)	<0.001	3 (3–4)	5 (4–6)	3 (3–4)	5 (4–6)
Genomic Alterations						
EGFR mutation, n (%)	498 (33.5)	0.452	58 (38.4)	158 (34.6)	79 (32.4)	203 (32.0)
KRAS mutation, n (%)	346 (23.3)	0.258	26 (17.2)	104 (22.8)	60 (24.6)	156 (24.6)
TP53 mutation, n (%)	852 (57.3)	0.281	77 (51.0)	274 (60.0)	138 (56.6)	363 (57.3)
STK11 mutation, n (%)	149 (10.0)	0.588	13 (8.6)	44 (9.6)	21 (8.6)	71 (11.2)
KEAP1 mutation, n (%)	158 (10.6)	0.110	8 (5.3)	46 (10.1)	29 (11.9)	75 (11.8)
ALK fusion, n (%)	62 (4.2)	0.261	10 (6.6)	22 (4.8)	8 (3.3)	22 (3.5)
ROS1 fusion, n (%)	29 (2.0)	0.986	3 (2.0)	8 (1.8)	5 (2.0)	13 (2.1)
RET fusion, n (%)	35 (2.4)	0.123	4 (2.6)	15 (3.3)	1 (0.4)	15 (2.4)
BRAF mutation, n (%)	70 (4.7)	0.630	7 (4.6)	18 (3.9)	15 (6.1)	30 (4.7)
Treatment (any line)						
Received immunotherapy, n (%)	737 (49.6)	0.039	62 (41.1)	231 (50.5)	111 (45.5)	333 (52.5)
Received targeted therapy, n (%)	542 (36.5)	0.248	58 (38.4)	182 (39.8)	85 (34.8)	217 (34.2)
Received chemotherapy, n (%)	847 (57.0)	0.607	84 (55.6)	250 (54.7)	142 (58.2)	371 (58.5)
First-line Treatment						
Chemo-IO, n (%)	541 (36.4)	<0.001	44 (29.1)	167 (36.5)	68 (27.9)	262 (41.3)
Targeted, n (%)	494 (33.2)	—	53 (35.1)	169 (37.0)	73 (29.9)	199 (31.4)
Chemo, n (%)	241 (16.2)	—	32 (21.2)	57 (12.5)	60 (24.6)	92 (14.5)
ICI, n (%)	90 (6.1)	—	6 (4.0)	33 (7.2)	12 (4.9)	39 (6.2)
Other, n (%)	98 (6.6)	—	14 (9.3)	29 (6.3)	23 (9.4)	32 (5.0)
Outcomes (from 6-month landmark)						
Median OS, months (95% CI)	21.5 (19.9–23.4)	—	32.0 (25.7–37.6)	21.2 (18.3–23.7)	38.0 (29.0–45.0)	16.2 (14.2–18.8)
Deaths, n (%)	958 (64.5)	—	90 (59.6)	296 (64.8)	135 (55.3)	437 (68.9)

Continuous variables: median (IQR), Kruskal-Wallis test. Categorical variables: n (%), chi-square test. OS: Kaplan-Meier median (95% CI). Significant P-values (< 0.05) in bold red.

Statistical Independence of Burden and Functional Trajectories

To determine whether functional decline is a downstream consequence of tumor progression, we tested the statistical independence of burden and ECOG trajectories in the combined cohort (N = 1,486). Observed cell counts for burden and ECOG worsening closely matched those expected under independence (Fig. 2a). The chi-square test was non-significant ($\chi^2 = 1.61$, df = 1, P = 0.20), and the phi (ϕ) coefficient was –0.033 (95% CI –0.084 to 0.018), indicating negligible correlation well below conventional thresholds for even weak association ($|\phi| < 0.1$). Because a non-significant chi-square test does not prove independence, we performed a two one-sided tests (TOST) equivalence procedure, which confirmed that the association is bounded within a negligible range (P = 0.005 for equivalence margin $|\phi|$

< 0.10). A Bayes factor analysis using the BIC approximation yielded $BF_{01} = 17.3$, providing strong evidence in favor of independence over association¹⁹. This lack of association resulted in high discordance: nearly half (47.2%) of patients had discordant trajectories, where 457 (30.8%) had burden worsening without ECOG decline, and 244 (16.4%) had ECOG worsening without burden increase (Fig. 2a). Continuous trajectory measures were consistent with independence, showing near-zero correlation between burden slope and ECOG slope ($\rho = 0.008$, $P = 0.75$) (Fig. 2b).

Prognostic Value of Individual and Combined Trajectories

Consistent with their statistical independence, each trajectory proved to be an independent prognostic factor. Among 2,387 Stage IV patients at the 6-month landmark, burden trajectory features improved cross-validated discrimination beyond static baseline measures ($\Delta C = +0.033$) (Supplementary Figure S1). Oligo-to-polymetastatic transition by 6 months identified a high-risk subgroup (median post-landmark OS 15.2 vs. 29.5 months, log-rank $P < 0.001$) (Fig. 3a). Any burden increase was independently prognostic (HR 1.43, 95% CI 1.26–1.62, $P = 1.5 \times 10^{-8}$). Similarly, in the ECOG cohort ($N = 3,026$), trajectory features improved discrimination beyond static baseline ($\Delta C = +0.032$) (Supplementary Figure S2). ECOG slope was an independent predictor (HR 1.36, 95% CI 1.18–1.57, $P = 2.3 \times 10^{-5}$) and patients reaching ECOG ≥ 2 ($N = 428$; 15.0% of 2,860 with baseline 0–1) had markedly worse outcomes (median 14.4 vs. 36.5 months, $P = 2.3 \times 10^{-25}$; Fig. 3b).

To quantify the additive prognostic utility of these independent trajectories, we constructed progressive Cox models in the combined cohort (Fig. 4a). Adding ECOG trajectory to the burden-augmented model improved discrimination by $\Delta C = +0.022$ (bootstrap 95% CI 0.011–0.035) and adding burden trajectory to the ECOG-augmented model improved discrimination by $\Delta C = +0.021$ (bootstrap 95% CI 0.010–0.033). These symmetric, nearly equal increments demonstrate that each trajectory captures prognostic information not contained in the other. Consequently, the combined ΔC of +0.054 is approximately the sum of the individual contributions. Furthermore, a model with trajectory features alone (without baseline ECOG or burden) substantially outperformed the baseline-only model $\Delta C = +0.048$, bootstrap 95% CI 0.025–0.064, $P < 0.001$, demonstrating that the dynamic information in trajectories is not merely a recapitulation of baseline status (Fig. 4b). In the combined Cox model, the strongest trajectory predictors were any burden increase (HR 1.43), oligo-to-poly transition (HR 1.38), and ECOG slope (HR 1.36), all of which remained significant after mutual adjustment (Supplementary Table S1).

Trajectory Discordance Defines Distinct Clinical Phenotypes

Based on trajectory discordance, patients were classified into four distinct risk groups (Fig. 5a-b). After adjustment for baseline characteristics (ECOG, burden, age, sex, smoking), a clear risk hierarchy emerged: Both Stable (reference), ECOG Worsened Only (adjusted HR 1.25, 95% CI 0.95–1.65, $P = 0.12$), Burden Worsened Only (adjusted HR 1.32, 95% CI 1.06–1.66, $P = 0.014$), and Both Worsened (adjusted

HR 2.16, 95% CI 1.71–2.74, $P < 0.001$) (Fig. 5a-b). Isolated ECOG worsening had unadjusted HR 0.84. This was driven by a baseline imbalance, resolved after adjustment for baseline ECOG (adjusted HR 1.25, 95% CI 0.95–1.65). Despite this adjustment, among 184 patients with post-landmark ECOG assessments, 157 (85.3%) showed subsequent ECOG recovery (Table 2). This group also had the highest rate of chemotherapy receipt (24.6%), consistent with the hypothesis that ECOG worsening in this group might reflect transient, reversible causes such as treatment toxicity rather than irreversible disease progression. In contrast, concordant worsening (Both Worsened) identified the highest-risk patients, with a median survival of 16.2 months.

Risk group characterization further supported these distinctions. While age differed modestly across groups ($F = 2.71$, $P = 0.044$), driver mutation frequencies did not differ significantly ($P > 0.10$), indicating that these phenotypes are not confounded by genomic subtype (Supplementary Fig. S3a). Baseline burden was significantly higher in the Both Stable group (mean 2.9 sites, $P < 0.001$), reflecting the finding that patients with higher starting burden who remain stable have favorable outcomes (Supplementary Fig. S3b). Finally, baseline ECOG showed the strongest inter-group difference ($F = 506.8$, $P < 0.001$), confirming that trajectory discordance is largely driven by baseline status: ECOG-worsened groups began with near-normal function (means 0.16–0.26), whereas stable groups had established impairment (means 1.11–1.17).

Table 2
ECOG-Only Group Characterization (N = 244)

Characteristic	Value
Baseline ECOG 0	206 (84.4%)
Baseline ECOG 1	36 (14.8%)
Baseline ECOG 2	2 (0.8%)
Current ECOG at landmark: 0	78 (32.0%)
Current ECOG at landmark: 1	143 (58.6%)
Current ECOG at landmark: ≥2	23 (9.4%)
ECOG recovery post-landmark*	157/184 (85.3%)
Treatment pattern: chemotherapy	60 (24.6%)
Treatment pattern: targeted	73 (29.9%)

*Among those with post-landmark ECOG assessments, percentage whose ECOG improved below their pre-landmark maximum.

Sensitivity Analyses

Findings were robust across landmark times, clinical subgroups, and definitions. The independence of burden and ECOG trajectories was consistent at 3-, 6-, and 12-month landmarks ($\phi = -0.025$, -0.033 , and -0.022 respectively; all $|\phi| < 0.035$, TOST $P < 0.01$), with the combined prognostic value (ΔC) increasing with longer observation windows (Supplementary Figure S4a). Subgroup analyses confirmed independence across genomic and treatment subgroups: EGFR-mutant ($\phi = -0.013$, $P = 0.77$, $N = 498$), KRAS-mutant ($\phi = -0.087$, $P = 0.11$, $N = 346$), EGFR/KRAS wild-type ($\phi = -0.024$, $P = 0.54$, $N = 652$), first-line ICI ($\phi = -0.101$, $P = 0.34$, $N = 90$), first-line targeted therapy ($\phi = -0.034$, $P = 0.45$, $N = 494$), and first-line chemotherapy ($\phi = -0.035$, $P = 0.59$, $N = 241$), with discordance rates remaining approximately 47–50% and $|\phi|$ remained below 0.10 in all groups (Supplementary Figure S4b). Finally, the high base rate of burden worsening (73.4% using any-increase threshold) raised the possibility that the definition was overly sensitive. Independence held with stricter thresholds: oligo-to-polymetastatic transition only ($\phi = -0.038$, $P = 0.14$, 55.9% discordant) and burden change ≥ 2 sites ($\phi = +0.026$, $P = 0.32$, 55.7% discordant). Using the stricter definition of ECOG worsening as reaching ECOG ≥ 2 (rather than any deterioration), 340 patients (22.9%) met the threshold. With this definition, a weak positive association emerged ($\phi = +0.085$, $P = 0.001$), consistent with the expectation that clinically significant functional decline is more likely to co-occur with disease progression than minor ECOG fluctuation. The four-group HRs with this stricter definition were: burden-only 1.53 ($P < 0.001$), ECOG ≥ 2 -only 1.62 ($P = 0.004$), and both-worsened 2.62 ($P < 0.001$), demonstrating approximate additivity (Supplementary Figure S4c). This pattern suggests a dose-response relationship: minor ECOG fluctuations ($0 \rightarrow 1$) are independent of burden trajectory, while clinically significant functional decline (reaching ECOG ≥ 2) shows a modest positive association. To assess whether treatment during the pre-landmark period confounded the trajectory-outcome relationship, we additionally adjusted for first-line treatment class (chemo-immunotherapy [reference, $N = 541$], targeted therapy, chemotherapy, immunotherapy, other). Cross-validated ΔC for trajectories over baseline + treatment was $+0.047$, comparable to the $+0.054$ without treatment adjustment, confirming that the prognostic value is not confounded by treatment selection (Supplementary Figure S4d).

Discussion

This study addresses a fundamental question in NSCLC prognostication: is performance status decline simply a downstream consequence of disease progression, or does it capture an independent dimension of clinical deterioration? We provide three lines of evidence that burden trajectory and ECOG trajectory are largely independent prognostic dimensions in stage IV NSCLC.

First, the two trajectories are statistically independent. The phi coefficient between burden worsening and ECOG worsening was near zero, and a TOST equivalence test confirmed that the association is bounded within a negligible range. The correlation between continuous trajectory slopes was also near zero. Nearly half (47.2%) of patients had discordant trajectories. If ECOG decline (deterioration of performance status) were principally caused by metastatic spread, we would expect a strong positive association and rare discordance.

Second, the two trajectories are additively prognostic. Each adds c-index delta after the other, and the combined improvement is approximately the sum of individual contributions. This additivity is the hallmark of independent predictors²⁰. If the two trajectories measured the same construct, adding the second after the first would provide negligible incremental value. Notably, a model with trajectory features alone (without baseline ECOG or burden) outperformed the static baseline model, demonstrating that the dynamic information in trajectories captures prognostic signal.

Third, discordant trajectories reveal distinct clinical phenotypes. The four-group risk stratification reveals a clear hierarchy after baseline adjustment. The apparent of ECOG-only worsening having better unadjusted survival than both-stable was explained by a floor effect: this group's low baseline ECOG (84.4% at ECOG 0) meant they were inherently healthier at baseline. After adjustment, the expected ordering was restored, though ECOG worsening without burden increase remained a qualitatively different and more favorable event than burden worsening.

This difference has a mechanistic basis. Among ECOG-only patients with post-landmark ECOG (n = 184), 85.3% showed subsequent ECOG recovery. Their functional decline was mostly ECOG 0→1, and they had the highest chemotherapy rate (24.6%). These features are compatible with transient treatment-related toxicity rather than progressive disease, although causal inference is precluded by the retrospective design^{21,22}. Conversely, isolated burden worsening without ECOG decline (median 21.2 months) may represent silent radiographic progression detectable on imaging before functional impairment, a group that may benefit from early treatment adaptation while performance status still permits aggressive intervention.

Our data support a model in which burden trajectory reflects the rate and pattern of metastatic spread, whereas ECOG trajectory reflects host fitness, representing a composite of treatment toxicity, comorbidity burden, physiological reserve, and only partial tumor-related functional decline²³. Because only the last component is shared, the overall correlation is near zero. This distinction has practical implications: isolated burden worsening calls for oncologic treatment modification (systemic therapy change, local therapy). While isolated ECOG worsening should prompt evaluation for reversible causes (for example: toxicity, infection, depression, anemia) and should call for supportive care optimization (dose reduction, rehabilitation, symptom management)¹³.

The prognostic value of baseline ECOG is well-established, and the oligometastatic paradigm has gained traction through randomized trials. However, longitudinal trajectories of these measures have not been previously studied at this scale. Prior work on dynamic risk scores has focused on laboratory markers or tumor response (RECIST)²⁴. Our study extends this concept to NLP-derived clinical phenotypes, demonstrating that automated extraction of trajectory features from clinical records can complement structured outcome assessments.

Several limitations should be acknowledged. This is a single-institution retrospective study from a tertiary referral center. NLP extraction has inherent error; however, validation studies demonstrate high

accuracy. The binary any-ECOG-deterioration variable has baseline-dependent sensitivity by construction and thus subject to a floor effect: patients starting at ECOG 0 are more likely to trigger it from clinically minor changes (ECOG 0→1). This asymmetric detection inflates the apparent ECOG-worsened share among baseline-ECOG-0 patients; we addressed it through baseline-ECOG-adjusted Cox models and through the stricter ECOG ≥ 2 sensitivity analysis (Supp. Figure 1c). The high base rate of burden worsening (73.4%) may reflect overly sensitive detection, though independence held with stricter thresholds (oligo-to-poly, ≥ 2 site increase; all $|p| < 0.05$). The enrichment for younger never-smokers in the combined cohort likely reflects overrepresentation of EGFR/ALK-driven adenocarcinoma with distinct biology and treatment patterns, hence generalizability to older, smoking-associated NSCLC populations warrants external validation. Finally, metastatic burden was measured as the count of distinct anatomical sites rather than volumetric assessment, and the absence of RECIST data precluded capturing within-site progression or providing an independent measure of tumor dynamics.

These findings support integration of dual-trajectory monitoring. Isolated burden increase may warrant consideration of treatment adaptation in prospective studies while performance status remains favorable. Isolated ECOG worsening may prompt evaluation for reversible causes rather than immediate treatment escalation. Concordant worsening identifies the highest-risk patients. Prospective evaluation of the clinical utility of trajectory-guided evaluation is needed.

Declarations

Data Availability

The MSK-CHORD dataset is available through cBioPortal (https://www.cbioportal.org/study/summary?id=msk_chord_2024).

Code Availability

Analysis code is available at <https://github.com/adarya/nsclc-trajectories>

Ethics Statement

This study used de-identified, publicly available data from the MSK-CHORD cohort, accessed through cBioPortal (https://www.cbioportal.org/study/summary?id=msk_chord_2024). The primary MSK-CHORD study was approved by the Memorial Sloan Kettering Cancer Center Institutional Review Board for retrospective analysis of de-identified clinical data¹⁴. Secondary analysis of this publicly available, de-identified dataset did not require additional institutional review board approval at the authors' institutions, in accordance with local regulations governing research on de-identified public data.

Acknowledgments

We thank the MSK-CHORD team for making the dataset available through cBioPortal.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Claude Opus 4.0 and 4.7 in order to assist with analyses and manuscript editing. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contributions

A.P.: Conceptualization; Investigation; Data curation; Writing – original draft; Writing – review & editing. A.Z.: Investigation; Writing – review & editing. R.G.: Methodology; Validation; Writing – review & editing. R.B: Investigation; Writing – review & editing. I.D.D.: Investigation; Writing – review & editing. N.A: Investigation; Writing – review & editing. L.L.: Investigation; Writing – review & editing. N.P.: Investigation; Writing – review & editing. A.Y.: Conceptualization; Methodology; Software; Formal analysis; Data curation; Visualization; Writing – original draft; Writing – review & editing; Supervision. All authors approved the final manuscript.

Competing Interests

N.P. reports Advisor & Honorarium from & Research with AstraZeneca, Bayer, Boehringer Ingelheim, Bristol-Myers Squibb, Eli Lilly, Imagine, Guardant360, Imagene, Genesis, Merck, MSD, Novartis, Pfizer, Roche, Rhenium, and Takeda. The other authors declare no competing interests.

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A (2023) Cancer statistics, 2023. *CA Cancer J Clin* 73:17–48
2. Ganti AK, Klein AB, Cotala I, Seal B, Chou E (2021) Update of Incidence, Prevalence, Survival, and Initial Treatment in Patients With Non-Small Cell Lung Cancer in the US. *JAMA Oncol* 7:1824–1832
3. Goldstraw P et al (2016) The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer* 11:39–51
4. Planchard D et al (2018) Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol Off J Eur Soc Med Oncol* 29:iv192–iv237
5. Sehgal K et al (2021) Association of Performance Status With Survival in Patients With Advanced Non–Small Cell Lung Cancer Treated With Pembrolizumab Monotherapy. *JAMA Netw Open* 4:e2037120
6. Lobefaro R et al (2021) Immunotherapy in advanced Non-Small Cell Lung Cancer patients with poor performance status: The role of clinical-pathological variables and inflammatory biomarkers. *Lung*

7. Kawaguchi T et al (2010) Performance Status and Smoking Status Are Independent Favorable Prognostic Factors for Survival in Non-small Cell Lung Cancer: A Comprehensive Analysis of 26,957 Patients with NSCLC. *J Thorac Oncol* 5:620–630
8. Hellman S, Weichselbaum RR, Oligometastases. (1995) *J Clin Oncol Off J Am Soc Clin Oncol* 13:8–10
9. Gomez DR et al (2019) Local Consolidative Therapy Vs. Maintenance Therapy or Observation for Patients With Oligometastatic Non-Small-Cell Lung Cancer: Long-Term Results of a Multi-Institutional, Phase II, Randomized Study. *J Clin Oncol Off J Am Soc Clin Oncol* 37:1558–1565
10. Palma DA et al (2019) Stereotactic ablative radiotherapy versus standard of care palliative treatment in patients with oligometastatic cancers (SABR-COMET): a randomised, phase 2, open-label trial. *Lancet* 393:2051–2058
11. Marschner N et al (2020) Association of Disease Progression With Health-Related Quality of Life Among Adults With Breast, Lung, Pancreatic, and Colorectal Cancer. *JAMA Netw Open* 3:e200643
12. Kurtz DM et al (2019) Dynamic Risk Profiling Using Serial Tumor Biomarkers for Personalized Outcome Prediction. *Cell* 178:699–713e19
13. Temel JS et al (2010) Early Palliative Care for Patients with Metastatic Non–Small-Cell Lung Cancer. *N Engl J Med* 363:733–742
14. Jee J et al (2024) Automated real-world data integration improves cancer outcome prediction. *Nature* 636:728–736
15. Cheng DT et al (2015) Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT). *J Mol Diagn* 17:251–264
16. Dingemans A-MC et al (2019) Definition of Synchronous Oligometastatic Non–Small Cell Lung Cancer—A Consensus Report. *J Thorac Oncol* 14:2109–2119
17. Kehl KL et al (2020) Natural Language Processing to Ascertain Cancer Outcomes From Medical Oncologist Notes. *JCO Clin Cancer Inf* 4:680–690
18. Lakens D, Equivalence Tests (2017) A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Soc Psychol Personal Sci* 8:355–362
19. Kass RE, Raftery AE (1995) Bayes Factors. *J Am Stat Assoc* 90:773–795
20. Pencina MJ, D’Agostino RB, D’Agostino RB, Vasan RS (2008) Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat. Med.* 27, 157–172; discussion 207–212
21. Schneider BJ et al (2021) Management of Immune-Related Adverse Events in Patients Treated With Immune Checkpoint Inhibitor Therapy: ASCO Guideline Update. *J Clin Oncol Off J Am Soc Clin Oncol* 39:4073–4126
22. de Naurois J et al (2010) Management of febrile neutropenia: ESMO Clinical Practice Guidelines. *Ann Oncol Off J Eur Soc Med Oncol* 21(5):v252–256

23. Wildiers H et al (2014) International Society of Geriatric Oncology consensus on geriatric assessment in older patients with cancer. J Clin Oncol Off J Am Soc Clin Oncol 32:2595–2603

24. Eisenhauer EA et al (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 45:228–247

Figures

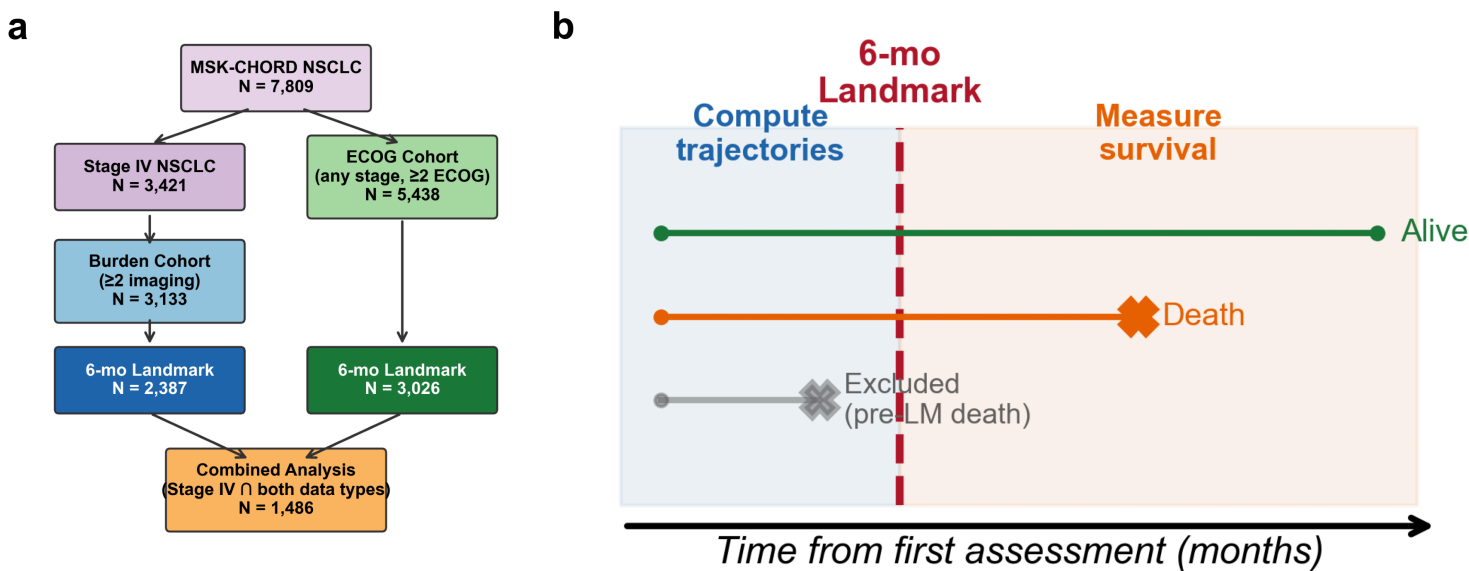


Figure 1

Study overview and landmark analysis framework. (a) CONSORT diagram showing patient flow for the metastatic burden cohort (N = 3,133 Stage IV), ECOG trajectory cohort (N = 5,438), and combined analysis cohort (N = 1,486 at the 6-month landmark). (b) Schematic of the landmark analysis framework: trajectory features are computed using only pre-landmark data, and survival is measured from the landmark forward, preventing immortal time bias.

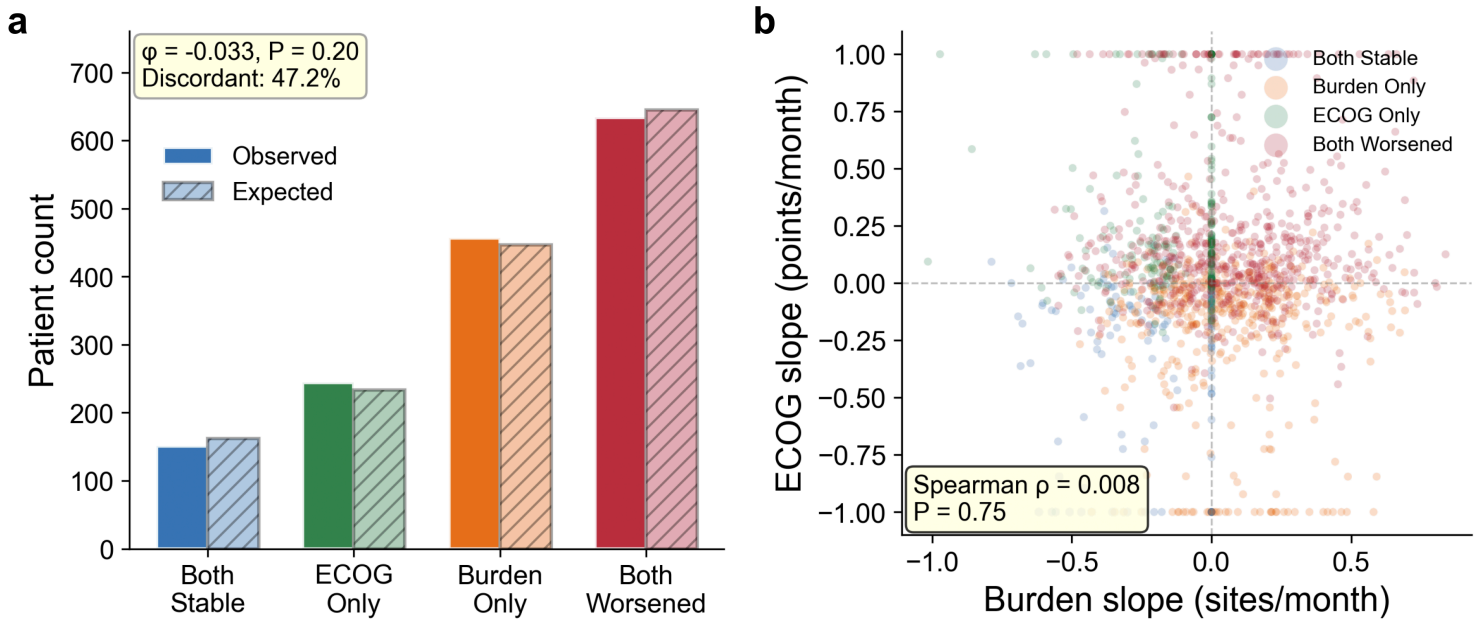


Figure 2

Burden and ECOG trajectories show no meaningful association. (a) Observed versus expected patient counts under the assumption of independence for the 2×2 contingency table (burden worsened $\times$ ECOG worsened). Observed counts closely match expected counts ($\phi = -0.033, P = 0.20$). Discordant patients comprise 47.2% of the cohort. (b) Scatter plot of burden slope versus ECOG slope for 1,486 combined patients, colored by risk group. Spearman $\rho = 0.008$ ($P = 0.75$), confirming near-zero correlation between continuous trajectory measures.

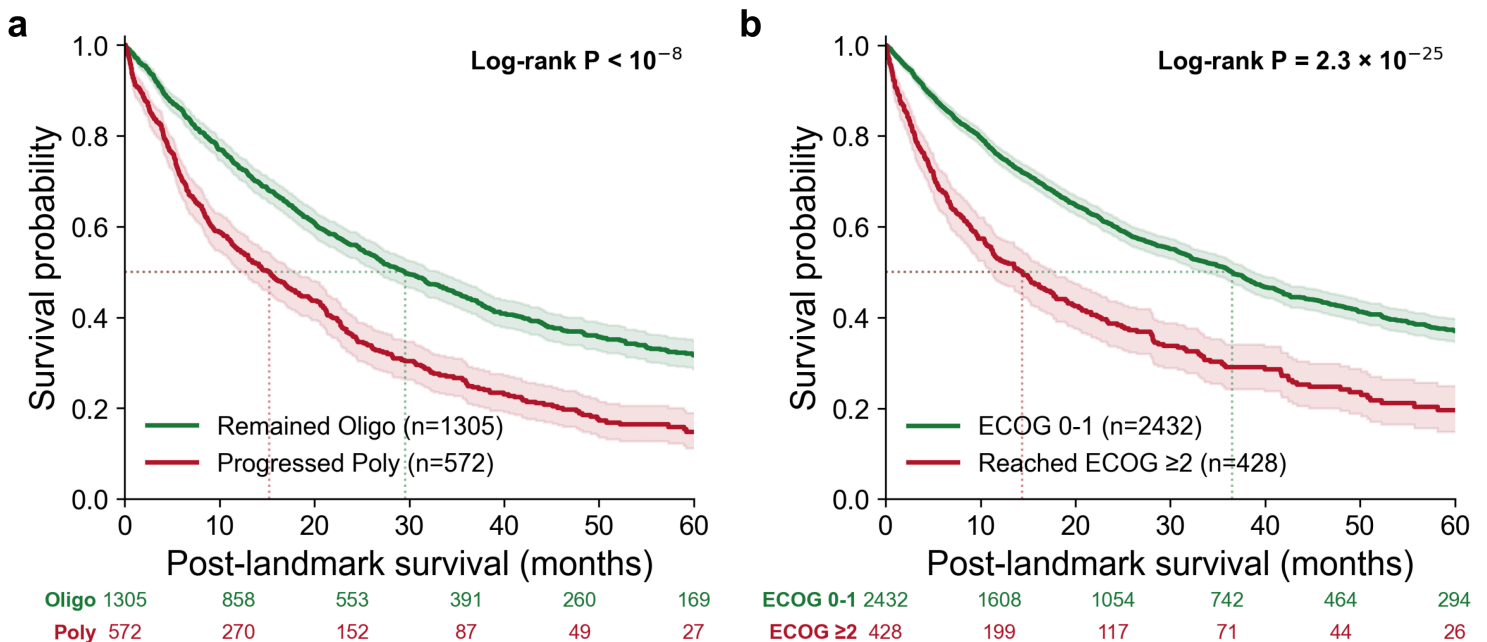


Figure 3

Individual trajectory results. (a) Kaplan-Meier curves for post-landmark overall survival by oligo-to-polymetastatic transition status at the 6-month landmark (N = 2,387). (b) Kaplan-Meier curves for post-landmark overall survival among baseline ECOG 0–1 patients stratified by whether they reached ECOG ≥ 2 by the 6-month landmark (N = 3,026).

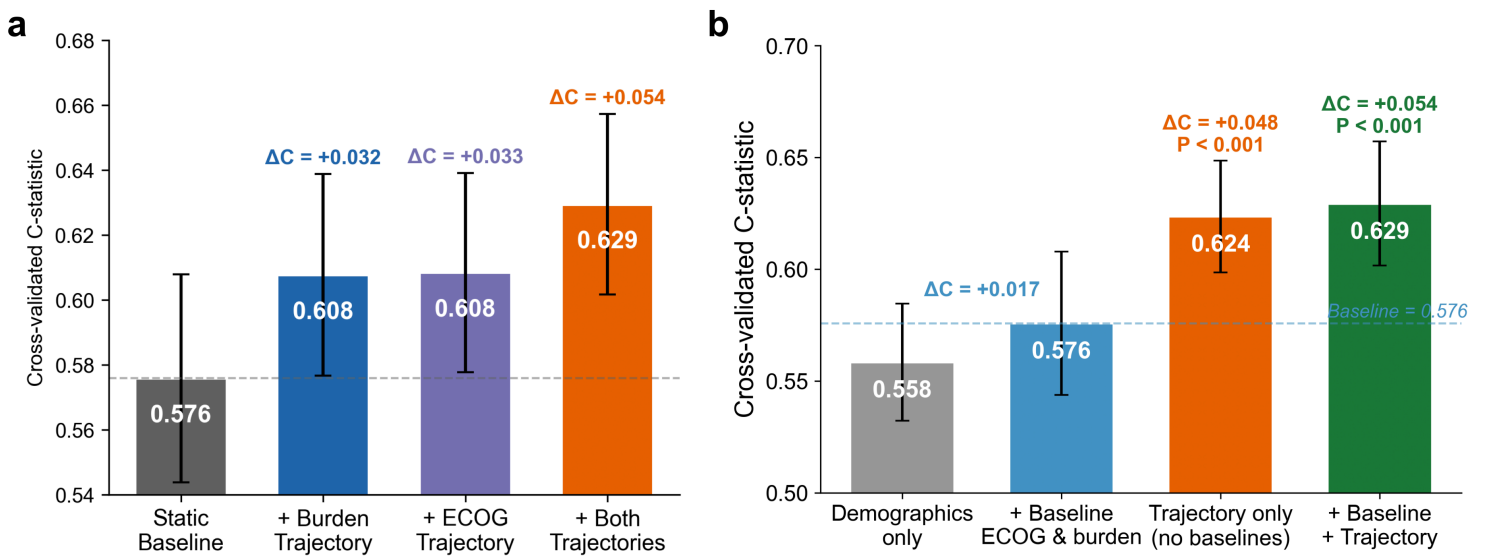


Figure 4

Additive prognostic value and four-group risk stratification. (a) Progressive model comparison showing cross-validated C-statistics. (b) Comparison of four model families: demographics alone, +baseline ECOG/burden, trajectory features alone (no baselines), and the full combined model.

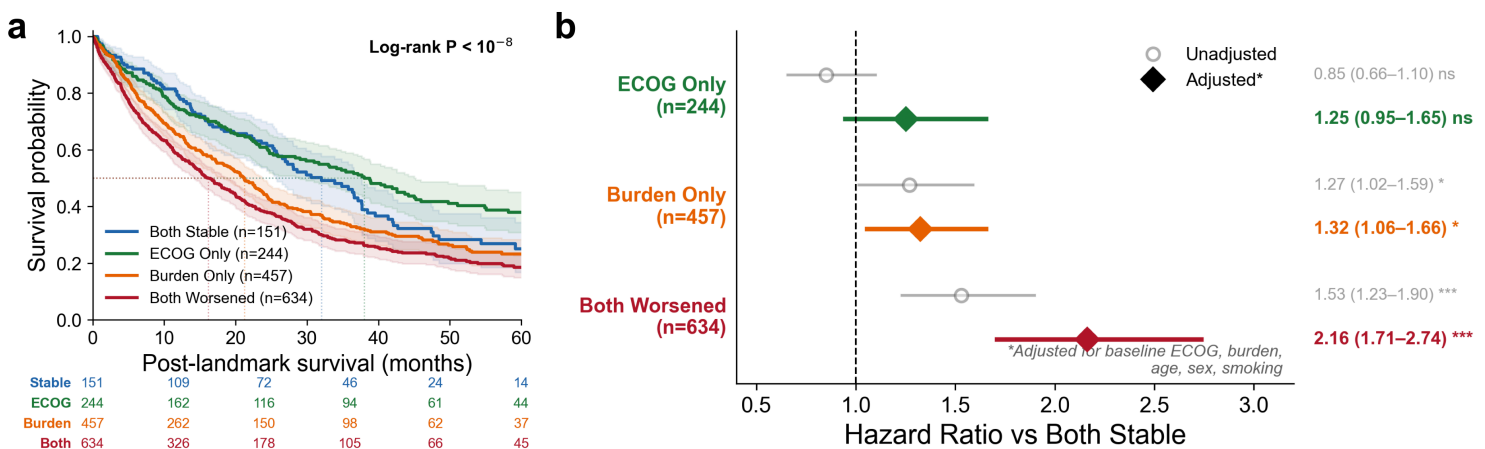


Figure 5

Hazard ratios for risk groups versus Both Stable reference.(a) Kaplan-Meier curves for the four risk groups defined by concordant vs. discordant trajectory worsening. (b) Forest plot showing baseline-adjusted HRs with 95% CIs for each risk group compared to Both Stable (adjusted for baseline ECOG, baseline burden, age, sex, and smoking status).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTables.docx](#)