Supplementary Materials

Tumor Characteristics and Cardiovascular Mortality in 19,228 Thyroid Cancer Patients: A Population-Based Competing Risk Study

Yumei Chen [1,†], Chang Liu [2,†], Tianwang Guan [3], Taohua Chen [2], Zeru Chen [2], Yaosheng Luo [1], Aihua Cai [2], Xiaoying Huangfang [1], Min Yi [1,*] and Ming Hu [4,*]

[1] Department of Endocrinology, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou 510260, China
[2] Department of Clinical Medicine, The Second Clinical College of Guangzhou Medical University, Guangzhou 511436, China
[3] Department of Cardiovascular Medicine, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Southern Medical University, Guangzhou 510080, China
[4] Guangdong Key Laboratory of Vascular Diseases, Guangzhou Institute of Cardiovascular Disease, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou 510260, China
* Correspondence: minyi@gzhmu.edu.cn or smu_min@126.com (M.Y.); huming1989@126.com (M.H.)
† These authors contributed equally to this work.

Contents

Supplementary Method S1. Study population and variable definitions.

Supplementary Method S2. Multivariable regression and subgroup analyses.

Supplementary Method S3. Time-varying associations and standardized absolute risks.

Supplementary Method S4. Exploratory nomogram and competing-risk prediction.

Supplementary Method S5. Treatment-related sensitivity analyses and diagnostics.

Table S1. Morphological codes of histological types of thyroid cancer.

Table S2. Classification and detailed description of Surveillance, Epidemiology, and End Results summary stage.

Table S3. Specific causes of cardiovascular death.

Table S4. Baseline characteristics of training and validation cohorts.

Table S5. Primary multivariable associations with cardiovascular death.

Table S6. Additional covariate-adjusted associations with cardiovascular death.

Table S7. Sex-conditional stage associations and formal interaction tests.

Table S8. Age-stratified stage and outcome counts in males.

Table S9. Age-stratified stage and outcome counts in females.

Table S10. Specific scores of prognostic factors in the exploratory nomogram.

Table S11. Time-dependent stage associations with cardiovascular death.

Table S12. Five-year absolute risks and stage risk differences.

Table S13. Broader eligibility sensitivity analyses.

Table S14. Competing-risk prediction models and validation performance.

Table S15. Surgery-restricted treatment-adjusted sensitivity analyses.

Table S16. Model diagnostics and sensitivity checks.

Table S17. Exploratory cumulative incidence by reconstructed nomogram score group.

Figure S1. Cause-specific cumulative hazard of cardiovascular death by sex. Nelson–Aalen cumulative hazard curves in the primary cohort (n = 19,228), stratified by sex. Non-cardiovascular deaths are censored at their occurrence. The vertical axis shows cause-specific cumulative hazard, not cumulative incidence or percentage risk.

Figure S2. Time-varying stage associations with cardiovascular mortality. Fine–Gray subdistribution hazard ratios for regional and distant versus localized SEER summary stage. Shading represents pointwise 95% confidence intervals. The joint stage-by-time test yielded $p = 7.81 \times 10^{-5}$ (Table S11).

Figure S3. Cardiovascular cumulative incidence by exploratory nomogram score group. Cumulative-incidence curves for reconstructed low (≤125), intermediate (126–221) and high (≥222) score groups in training and internal validation cohorts. Non-cardiovascular deaths were competing events; shading indicates pointwise 95% confidence intervals. Global Gray's tests compare curves through 120 months (Table S17).

Supplementary Methods

1. Study population and variable definitions

The source was the SEER 17 Registries Research Plus database, November 2022 submission (database coverage: 2000–2020). The analytic subset diagnosed in 2004–2015 provided a common period for the selected tumor variables and a potential administrative observation window of at least five years; individual follow-up could end earlier because of death or loss to follow-up. Eligible patients had a single microscopically confirmed primary thyroid cancer (site C73.9) diagnosed in 2004–2015, with papillary, follicular, medullary or anaplastic histology. Exclusion criteria were absence of a known localized, regional or distant "Combined Summary Stage (2004+)" category; nonpositive or missing survival time; unknown vital status or,

among decedents, missing cause of death; or incomplete data for age, sex, race, marital status, area-level income, histology, tumor size, grade, multifocality, domicile, diagnosis year or diagnosis-to-treatment interval. These restrictions yielded the same complete-record primary cohort (n = 19,228) for the primary and additional adjusted association models. Microscopic confirmation included positive histology (n = 18,976), exfoliative cytology (n = 249) and unspecified positive microscopic confirmation (n = 3). There were 316 patients younger than 18 years; no adult-only restriction was applied.

Histological code groups are specified in Table S1. Grade I–II was classified as low grade and III–IV as high grade. Tumor size was grouped as <20, 20–40 and >40 mm; multifocality as solitary or multifocal; and diagnosis-to-treatment interval as 0 or 1–24 months. Median household income is an area-level measure, grouped as <USD 50,000, USD 50,000–<75,000 and ≥USD 75,000. Other race combines American Indian/Alaska Native and Asian/Pacific Islander categories. Unmarried combines single, domestic partner, widowed, separated and divorced categories. Recorded radiation and chemotherapy were coded yes versus no/unknown. These indicators do not distinguish dose, medication, treatment intensity or the timing of administration.

The exact exposure field was "Combined Summary Stage (2004+)." Its localized, regional and distant categories describe registry disease extent and are not equivalent to AJCC stage groups or metastatic M1 status. In the historical thyroid summary-stage scheme, distant includes specified contiguous extension as well as distant sites or nodes (Table S2). Cause of death was assigned using the SEER cause-of-death recode categories in Table S3; other deaths formed the competing endpoint. Administrative follow-up ended in 2020, with earlier censoring where follow-up ended before that date.

The primary cohort contained 19,228 patients meeting the complete-record criteria above. Broader eligibility sensitivities comprised 90,105 patients meeting the main model covariate requirements without requiring size, grade, multifocality, domicile or treatment interval, and 84,049 patients meeting the latter completeness criteria while leaving grade optional (Table S13). These populations overlap and do not constitute independent validation. The broader analyses used their specified continuous-age spline bases and adjustment sets; their estimands are comparable in type but their numerical coefficients are not interchangeable with those from the primary cohort.

2. Multivariable regression and subgroup analyses

Fine–Gray models estimated CVD subdistribution hazards in the presence of all non-CVD deaths. Cause-specific Cox models censored non-CVD death at its occurrence. The primary model included stage, sex, a four-degree-of-freedom natural spline for age, race, marital status, income, histology and diagnosis year centered at 2010. Age spline interior knots were 36, 47 and 58 years, with boundaries at 3 and 90. Reference groups were localized stage, female sex, other race, married status, highest income and follicular histology. The additional adjustment set included tumor size, grade, multifocality, domicile and diagnosis-to-treatment interval. Both adjustment sets used the same 19,228 patients and 321 CVD deaths. The additional set assessed robustness to correlated measures of tumor extent/aggressiveness (size, grade and multifocality), geographic context (domicile) and recorded care timing (diagnosis-to-treatment interval). Because the interval is postdiagnosis and may reflect disease severity or care pathways, its inclusion is a

sensitivity adjustment and does not establish that treatment-related confounding has been removed. Age-basis coefficients are not clinical age-category effects.

Sex interaction models added both sex-by-stage terms. A joint two-degree-of-freedom Wald test evaluated effect modification; individual interaction coefficients and their confidence intervals are also reported. Sex-conditional stage effects were derived from the same interaction model and covariance matrix, not inferred from differences in subgroup significance. Age 55 was used for descriptive summaries and the categorical-age nomogram, reflecting its DTC prognostic context. Age 50 was used as an exploratory midlife cutoff reflecting the period of hormonal and cardiovascular risk changes; individual hormone levels were not measured. Tables S8 and S9 report event support; zero-event strata do not identify a protective association or a clinical threshold.

3. Time-varying associations and standardized absolute risks

The Fine–Gray time model interacted the regional and distant indicators with log(1+t/12), where t is months since diagnosis. The two time coefficients were tested jointly. Time-specific sHRs have pointwise, not simultaneous, 95% confidence intervals. The time model evaluates the proportional subdistribution hazard assumption for stage; it does not constitute a global test for every covariate. A separate fit stratified the censoring distribution as a sensitivity analysis (Table S16). A late sHR below one does not imply decreasing cumulative incidence.

Cause-specific stage effects were compared across fixed intervals of 0–12, >12–36, >36–60 and >60 months. The joint interval interaction tests had six degrees of freedom. Five-year standardized risks combined the fixed-stage CVD model with the interval-varying non-CVD model and averaged predictions over the full empirical primary cohort under each stage category. Risks therefore refer to a common covariate distribution; the contrast is model-based standardization, not the effect of changing tumor stage. A sensitivity calculation allowed stage effects to vary by interval for both causes. Observed Aalen–Johansen cumulative incidence uses the actual patients in each stage group and is reported separately.

Patient-level bootstrap resampling repeated model fitting and standardization for 300 draws (seed 20261016). Of these, 286 were valid and 14 failed; percentile intervals use the valid draws. Stage risk-difference intervals were calculated from paired within-draw contrasts, not by subtracting marginal confidence limits. Failed draws were omitted rather than replaced. These intervals capture sampling and model-fitting variation under the specified models; sensitivity to the time structure is shown separately in Table S12 and Figure 5. Time-varying Fine–Gray stage associations are shown in Figure S2.

4. Exploratory nomogram and competing-risk prediction

The categorical-age nomogram point scale and the previously selected cutoffs of 125 and 221 are retained as exploratory components (Table S10). The historical X-tile display and the original unrounded scores were not available for re-verification; consequently, the cutoff values are retained as previously selected exploratory boundaries, not presented as independently reproduced thresholds. The point score is distinct from the separately

evaluated continuous-age competing-risk prediction models (Table S14). Figure S1 shows Nelson–Aalen cause-specific cumulative hazard, which is not a cumulative-incidence curve. No new absolute-risk calibration of the categorical-age point score is adopted in the present revision.

For the current score-group analysis, total scores were reconstructed by summing the displayed point assignments in Table S10. Original unrounded patient-level nomogram scores and original X-tile membership were not available for verification; these groups are therefore described as reconstructed rather than reproduced original assignments. The fixed 125 and 221 cutoffs were applied without refitting the score or searching for new cutoffs. Cardiovascular and other-death cumulative incidence were estimated separately within the frozen training and validation cohorts using competing-risk methods (cmprsk). Observations extending beyond 120 months were administratively censored at 120 months; events exactly at that time were retained. Global Gray's tests used rho = 0 over this same restricted follow-up. All three pairwise comparisons were reported, with Holm adjustment separately within each cohort and cause. Five-, eight- and ten-year cumulative incidence estimates used pointwise complementary-log-log intervals based on Aalen variance, not simultaneous bands. Point estimates were checked against an independent Aalen–Johansen recursion. Training comparisons were exploratory because the cutoffs originated in development data; the same-source validation cohort provides an internal assessment rather than external validation. These unadjusted comparisons do not establish incremental value beyond the score components. Cardiovascular risk-group curves are shown in Figure S3; full estimates, event support, and risk counts are in Table S17.

Continuous-age prediction used a fixed random split (seed 20221008): training n=13,453, with 222 CVD deaths and 972 competing deaths; validation n=5,775, with 99 CVD deaths and 420 competing deaths. There were 47 CVD deaths by five years in validation. All four models were fitted in training and applied without refitting or recalibration in validation. Age was modeled using a three-degree-of-freedom natural spline with training knots 40 and 54 years and boundary knots 3 and 90. This basis differs from the whole-cohort association analysis. The demographic predictor set comprised age, sex, race, marital status, income, domicile and diagnosis-year group (2004–2009 versus 2010–2015). The nomogram-related predictor set comprised age, sex, race, marital status, income and histology. Stage-containing versions added two stage main effects and two interactions with log(1+t)−log(61), where t is months since diagnosis; the stage main coefficients are centered at 60 months.

Competing-risk time-dependent AUC and Brier scores were evaluated at 60, 96 and 120 months using Cox-model estimates of the censoring distribution. Within each paired comparison, models used the same patients and censoring weights. Paired contrast intervals describe validation uncertainty conditional on the trained models; they do not repeat the model-development process. Five years was the primary horizon, with eight and ten years secondary. Training stage-block Wald tests evaluated association, not improvement in validation performance. Calibration used quintiles of predicted risk with observed CVD cumulative incidence; no calibration slope, calibration-in-the-large or uncertainty intervals for these points were available. Full comparisons, including secondary findings, are reported in Table S14.

5. Treatment-related sensitivity analyses and diagnostics

Post hoc surgery-restricted analyses evaluated all eligible histologies and a fixed papillary/follicular DTC subset. The primary surgical population contained 19,047 patients, with 319 CVD and 1,222 competing deaths. Its DTC subset contained 18,675 patients, with 312 CVD and 973 competing deaths. The paired broader treatment-analysis population contained 84,040 patients before surgical restriction, distinct from the 84,049-patient eligibility sensitivity population. Its surgical all-histology and DTC subsets contained 83,713 and 82,174 patients, respectively (Table S15).

Treatment-adjusted models included stage, a three-degree-of-freedom natural age spline, sex, race, marital status, income, domicile, diagnosis-year group, histology, size, multifocality, recorded radiation and chemotherapy; grade was included in the primary surgical population and unavailable by design in the broader population. Cause-specific models additionally stratified by five-year age bands. Surgery was an eligibility criterion, not a varying covariate. Fine–Gray models used log(1+t)−log(61), with time in months, so main stage coefficients refer to 60 months. One specification allowed stage effects to vary; another allowed all modeled covariate effects to vary. Four cause-specific joint stage tests and eight Fine–Gray joint stage-time tests were adjusted in separate Holm families. Individual confidence intervals are pointwise and nominal. All specified fits are reported, without selecting models by significance.

Restriction to surgery and adjustment for postdiagnosis treatment can introduce selection, collider and treatment-timing bias. These fits are sensitivity analyses of associations, not causal effects of treatment or deployable baseline prediction models. Four separate alternative non-CVD models retained residual nonproportionality after attempted adjustment and were withheld from inferential interpretation (Table S16). They used a different restricted-time specification from the models producing Table S12; their diagnostic failure does not invalidate or replace the reported primary standardized-risk analysis. Detailed Cox proportional-hazards diagnostics are provided in Table S16.

Table S1. Morphological codes of histological types of thyroid cancer

Histological types	Morphological codes
PTC	8050, 8052, 8260, 8340, 8341, 8342, 8343, 8344, 8350, 8450
FTC	8290, 8330, 8331, 8332, 8335, 8337, 8339
MTC	8345, 8510
ATC	8004, 8012, 8020, 8021, 8022, 8030, 8031, 8032, 8033, 8034, 8035

Histological types	Morphological codes
Other histologies (excluded)	8000, 8001, 8010, 8011, 8013, 8015, 8041, 8046, 8070, 8071, 8072, 8074, 8075, 8120, 8130, 8140, 8141, 8190, 8201, 8230, 8240, 8246, 8255, 8262, 8263, 8265, 8310, 8323, 8346, 8347, 8408, 8430, 8452, 8460, 8480, 8481, 8502, 8504, 8507, 8560, 8570, 8574, 8582, 8586, 8588, 8589, 8680, 8800, 8801, 8802, 8805, 8810, 8811, 8815, 8890, 8891, 8980, 8982, 9040, 9041, 9071, 9080, 9085, 9120, 9250, 9260, 9364

The four eligible histological groups are shown above; the other codes were excluded. TC, thyroid cancer; PTC, papillary thyroid carcinoma; FTC, follicular thyroid carcinoma; MTC, medullary thyroid carcinoma; ATC, anaplastic thyroid carcinoma.

Table S2. Classification and detailed description of Surveillance, Epidemiology, and End Results summary stage

SEER stage	Definition
Localized	Single or multifocal invasive tumor confined to the thyroid, including capsular involvement without extension beyond the capsule.
Regional	Regional direct extension and/or regional lymph node involvement according to the thyroid summary-stage scheme.
Distant	Distant sites or distant lymph nodes, or specified extensive contiguous extension, including trachea, bone, mediastinal tissue and certain skeletal muscles in the historical thyroid scheme. This is not synonymous with AJCC M1 disease.

Exact variable: Combined Summary Stage (2004+). Source: National Cancer Institute, SEER Summary Staging Manual 2000, thyroid gland chapter, printed pp. 270–271. https://seer.cancer.gov/tools/ssm/ssm2000/SSSM2000-122012.pdf (accessed 17 October 2026).

Table S3. Specific causes of cardiovascular death

ICD-10 codes	Causes of cardiovascular death
I00–I09, I11, I13, I20–I51	Diseases of heart
I10, I12	Hypertension without heart disease
I60–I69	Cerebrovascular disease
I70	Atherosclerosis
I71	Aortic aneurysm and dissection

ICD-10 codes	Causes of cardiovascular death
I72–I78	Other diseases of arteries, arterioles and capillaries

The endpoint uses these six SEER cause-of-death recode groups. ICD-10 ranges describe the corresponding causes; all other deaths are competing events. Source: National Cancer Institute, SEER Cause of Death Recode 1969+ (03/01/2018). https://seer.cancer.gov/codrecode/1969_d03012018/index.html

Table S4. Baseline characteristics of training and validation cohorts

Variable	Category	Training n=13,453	Validation n=5,775
Age at diagnosis, years	<55	9,159 (68.1)	3,929 (68.0)
Age at diagnosis, years	≥55	4,294 (31.9)	1,846 (32.0)
Year of diagnosis	2004–2009	3,965 (29.5)	1,719 (29.8)
Year of diagnosis	2010–2015	9,488 (70.5)	4,056 (70.2)
Sex	Female	10,210 (75.9)	4,371 (75.7)
Sex	Male	3,243 (24.1)	1,404 (24.3)
Race	Other	1,724 (12.8)	797 (13.8)
Race	White	10,873 (80.8)	4,594 (79.5)
Race	Black	856 (6.4)	384 (6.6)
Marital status	Married	8,610 (64.0)	3,746 (64.9)
Marital status	Unmarried	4,843 (36.0)	2,029 (35.1)
Median household income, USD thousands	≥75	5,618 (41.8)	2,417 (41.9)
Median household income, USD thousands	50–<75	6,434 (47.8)	2,772 (48.0)

Variable	Category	Training n=13,453	Validation n=5,775
Median household income, USD thousands	<50	1,401 (10.4)	586 (10.1)
Residence	Metropolitan	12,018 (89.3)	5,167 (89.5)
Residence	Nonmetropolitan	1,435 (10.7)	608 (10.5)
Histological type	Papillary	11,825 (87.9)	5,070 (87.8)
Histological type	Follicular	1,265 (9.4)	545 (9.4)
Histological type	Medullary	77 (0.6)	35 (0.6)
Histological type	Anaplastic	286 (2.1)	125 (2.2)
Tumor size, mm	<20	7,661 (56.9)	3,315 (57.4)
Tumor size, mm	20–40	3,977 (29.6)	1,716 (29.7)
Tumor size, mm	>40	1,815 (13.5)	744 (12.9)
Tumor grade	Low	12,474 (92.7)	5,356 (92.7)
Tumor grade	High	979 (7.3)	419 (7.3)
SEER summary stage	Localized	8,691 (64.6)	3,778 (65.4)
SEER summary stage	Regional	4,080 (30.3)	1,743 (30.2)
SEER summary stage	Distant	682 (5.1)	254 (4.4)
Tumor focality	Solitary	7,951 (59.1)	3,388 (58.7)
Tumor focality	Multifocal	5,502 (40.9)	2,387 (41.3)
Diagnosis to treatment, months	0	8,114 (60.3)	3,461 (59.9)

Variable	Category	Training n=13,453	Validation n=5,775
Diagnosis to treatment, months	1–24	5,339 (39.7)	2,314 (40.1)
Recorded radiation	No/unknown	6,379 (47.4)	2,809 (48.6)
Recorded radiation	Yes	7,074 (52.6)	2,966 (51.4)
Recorded chemotherapy	No/unknown	13,190 (98.0)	5,676 (98.3)
Recorded chemotherapy	Yes	263 (2.0)	99 (1.7)
Surgery	No	123 (0.9)	58 (1.0)
Surgery	Yes	13,330 (99.1)	5,717 (99.0)
Follow-up status	Censored	12,259 (91.1)	5,256 (91.0)
Follow-up status	Cardiovascular death	222 (1.7)	99 (1.7)
Follow-up status	Other death	972 (7.2)	420 (7.3)

Values are n (%), using each split denominator. Category definitions follow Supplementary Methods 1. Baseline significance tests of a random split were not used. Low-grade validation tumors: 5,356/5,775 (92.7%).

Table S5. Primary multivariable associations with cardiovascular death

Contrast or coefficient	Fine–Gray sHR (95% CI)	p	Cause-specific csHR (95% CI)	p
Regional vs localized	1.16 (0.90–1.50)	0.247	1.32 (1.03–1.71)	0.029
Distant vs localized	1.10 (0.74–1.62)	0.647	2.46 (1.71–3.55)	<0.001
Male vs female	2.19 (1.73–2.78)	<0.001	2.30 (1.82–2.90)	<0.001
Age spline basis 1 (log coefficient)	2.5767 (SE 2.3385)	0.271	2.5991 (SE 2.4485)	0.288
Age spline basis 2 (log coefficient)	4.3266 (SE 1.3663)	0.002	4.2748 (SE 1.4280)	0.003

Contrast or coefficient	Fine–Gray sHR (95% CI)	p	Cause-specific csHR (95% CI)	p
Age spline basis 3 (log coefficient)	6.7278 (SE 4.9740)	0.176	6.8672 (SE 5.2277)	0.189
Age spline basis 4 (log coefficient)	6.3659 (SE 0.9914)	<0.001	6.9757 (SE 1.0279)	<0.001
White vs other race	1.35 (0.91–2.02)	0.140	1.37 (0.92–2.03)	0.122
Black vs other race	2.91 (1.71–4.95)	<0.001	3.00 (1.81–4.98)	<0.001
Unmarried vs married	1.63 (1.28–2.09)	<0.001	1.63 (1.29–2.07)	<0.001
Income USD 50,000–<75,000 vs ≥75,000	0.98 (0.77–1.25)	0.873	1.02 (0.80–1.31)	0.843
Income <USD 50,000 vs ≥75,000	1.39 (0.98–1.98)	0.066	1.53 (1.08–2.15)	0.016
Papillary vs follicular	0.95 (0.70–1.28)	0.726	0.91 (0.67–1.22)	0.512
Medullary vs follicular	0.41 (0.09–1.77)	0.232	0.56 (0.14–2.31)	0.423
Anaplastic vs follicular	0.19 (0.08–0.43)	<0.001	0.90 (0.40–2.02)	0.806
Diagnosis year (per year)	0.96 (0.93–0.99)	0.021	0.98 (0.95–1.02)	0.388

N=19,228, with 321 cardiovascular deaths. The primary adjustment set includes stage, sex, a four-degree-of-freedom natural age spline, race, marital status, area-level income, histology and diagnosis year. Localized stage and follicular histology are the corresponding reference categories. Non-cardiovascular deaths are competing events in Fine–Gray models and are censored in cause-specific Cox models. Age-basis rows give log coefficients and standard errors, not age-category hazard ratios. Time-dependent stage estimates are reported in Table S11. sHR, subdistribution hazard ratio; csHR, cause-specific hazard ratio; CI, confidence interval; SE, standard error.

Table S6. Additional covariate-adjusted associations with cardiovascular death

Contrast or coefficient	Fine–Gray sHR (95% CI)	p	Cause-specific csHR (95% CI)	p
Regional vs localized	1.21 (0.93–1.57)	0.157	1.28 (0.98–1.67)	0.073
Distant vs localized	1.23 (0.79–1.92)	0.355	2.32 (1.57–3.43)	<0.001

Contrast or coefficient	Fine–Gray sHR (95% CI)	p	Cause-specific csHR (95% CI)	p
Male vs female	2.22 (1.74–2.83)	<0.001	2.26 (1.78–2.86)	<0.001
Age spline basis 1 (log coefficient)	2.5991 (SE 2.3103)	0.261	2.5839 (SE 2.4336)	0.288
Age spline basis 2 (log coefficient)	4.3400 (SE 1.3509)	0.001	4.2532 (SE 1.4197)	0.003
Age spline basis 3 (log coefficient)	6.7755 (SE 4.9157)	0.168	6.8401 (SE 5.1961)	0.188
Age spline basis 4 (log coefficient)	6.4542 (SE 0.9810)	<0.001	6.9569 (SE 1.0231)	<0.001
White vs other race	1.35 (0.91–2.02)	0.140	1.39 (0.94–2.06)	0.103
Black vs other race	2.83 (1.66–4.84)	<0.001	2.90 (1.74–4.82)	<0.001
Unmarried vs married	1.64 (1.28–2.10)	<0.001	1.60 (1.26–2.03)	<0.001
Income USD 50,000–<75,000 vs ≥75,000	1.00 (0.78–1.27)	0.972	1.03 (0.81–1.32)	0.794
Income <USD 50,000 vs ≥75,000	1.72 (1.10–2.69)	0.018	1.94 (1.25–3.01)	0.003
Papillary vs follicular	0.93 (0.67–1.29)	0.665	0.96 (0.70–1.34)	0.824
Medullary vs follicular	0.48 (0.11–2.06)	0.322	0.61 (0.15–2.52)	0.493
Anaplastic vs follicular	0.23 (0.10–0.55)	<0.001	0.76 (0.33–1.76)	0.522
Diagnosis year (per year)	0.96 (0.93–0.99)	0.012	0.99 (0.95–1.03)	0.565
Tumor size 20–40 vs <20 mm	1.11 (0.85–1.45)	0.457	1.11 (0.85–1.46)	0.445
Tumor size >40 vs <20 mm	1.02 (0.73–1.43)	0.894	1.13 (0.81–1.58)	0.461
High vs low grade	0.74 (0.50–1.10)	0.141	1.31 (0.91–1.88)	0.145
Multifocal vs solitary	1.16 (0.92–1.46)	0.214	1.16 (0.92–1.46)	0.224
Nonmetropolitan vs metropolitan	0.68 (0.44–1.06)	0.087	0.67 (0.43–1.03)	0.068

Contrast or coefficient	Fine–Gray sHR (95% CI)	p	Cause-specific csHR (95% CI)	p
Treatment interval 0 vs 1–24 months	1.17 (0.92–1.48)	0.210	1.17 (0.92–1.50)	0.208

N=19,228, with 321 cardiovascular deaths, using the same patients as Table S5. This sensitivity model adds tumor size, grade, multifocality, domicile and diagnosis-to-treatment interval to the primary adjustment set. Localized stage and follicular histology are the corresponding reference categories. Age-basis rows give log coefficients and standard errors, not age-category hazard ratios. The primary and additional adjustment sets are specified in Supplementary Methods 2. sHR, subdistribution hazard ratio; csHR, cause-specific hazard ratio; CI, confidence interval; SE, standard error.

Table S7. Sex-conditional stage associations and formal interaction tests

Adjustment / model	Sex	Stage vs localized	Estimate (95% CI)	p
Primary / sex interaction / Fine–Gray	Female	Regional vs Localized	1.33 (0.96–1.85)	0.086
Primary / sex interaction / Fine–Gray	Female	Distant vs Localized	0.91 (0.52–1.60)	0.752
Primary / sex interaction / Fine–Gray	Male	Regional vs Localized	0.98 (0.67–1.43)	0.906
Primary / sex interaction / Fine–Gray	Male	Distant vs Localized	1.26 (0.75–2.12)	0.381
Primary / sex interaction / Cause-specific Cox	Female	Regional vs Localized	1.50 (1.09–2.08)	0.014
Primary / sex interaction / Cause-specific Cox	Female	Distant vs Localized	2.23 (1.30–3.83)	0.003
Primary / sex interaction / Cause-specific Cox	Male	Regional vs Localized	1.12 (0.76–1.65)	0.558
Primary / sex interaction / Cause-specific Cox	Male	Distant vs Localized	2.60 (1.60–4.24)	<0.001
Additional adjustment / sex interaction / Fine–Gray	Female	Regional vs Localized	1.37 (0.98–1.90)	0.065
Additional adjustment / sex interaction / Fine–Gray	Female	Distant vs Localized	1.02 (0.55–1.91)	0.944
Additional adjustment / sex interaction / Fine–Gray	Male	Regional vs Localized	1.03 (0.70–1.51)	0.887
Additional adjustment / sex interaction / Fine–Gray	Male	Distant vs Localized	1.41 (0.82–2.43)	0.209

Adjustment / model	Sex	Stage vs localized	Estimate (95% CI)	p
Additional adjustment / sex interaction / Cause-specific Cox	Female	Regional vs Localized	1.44 (1.02–2.01)	0.037
Additional adjustment / sex interaction / Cause-specific Cox	Female	Distant vs Localized	2.12 (1.21–3.70)	0.008
Additional adjustment / sex interaction / Cause-specific Cox	Male	Regional vs Localized	1.10 (0.74–1.63)	0.644
Additional adjustment / sex interaction / Cause-specific Cox	Male	Distant vs Localized	2.45 (1.48–4.05)	<0.001

Direct multiplicative interaction contrasts

Adjustment / model	Interaction term	Ratio of hazard ratios (95% CI)	p
Primary / sex interaction / Fine–Gray	Regional vs localized:Male vs female	0.73 (0.45–1.20)	0.218
Primary / sex interaction / Fine–Gray	Distant vs localized:Male vs female	1.38 (0.66–2.88)	0.390
Primary / sex interaction / Cause-specific Cox	Regional vs localized:Male vs female	0.75 (0.45–1.23)	0.251
Primary / sex interaction / Cause-specific Cox	Distant vs localized:Male vs female	1.17 (0.57–2.38)	0.673
Additional adjustment / sex interaction / Fine–Gray	Regional vs localized:Male vs female	0.75 (0.46–1.23)	0.258
Additional adjustment / sex interaction / Fine–Gray	Distant vs localized:Male vs female	1.38 (0.66–2.90)	0.389
Additional adjustment / sex interaction / Cause-specific Cox	Regional vs localized:Male vs female	0.77 (0.46–1.26)	0.295
Additional adjustment / sex interaction / Cause-specific Cox	Distant vs localized:Male vs female	1.16 (0.57–2.37)	0.688

Joint tests

Model	Test	χ^2	df	p
Primary / Fine–Gray	Joint stage test	1.355	2	0.508

Model	Test	χ^2	df	p
Primary / Cause-specific Cox	Joint stage test	24.116	2	<0.001
Primary / sex interaction / Fine–Gray	Joint sex-by-stage interaction test	2.917	2	0.233
Primary / sex interaction / Cause-specific Cox	Joint sex-by-stage interaction test	1.847	2	0.397
Additional adjustment / Fine–Gray	Joint stage test	2.197	2	0.333
Additional adjustment / Cause-specific Cox	Joint stage test	18.088	2	<0.001
Additional adjustment / sex interaction / Fine–Gray	Joint sex-by-stage interaction test	2.593	2	0.274
Additional adjustment / sex interaction / Cause-specific Cox	Joint sex-by-stage interaction test	1.562	2	0.458

Sex-conditional estimates are sHRs for Fine–Gray models and csHRs for cause-specific Cox models, derived from a common sex-by-stage interaction model. Direct interaction contrasts are ratios of the male and female stage hazard ratios. Joint sex-by-stage Wald tests assess both interaction terms simultaneously; the joint stage tests shown for models without sex-by-stage terms assess the two stage coefficients and address a separate question. Individual within-sex significance does not establish effect modification. CI, confidence interval; df, degrees of freedom.

Table S8. Age-stratified stage and outcome counts in males

Age, years	Stage	N	Cardiovascular deaths	Other deaths	Censored
<50	Localized	1227	8	30	1189
<50	Regional	931	6	46	879
<50	Distant	85	0	30	55
≥50	Localized	1360	65	124	1171
≥50	Regional	732	35	143	554
≥50	Distant	312	22	230	60

Exploratory age cutoff. Zero-event distant-stage strata below age 50 do not support stable adjusted stage estimates. Hormonal and menopausal status were not recorded; no menopausal classification or biological threshold is assigned.

Table S9. Age-stratified stage and outcome counts in females

Age, years	Stage	N	Cardiovascular deaths	Other deaths	Censored
<50	Localized	5716	10	51	5655
<50	Regional	2793	3	30	2760
<50	Distant	122	0	19	103
≥50	Localized	4166	103	200	3863
≥50	Regional	1367	53	191	1123
≥50	Distant	417	16	298	103

Exploratory age cutoff. Zero-event distant-stage strata below age 50 do not support stable adjusted stage estimates. Hormonal and menopausal status were not recorded; no menopausal classification or biological threshold is assigned.

Table S10. Specific scores of prognostic factors in the exploratory nomogram

Variables	Nomogram Point
Age at diagnosis (years)	
< 55	0.0
≥ 55	100.0
Race	
Other[a]	0.0
White	12.0

Variables	Nomogram Point
Black	37.0
Sex	
Female	0.0
Male	35.0
Marital status	
Married	0.0
Unmarried[b]	30.0
Median household income (USD thousands)	
≥75	0.0
50–<75	8.0
< 50	25.0
Histologic type	
Anaplastic	0.0
Medullary	36.0
Papillary	43.0
Follicular	52.0
SEER stage	
Localized	0.0

Variables	Nomogram Point
Regional	11.0
Distant	23.0

aOther includes American Indian/AK Native, Asian/Pacific Islander;
bUnmarried includes Single (never married)/Unmarried or Domestic Partner/Widowed/Separated/Divorced.
SEER indicates Surveillance, Epidemiology, and End Results.

Point assignments are retained from the exploratory categorical-age nomogram and are distinct from coefficients of the continuous-age competing-risk models. The displayed points were used to reconstruct the current score groups in the Table S17 analyses; original unrounded scores and original X-tile membership were not verified. The lower anaplastic point assignment must not be interpreted as cardiovascular protection. Cutoffs of 125 and 221 are previously selected exploratory boundaries, not clinical decision thresholds.

Table S11. Time-dependent stage associations with cardiovascular death

Stage vs localized	Time, months	sHR (95% CI)
Regional	6	1.34 (0.77–2.34)
Regional	12	1.30 (0.82–2.08)
Regional	24	1.25 (0.88–1.77)
Regional	36	1.21 (0.91–1.62)
Regional	60	1.16 (0.90–1.50)
Distant	6	3.85 (2.00–7.41)
Distant	12	2.79 (1.61–4.83)
Distant	24	1.78 (1.15–2.75)
Distant	36	1.29 (0.85–1.94)

Stage vs localized	Time, months	sHR (95% CI)
Distant	60	0.82 (0.52–1.30)

Primary covariate adjustment, N=19,228. Stage-by-log(1+t/12) joint test: χ^2=18.9139, df=2, p=7.81 × 10^{-5}, with t measured in months. Estimates are time-specific subdistribution hazard ratios (sHRs) with pointwise 95% confidence intervals (CIs), not absolute risks. The 12- and 60-month estimates correspond to Results 3.2. Cause-specific interval tests and the censoring-stratified sensitivity fit are reported in Table S16.

Table S12. Five-year absolute risks and stage risk differences

Stage	Measure at five years	Estimate (95% CI)
Localized	Cardiovascular CIF, %	0.770 (0.635 to 0.925)
Regional	Cardiovascular CIF, %	0.868 (0.671 to 1.112)
Distant	Cardiovascular CIF, %	1.076 (0.704 to 1.467)
Localized	Cardiovascular RD, percentage points	0.000 (0.000 to 0.000)
Regional	Cardiovascular RD, percentage points	0.098 (-0.077 to 0.309)
Distant	Cardiovascular RD, percentage points	0.306 (-0.055 to 0.694)
Localized	Other-death CIF, %	2.129 (1.801 to 2.483)
Regional	Other-death CIF, %	5.869 (5.270 to 6.509)
Distant	Other-death CIF, %	17.895 (16.371 to 20.175)
Localized	Other-death RD, percentage points	0.000 (0.000 to 0.000)
Regional	Other-death RD, percentage points	3.740 (3.122 to 4.393)
Distant	Other-death RD, percentage points	15.765 (14.062 to 18.044)

CIF, cumulative incidence function; RD, risk difference versus localized disease; CI, confidence interval. CIFs are percentages; RDs are percentage points. Risks are standardized over the empirical primary cohort using the cause-specific models in Supplementary Methods 3. The five-year cardiovascular CIFs of 0.770%, 0.868% and 1.076% are the estimates reported in Results 3.2. Bootstrap percentile confidence intervals use 286 valid draws of 300; 14 failed. Differences are paired within draws. Reference differences are defined as zero, not estimated contrasts.

Observed unadjusted cumulative incidence

Stage	Endpoint	Observed CIF, % (95% pointwise CI)
Localized	Cardiovascular death	0.646 (0.504–0.788)
Localized	Other death	1.572 (1.352–1.793)
Regional	Cardiovascular death	0.864 (0.623–1.106)
Regional	Other death	4.735 (4.185–5.286)
Distant	Cardiovascular death	2.709 (1.661–3.757)
Distant	Other death	54.592 (51.378–57.806)

Observed and standardized risks have different covariate distributions and are not interchangeable.

Alternative time structure for standardized risk

Stage	CVD CIF, %	Other-death CIF, %	CVD RD, pp	Other-death RD, pp
Localized	0.735	2.132	0.000	0.000
Regional	0.912	5.885	0.176	3.753
Distant	1.078	17.835	0.343	15.703

This sensitivity specification allows interval-varying stage effects for both causes. It differs from the primary standardized-risk specification above. Point estimates only are available; no confidence intervals are imputed. CVD, cardiovascular disease; CIF, cumulative incidence function; RD, risk difference versus localized disease; pp, percentage points.

Table S13. Broader eligibility sensitivity analyses

Eligibility population	N	CVD deaths	Other deaths	Censored	Missing grade
Broader model-covariate eligibility	90,105	1,781	5,232	83,092	69,931
Grade-optional, other completeness retained	84,049	1,584	4,250	78,215	64,818

Population N	Model / endpoint	Stage vs localized	Estimate (95% CI)	p
90,105	Fine–Gray / cardiovascular death	Regional vs localized	1.002 (0.894–1.123)	0.976
90,105	Fine–Gray / cardiovascular death	Distant vs localized	0.715 (0.582–0.879)	0.001
90,105	Cause-specific Cox / cardiovascular death	Regional vs localized	1.131 (1.011–1.265)	0.031
90,105	Cause-specific Cox / cardiovascular death	Distant vs localized	1.701 (1.412–2.050)	<0.001
84,049	Fine–Gray / cardiovascular death	Regional vs localized	1.016 (0.902–1.146)	0.790
84,049	Fine–Gray / cardiovascular death	Distant vs localized	0.794 (0.629–1.003)	0.053
84,049	Cause-specific Cox / cardiovascular death	Regional vs localized	1.124 (0.999–1.265)	0.053
84,049	Cause-specific Cox / cardiovascular death	Distant vs localized	1.612 (1.298–2.000)	<0.001

Adjustment includes stage, sex, a continuous-age natural spline, race, marital status, income, histology and diagnosis year. Fine–Gray entries are subdistribution hazard ratios; cause-specific Cox entries are cause-specific hazard ratios, each with 95% confidence intervals. The broader populations relax completeness requirements and differ in composition from the primary cohort; they overlap and are not independent validation samples. The 90,105-patient distant-stage comparison is reported in Results 3.2. Proportional subdistribution hazards were not formally assessed in these broader eligibility fits; their constant estimates are sensitivity summaries, not validated temporal effects. CVD, cardiovascular disease.

Table S14. Competing-risk prediction models and validation performance

Predictor set	Base covariates	Stage terms	Parameters

Predictor set	Base covariates	Stage terms	Parameters
Demographic	Age spline, sex, race, marital status, income, domicile, year group	None	11
Demographic + stage	Same demographic set	Two stage main + two stage-time terms	15
Nomogram-related	Age spline, sex, race, marital status, income, histology	None	12
Nomogram-related + stage	Same histology-containing set	Two stage main + two stage-time terms	16

All models use training n=13,453 (222 cardiovascular deaths) and validation n=5,775 (99 cardiovascular deaths). The demographic and nomogram-related base models use different covariate sets, as listed above: the latter contains histology and omits domicile and diagnosis-year group. Thus, the two bases are not a nested comparison obtained solely by adding histology. Within each base-model family, the stage-containing version adds two stage main effects and two stage-by-time terms. Age uses a three-degree-of-freedom natural spline, with knots 40 and 54 and boundaries 3 and 90. The stage time function is log(1+t)−log(61), with t in months and reference time 60 months. These continuous-age models are distinct from the categorical-age point scale in Table S10.

Stage-containing model	Training stage-block χ^2	df	p
Demographic + stage	21.997	4	<0.001
Nomogram-related + stage	21.345	4	<0.001

Training stage-block tests assess four coefficients jointly; they are not tests of out-of-sample improvement.

Model	Years	AUC (95% CI)	Brier score (95% CI)
Demographic	5	0.8584 (0.8155–0.9013)	0.008152 (0.005937–0.010366)
Demographic	8	0.8772 (0.8444–0.9100)	0.013292 (0.010449–0.016136)
Demographic	10	0.8942 (0.8616–0.9268)	0.017427 (0.013059–0.021795)
Demographic + stage	5	0.8584 (0.8152–0.9016)	0.008145 (0.005936–0.010354)
Demographic + stage	8	0.8768 (0.8441–0.9095)	0.013295 (0.010455–0.016135)

Model	Years	AUC (95% CI)	Brier score (95% CI)
Demographic + stage	10	0.8945 (0.8618–0.9272)	0.017376 (0.013044–0.021708)
Nomogram-related	5	0.8543 (0.8120–0.8966)	0.008180 (0.005971–0.010390)
Nomogram-related	8	0.8760 (0.8433–0.9087)	0.013214 (0.010394–0.016034)
Nomogram-related	10	0.8948 (0.8619–0.9277)	0.017163 (0.012879–0.021447)
Nomogram-related + stage	5	0.8541 (0.8110–0.8972)	0.008142 (0.005942–0.010342)
Nomogram-related + stage	8	0.8765 (0.8438–0.9093)	0.013186 (0.010371–0.016002)
Nomogram-related + stage	10	0.8951 (0.8624–0.9278)	0.017152 (0.012870–0.021434)

AUC, area under the time-dependent receiver operating characteristic curve; CI, confidence interval. Higher AUC and lower Brier score indicate better performance. Confidence intervals describe validation uncertainty conditional on fitted training models. Five years is the primary horizon; eight and ten years are secondary. The nomogram-related model with stage yields validation AUCs of 0.8541, 0.8765 and 0.8951 at 5, 8 and 10 years, respectively; these are rounded to 0.854, 0.877 and 0.895 in Results 3.4 and the Figure 4 legend. They are not performance estimates for the original categorical-age nomogram.

Paired changes after adding stage

Predictor set	Years	Δ AUC (95% CI)	p	Δ Brier (95% CI)	p
Demographic	5	-0.000031 (-0.001141, +0.001079)	0.956	-0.00000663 (-0.00002849, +0.00001524)	0.552
Nomogram-related	5	-0.000201 (-0.002973, +0.002572)	0.887	-0.00003788 (-0.00009682, +0.00002105)	0.208
Demographic	8	-0.000415 (-0.002615, +0.001785)	0.712	+0.00000267 (-0.00006311, +0.00006846)	0.937
Nomogram-related	8	+0.000536 (-0.000441, +0.001514)	0.282	-0.00002729 (-0.00004744, -0.00000714)	0.008
Demographic	10	+0.000292 (-0.002341, +0.002925)	0.828	-0.00005115 (-0.00017282, +0.00007051)	0.410
Nomogram-related	10	+0.000332 (-0.000694, +0.001358)	0.526	-0.00001120 (-0.00004084, +0.00001844)	0.459

Paired changes are stage-containing minus base model within the same predictor-set family. Negative Brier changes and positive AUC changes favor adding stage. All horizons and both predictor-set families are retained. The nomogram-related eight-year comparison had a small lower Brier score (nominal p=0.008); this secondary result does not establish a clear five-year incremental benefit or a clinical decision threshold. AUC, area under the time-dependent receiver operating characteristic curve; CI, confidence interval.

Grouped validation calibration

Model	Years	Quintile	N	CVD deaths by horizon	Mean predicted risk, %	Observed CIF, %
Demographic	5	1	1156	0	0.0383	0.0000
Demographic	5	2	1159	1	0.0866	0.0890
Demographic	5	3	1150	2	0.2023	0.1797
Demographic	5	4	1155	10	0.5419	0.8752
Demographic	5	5	1155	34	3.1697	2.9878
Demographic	8	1	1156	0	0.0717	0.0000
Demographic	8	2	1159	1	0.1621	0.0890
Demographic	8	3	1150	3	0.3785	0.2944
Demographic	8	4	1155	13	1.0124	1.2423
Demographic	8	5	1155	57	5.7677	5.5575
Demographic	10	1	1156	0	0.0871	0.0000
Demographic	10	2	1159	1	0.1970	0.0890
Demographic	10	3	1150	3	0.4598	0.2944
Demographic	10	4	1155	15	1.2287	1.6380
Demographic	10	5	1155	62	6.9186	6.6534

Model	Years	Quintile	N	CVD deaths by horizon	Mean predicted risk, %	Observed CIF, %
Demographic + stage	5	1	1155	0	0.0351	0.0000
Demographic + stage	5	2	1155	1	0.0796	0.0891
Demographic + stage	5	3	1158	2	0.1878	0.1784
Demographic + stage	5	4	1152	10	0.5139	0.8775
Demographic + stage	5	5	1155	34	3.2015	2.9890
Demographic + stage	8	1	1155	0	0.0701	0.0000
Demographic + stage	8	2	1155	1	0.1583	0.0892
Demographic + stage	8	3	1154	3	0.3722	0.2925
Demographic + stage	8	4	1156	13	1.0028	1.2429
Demographic + stage	8	5	1155	57	5.8219	5.5483
Demographic + stage	10	1	1157	0	0.0868	0.0000
Demographic + stage	10	2	1153	1	0.1956	0.0894
Demographic + stage	10	3	1155	3	0.4592	0.2921
Demographic + stage	10	4	1152	17	1.2287	1.8325
Demographic + stage	10	5	1158	60	6.9870	6.4312
Nomogram-related	5	1	1155	0	0.0377	0.0000
Nomogram-related	5	2	1152	1	0.0847	0.0894
Nomogram-related	5	3	1150	2	0.1953	0.1799
Nomogram-related	5	4	1159	12	0.5117	1.0464

Model	Years	Quintile	N	CVD deaths by horizon	Mean predicted risk, %	Observed CIF, %
Nomogram-related	5	5	1159	32	3.2079	2.8011
Nomogram-related	8	1	1155	0	0.0709	0.0000
Nomogram-related	8	2	1152	1	0.1593	0.0894
Nomogram-related	8	3	1150	3	0.3670	0.2936
Nomogram-related	8	4	1159	15	0.9603	1.3704
Nomogram-related	8	5	1159	55	5.8357	5.3928
Nomogram-related	10	1	1155	0	0.0863	0.0000
Nomogram-related	10	2	1152	1	0.1941	0.0894
Nomogram-related	10	3	1150	3	0.4470	0.2936
Nomogram-related	10	4	1159	17	1.1687	1.7706
Nomogram-related	10	5	1159	60	7.0051	6.4634
Nomogram-related + stage	5	1	1155	0	0.0348	0.0000
Nomogram-related + stage	5	2	1154	1	0.0789	0.0891
Nomogram-related + stage	5	3	1155	2	0.1828	0.1794
Nomogram-related + stage	5	4	1156	13	0.4871	1.1349
Nomogram-related + stage	5	5	1155	31	3.2201	2.7263
Nomogram-related + stage	8	1	1157	0	0.0697	0.0000
Nomogram-related + stage	8	2	1153	1	0.1569	0.0892
Nomogram-related + stage	8	3	1157	3	0.3635	0.2917

Model	Years	Quintile	N	CVD deaths by horizon	Mean predicted risk, %	Observed CIF, %
Nomogram-related + stage	8	4	1155	16	0.9606	1.4620
Nomogram-related + stage	8	5	1153	54	5.8694	5.3367
Nomogram-related + stage	10	1	1150	0	0.0860	0.0000
Nomogram-related + stage	10	2	1162	1	0.1937	0.0886
Nomogram-related + stage	10	3	1153	3	0.4490	0.2928
Nomogram-related + stage	10	4	1155	18	1.1797	1.8533
Nomogram-related + stage	10	5	1155	59	7.0530	6.4155

Quintiles are defined separately for each model and horizon. CVD, cardiovascular disease; CIF, cumulative incidence function. Observed cardiovascular CIF accounts for competing death and censoring. No confidence intervals for calibration points, calibration slope or calibration-in-the-large are inferred. Nonuniform agreement remains apparent; grouped calibration does not establish clinical utility.

Table S15. Surgery-restricted treatment-adjusted sensitivity analyses

Source population	Histology restriction	Surgical N	CVD deaths	Other deaths
Primary	All histologies	19,047	319	1,222
Primary	DTC	18,675	312	973
Broader treatment population	All histologies	83,713	1,571	3,996
Broader treatment population	DTC	82,174	1,533	3,577

The broader treatment source contains 84,040 patients before surgical restriction and differs from the 84,049-person grade-optional sensitivity cohort. Primary DTC contains 18,705 patients before surgical restriction. DTC comprises papillary and follicular carcinoma.

Population / model	Stage vs localized	Estimate (95% CI)	p

Population / model	Stage vs localized	Estimate (95% CI)	p
Primary surgical / all histologies / cause-specific	Regional vs localized	1.37 (1.05–1.81)	0.023
Primary surgical / all histologies / cause-specific	Distant vs localized	2.58 (1.73–3.83)	<0.001
Primary surgical / all histologies / stage-time	Regional vs localized	1.24 (0.94–1.63)	0.122
Primary surgical / all histologies / stage-time	Distant vs localized	0.99 (0.61–1.61)	0.967
Primary surgical / all histologies / all-covariate time	Regional vs localized	1.25 (0.95–1.64)	0.115
Primary surgical / all histologies / all-covariate time	Distant vs localized	1.05 (0.64–1.72)	0.850
Primary surgical / DTC / cause-specific	Regional vs localized	1.34 (1.02–1.77)	0.039
Primary surgical / DTC / cause-specific	Distant vs localized	2.46 (1.64–3.70)	<0.001
Primary surgical / DTC / stage-time	Regional vs localized	1.22 (0.93–1.61)	0.157
Primary surgical / DTC / stage-time	Distant vs localized	1.02 (0.62–1.68)	0.930
Primary surgical / DTC / all-covariate time	Regional vs localized	1.22 (0.93–1.61)	0.149
Primary surgical / DTC / all-covariate time	Distant vs localized	1.07 (0.65–1.76)	0.791
Broader surgical / all histologies / cause-specific	Regional vs localized	1.22 (1.07–1.38)	0.003
Broader surgical / all histologies / cause-specific	Distant vs localized	1.87 (1.49–2.35)	<0.001
Broader surgical / all histologies / stage-time	Regional vs localized	1.09 (0.96–1.25)	0.168
Broader surgical / all histologies / stage-time	Distant vs localized	0.92 (0.72–1.18)	0.511
Broader surgical / all histologies / all-covariate time	Regional vs localized	1.10 (0.97–1.25)	0.149
Broader surgical / all histologies / all-covariate time	Distant vs localized	0.93 (0.72–1.19)	0.560
Broader surgical / DTC / cause-specific	Regional vs localized	1.20 (1.06–1.37)	0.005

Population / model	Stage vs localized	Estimate (95% CI)	p
Broader surgical / DTC / cause-specific	Distant vs localized	1.87 (1.48–2.35)	<0.001
Broader surgical / DTC / stage-time	Regional vs localized	1.09 (0.95–1.24)	0.208
Broader surgical / DTC / stage-time	Distant vs localized	0.96 (0.75–1.24)	0.781
Broader surgical / DTC / all-covariate time	Regional vs localized	1.09 (0.96–1.25)	0.185
Broader surgical / DTC / all-covariate time	Distant vs localized	0.97 (0.76–1.25)	0.820

Cause-specific entries are cause-specific hazard ratios (csHRs). Fine–Gray entries are subdistribution hazard ratios (sHRs) at 60 months, not constant follow-up effects. All models adjust for recorded radiation and chemotherapy and the covariates specified in Supplementary Methods 5. Surgery defines the restricted population and is not a varying covariate. CI, confidence interval; DTC, differentiated thyroid cancer, comprising papillary and follicular carcinoma. Intervals and individual p values are nominal, not multiplicity-adjusted.

Joint tests with fixed-family multiplicity adjustment

Population / model	Test	χ^2	df	Nominal p	Holm p
Primary surgical / all histologies / cause-specific	Stage block	22.390	2	<0.001	<0.001
Primary surgical / DTC / cause-specific	Stage block	19.087	2	<0.001	<0.001
Broader surgical / all histologies / cause-specific	Stage block	32.594	2	<0.001	<0.001
Broader surgical / DTC / cause-specific	Stage block	30.939	2	<0.001	<0.001
Primary surgical / all histologies / stage-time	Stage-time block	20.468	2	<0.001	<0.001
Primary surgical / all histologies / all-covariate time	Stage-time block	8.405	2	0.015	0.019
Primary surgical / DTC / stage-time	Stage-time block	18.522	2	<0.001	<0.001
Primary surgical / DTC / all-covariate time	Stage-time block	9.324	2	0.009	0.019
Broader surgical / all histologies / stage-time	Stage-time block	22.967	2	<0.001	<0.001

Population / model	Test	χ^2	df	Nominal p	Holm p
Broader surgical / all histologies / all-covariate time	Stage-time block	14.170	2	<0.001	0.003
Broader surgical / DTC / stage-time	Stage-time block	20.819	2	<0.001	<0.001
Broader surgical / DTC / all-covariate time	Stage-time block	14.969	2	<0.001	0.002

Holm adjustment is applied to separate families of four cause-specific stage tests and eight Fine–Gray stage-time tests. df, degrees of freedom. These joint tests are distinct from individual coefficient tests. Statistical significance does not remove bias introduced by treatment selection or measurement limitations.

Table S16. Model diagnostics and sensitivity checks

Cause-specific model	Term	χ^2	df	p
Primary / Cause-specific Cox	stage	2.210	2	0.331
Primary / Cause-specific Cox	sex	1.470	1	0.225
Primary / Cause-specific Cox	age_ns1	5.580	1	0.018
Primary / Cause-specific Cox	age_ns2	0.097	1	0.755
Primary / Cause-specific Cox	age_ns3	5.446	1	0.020
Primary / Cause-specific Cox	age_ns4	5.700	1	0.017
Primary / Cause-specific Cox	race	0.747	2	0.688
Primary / Cause-specific Cox	marital	2.752	1	0.097
Primary / Cause-specific Cox	income	0.409	2	0.815
Primary / Cause-specific Cox	histology	2.039	3	0.564
Primary / Cause-specific Cox	Diagnosis year (per year)	0.152	1	0.697
Primary / Cause-specific Cox	GLOBAL	16.963	16	0.388

Cause-specific model	Term	χ^2	df	p
Primary / sex interaction / Cause-specific Cox	stage	2.196	2	0.333
Primary / sex interaction / Cause-specific Cox	sex	1.378	1	0.240
Primary / sex interaction / Cause-specific Cox	age_ns1	5.534	1	0.019
Primary / sex interaction / Cause-specific Cox	age_ns2	0.086	1	0.770
Primary / sex interaction / Cause-specific Cox	age_ns3	5.384	1	0.020
Primary / sex interaction / Cause-specific Cox	age_ns4	5.561	1	0.018
Primary / sex interaction / Cause-specific Cox	race	0.762	2	0.683
Primary / sex interaction / Cause-specific Cox	marital	2.672	1	0.102
Primary / sex interaction / Cause-specific Cox	income	0.405	2	0.817
Primary / sex interaction / Cause-specific Cox	histology	2.021	3	0.568
Primary / sex interaction / Cause-specific Cox	Diagnosis year (per year)	0.181	1	0.671
Primary / sex interaction / Cause-specific Cox	stage:sex	0.601	2	0.740
Primary / sex interaction / Cause-specific Cox	GLOBAL	18.119	18	0.448
Additional adjustment / Cause-specific Cox	stage	2.115	2	0.347
Additional adjustment / Cause-specific Cox	sex	1.478	1	0.224
Additional adjustment / Cause-specific Cox	age_ns1	5.575	1	0.018
Additional adjustment / Cause-specific Cox	age_ns2	0.057	1	0.811
Additional adjustment / Cause-specific Cox	age_ns3	5.354	1	0.021
Additional adjustment / Cause-specific Cox	age_ns4	5.277	1	0.022

Cause-specific model	Term	χ^2	df	p
Additional adjustment / Cause-specific Cox	race	0.795	2	0.672
Additional adjustment / Cause-specific Cox	marital	2.495	1	0.114
Additional adjustment / Cause-specific Cox	income	0.507	2	0.776
Additional adjustment / Cause-specific Cox	histology	2.127	3	0.546
Additional adjustment / Cause-specific Cox	Diagnosis year (per year)	0.148	1	0.700
Additional adjustment / Cause-specific Cox	tumor_size	4.801	2	0.091
Additional adjustment / Cause-specific Cox	grade	0.042	1	0.837
Additional adjustment / Cause-specific Cox	multifocal	0.660	1	0.416
Additional adjustment / Cause-specific Cox	residence	0.600	1	0.439
Additional adjustment / Cause-specific Cox	tx_lag	0.325	1	0.569
Additional adjustment / Cause-specific Cox	GLOBAL	22.208	22	0.448
Additional adjustment / sex interaction / Cause-specific Cox	stage	2.123	2	0.346
Additional adjustment / sex interaction / Cause-specific Cox	sex	1.384	1	0.239
Additional adjustment / sex interaction / Cause-specific Cox	age_ns1	5.518	1	0.019
Additional adjustment / sex interaction / Cause-specific Cox	age_ns2	0.046	1	0.829
Additional adjustment / sex interaction / Cause-specific Cox	age_ns3	5.277	1	0.022
Additional adjustment / sex interaction / Cause-specific Cox	age_ns4	5.118	1	0.024
Additional adjustment / sex interaction / Cause-specific Cox	race	0.820	2	0.664
Additional adjustment / sex interaction / Cause-specific Cox	marital	2.415	1	0.120

Cause-specific model	Term	χ^2	df	p
Additional adjustment / sex interaction / Cause-specific Cox	income	0.500	2	0.779
Additional adjustment / sex interaction / Cause-specific Cox	histology	2.117	3	0.548
Additional adjustment / sex interaction / Cause-specific Cox	Diagnosis year (per year)	0.177	1	0.674
Additional adjustment / sex interaction / Cause-specific Cox	tumor_size	4.943	2	0.084
Additional adjustment / sex interaction / Cause-specific Cox	grade	0.065	1	0.798
Additional adjustment / sex interaction / Cause-specific Cox	multifocal	0.672	1	0.412
Additional adjustment / sex interaction / Cause-specific Cox	residence	0.578	1	0.447
Additional adjustment / sex interaction / Cause-specific Cox	tx_lag	0.309	1	0.578
Additional adjustment / sex interaction / Cause-specific Cox	stage:sex	0.626	2	0.731
Additional adjustment / sex interaction / Cause-specific Cox	GLOBAL	23.457	24	0.493

Schoenfeld-residual proportional-hazards diagnostics are shown for all reported main and sex-interaction cause-specific models. These diagnostics are separate from the Fine–Gray stage-time test in Table S11. Individual age-basis diagnostics do not define an age threshold. df, degrees of freedom.

Endpoint	Stage-by-interval likelihood-ratio χ^2	df	p
Other death	92.505	6	<0.001
Cardiovascular death	9.423	6	0.151

Four fixed intervals: 0–12, >12–36, >36–60 and >60 months. The non-CVD stage process showed substantial temporal heterogeneity; the CVD interval comparison was not significant. Fine–Gray log-time and cause-specific interval tests assess different estimands and are not expected to coincide.

Stage-time test with stratified estimation of the censoring distribution: χ^2=23.1143, df=2, p=9.6e-06. This sensitivity did not replace the primary model.

Treatment-analysis cause-specific diagnostics

Surgical model	Term	χ^2	df	p
Primary surgical / all histologies / cause-specific	stage	2.678	2	0.262
Primary surgical / all histologies / cause-specific	GLOBAL	24.212	22	0.336
Primary surgical / DTC / cause-specific	stage	3.332	2	0.189
Primary surgical / DTC / cause-specific	GLOBAL	23.748	20	0.254
Broader surgical / all histologies / cause-specific	stage	1.104	2	0.576
Broader surgical / all histologies / cause-specific	GLOBAL	23.613	21	0.312
Broader surgical / DTC / cause-specific	stage	1.151	2	0.562
Broader surgical / DTC / cause-specific	GLOBAL	25.600	19	0.142

Alternative non-CVD diagnostic assessment

Alternative other-death model	Global χ^2	df	p	Interpretation
Primary all covariates piecewise	133.387	46	<0.001	Inference withheld
Primary hist grade stratified	83.843	38	<0.001	Inference withheld
Expanded all covariates piecewise	151.038	44	<0.001	Inference withheld
Expanded hist grade stratified	135.747	38	<0.001	Inference withheld

All four alternative non-CVD fits retained global nonproportionality. Their coefficients and stage-interval joint tests were withheld from substantive interpretation. These models used restricted intervals 0–36 and >36–120 months and a different analysis frame from the primary absolute-risk calculation. They are not used to support Table S12.

Table S17. Exploratory cumulative incidence by reconstructed nomogram score group

Groups were reconstructed from the displayed points in Table S10 using fixed cutoffs of 125 and 221. Original unrounded nomogram scores and original X-tile assignments were not verified. No score refitting or cutoff optimization was performed. All current comparisons use the frozen training/validation partition; deaths from the other cause remain competing events. These analyses describe score-group separation, not original-nomogram probability calibration, stage-specific incremental value or established clinical utility.

Group size and event support

Cohort	Score group	N	CVD deaths full follow-up	Other deaths full follow-up	CVD deaths by 120 months	Other deaths by 120 months	Censored by 120 months
Training	Low	8,448	21	190	18	175	8,255
Training	Intermediate	4,644	163	693	140	644	3,860
Training	High	361	38	89	34	84	243
Validation	Low	3,638	11	81	9	74	3,555
Validation	Intermediate	1,975	75	288	59	267	1,649
Validation	High	162	13	51	13	46	103

Administrative restriction to 120 months yields 192 cardiovascular and 903 other deaths in training, and 81 cardiovascular and 387 other deaths in validation. Observations after 120 months are censored at that horizon; events exactly at 120 months are retained. The complete-follow-up totals remain 222/972 and 99/420, respectively.

Cumulative incidence at five eight and ten years

Cohort	Score group	Endpoint	Five years % (95% CI)	Eight years % (95% CI)	Ten years % (95% CI)
Training	Low	Cardiovascular death	0.13 (0.07–0.24)	0.25 (0.15–0.39)	0.25 (0.15–0.39)
Training	Low	Other death	1.50 (1.26–1.79)	2.06 (1.77–2.41)	2.41 (2.07–2.80)
Training	Intermediate	Cardiovascular death	1.62 (1.29–2.03)	3.06 (2.56–3.65)	3.86 (3.25–4.58)

Cohort	Score group	Endpoint	Five years % (95% CI)	Eight years % (95% CI)	Ten years % (95% CI)
Training	Intermediate	Other death	10.59 (9.73–11.52)	13.82 (12.81–14.91)	15.98 (14.82–17.21)
Training	High	Cardiovascular death	5.84 (3.84–8.82)	9.84 (7.00–13.72)	11.23 (7.97–15.71)
Training	High	Other death	17.23 (13.70–21.55)	23.53 (19.24–28.59)	26.63 (21.81–32.28)
Validation	Low	Cardiovascular death	0.23 (0.11–0.45)	0.26 (0.13–0.50)	0.26 (0.13–0.50)
Validation	Low	Other death	1.71 (1.33–2.19)	2.07 (1.64–2.61)	2.24 (1.78–2.83)
Validation	Intermediate	Cardiovascular death	1.79 (1.29–2.49)	3.05 (2.33–4.00)	3.78 (2.90–4.93)
Validation	Intermediate	Other death	10.20 (8.94–11.63)	13.82 (12.29–15.51)	15.35 (13.65–17.23)
Validation	High	Cardiovascular death	2.52 (0.95–6.61)	8.22 (4.73–14.09)	9.81 (5.61–16.85)
Validation	High	Other death	21.16 (15.59–28.35)	27.03 (20.67–34.87)	31.38 (24.26–39.98)

CIF, cumulative incidence function; CI, confidence interval. Estimates and pointwise intervals are percentages. Intervals use the complementary-log-log delta method with Aalen variance. The validation high-risk group has only four cardiovascular deaths by five years; the associated uncertainty should be considered when interpreting five-year separation.

Global Gray tests through 120 months

Cohort	Endpoint	χ^2	df	p
Training	Cardiovascular death	343.755	2	<0.001
Training	Other death	840.608	2	<0.001
Validation	Cardiovascular death	120.786	2	<0.001
Validation	Other death	392.448	2	<0.001

Tests use rho = 0 and compare entire cumulative-incidence curves through 120 months, not risk differences specifically at five years. Training CVD comparison is exploratory; validation and other-death comparisons are secondary. Global p values are nominal. Very small p values are reported as <0.001, not zero.

Pairwise Gray comparisons with Holm adjustment

Cohort	Endpoint	Comparison	Nominal p	Holm p
Training	Cardiovascular death	Low vs Intermediate	<0.001	<0.001
Training	Other death	Low vs Intermediate	<0.001	<0.001
Training	Cardiovascular death	Low vs High	<0.001	<0.001
Training	Other death	Low vs High	<0.001	<0.001
Training	Cardiovascular death	Intermediate vs High	<0.001	<0.001
Training	Other death	Intermediate vs High	<0.001	<0.001
Validation	Cardiovascular death	Low vs Intermediate	<0.001	<0.001
Validation	Other death	Low vs Intermediate	<0.001	<0.001
Validation	Cardiovascular death	Low vs High	<0.001	<0.001
Validation	Other death	Low vs High	<0.001	<0.001
Validation	Cardiovascular death	Intermediate vs High	0.001	0.001
Validation	Other death	Intermediate vs High	<0.001	<0.001

Holm correction covers all three fixed pairwise comparisons separately within each cohort and endpoint. Pairwise tests concern the entire 120-month comparison period and must not be interpreted as five-year pairwise tests. These exploratory comparisons do not correct for the historical data-dependent selection of the cutoffs.

Risk counts at five eight and ten years

Cohort	Score group	Five years	Eight years	Ten years
Training	Low	7,946	5,042	3,162

Cohort	Score group	Five years	Eight years	Ten years
Training	Intermediate	3,947	2,219	1,273
Training	High	274	130	69
Validation	Low	3,396	2,090	1,343
Validation	Intermediate	1,698	915	525
Validation	High	122	69	36

Risk counts are patients event-free and uncensored immediately before each time point, not the augmented Fine–Gray risk set. CVD, cardiovascular disease. Curve estimates were checked by an independent Aalen–Johansen recursion; maximum absolute differences were below 1×10^{-15}.

Supplementary Figures

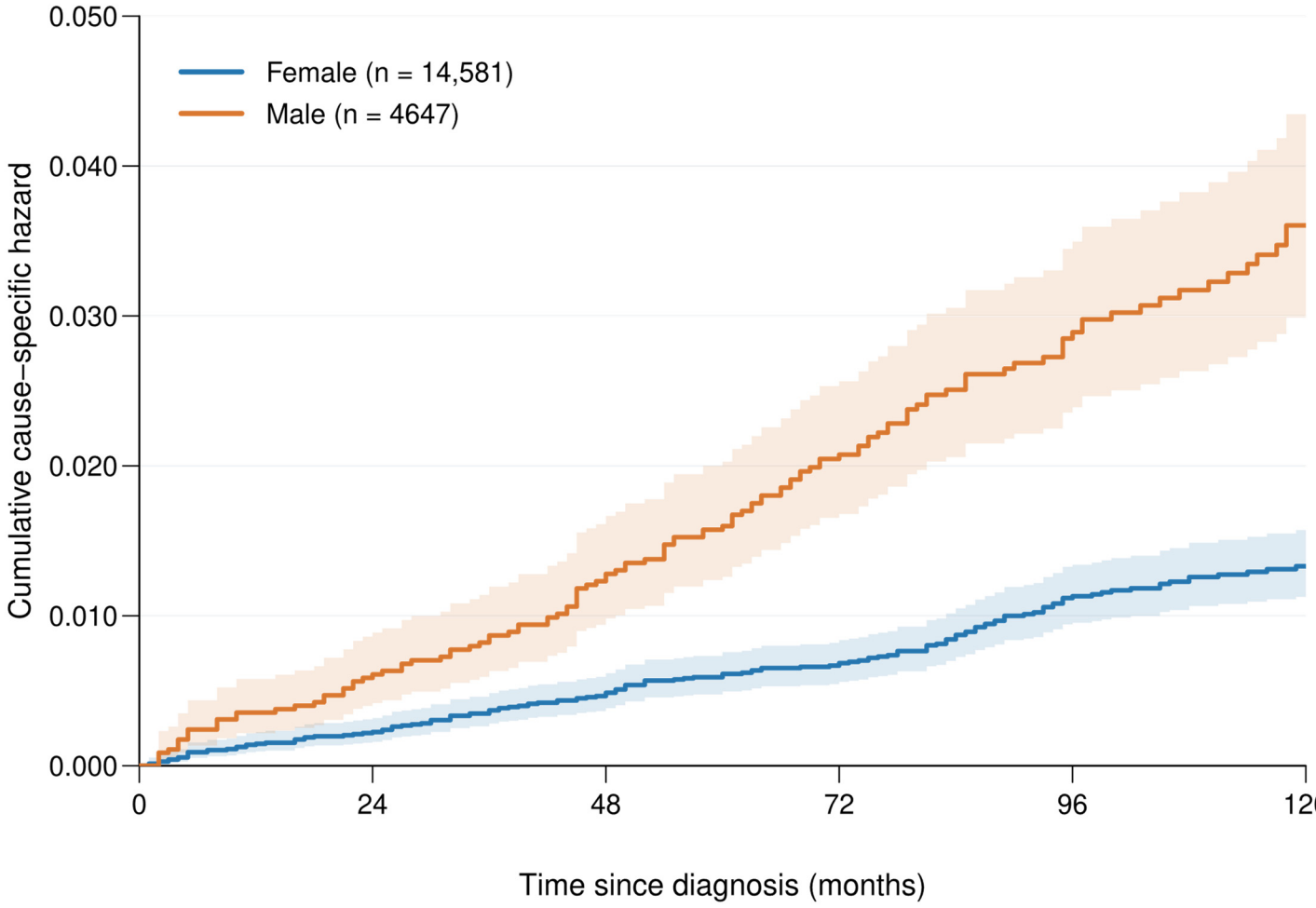


Figure S1. Cause-specific cumulative hazard of cardiovascular death by sex

Nelson–Aalen cumulative hazard curves for cardiovascular death, stratified by sex, in the 19,228-patient primary cohort (female, n = 14,581, 185 cardiovascular deaths; male, n = 4,647, 136 cardiovascular deaths). Non-cardiovascular deaths are censored at their occurrence. The vertical axis reports cumulative cause-specific hazard, not cumulative incidence or percentage risk.

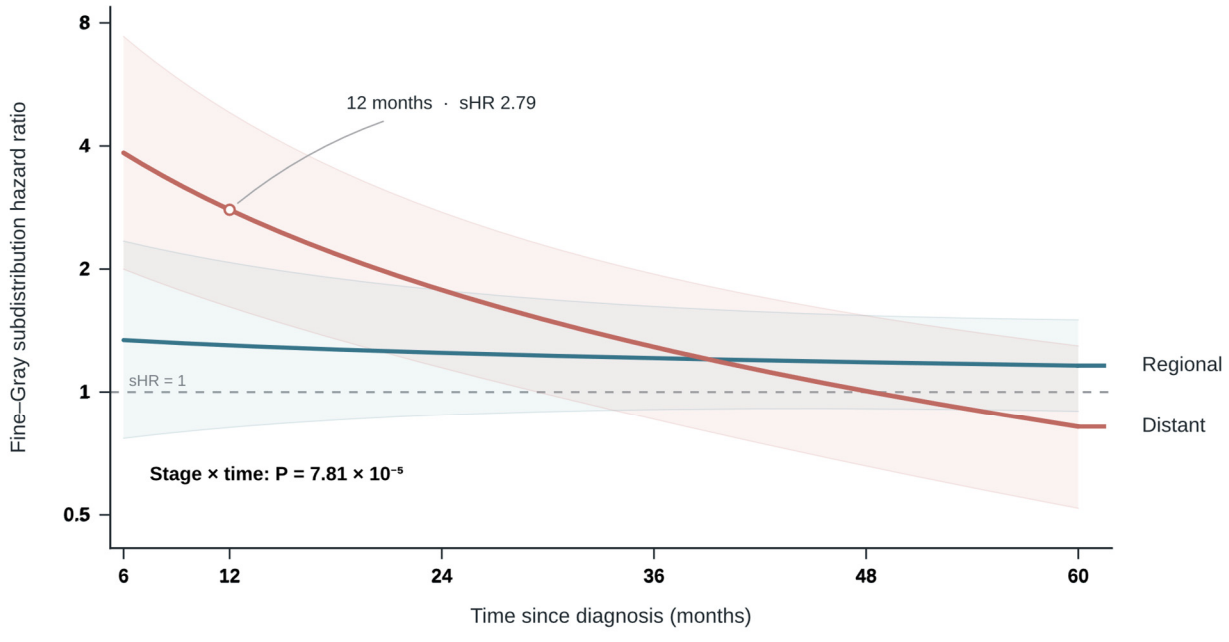


Figure S2. Time-varying stage associations with cardiovascular mortality. Fine–Gray subdistribution hazard ratios for regional and distant versus localized SEER summary stage. Shading represents pointwise 95% confidence intervals. The joint stage-by-time test yielded p = 7.81 × 10⁻⁵ (Table S11).

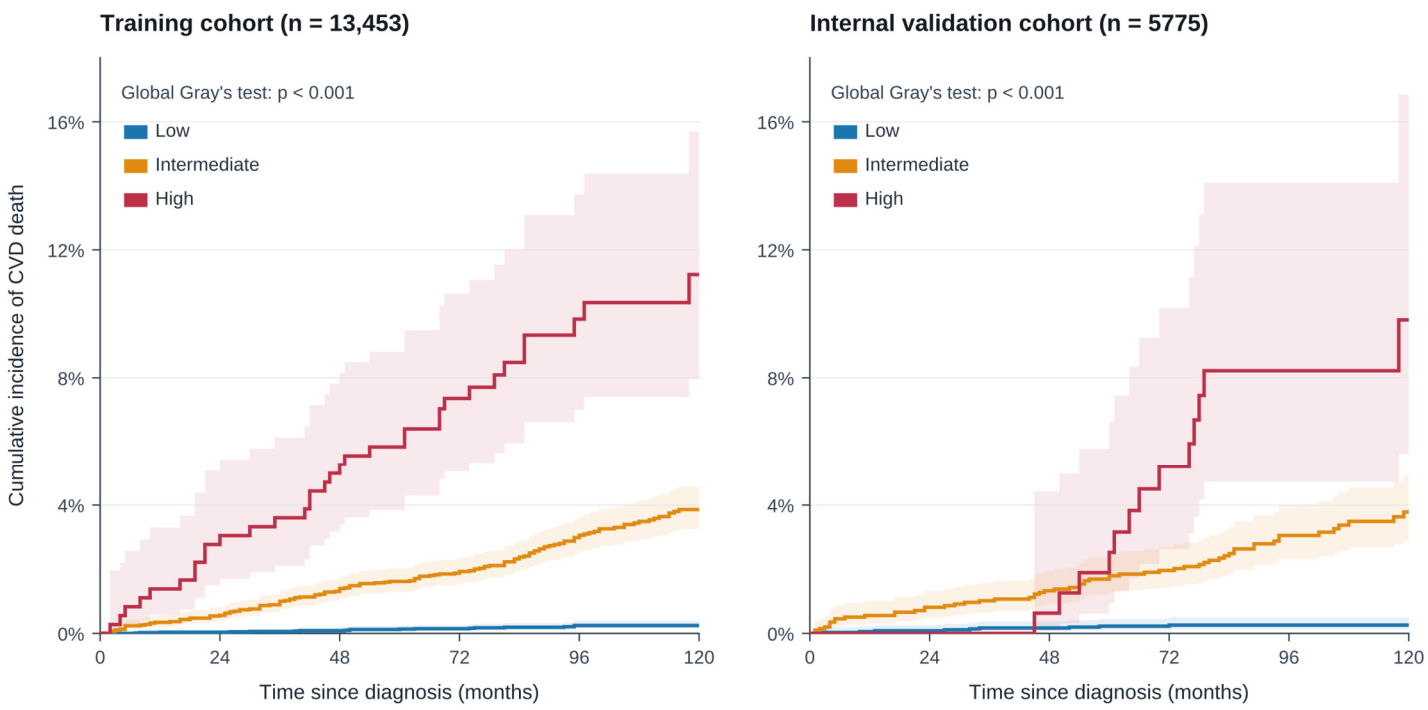


Figure S3. Cardiovascular cumulative incidence by exploratory nomogram score group. Cumulative-incidence curves for reconstructed low (≤125), intermediate (126–221) and high (≥222) score groups in training and internal validation cohorts. Non-cardiovascular deaths were competing events; shading indicates pointwise 95% confidence intervals. Global Gray's tests compare curves through 120 months (Table S17).