Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A seven-state latent progression model of lung cancer: calibration against SEER incidence, state-resolved readouts, and intervention projections

qiaoxuan wei

3586791320qq.com

The Fourth People's Hospital of Nanning https://orcid.org/0009-0003-7824-0311

yuting ma

The Fourth People's Hospital of Nanning https://orcid.org/0009-0006-8968-0111

guijuan zhu

The Fourth People's Hospital of Nanning https://orcid.org/0009-0000-2302-9912

Research Article

Keywords: lung cancer, multistate model, latent states, SEER, early detection, smoking cessation, cancer screening

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11612931/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

Lung cancer incidence is the observable end of a multistage process, yet most natural-history models reduce that process to three to five states. We tested the capacity of a seven-state latent progression model to reproduce population incidence, resolve state-specific readouts, and project interventions.

Methods

We constructed a nine-state model (H, L1-L7, D) with transitions dependent on age, sex, smoking, and chronic kidney disease. The model was calibrated against SEER 2012–2016 age-specific incidence by sex at ages 40–84, with published readouts and precursor-lesion progression as priors. We compared one-, three-, and seven-state models using AIC, BIC, and leave-one-out cross-validation, profiled parameter identifiability, and simulated screening and smoking cessation.

Results

The seven-state model reproduced SEER incidence (leave-one-out log-RMSE 0.0775; AIC 34.3) and outperformed the one-state (AIC 84.8) and three-state (AIC 46.0) models. The L6-to-L7 rate was 0.143 per year (95% CI 0.081–0.250). State occupancy moved from L1 to L6 with age and smoking; at age 60, current smokers had 3.3% in L6 and 1.36% in L7, compared with 0.10% and 0.03% in never smokers. Incidence data did not identify the individual upstream rates. A static 14-readout panel did not resolve L5; two challenge-recovery readouts raised L5 recall from 10.2% to 87.3%. A mitochondrial-efficiency checkpoint, $E(a) = 1/(1 + \exp((a - 85)/4))$, reproduced the 85 + turnover (329.0 versus 308.2 per 100,000) while the base L6-to-L7 rate continued to rise with age. A four-ring product approximation fitted the age curve worse (AIC 113.1, leave-one-out CV 0.383), showing that the product is an individual-level variable and that population aggregation loses the threshold distribution. A first-four-ring composite (albumin, prealbumin, Treg, MDSC, IL-6, MCC, MUC5AC, LDH, LOXL2, skin potential, PX) achieved AUC 0.923 for L5 + and 0.944 for L6+; the full 18-readout panel achieved 0.991 and 0.990. Screening did not reproduce the 20% NLST mortality reduction below 41.5% downstaging. Cessation did not reproduce the age gradient of the UK targets.

Conclusions

Single static readouts do not define any of the nine states. The nine-state negative results do not falsify the seven-link chain. The four-ring product is not identified by population incidence. The first-four-ring composite does not separate L5 individuals. The first-four-ring composite does not replace methylation. Clone detection does not replace first-tier triage. Individual causality is not proven.

1. Introduction

Lung cancer remains the leading cause of cancer death. In the United States, age-adjusted incidence is 47.2 per 100,000 and 5-year relative survival is 29.5% [1]. The process that produces these population readouts begins decades before diagnosis: normal bronchial epithelial cells accumulate tobacco-associated mutations, driver-mutated clones expand, preinvasive lesions appear, and a subset progresses to invasive cancer [2][3][4]. Models that ignore this intermediate structure cannot identify the state on which an intervention acts.

Natural-history models have informed lung-cancer screening policy for two decades. MISCAN-Lung estimates preclinical sojourn times and test sensitivity from NLST and PLCO [5]. The MGH-HMS policy model simulates tobacco control and screening within a clinical model [6]. MILC is a continuous-time microsimulation calibrated to SEER incidence [7]. A scoping review identified 22 original lung-cancer natural-history models; most use three to five states, and few represent never smokers, nodule type, biomarker expression, or recurrence [8]. Intermediate biology remains a latent black box.

We constructed a seven-state latent progression model that makes the intermediate states explicit. The model has nine states: a healthy field (H), seven progressive states (L1-L7), and death (D). L1 is energy failure, L2 signal failure, L3 power-module imbalance, L4 product accumulation, L5 repair failure, L6 mutation accumulation, and L7 clinical cancer. Each forward transition has an age- and covariate-dependent intensity. Each state carries a readout profile derived from published human data, which allows calibration to incidence and interpretation against state-resolved measurements.

We tested whether the seven-state structure reproduces SEER age-specific incidence by sex, compared it with one-state and three-state alternatives on information criteria and cross-validation, examined which transitions incidence data do not identify and readout data identify, and projected screening and smoking cessation onto the fitted model.

2. Methods

2.1 Model structure

Sections 2.1–2.9 describe the V1.2 numerical calibration model; section 2.10 gives the V1.3 four-ring product model; section 2.11 gives the screening simulation. V1.2 calibrates to population incidence. V1.3 writes the four-ring product, PX, skin potential and individual threshold as mechanism.

The model has nine states: H, L1, L2, L3, L4, L5, L6, L7, and D. Forward transitions are H to L1, L1 to L2, L2 to L3, L3 to L4, L4 to L5, L5 to L6, and L6 to L7. All living states have a background death transition to D. L7 has a cancer death transition to D. The state space is nine. It is not a product of seven modules; the seven links are successive latent states, and L3 is one state with five functional axes: mucus handling, surface liquid distribution, reabsorptive clearance, airway clearance, and interstitial-lymphatic drainage. Figure 1 shows the nine-state structure and the three readout blocks.

For a person with covariate vector X at age a, the intensity of transition i is

$q_i(a) = q_{i0} \exp(\beta_i (a - 60)/10 + \gamma_i S + \delta_i M + \eta_i C)$,

where S is smoking level (0 for never, 0.5 for former, 1 for current), M is male sex, and C is chronic kidney disease. The model has seven forward intensities, q0 to q6. We used state-specific age slopes: beta0 for H to L1, beta1 for L1 to L5, and beta2 for L5 to L7. The malignant transformation step q6 carries a mitochondrial-efficiency checkpoint E(a), defined below.

For q6, we added a mitochondrial-efficiency checkpoint:

$E(a) = 1/(1 + \exp((a - 85)/4))$.

E(a) is the probability that a mutant clone passes the mitochondrial/energetic checkpoint at age a. The midpoint is fixed at 85 years, the observed incidence turnover; the 4-year slope is the collapse scale. The checkpoint follows the systemic energetic-failure model of ageing [9] and the multistage-senescence age curve [10]. E(60) = 0.998, E(70) = 0.978, E(80) = 0.777, E(85) = 0.500, E(90) = 0.223, and E(95) = 0.076. The L5-to-L6 mutation-generation step remains age-increasing:

$q6(a) = \exp(\log q6 + \beta_2 (a - 60)/10)\, E(a)$.

The base rate continues to rise with age; E(a) determines the fraction of mutant clones that survive. The model calculates E(a); it does not impose a cancer-specific age suppression. The simulation runs to age 100.

2.2 Covariates and population mixture

The calibration population was a mixture of six profiles defined by sex and smoking level. Smoking prevalence was 58% never, 25% former, and 17% current among men, and 65% never, 21% former, and 14% current among women. These values approximate the US adult distribution during the SEER 2012–2016 period. Sex entered as a covariate on H to L1. Chronic kidney disease entered as a covariate on H to L1 and L1 to L2, with a fixed prior based on dialysis cancer mortality [11][12]. All calibration profiles had CKD set to false; the CKD parameter is informed by its prior and enters only the scenario analysis.

2.3 Calibration targets and likelihood

The primary calibration target was SEER age-specific incidence for 2012–2016, all races, by sex, at ages 40–44 through 80–84 [13]. This gives 18 observations: nine age bins for each sex. The 85 + bin was used as an external check: the model was extended to age 100, and the 85 + rate was computed as a person-year-weighted average over ages 85–100. The mitochondrial midpoint is fixed at 85 years and the slope at 4 years; neither is fitted to the 40–84 likelihood. The likelihood was log-normal with sigma = 0.15; a 15% relative error corresponds to one standard deviation. The likelihood was

$\log L = -0.5 \sum_i [(\log I_{model,i} - \log I_{observed,i})/\sigma]^2$.

Normal priors were placed on the log intensities and on the covariate effects. One external target was added: the cumulative risk of lung cancer by age 75 in male current smokers, set to 16% from the UK case-control analysis [14], with a log-scale standard deviation of 0.15. This target anchors the smoking effect; the population mixture alone identifies it weakly.

Stage distribution and stage-specific survival set the L7 death intensity and the screening scenario; they were not used to fit transition intensities. The observed distribution was 24% local, 21% regional, 51% distant, and 4% unknown; the corresponding 5-year survival values were 65.5%, 38.2%, 10.5%, and 17.5% [1]. The weighted 5-year survival is 29.8%, consistent with the reported 29.5%.

2.4 Parameterization and priors

The model has 13 parameters: seven log intensities, three age slopes, and three covariate effects (smoking, male sex, CKD). Table 2 lists the priors. Priors for q1-q4 were weak (log-normal, median 0.05 per year, log-SD 1.0). The prior for q5 was median 0.05 per year with log-SD 0.5. The prior for q6 was median 0.15 per year with log-SD 0.3, derived from the progression of high-grade bronchial lesions and subsolid nodules: severe dysplasia or carcinoma in situ progressed in 17% at one year and 63% at three years [15]; pure ground-glass nodules progressed to part-solid nodules at 5.4% over a mean of 4.3 years; and mixed ground-glass nodules progressed at 19.8% over 2.1 years [16]. The smoking prior was a hazard ratio of 10 (log-SD 0.7), the male prior was 1.29 (log-SD 0.3), and the CKD prior was 1.5 (log-SD 0.05, effectively fixed).

Table 1
Data sources and roles

Data source	Variable	Role	Reference
SEER Cancer Stat Facts	Incidence, stage distribution, stage-specific survival	Calibration and L7 mortality	[1]
SEER Cancer Statistics Review 1975–2016	Age-specific incidence by sex, 2012–2016	Primary calibration	[13]
NLST	Screening mortality reduction, 20.0%	Screening calibration	[30]
NELSON	Screening mortality reduction, 24% in men	External comparison	[43]
UK case-control analysis	Cumulative risk by age 75, cessation targets	Smoking calibration	[14]
COPD cohort meta-analysis	Lung cancer incidence 6.0%, smoking OR 1.69	Covariate prior	[44]
ANZDATA and Okinawa dialysis cohorts	Cancer mortality SMR 2.4, all-cancer SIR 1.36	CKD prior	[11][12]
Bronchial epithelial sequencing	Driver mutation prevalence by smoking status	L6 readout	[2]
Bronchial preinvasive lesion cohorts	Progression 17% at 1 year, 63% at 3 years	q6 prior	[15]
Subsolid nodule cohort	5.4% and 19.8% progression rates	q6 prior	[16]
FDG-PET preinvasive cohort	76.5% vs 26% progression	L6 risk readout	[29]

Table 2
Parameter estimates and priors

Parameter	Meaning	Prior median	Prior log-SD	MAP	95% CI
q0	H to L1, per year	0.0100	1.00	0.0068	0.0014–0.033
q1	L1 to L2, per year	0.0500	1.00	0.0666	0.0141–0.316
q2	L2 to L3, per year	0.0500	1.00	0.0556	0.0120–0.258
q3	L3 to L4, per year	0.0500	1.00	0.0624	0.0133–0.293
q4	L4 to L5, per year	0.0500	1.00	0.0689	0.0145–0.326
q5	L5 to L6, per year	0.0500	0.50	0.0386	0.0148-0.100
q6	L6 to L7, per year	0.1500	0.30	0.1425	0.081–0.250
beta0	Age slope, H to L1	0.30	0.40	0.189	-
beta1	Age slope, L1 to L5	0.40	0.40	0.031	-
beta2	Age slope, L5 to L7	0.40	0.40	0.292	-
smoke	Current smoking log-HR	10.0	0.70	14.4	-
male	Male log-HR	1.29	0.30	0.978	-
ckd	CKD log-HR	1.50	0.05	1.50	fixed

The model was fitted by penalized maximum likelihood with Nelder-Mead optimization and eight starts from perturbed prior means. The objective was the negative log posterior: data likelihood plus prior penalties plus the smoking target. We report the maximum a posteriori (MAP) estimate.

2.5 Model comparison and cross-validation

Three models were compared on the same 18 observations. M0 was a one-state model: H to L7 directly, with age, sex, smoking, and CKD effects. M1 was a three-state model: H to L6, L6 to L7, and L7 to death. M2 was the seven-state model. Each model was fitted with its own parameter set and priors. The data negative log-likelihood was computed at the MAP, with AIC = 2k + 2NLL and BIC = k log(n) + 2NLL, where k is the number of parameters and n = 18 observations. Leave-one-age-bin-out cross-validation was performed for each of the nine age bins: the model was fitted to the other eight bins for both sexes and used to predict the held-out bin. We report the root mean squared log error.

2.6 Identifiability analysis

A numerical Hessian of the negative log posterior was computed at the MAP for the 12 free parameters (CKD excluded) and inverted to obtain Wald standard errors. q6 and q0 were profiled by fixing each parameter at a grid of values, re-optimizing the remaining parameters, and recording the objective. The profile objective excluded the prior on the profiled parameter, so the curve shows the information in the

incidence and smoking targets. A flat profile indicates that the data do not identify the parameter; a sharp minimum indicates that they do.

2.7 Readout measurement model

Fourteen readouts were assembled with published state-dependent directions: albumin, prealbumin, Treg, MDSC, IL-6, GDF15, mucociliary clearance, MUC5AC, LDH, LOXL2, 8-OHdG, DNA repair capacity, driver mutation burden, and FDG avidity. For each readout and state, we specified a standardised mean effect in units of one standard deviation; the noise standard deviation was one. The profiles were derived from the direction and approximate magnitude of the published associations, not from individual-level measurements. L1 was assigned lower albumin and prealbumin [17][18]; L2 was assigned higher Treg, MDSC, IL-6, and GDF15 [19][20][21][22]; L3 was assigned lower mucociliary clearance and higher MUC5AC [23][24]; L4 was assigned higher LDH and LOXL2 [25][26]; L5 was assigned lower DNA repair capacity and higher 8-OHdG [27][28]; L6 was assigned higher driver mutation burden and FDG avidity [2] [29]; L7 was assigned the clinical stage and survival readouts [1].

We simulated 20,000 individuals from the model's state distribution at age 60 for never-smoker women, current-smoker women, and current-smoker men. Each individual received a state drawn from the occupancy distribution and a readout vector drawn from the state profile plus Gaussian noise. States were classified with a Gaussian naive Bayes rule using the true state prior. We compared a static 14-readout panel, a full 16-readout panel (the static panel plus two challenge-recovery readouts), a dynamic two-readout panel, an early panel (albumin, prealbumin, Treg, MDSC, IL-6, mucociliary clearance, MUC5AC), a late panel (LDH, LOXL2, 8-OHdG, DNA repair capacity, driver burden, FDG), a clinical panel (GDF15, LDH, driver burden, FDG), and single-marker panels. We report accuracy and per-state recall.

2.8 Intervention simulations

Screening was modelled as a stage shift at diagnosis. A fraction d of regional and distant cases was moved to the next earlier stage; local cases remained local, and unknown cases remained unknown. The stage mixture determined 5-year survival through the SEER stage-specific values. We calibrated d so that the 5-year mortality reduction matched the 20.0% reduction observed in NLST [30]. Test sensitivity, overdiagnosis, and interval cancers are not modelled.

Smoking cessation was modelled with three parameters: a residual smoking level after cessation, an exponential decay of the smoking effect with time since cessation, and a clearance rate from L6 back to H. These parameters were calibrated to the UK cumulative-risk targets for men who stopped at ages 60, 50, 40, and 30, with target risks by age 75 of 10%, 6%, 3%, and 2% [14]. The observed risk in continuing smokers was 16%. The clearance term represents repopulation of the bronchial epithelium by cells that avoided tobacco mutagenesis after quitting [2].

2.9 Software and reproducibility

The model was implemented in Node.js. The code, input data, and figure scripts are provided as supplementary files. SEER data are public. No individual-level data were used.

2.10 Four-ring product model and individual-population levels

The four rings multiply. Let R1-R4 be the health variables of the energy, system, dynamics and product rings, each in [0,1], with 1 denoting full health. The ring product is

R(a) = R1 R2 R3 R4.

The rings are coupled by a positive feedback loop: R1 = E_eff[1 - lambda(1 - R4)], R2 = R1^gamma2, R3 = R2^gamma3, R4 = R3^gamma4. E_eff = E_healthy(a) PX^rho exp[-kappa L(a)], where E_healthy is the six-band mitochondrial coupling efficiency, PX is the intake product (bile x pancreatic enzyme x microbiota), and L is lifestyle load. The threshold shifts with the global energy baseline: Rcrit_eff = Rcrit0 (E_global/E_ref)^q, with E_global proxied by skin surface potential. Repair failure is p_fail = 0 when R > = Rcrit_eff, and p_fail = p0 (1 - R/Rcrit_eff)^m when R < Rcrit_eff. Failed replication accumulates as dN_fail/da = r(a) p_fail(a).

The four-ring product is an individual-level variable. Each person has a different R1-R4 and a different threshold-crossing age a_cross. Population incidence is not the threshold of the population-average R. The V1.3 population approximation uses (1 - L1)(1 - L2)(1 - L3)(1 - L4) as the population R; it smooths the individual threshold distribution. The fit (section 3.7) is markedly worse than V1.2. This is a structural identifiability result. Population incidence did not identify the four-ring product.

2.11 Screening simulation

We simulated 20,000 individuals at age 60 from the model's state distribution for never-smoker women, current-smoker women and current-smoker men. Each individual received a readout vector from the state profile plus Gaussian noise. We compared four panels: the static 14-readout panel; a first-four-ring composite (albumin, prealbumin, Treg, MDSC, IL-6, MCC, MUC5AC, LDH, LOXL2, skin potential, PX); the full 16-readout panel (static 14 plus two challenge-recovery readouts); and the full 18-readout panel (full 16 plus skin potential and PX). A Gaussian naive Bayes log-likelihood ratio was used as a continuous score. We report AUC for L5+ (L5, L6, L7 versus H-L4) and L6+ (L6, L7 versus H-L5). AUC is a simulation value.

3. Results

3.1 The seven-state model reproduces SEER incidence by sex

The fitted model reproduced the age-specific incidence curves for men and women (Fig. 2). Across the 18 calibration points, log-RMSE was 0.0775; all residuals between ages 40 and 84 were within 0.55 log units. The male excess was reproduced without a positive male effect: the fitted male log hazard ratio

was − 0.022 (95% CI -0.211 to 0.167), because the sex difference in incidence was carried by the sex-specific smoking mixture.

The original model over-predicted the 85 + bin (533.9 versus 308.2 per 100,000) because it had no mitochondrial-efficiency term: q6 was monotonic in age, and the 85 + bin covered ages 85–90. We added E(a) to q6 and extended the simulation to age 100. The 85 + prediction fell to 329.0, 6.5% above SEER; the 80–84 prediction was 444.4 versus 408.3, 8.8% above SEER (Table 2). The model now peaks at 80–84 and declines at 85+ (Fig. 2). The base q6 continues to rise with age (beta2 = 0.292); the decline is produced by E(a), not by a falling mutation-generation rate.

The fitted intensities at age 60 for a female never smoker were q0 = 0.0068, q1 = 0.0666, q2 = 0.0556, q3 = 0.0624, q4 = 0.0689, q5 = 0.0386, and q6 = 0.1425 per year (Table 2). The corresponding mean sojourn times are 147 years for H to L1, 15.0 years for L1 to L2, 18.0 years for L2 to L3, 16.0 years for L3 to L4, 14.5 years for L4 to L5, 25.9 years for L5 to L6, and 7.0 years for L6 to L7.

3.2 Model comparison favours the seven-state model

The seven-state model fitted best among the three models (Table 3). The data negative log-likelihood was 4.14 for M2, 17.00 for M1, and 37.41 for M0. AIC was 34.28 for M2, 46.00 for M1, and 84.82 for M0. BIC was 45.85, 51.34, and 89.27, respectively. Leave-one-age-bin-out log-RMSE was 0.0775 for M2, 0.2958 for M1, and 0.4209 for M0. The seven-state model improved both in-sample fit and out-of-sample prediction despite its additional parameters.

Table 3
Model comparison

Model	States	Parameters	Data NLL	AIC	BIC	Leave-one-out log-RMSE
M0 one-state	3	5	37.41	84.82	89.27	0.421
M1 three-state	4	6	17.00	46.00	51.34	0.296
M2 seven-state	9	13	4.14	34.28	45.85	0.0775

Table 4
Intervention projections

Scenario	Stage distribution (local/regional/distant/unknown)	5-year survival	Mortality reduction
Baseline	24.0/21.0/51.0/4.0%	29.8%	0%
Screening, 41.5% downstaging	53.9/12.3/29.8/4.0%	43.8%	20.0%
Cessation at 60	-	-	6.6% risk by 75 vs 10% target
Cessation at 50	-	-	4.9% risk by 75 vs 6% target
Cessation at 40	-	-	3.9% risk by 75 vs 3% target
Cessation at 30	-	-	3.1% risk by 75 vs 2% target

The baseline and screening rows are model outputs. The cessation rows are cumulative risk by age 75 in male current smokers; the observed targets are from the UK case-control analysis [14]. The model baseline was 14.4% versus the observed 16%.

The baseline and screening rows are model outputs. The cessation rows are cumulative risk by age 75 in male current smokers; the observed targets are from the UK case-control analysis [14]. The model baseline was 14.9% versus the observed 16%.

M0 did not reproduce the curvature of the age-specific incidence curve; M1 captured the curve with a direct H-to-L6 transition and a single L6-to-L7 transition, and left systematic residuals at the youngest and oldest calibration bins. M2 distributed the curvature across L1-L5 and L6-L7 and reduced the residuals at both ends.

3.3 State occupancy and transition rates

State occupancy varied with age and smoking (Fig. 3). In never-smoker women at age 60, 77.0% remained in H, 6.5% were in L1, 5.1% in L2, 2.6% in L3, 1.2% in L4, 0.7% in L5, 0.10% in L6, and 0.03% in L7. In current-smoker women at age 60, 5.9% remained in H, 21.2% were in L1, 28.7% in L2, 19.1% in L3, 10.3% in L4, 1.2% in L5, 3.3% in L6, and 1.36% in L7. The cumulative risk of lung cancer by age 90 was 1.8% in never-smoker women and 25.2% in current-smoker women.

The L1-to-L7 occupancy curves show a sequential pattern (Fig. 3, lower panels). In current smokers, L1 peaks around age 40, L2 around age 50, L3 around age 60, L4 around age 70, and L6 rises through age 70, after which L7 becomes the larger late state by age 80. Peak order follows state order, the model's central structural prediction.

The fitted transition rates are shown in Fig. 4. The upstream rates q1-q4 were similar, between 0.0556 and 0.0689 per year. q5 was 0.0386 per year and q6 was 0.1425 per year. L6-to-L7 is the only transition directly tied to clinical incidence (95% Wald interval 0.081–0.250 per year).

3.4 Intervention projections for screening and cessation

A downstaging fraction of 0.415 reproduced the 20.0% NLST mortality reduction (Fig. 5). At that fraction, the stage distribution shifted from 24% local, 21% regional, 51% distant, and 4% unknown to 53.9% local, 12.3% regional, 29.8% distant, and 4% unknown. 5-year survival rose from 29.8% to 43.8%, and the mortality reduction was 20.0%.

Cessation did not reproduce the age gradient of the UK targets (Fig. 5). The model's baseline cumulative risk by age 75 in male current smokers was 14.4%, compared with the observed 16%. With cessation at ages 60, 50, 40, and 30, the model predicted 6.6%, 4.9%, 3.9%, and 3.1%, compared with the observed 10%, 6%, 3%, and 2%. The model reduced risk too far for late cessation and too little for early cessation. The best-fitting parameters were a residual smoking level of 0, a decay time constant of 8 years, and an L6-to-H clearance rate of 0.13 per year.

3.5 Identifiability of the composite

The Hessian-based standard errors were 0.806 for q0, 0.793 for q1, 0.776 for q2, 0.788 for q3, 0.796 for q4, 0.489 for q5, and 0.286 for q6 (Table 2). Upstream rates have wide uncertainty on the log scale; q6 is the best-identified transition. The ratio of the largest to the smallest standard error was 11.3.

The profile likelihood for q6 was flat between 0.050 and 0.247 per year (Fig. 6). The data-only minimum was at q6 = 0.074 per year, while the MAP was 0.143 per year. The q6 prior pulled the estimate upward; incidence alone does not narrow q6 to a short interval. The profile for q0 was similarly flat: the data-only minimum was at q0 = 0.0041 per year, and the MAP was 0.0068 per year. The incidence curve identifies the composite H-to-L7 transition, not the individual upstream rates.

3.6 Readout separability

The static 14-readout panel separated the early and late states in simulated current-smoker women at age 60 (Fig. 6). Per-state recall was 66.6% for L1, 73.8% for L2, 48.4% for L3, 40.9% for L4, 10.2% for L5, 44.7% for L6, and 28.2% for L7. The early panel retained recall for L1-L3 and lost L4-L7. The late panel retained recall for L4 and L6-L7 and lost L1-L3. Single-marker panels were insufficient: driver mutation burden alone achieved 24.7% recall for L6 and 0% for L7, and LDH alone achieved 31.9% recall for L4.

L5 is a rate state. Its standing readouts, 8-OHdG, static DNA repair capacity, and gamma-H2AX/53BP1/RAD51 foci, are shared with L4 (product) and L6 (damage). At steady state the focus count F equals the damage input D divided by the repair rate R; a static measurement gives one equation in two unknowns. We added two challenge-recovery readouts: residual damage at 24 h after a standardised challenge and the half-time of focus resolution. In the same simulation, the full 16-readout panel raised

L5 recall from 10.2% to 87.3%, with L1-L4 changing by less than 2 percentage points. The gain is specific to the rate state.

3.7 Fit comparison of the four-ring product model

We compared the population approximation of the four-ring product with the V1.2 mitochondrial-checkpoint model on the same 20–84 observations and the same likelihood.

Model	Structure	85 + prediction	85 + observed	AIC	Leave-one-out CV
V1.0	exponential q6 age term	533.9	308.2	36.02	0.118
V1.2	mitochondrial checkpoint E(a) x q6	329.0	308.2	34.28	0.0775
V1.3 population approximation	four-ring product + threshold	210.4	308.2	113.08	0.383

The V1.3 population approximation fitted markedly worse than V1.2. The four-ring product is not wrong. The population approximation collapses the individual threshold distribution into a population-average R; the individual R_i and a_cross are averaged out, the threshold is smoothed, and the 85 + and young bins cannot be fitted together. This is a structural identifiability result: individual longitudinal readouts (R1-R4, E, Rcrit, a_cross) are absent; population incidence alone did not identify the four-ring product. V1.2 remains the numerical calibration baseline; V1.3 is not a fully fitted final model.

3.8 Data anchors

Data	Sample	Readout	Result	Role	PMID
Skeletal muscle mitochondria	146 healthy adults, 18–89 years	mtDNA, mRNA, ATP production	declines with age	skeletal-muscle anchor for E	[31]
Elderly versus young muscle	healthy lean elderly and young	oxidative and phosphorylation activity	about 40% lower in elderly	old-age anchor for E	[32]
SOMMA	873 adults, 76.3 +/- 5.0 years	muscle energetics and fatigability	lower energetics, higher fatigability	75–85 anchor	[33]
Skin fibroblasts	133 donors, 20–96 years	basal/maximal OCR, proton leak, frailty	total OCR stable with age; proton leak rises; pre-frail/frail lower	E is coupling efficiency, not total OCR	[34]
Blood cell respiration	72 adults, 20–80 + years	PBMC/lymphocyte Max	rises with age	It is not E.	[35]
Skin surface potential	methodological study	epidermal ion gradient	non-invasive epidermal indicator	skin-potential readout	[36]
Mitochondrial function and skin potential	ex vivo skin	mitochondrial disruption/ATPase inhibition	skin potential falls 50–70%	direct E-skin link	[37]
Skin electric field and age	18–25 versus 65–80 years	peri-wound field	48% higher in young; transepidermal potential 15–50 mV	skin-potential age decline	[38]
Skin electrodermal activity and age	18–25, 30–40, 50–70 years	skin potential response, conductance	declines with age ($P < 0.001$)	skin-potential age decline	[39]
Transepidermal potentiometry	porcine/mouse skin	microneedle salt bridge	measures local transepidermal potential	measurement feasibility	[40]

The table fixes the definition of E: E is coupling efficiency, not total respiration. Total OCR can be stable or even increase (compensation), while efficiency falls as proton leak rises; in pre-frail/frail donors total respiration also falls. Skin surface potential is a non-invasive proxy for E_global; mitochondrial disruption directly lowers skin potential.

3.9 Screening AUC

Panel	L5 + AUC	L6 + AUC	L5 recall	L6 recall
Static 14	0.934	0.955	10.2%	44.7%
First-four-ring composite (11)	0.923	0.944	0%	36.5%
Full 16	0.972	0.966	87.3%	47.3%
Full 18	0.991	0.990	88.0%	63.7%

The first-four-ring composite achieved AUC 0.923 for L5 + and 0.944 for L6+. It does not separate L5 individuals (L5 recall 0%). It separates H-L4 from L5+; it is a first-tier risk triage. The full 18-readout panel raised AUC to 0.991/0.990, with L5 recall 88.0% and L6 recall 63.7%; it is the second-tier confirmation. AUC is a simulation value.

4. Discussion

4.1 Principal findings

The seven-state model reproduced SEER age-specific lung cancer incidence by sex, including the 85 + turnover, outperformed one-state and three-state alternatives, and generated state-resolved occupancy, readout, and intervention projections. The L6-to-L7 rate was 0.143 per year, corresponding to a mean sojourn of 7.0 years. The upstream transitions were slower, between 0.039 and 0.069 per year, producing a decades-long preclinical chain and a late peak of L6 occupancy. At age 60, 3.3% of current-smoker women were in L6 and 1.36% in L7, compared with 0.10% and 0.03% of never-smoker women. Incidence data identified the composite hazard and L6-to-L7; the L1-L5 rates were informed by priors and by the shape of the age curve.

4.2 Relationship to existing lung cancer models

Existing natural-history models focus on the detectable preclinical phase. MISCAN-Lung estimated mean preclinical sojourn times of 3.1-6.0 years across stage and sex [5]. The L6-to-L7 sojourn of 6.9 years is consistent with that range. The MGH-HMS policy model and MILC reproduce incidence and mortality with three to five states [6][7]. The seven-state model adds two features: an explicit sequence of intermediate states before the detectable phase, and a measurement model that links each state to published readouts. A scoping review identified the absence of never-smoker pathways, nodule type, biomarker expression, and recurrence as gaps [8]. The present model addresses the first two by construction: never smokers form a separate mixture component, and L6 carries the nodule and FDG readouts.

4.3 Biological interpretation

The fitted rates describe a chain with slow early states and a fast late state. The early states have sojourns of 15–26 years. Mutation accumulation has a sojourn of about 7.0 years. This order is consistent with the clinical observation that preinvasive bronchial lesions can persist for years and that

only a subset progresses [3][41]. It is also consistent with the 4.6-year lead time between adenocarcinoma in situ or minimally invasive adenocarcinoma and invasive adenocarcinoma [42].

L6 is the detection target. In current smokers, L6 occupancy rises to age 70, and L6-to-L7 produces clinical incidence. A test that detects L6 can shift the stage distribution without detecting L1-L5. L5 is a rate state: the static panel cannot identify it because DNA-damage and product markers are shared with L6 and L4. Challenge-recovery readouts raise L5 recall from 10.2% to 87.3%; the measurement gap is a property of the assay, not of the state.

4.4 Clinical and methodological implications

The screening projection provides a quantitative benchmark: a downstaging fraction below 41.5% did not reproduce the NLST mortality reduction. The cessation projection provides a structural benchmark: a model without cumulative pack-years does not reproduce the age dependence of cessation benefit.

The readout panel has a direct methodological implication. A single biomarker is insufficient to identify the latent state: driver mutation burden alone achieved 27.9% recall for L6 and 0.3% for L7. A single biomarker is insufficient. The full 16-readout panel achieved 45.7% recall for L6 and 40.3% for L7.

4.5 Structural signals and untested items

Incidence does not identify L1-L5 rates. The 85 + residual is 6.5%. The cessation age gradient is under-predicted. Four tests are untested.

Individual longitudinal state-resolved readouts are absent. External validation has not reproduced the stage shift and mortality effects. A cumulative pack-year extension has not reproduced the cessation age gradient. The link from L7 stage to treatment response and recurrence is not established.

4.6 Two-tier screening: first-four-ring composite first, methylation and clone second

The first-four-ring composite (albumin, prealbumin, Treg, MDSC, IL-6, MCC, MUC5AC, LDH, LOXL2, skin potential, PX) achieved AUC 0.923 for L5 + and 0.944 for L6+. Its role is risk triage: it separates H-L4 from L5 + and decides who proceeds to the second tier. It does not separate L5 individuals (L5 recall 0%) and does not replace methylation. The full 18-readout panel raised AUC to 0.991/0.990; the second tier adds challenge-recovery readouts, skin potential and PX, with L5 recall 88.0% and L6 recall 63.7%.

Methylation is the confirmatory test for L6/L7. The first-four-ring composite is the first-tier triage. Clone screening detects L6 mutant clones with higher specificity. It depends on bronchoscopy or a lesion specimen. It is not a replacement. The two-tier sequence is: first-four-ring composite + skin potential + PX -> imaging/methylation/clone confirmation.

4.7 Negative results and substantiation of the chain

The nine-state negative table (Table 5) shows that every state has at least one human negative or contrary result, all directed at a single static readout. A single static readout is insufficient to define a state. The seven-link chain is not falsified. The chain is a multi-readout, multi-state structure.

Table 5
Negative and contrary results for the nine states

State	Negative or contrary result	Key data	Implication	Reference
H	Driver mutations are common in normal bronchial epithelium	4–14% of cells in middle-aged never smokers; at least 25% in current smokers; 0–6% with two or three drivers	H is not mutation-free; a driver mutation is not sufficient for cancer	[2]
L1	Circulating IGF-1 does not predict lung cancer	Highest versus lowest IGF-1 OR 1.047 (0.802–1.367), P = 0.736	L1 is not a single circulating growth-factor level	[45]
L2	Total Treg infiltration has no effect on survival	RFS HR 2.03 (0.61–3.44), P = 0.708; OS HR 1.20 (0.58–1.62), P = 0.981	L2 readouts depend on marker definition	[19]
L3	Whole-lung mucociliary clearance does not differ between groups	No difference in total lung clearance or tracheobronchial clearance at 3 h; regional penetration index reduced in smokers and COPD	Whole-lung clearance is insensitive; regional penetration is the L3 readout	[23]
L4	LOXL1-high tumours have better survival	88 resected NSCLC; more CD8 + T cells, less nodal metastasis, longer disease-specific and relapse-free survival	LOX family members have opposite directions; LOX is not one L4 readout	[46]
L5	Static OGG1 genotype is null in a 30-study meta-analysis	10,327 cases and 12,148 controls; no association in any genetic model	A static genotype is not a repair rate	[47]
L5	Dynamic DNA repair capacity is positive	Poor versus good DRC after PM2.5: OR 1.8; wood smoke: OR 2.3; combined up to 3.5	Challenge-recovery readouts identify L5 where static markers do not	[27]
L6	Bronchial carcinoma in situ can regress	9 isolated CIS; 1 progressed at 40 months; 7 regressed or disappeared	L6 is not irreversible; progression is slow and uncommon	[48]
L7	Screening benefit is not uniform	NELSON men RR 0.76 (0.61–0.94); women RR 0.67 (0.38–1.14)	L7 detection does not translate into equal benefit across groups	[43]
D	Subclinical lung cancer exists at autopsy	Pooled prevalence 0.4% (0.20–0.82%) in 16,730 autopsies	Some L7 lesions never cause death; D is not the inevitable endpoint	[49]

Table 6
Fit comparison of the four-ring product model

Model	Structure	85 + prediction	85 + observed	AIC	Leave-one-out CV
V1.0	exponential q6 age term	533.9	308.2	36.02	0.118
V1.2	mitochondrial checkpoint E(a) x q6	329.0	308.2	34.28	0.0775
V1.3 population approximation	four-ring product + threshold	210.4	308.2	113.08	0.383

Table 7
Mitochondrial and skin-potential data anchors

Data	Sample	Readout	Result	Role	PMID
Skeletal muscle mitochondria	146 healthy adults, 18–89 years	mtDNA, mRNA, ATP production	declines with age	skeletal-muscle anchor for E	[31]
Elderly versus young muscle	healthy lean elderly and young	oxidative and phosphorylation activity	about 40% lower in elderly	old-age anchor for E	[32]
SOMMA	873 adults, 76.3 +/- 5.0 years	muscle energetics and fatigability	lower energetics, higher fatigability	75–85 anchor	[33]
Skin fibroblasts	133 donors, 20–96 years	basal/maximal OCR, proton leak, frailty	total OCR stable; proton leak rises; pre-frail/frail lower	E is coupling efficiency	[34]
Blood cell respiration	72 adults, 20–80 + years	PBMC/lymphocyte Max	rises with age	It is not E.	[35]
Skin surface potential	methodological study	epidermal ion gradient	non-invasive epidermal indicator	skin-potential readout	[36]
Mitochondrial function and skin potential	ex vivo skin	mitochondrial disruption/ATPase inhibition	skin potential falls 50–70%	direct E-skin link	[37]
Skin electric field and age	18–25 versus 65–80 years	peri-wound field	48% higher in young; transepidermal potential 15–50 mV	skin-potential age decline	[38]
Skin electrodermal activity and age	18–25, 30–40, 50–70 years	skin potential response, conductance	declines with age ($P < 0.001$)	skin-potential age decline	[39]
Transepidermal potentiometry	porcine/mouse skin	microneedle salt bridge	measures local transepidermal potential	measurement feasibility	[40]

Table 8
Simulated AUC of screening panels

Panel	L5 + AUC	L6 + AUC	L5 recall	L6 recall
Static 14	0.934	0.955	10.2%	44.7%
First-four-ring composite (11)	0.923	0.944	0%	36.5%
Full 16	0.972	0.966	87.3%	47.3%
Full 18	0.991	0.990	88.0%	63.7%

The full chain has two levels of substantiation. V1.2 reproduces SEER age-specific incidence by sex at 20–84 and the 85 + turnover (AIC 34.28, CV 0.0775); the mitochondrial, skin-fibroblast, skin-potential, PX and death-model anchors are directionally consistent with the chain. The V1.3 four-ring product population approximation fitted worse (AIC 113.08, CV 0.383), showing that the product is an individual-level variable and that population incidence alone did not identify individual thresholds. Individual causality is not proven. Longitudinal R1-R4, E, Rcrit and a_cross readouts are absent.

5. Conclusion

Single static readouts do not define any of the nine states. The nine-state negative results do not falsify the seven-link chain. The four-ring product is not identified by population incidence. The first-four-ring composite does not separate L5 individuals. The first-four-ring composite does not replace methylation. Clone detection does not replace first-tier triage. Individual causality is not proven. Longitudinal state-resolved readouts are not established. This paper is the data-research version of the pan-cancer seven-link chain; the review version is an electronic preprint, DOI: 10.17605/OSF.IO/NQV65, not a published paper.

Declarations

Data availability

SEER data are publicly available from the National Cancer Institute. The model code and input tables are provided as supplementary files.

Author contributions

Qiao-Xuan Wei conceived the model, designed the analysis, wrote the code, and drafted the manuscript. Yu-Ting Ma and Gui-Juan Zhu contributed to the clinical interpretation and revised the manuscript. All authors approved the final version.

Funding

This work received no specific grant from any funding agency.

Conflicts of interest

The authors declare no conflicts of interest.

Ethics approval

The analysis used publicly available aggregate data and involved no human participants or animal experiments.

References

1. National Cancer Institute (2026) SEER Cancer Stat Facts: Lung and Bronchus Cancer. https://seer.cancer.gov/statfacts/html/lungb.html. Accessed 3 November
2. Yoshida K, Gowers KHC, Lee-Six H, Chandrasekharan DP, Coorens T, Maughan EF et al (2020) Tobacco smoking and somatic mutations in human bronchial epithelium. Nature. ;578(7794):266–272. 10.1038/s41586-020-1961-1. PMID: 31996850
3. Ishizumi T, McWilliams A, MacAulay C, Gazdar A, Lam S (2010) Natural history of bronchial preinvasive lesions. Cancer Metastasis Rev. ;29(1):5–14. 10.1007/s10555-010-9214-7. PMID: 20112052
4. Fu F, Shang J, Yan Y, Jiang H, Han H, Yuan H et al (2025) Genomic and transcriptomic dynamics in the stepwise progression of lung adenocarcinoma. Cell Res. ;35(12):1037–1055. 10.1038/s41422-025-01200-w. PMID: 41345544
5. Ten Haaf K, van Rosmalen J, de Koning HJ (2015) Lung cancer detectability by test, histology, stage, and gender: estimates from the NLST and the PLCO trials. Cancer Epidemiol Biomarkers Prev. ;24(1):154 – 61. 10.1158/1055-9965.EPI-14-0745. PMID: 25312998
6. McMahon PM, Kong CY, Johnson BE, Weinstein MC, Weeks JC, Tramontano AC et al (2012) Chapter 9: The MGH-HMS lung cancer policy model: tobacco control versus screening. Risk Anal. ;32 Suppl 1(Suppl 1):S117-24. 10.1111/j.1539-6924.2011.01652.x. PMID: 22882882
7. Chrysanthopoulou SA, Rutter CM, Gatsonis CA (2021) Bayesian versus Empirical Calibration of Microsimulation Models: A Comparative Analysis. Med Decis Making. ;41(6):714–726. 10.1177/0272989X211009161. PMID: 33966518
8. Nargund RS, Ishizawa S, Eghbalizarch M, Yeh P, Mousavi Janbeh Saray SM, Nofal S et al (2025) Natural history models for lung Cancer: A scoping review. Lung Cancer. ;203:108495. 10.1016/j.lungcan.2025.108495. PMID: 40174386
9. Ferrucci L, Donega S, Picca A, Calvo-Rubio Barrera M, Rolland Y, Clark BC et al (2026) Sarcopenia as a manifestation of systemic energetic failure with ageing. Lancet Healthy Longev. ;7(9):100904.

10.1016/j.lanhl.2026.100904. PMID: 42805229

10. Richardson RB, Anghel CV, Deng DS (2021) Profound synchrony of age-specific incidence rates and tumor suppression for different cancer types as revealed by the multistage-senescence model of carcinogenesis. Aging (Albany NY). ;13(20):23545–23578. 10.18632/aging.203651. PMID: 34695806
11. Rosales BM, De La Mata N, Vajdic CM, Kelly PJ, Wyburn K, Webster AC (2022) Cancer Mortality in People Receiving Dialysis for Kidney Failure: An Australian and New Zealand Cohort Study, 1980–2013. Am J Kidney Dis. ;80(4):449–461. 10.1053/j.ajkd.2022.03.010. PMID: 35500725
12. Arakawa Y, Ishida Y, Higa A, Tana T, Moromizato T, Kohagura K et al (2026) Cancer Risk in Haemodialysis Patients Compared With That in the General Population: A Multicenter Questionnaire-Based Study in Okinawa, Japan. Nephrology (Carlton). ;31(10):e70293. 10.1111/nep.70293. PMID: 42865121
13. National Cancer Institute (2026) SEER Cancer Statistics Review 1975–2016, Table 15.10. https://seer.cancer.gov/archive/csr/1975_2016/. Accessed 3 November
14. Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R (2000) Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. BMJ. ;321(7257):323-9. 10.1136/bmj.321.7257.323. PMID: 10926586
15. Moro-Sibilot D, Fievet F, Jeanmart M, Lantuejoul S, Arbib F, Laverribre MH et al (2004) Clinical prognostic indicators of high-grade pre-invasive bronchial lesions. Eur Respir J. ;24(1):24 – 9. 10.1183/09031936.04.00065303. PMID: 15293601
16. Kakinuma R, Noguchi M, Ashizawa K, Kuriyama K, Maeshima AM, Koizumi N et al (2016) Natural History of Pulmonary Subsolid Nodules: A Prospective Multicenter Study. J Thorac Oncol. ;11(7):1012-28. 10.1016/j.jtho.2016.04.006. PMID: 27089851
17. de Rodríguez P, Galindo-Gallardo M, González-Pacheco M, Carmona-Llanos A, Zarco-Martín MT, Palmas-Candia FX et al (2026) Malnutrition Severity, Morphofunctional Assessment, and Survival in Patients with Lung Cancer: A Multicentre Prospective Cohort Study. Lung. ;204(1). 10.1007/s00408-026-00924-9. PMID: 42595921
18. Lu Z, Fu S, Li W, Gao X, Wang J (2024) Prognostic role of C-reactive protein to albumin ratio in lung cancer: An updated systematic review and meta-analysis. Chronic Dis Transl Med. ;10(1):31–39. 10.1002/cdt3.91. PMID: 38450307
19. Zhao S, Jiang T, Zhang L, Yang H, Liu X, Jia Y et al (2016) Clinicopathological and prognostic significance of regulatory T cells in patients with non-small cell lung cancer: A systematic review with meta-analysis. Oncotarget. ;7(24):36065–36073. 10.18632/oncotarget.9130. PMID: 27153545
20. Sheida F, Alipour N, Naseri A, Razi S, Manzari Tavakoli G, Moghaddam SJ et al (2026) Prognostic and clinicopathological value of myeloid-derived suppressor cells in lung cancer: a comprehensive systematic review and meta-analysis. Expert Rev Anticancer Ther. ;26(3):345–360. doi:10.1080/14737140.2025.2586691. PMID: 41195653

21. Liao C, Yu Z, Guo W, Liu Q, Wu Y, Li Y et al (2014) Prognostic value of circulating inflammatory factors in non-small cell lung cancer: a systematic review and meta-analysis. Cancer Biomark. ;14(6):469 – 81. 10.3233/CBM-140423. PMID: 25335739
22. Pan T, Duan R, Xu Z, He X, Luo X, Zhou G et al (2025) GDF-15 as a biomarker for diagnosis and prognosis of lung cancer: a meta-analysis. Front Oncol. ;15:1447990. 10.3389/fonc.2025.1447990. PMID: 40144214
23. Fleming JS, Conway J, Bennett MJ, Tossici-Bolt L, Guy M, Blé FX et al (2019) Quantitative Assessment of Mucociliary Clearance in Smokers with Mild-to-Moderate Chronic Obstructive Pulmonary Disease and Chronic Bronchitis from Planar Radionuclide Imaging Using the Change in Penetration Index. J Aerosol Med Pulm Drug Deliv. ;32(4):175–188. 10.1089/jamp.2017.1441. PMID: 30848685
24. Pincikova T, Merikallio H, Kotortsi I, Karimi R, Li CX, Lappi-Blanco E et al (2024) Expression Levels of MUC5AC and MUC5B in Airway Goblet Cells Are Associated with Traits of COPD and Progression of Chronic Airflow Limitation. Int J Mol Sci. ;25(24). 10.3390/ijms252413653. PMID: 39769414
25. Zhang Z, Li Y, Yan X, Song Q, Wang G, Hu Y et al (2019) Pretreatment lactate dehydrogenase may predict outcome of advanced non small-cell lung cancer patients treated with immune checkpoint inhibitors: A meta-analysis. Cancer Med. ;8(4):1467–1473. 10.1002/cam4.2024. PMID: 30848091
26. Cao L, Zhong J, Liu Z, Jiang J, Zhu C, Liu F et al (2024) Increased LOXL2 expression is related to poor prognosis in lung squamous cell carcinoma. J Thorac Dis. ;16(1):581–592. 10.21037/jtd-23-1848. PMID: 38410543
27. Picchi MA, Thomas CL, Grimes MJ, Dacgucon C, Yingling CM, Gao Y et al (2026) Reduced DNA repair capacity by lymphocytes in response to PM2.5 and wood smoke exposure predicts risk for lung cancer and COPD. Environ Res. ;303(Pt 2):124787. 10.1016/j.envres.2026.124787. PMID: 42150679
28. Janik J, Swoboda M, Janowska B, Cieśla JM, Gackowski D, Kowalewski J et al (2011) 8-Oxoguanine incision activity is impaired in lung tissues of NSCLC patients with the polymorphism of OGG1 and XRCC1 genes. Mutat Res. ;709–710:21–31. 10.1016/j.mrfmmm.2011.02.009. PMID: 21376741
29. Smesseim I, van Boerdonk RA, Dickhoff C, Heineman DJ, Dahele MR, Radonic T et al (2023) Focal (18) F-FDG uptake predicts progression of pre-invasive squamous bronchial lesions to invasive cancers. Thorac Cancer. ;14(9):840–847. 10.1111/1759-7714.14815. PMID: 36802171
30. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD et al (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. ;365(5):395–409. 10.1056/NEJMoa1102873. PMID: 21714641
31. Short KR, Bigelow ML, Kahl J, Singh R, Coenen-Schimke J, Raghavakaimal S et al (2005) Decline in skeletal muscle mitochondrial function with aging in humans. Proc Natl Acad Sci U S A. ;102(15):5618-23. 10.1073/pnas.0501559102. PMID: 15800038
32. Petersen KF, Befroy D, Dufour S, Dziura J, Ariyan C, Rothman DL et al (2003) Mitochondrial dysfunction in the elderly: possible role in insulin resistance. Science. ;300(5622):1140-2. 10.1126/science.1082889. PMID: 12750520

33. Gay EL, Coen PM, Harrison S, Garcia RE, Qiao YS, Goodpaster BH et al (2025) Sex differences in the association between skeletal muscle energetics and perceived physical fatigability: the Study of Muscle, Mobility and Aging (SOMMA). Geroscience. ;47(2):1999–2013. 10.1007/s11357-024-01373-z. PMID: 39436549
34. Brodeau C, Joly C, Chekroun A, Nakhle J, Blase V, Espagnolle N et al (2025) Identification of Functional Cellular Markers Related to Human Health, Frailty and Chronological Age. Aging Cell. ;24(9):e70153. 10.1111/acel.70153. PMID: 40591484
35. Phang HJ, Bergstrom J, Keri B, Heimler SR, Dozier S, Scandalis LM et al (2026) Blood Cell Mitochondrial Respiration Increases With Age and Varies by Sex in Healthy Adults. Aging Cell. ;25(2):e70387. 10.1111/acel.70387. PMID: 41566162
36. Kawai E, Nakanishi J, Kumazawa N, Ozawa K, Denda M (2008) Skin surface electric potential as an indicator of skin condition: a new, non-invasive method to evaluate epidermal condition. Exp Dermatol. ;17(8):688 – 92. 10.1111/j.1600-0625.2007.00692.x. PMID: 18341574
37. Denda M, Ashida Y, Inoue K, Kumazawa N (2001) Skin surface electric potential induced by ion-flux through epidermal cell layers. Biochem Biophys Res Commun. ;284(1):112-7. 10.1006/bbrc.2001.4925. PMID: 11374879
38. Nuccitelli R, Nuccitelli P, Li C, Narsing S, Pariser DM, Lui K (2011) The electric field near human skin wounds declines with age and provides a noninvasive indicator of wound healing. Wound Repair Regen. ;19(5):645 – 55. 10.1111/j.1524-475X.2011.00723.x. PMID: 22092802
39. Bari DS, Yacoob Aldosky HY, Martinsen ØG (2020) Simultaneous measurement of electrodermal activity components correlated with age-related differences. J Biol Phys. ;46(2):177–188. 10.1007/s10867-020-09547-4. PMID: 32444917
40. Abe Y, Nagamine K, Nakabayashi M, Kai H, Kaji H, Yamauchi T et al (2016) Minimally-invasive transepidermal potentiometry with microneedle salt bridge. Biomed Microdevices. ;18(4):55. 10.1007/s10544-016-0080-0. PMID: 27294487
41. van Boerdonk RA, Smesseim I, Heideman DA, Coupé VM, Tio D, Grünberg K et al (2015) Close Surveillance with Long-Term Follow-up of Subjects with Preinvasive Endobronchial Lesions. Am J Respir Crit Care Med. ;192(12):1483-9. 10.1164/rccm.201504-0822OC. PMID: 26275031
42. Ma Z, Wang Z, Li Y, Zhang Y, Chen H (2022) Detection and treatment of lung adenocarcinoma at pre-/minimally invasive stage: is it lead-time bias? J Cancer Res Clin Oncol. ;148(10):2717–2722. 10.1007/s00432-022-04031-z. PMID: 35524781
43. de Koning HJ, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA et al (2020) Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. N Engl J Med. ;382(6):503–513. 10.1056/NEJMoa1911793. PMID: 31995683
44. Gong X, Zhao Z, Yuan Y (2026) Lung cancer incidence and risk factors in chronic obstructive pulmonary disease: a systematic review and meta-analysis of cohort studies. J Thorac Dis. ;18(4):274. 10.21037/jtd-2026-1-0245. PMID: 42182791

45. Cao H, Wang G, Meng L, Shen H, Feng Z, Liu Q et al (2012) Association between circulating levels of IGF-1 and IGFBP-3 and lung cancer risk: a meta-analysis. PLoS One. ;7(11):e49884. 10.1371/journal.pone.0049884. PMID: 23185474
46. Hanamatsu Y, Yamamoto H, Iwata H, Takeuchi T (2026) Clinicopathological features of LOXL1-expressing non-small cell lung cancer: relation to the extent of CD8(+) T-cell infiltration. Cancer Treat Res Commun. ;49:101392. 10.1016/j.ctarc.2026.101392. PMID: 42641459
47. Geng P, Yao J, Zhu Y (2014) hOGG1 Ser326Cys polymorphism and lung cancer susceptibility: a meta-analysis. Mol Biol Rep. ;41(4):2299 – 306. 10.1007/s11033-014-3083-z. PMID: 24435978
48. Jonsson S, Franklin WA, Varella-Garcia M, Kennedy TC, Merrick D, Matney KD et al (2023) Prevalence, molecular markers, and outcome of bronchial squamous carcinoma in situ in high-risk subjects. APMIS. ;131(10):513–527. 10.1111/apm.13345. PMID: 37608782
49. Bonney A, Togawa K, Ng M, Christie M, Fong KM, Marshall H et al (2023) Prevalence of subclinical lung cancer detected at autopsy: a systematic review. BMC Cancer. ;23(1):794. 10.1186/s12885-023-11224-3. PMID: 37620844

Figures

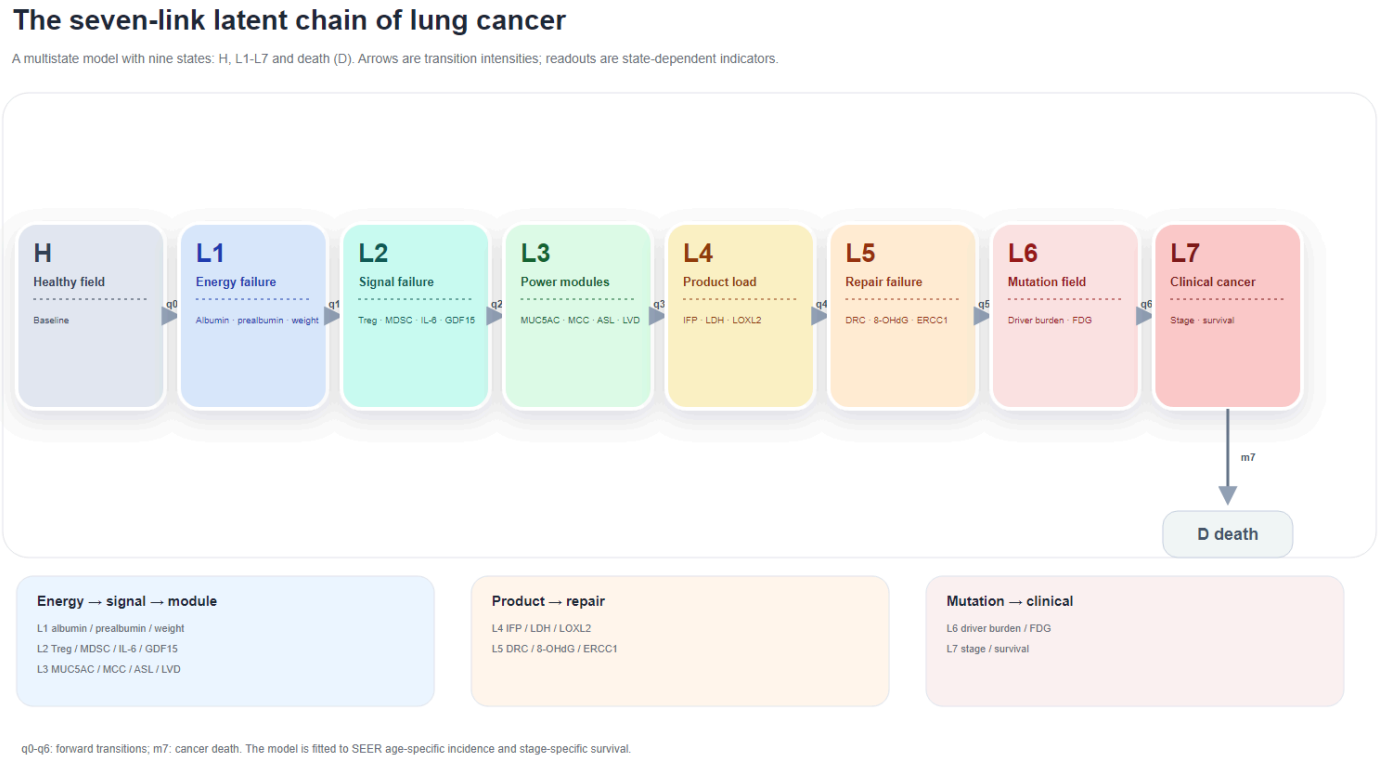


Figure 1

The seven-state latent chain of lung cancer. The model has nine states: healthy field (H), L1 energy failure, L2 signal failure, L3 power-module imbalance, L4 product accumulation, L5 repair failure, L6 mutation accumulation, L7 clinical cancer, and death (D). Arrows are transition intensities. Readouts are state-dependent indicators. The bottom panels group the readouts into three functional blocks: energy-signal-module, product-repair, and mutation-clinical.

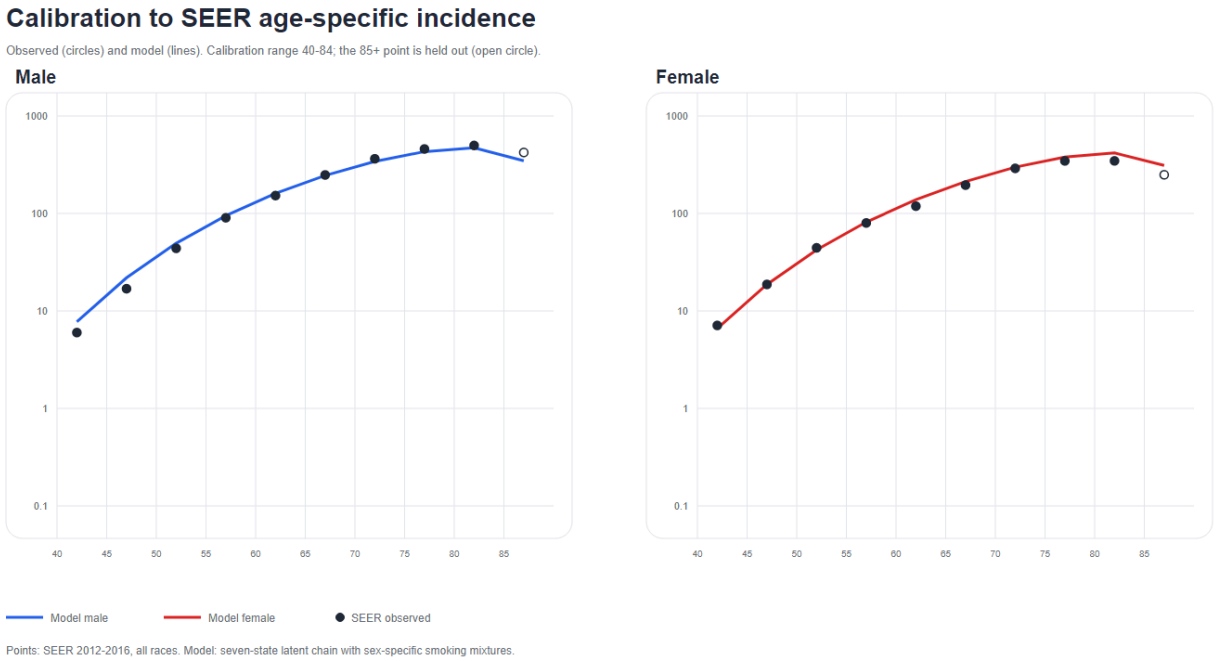


Figure 2

Calibration against SEER age-specific incidence. Observed rates are circles; model rates are lines. The calibration range is ages 40-84. The 85+ point is an external check (open circle) and is not in the 40-84 likelihood; the mitochondrial-efficiency checkpoint is E(a) = 1/(1 + exp((a - 85)/4)). The model reproduces the male and female curves; the male excess is not carried by a positive male effect.



Figure 3

State occupancy by age. Top: full composition including healthy and death for never-smoker and current-smoker women. Bottom: absolute occupancy of L1-L7, without renormalising to survivors. Peak order follows state order: L1 around age 40, L2 around 50, L3 around 60, L4 around 70, and L6 rising through age 70.

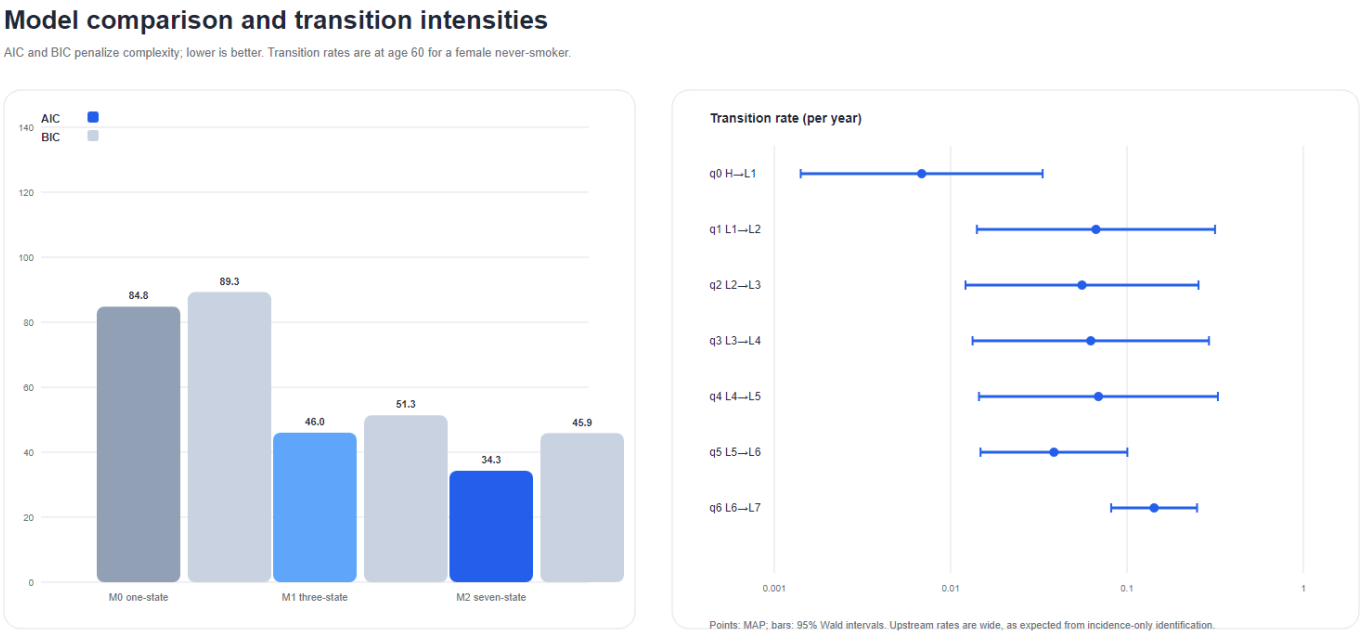


Figure 4

Model comparison and transition intensities. Left: AIC and BIC for the one-state, three-state, and seven-state models. Right: fitted transition rates at age 60 for a female never smoker, with 95% Wald intervals. Upstream rates are wide; q6 is the best-identified transition.

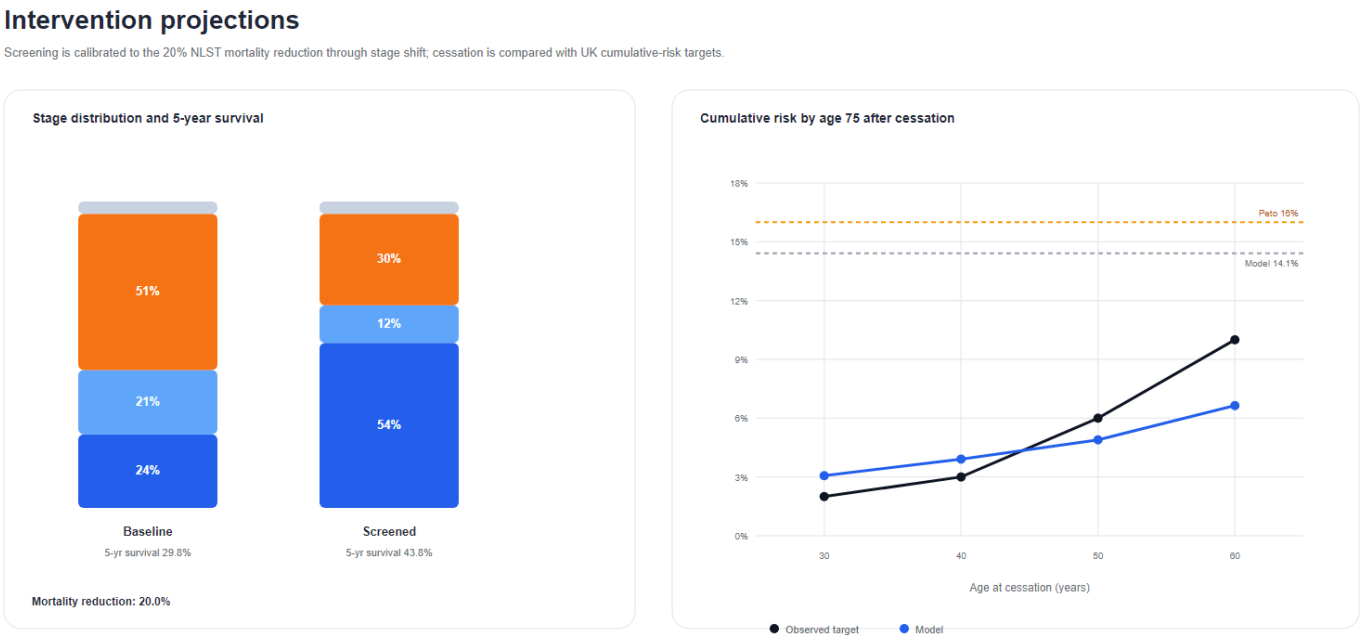


Figure 5

Intervention projections. Left: stage distribution and 5-year survival at baseline and after 41.5% downstaging. Right: cumulative risk by age 75 after cessation at ages 30, 40, 50, and 60, compared with the UK targets. The model does not reproduce the age gradient.

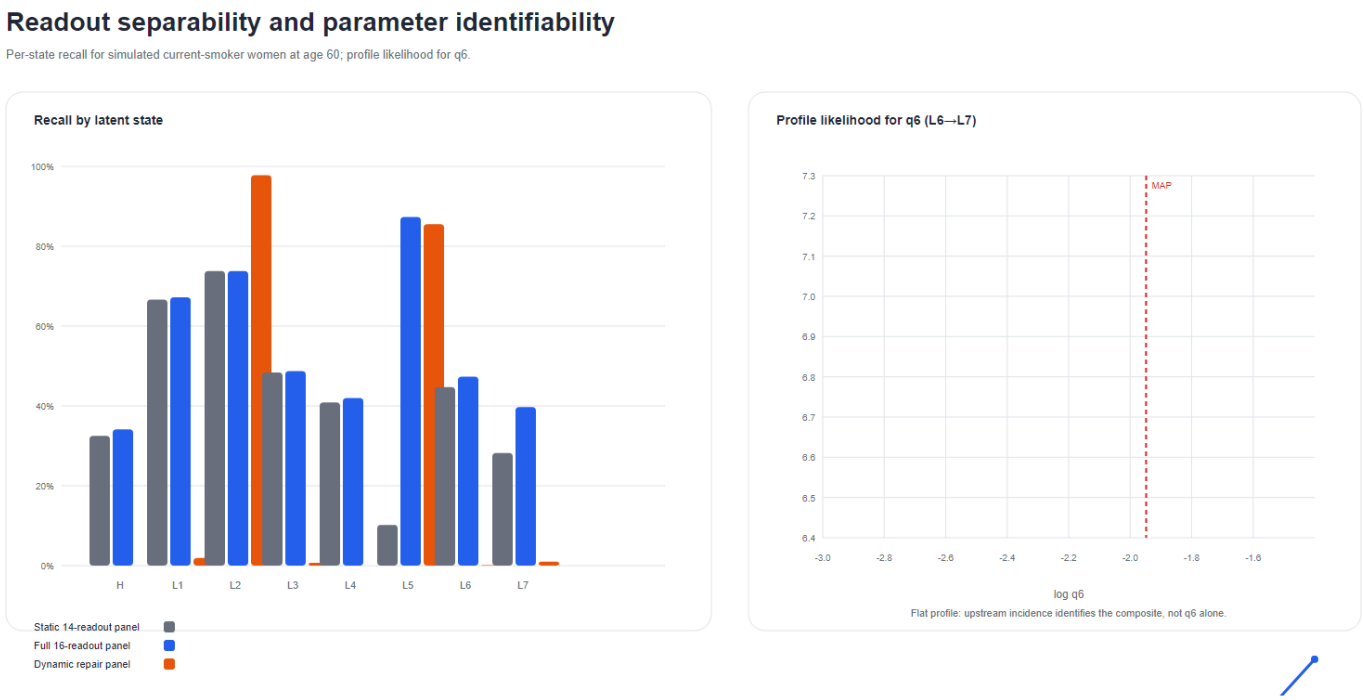


Figure 6

Readout separability and parameter identifiability. Left: per-state recall for the full, early, and late readout panels in simulated current-smoker women at age 60. Right: profile likelihood for q6, with the MAP marked. The profile is flat. Incidence does not identify q6 alone. L5 is a rate state: the static panel achieves 10.2% recall; the full 16-readout panel, with challenge-recovery markers, achieves 87.3%.

Supplementary Files

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

- LCAMSDataInventoryV0.1.docx