Predicting Real World Overall Survival in Patients with Non-Small Cell Lung Receiving Immune Checkpoint Inhibitor Therapy Using a Joint Model Integrating Serial Genetic and Epigenetic Biomarkers

MSR61

Christopher Pretz[1], Aaron Hardin[1], Sara Wienke[1], Carin Espenschied[1], Amar Das[1]

1. Guardant Health, Inc. Palo Alto, CA

Background

Methylation-based tumor fraction (TF) dynamics are strongly associated with outcomes in non–small cell lung cancer (NSCLC) treated with immune checkpoint inhibitors. *KRAS* mutations, common oncogenic drivers, are associated with tumor heterogeneity and may evolve under therapeutic pressure, reflecting clonal adaptation. Because TF captures tumor burden and *KRAS* mutations provide insight into genomic evolution, jointly modeling their trajectories may improve understanding of biomarker co-evolution and enhance outcome prediction. We developed a statistical framework that simultaneously characterizes serial *KRAS* mutation burden and TF, capturing both genetic adaptation and epigenetic quantification of tumor fraction.

Methods

- **Cohort:** Consists of 251 patients with NSCLC from the prospective RADIOHEAD study, a pan-cancer cohort of immunotherapy-naïve patients receiving immune checkpoint inhibitor inhibitor (ICI) regimens. Patients included had a baseline and ≥2 on-treatment samples.
- **Biomarker Assessment:** TF was measured using Guardant Reveal (Guardant Health, Palo Alto CA), a tissue-free liquid biopsy assay quantifying ctDNA-derived tumor signal in plasma. plasma. *KRAS* alterations were assessed using Guardant360 Liquid (Guardant Health, Palo Palo Alto CA). For each sample, the *KRAS* mutation with the highest variant allele frequency frequency (VAF) was selected to represent mutation burden. TF and *KRAS* VAF were logit-logit-transformed for modeling and back-transformed for interpretation.
- **Baseline Covariates:** Age, smoking status (never (Ref) vs. past/current), disease stage (III (Ref) vs. IV), and gender (Female (Ref) vs. Male).
- **Statistical Modeling:**
  A joint modeling (JM) framework was used to simultaneously analyze longitudinal biomarker biomarker data and time-to-event outcomes. Longitudinal TF and *KRAS* trajectories were were characterized using hierarchical cubic spline mixed-effects sub-models, while real-world world overall survival (rwOS) was modeled using a Cox regression sub-model. Covariates were included in both sub-models to improve prediction and adjust for confounding.

More formally:

Longitudinal Sub-Model captures the evolution of the biomarker over time:

$$Y_i(t) = m_i(t) + \varepsilon_i(t) \qquad (1)$$

$Y_i(t)$: Biomarker value for individual i at timepoint t
$m_i(t)$: Patient trajectory (hierarchical mixed-effects cubic spline model with baseline covariates)
$\varepsilon_i(t)$: Measurement error

Survival Sub-Model models the hazard of an event (in this case patient overall survival) given by:

$$h_i(t) = h_0(t)\exp(\gamma^T w_i + \alpha\, m_i(t)) \qquad (2)$$

$h_0(t)$: Baseline hazard function
w_i: Baseline covariates
$m_i(t)$: Latent longitudinal process at timepoint t
α: Association parameter linking the information from the longitudinal and survival sub-models

Dynamic Predictions: Within a Bayesian framework, the estimate for the probability of being "death free" for a new patient given a future prediction horizon (u > t) is:

$$\pi(t, u) = \int \Pr(T_i^* \geq u \mid T_i^* > t, \tilde{y}_i(t); \theta)\, p(\theta \mid D_n)\, d\theta \qquad (3)$$

$\tilde{y}_i$: A new patient with a set of biomarker measurements and unique baseline measurements
D_n: The sample for which the joint model was fitted
T_i^*: The patient's true event time
θ: A vector of global parameter estimates (i.e., longitudinal and survival outcomes)
$p(\theta \mid D_n)$: Posterior distribution of the parameters given the data.

An estimate of $\pi(t, u)$ is obtained via Markov Chain Monte Carlo, and the 95% credible interval can be extracted from the corresponding posterior sample percentiles. The sample size was sufficient to conduct the proposed analysis.[1]

Longitudinal Data & Model Results

Figure 1

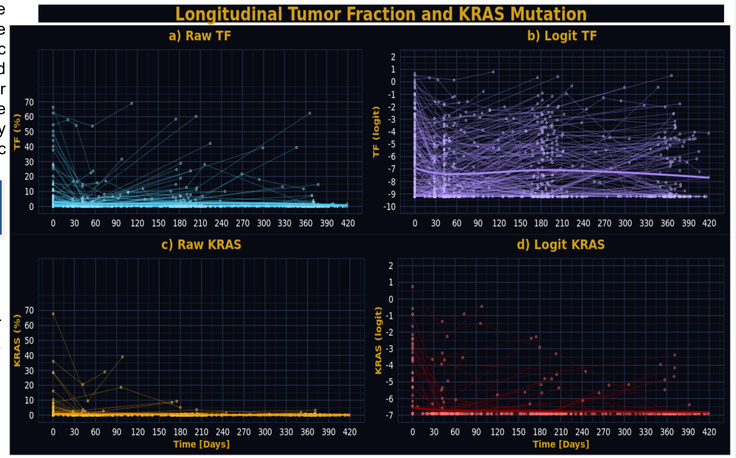


Figure 1 (a) Spaghetti plot of the raw TF data with cohort-level trend line. Dots represent the observed values and are connected by lines to map out a patient's progression. **(b)** Depicts the logit transomed TF upon which the JM is built. Similarly, **(c)** represents the raw *KRAS* VAF data over time, and **(d)** presents the longitudinal logit transformed *KRAS* data used to fit the JM.

Table 1

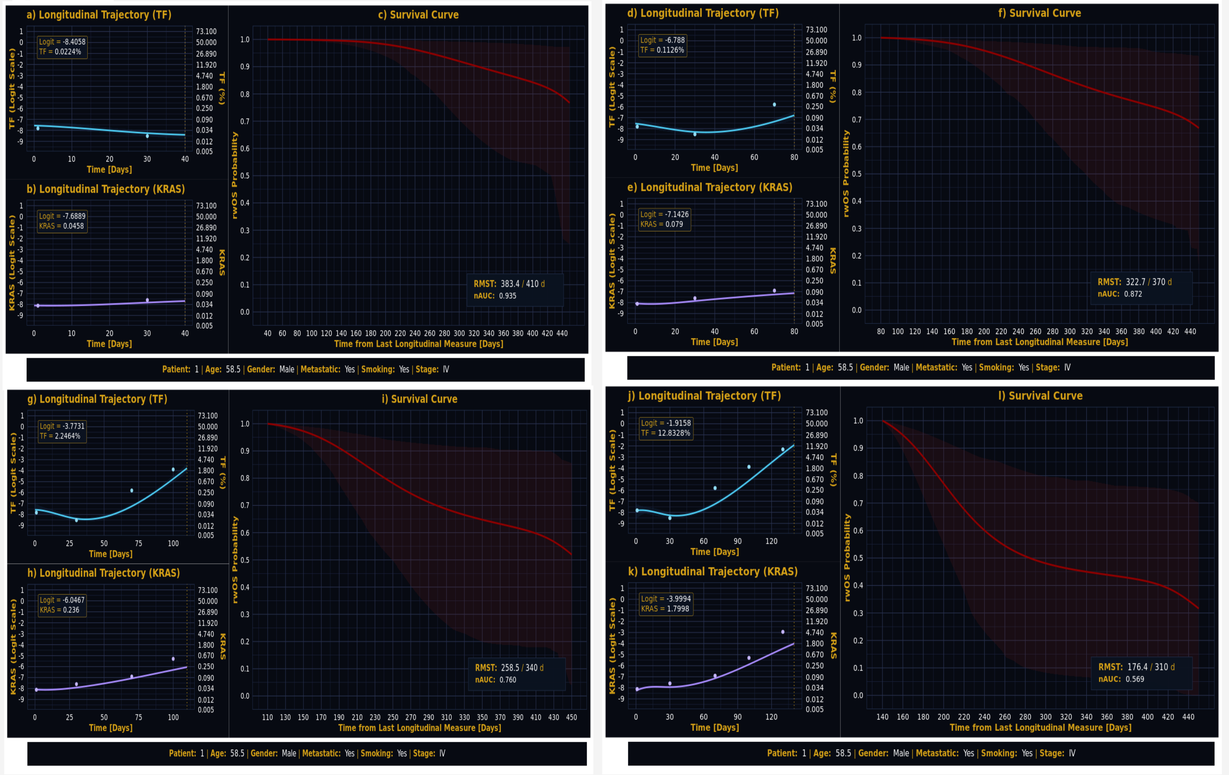
Covariate	Mean	SD	2.5% CrI	97.5% CrI	HR	95% CrI (HR)	P-value	R-hat
Baseline Age	0.0228	0.0199	-0.0161	0.0611	1.023	0.984, 1.063	0.2595	1.0006
Gender	-0.1401	0.2886	-0.7069	0.4236	0.869	0.493, 1.527	0.6244	1.0011
Disease Stage Baseline	0.4377	0.3103	-0.1633	1.0607	1.549	0.849, 2.888	0.1537	1.0003
Smoking Status	1.3893	0.8365	-0.0084	3.2635	4.012	0.992, 26.141	0.0517	1.0001
TF Current Value (logit)	0.3609	0.0574	0.2506	0.4757	1.435	1.285, 1.609	0.0000	1.0032
KRAS Current Value (logit)	0.2691	0.1244	0.0148	0.5026	1.309	1.015, 1.653	0.0394	1.0031

In the Bayesian joint model, the current values of logit-transformed tumor fraction and *KRAS* VAF were the primary dynamic predictors of survival, capturing time-updated effects across patients and longitudinal measurements. As this is a Bayesian framework, credible intervals (CrI) are reported.

Results (Table 1) demonstrate that higher TF is significantly associated with increased hazard across all patients and follow-up times (HR = 1.44, 95% CrI: 1.29–1.61, $p < 0.001$). Similarly, higher *KRAS* mutation burden was independently associated with worse survival (HR = 1.31, 95% CrI: 1.02–1.65, $p = 0.039$) across all patients and follow-up times. These effects represent instantaneous associations between biomarker levels and risk, reflecting the impact of evolving disease burden and clonal dynamics. Baseline covariates, primarily serving as statistical controls (e.g. age, gender, and disease stage), were not significantly associated with survival. Smoking status showed a strong trend toward increased hazard (HR = 4.01, 95% CrI: 0.99–26.14, $p = 0.052$), though with wide uncertainty. All parameters demonstrated excellent convergence (R-hat < 1.01), supporting the reliability of the estimates. These results give rise to dynamic predictions as illustrated in the following example.

Dynamic Prediction Results

Figure 2



Dynamic prediction results for a "new" patient presented in **Figure 2** illustrate updated survival probabilities at increasing landmark times (~70, 130, 220, and 310 days; panels c, f, l, l). As both TF (panels a, d, g, j) and *KRAS* (panels b, e, h, k) burden evolve over time, corresponding survival predictions are continuously updated. At earlier timepoints, relatively low and stable TF and *KRAS* levels are associated with favorable survival (panels c and f), reflected by higher normalized area under the curve (nAUC) values and higher restricted mean survival times (RMST)—the expected survival up to a fixed horizon. As biomarker burden increases over time, particularly with rising TF (panels g and j) and concurrent increases in *KRAS* VAF (panels h and k), survival curves shift downward and decline more rapidly, indicating elevated mortality risk (i.e. lower nAUC and RMST values). Persistently increasing TF and *KRAS* trajectories demonstrate the most pronounced reductions in predicted survival, while low and stable changes maintain comparatively favorable outcomes. These patterns highlight the joint contribution of tumor burden and clonal dynamics to risk stratification. Overall, the JM framework enables real-time, patient-specific survival prediction, dynamically integrating evolving TF and *KRAS* trajectories to quantify changes in prognosis over the disease course.

KEY FINDING: TF and KRAS Exhibit a Joint Inverse Relationship with Survival—Concurrent Increases in Both are Associated with Worsening rwOS, Reflected by Downward Shifts in Survival Curves and Consistent Declines in RMST and nAUC.

Conclusions

(1) Joint modeling of longitudinal TF and KRAS dynamics enables real-time, patient-specific survival prediction in NSCLC treated with ICIs.

(2) Integrating epigenetic tumor burden and genomic evolution improves risk stratification and can identify high-risk patients earlier.

(3) This approach supports incorporating serial liquid biopsy biomarkers into personalized, data-driven treatment and monitoring strategies.

References:
1. Rizopoulos D. Joint models for longitudinal and time-to-event data with applications in R. 2012; CRC Press.