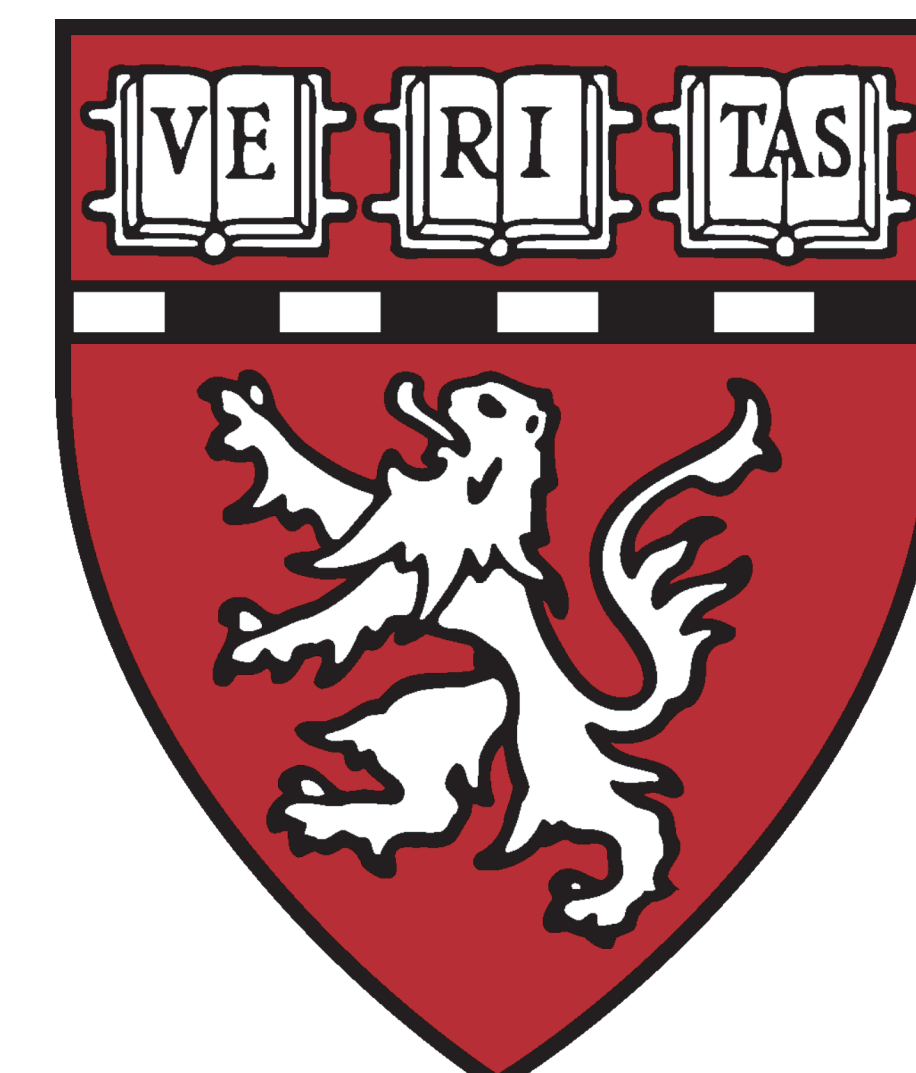




Assessing Sybil's AI lung cancer risk prediction performance on chest CT scans containing pulmonary nodules



Allison E. B. Chang^{1,2}, Florian J. Fintelmann^{2,3}, Jonathan C. Mueller³, Julian A. Westphal³, Ryan W. Trotzky³, Julius Chiang-Boeckmann¹, Chetan Shukla¹, Elizabeth Mitchell¹, Raiya Suliman^{1,4}, David Cheng⁵, Jennifer Clow¹, Abigail Moore¹, Angela Frank^{1,2,6}, PuiYee Chan¹, Jill Feldman⁷, Frank Weinberg⁸, Chi-Fu Jeffrey Yang^{2,9}, Alexandra Potter^{2,9}, Matthew P. Smeltzer¹⁰, William Mayfield¹¹, Ray U. Osarogiagbon¹², Mary Pasquinelli⁸, Erica T. Warner^{1,13}, Lecia V. Sequist^{1,2}

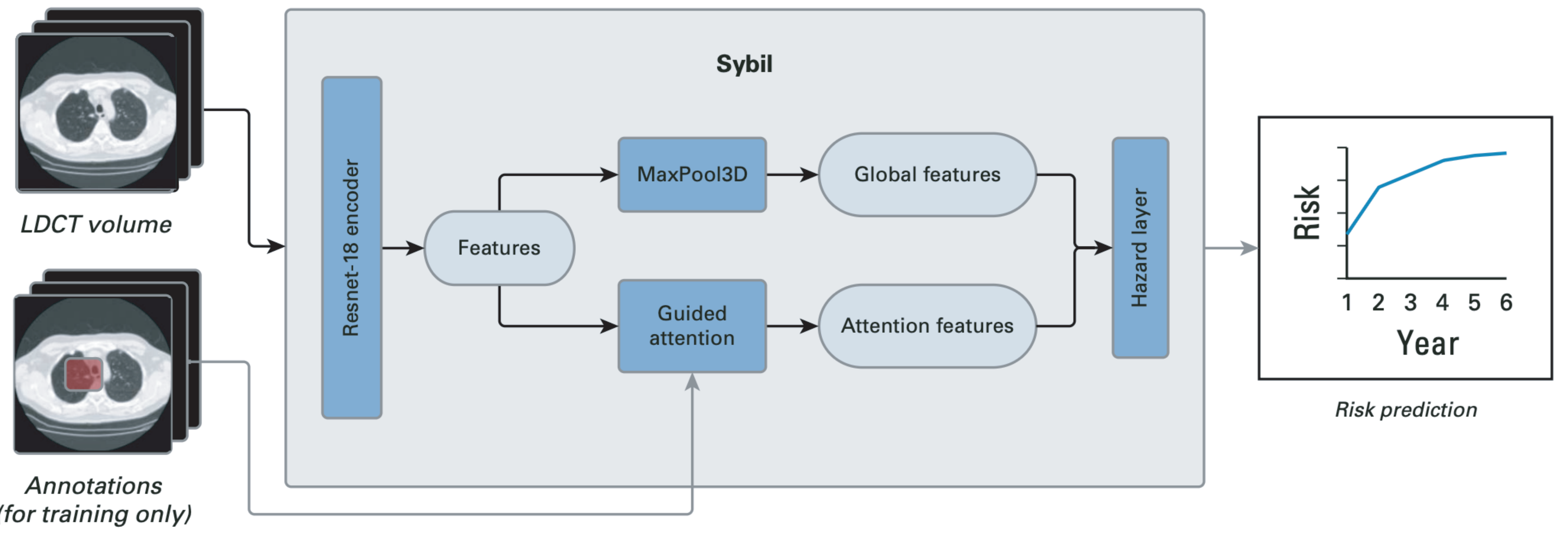
1. Mass General Brigham Cancer Institute, Boston, MA; 2. Harvard Medical School, Boston, MA; 3. Dept of Radiology, Mass General Brigham, Boston, MA; 4. UMass Chan Medical School, Worcester, MA; 5. MGH Biostatistics Center, Mass General Brigham, Boston, MA; 6. Dept of Pulmonary and Critical Care Medicine, Mass General Brigham, Boston, MA; 7. EGFR Resisters Patient Advocacy Group, Deerfield, IL, USA; 8. Dept of Hematology and Oncology, University of Illinois, Chicago, IL; 9. Division of Thoracic Surgery, Dept of Surgery, Mass General Brigham, Boston, MA; 10. School of Public Health, University of Memphis, Memphis, TN; 11. Division of of Thoracic Surgery, Department of Surgery, Wellstar Health System, Marietta, GA; 12. Multidisciplinary Thoracic Oncology Program, Baptist Cancer Center, Memphis, TN; 13. Clinical Translational Epidemiology Unit, Mongan Institute, Department of Medicine, Mass General Brigham, Boston, MA

Background:

•Pulmonary nodules (PNs) are identified on ~30% of the ~20 million chest CTs performed annually in the US, representing a key opportunity to intercept lung cancer early. Most PNs are not malignant, so risk stratification is critical.

•Sybil is an open-source deep learning model that accurately stratifies lung cancer risk using non-contrast chest CT images. It was trained on scans from the National Lung Screening Trial (NLST), some of which contained nodules and some of which did not. Sybil does not require visible nodules to make its predictions. It provides scan-level risk estimates, not nodule-level risk estimates. NB: Model training included attention directing to malignant lesions (Mikhael et al., JCO 2023).

•Questions remain regarding Sybil's performance in scans containing nodules of different types, with prior studies showing mixed results in different nodule cohorts.



Methods:

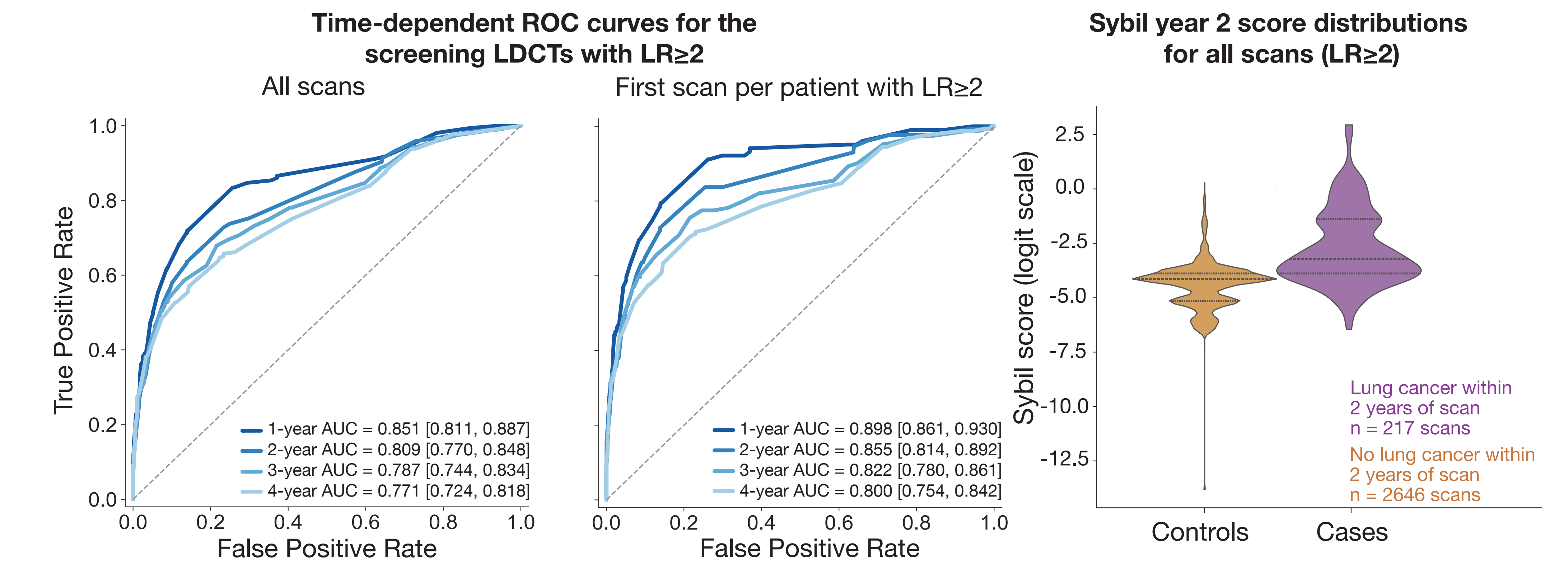
- We retrospectively assessed Sybil's performance in predicting lung cancer, with primary analyses focused on the 2-year horizon as a measure of Sybil's utility in short-term clinical decision-making.
- Patient demographics and lung cancer diagnoses were abstracted from the electronic health record.
- At each time horizon, cases were diagnosed by time t; controls were known cancer-free through t; observations with insufficient follow-up were considered unevaluable at that horizon.

Screen-Detected Nodule Cohort		High-Risk Pulmonary Nodule Cohort	
LDCTs from a US-based lung cancer screening program that had a Lung-RADS score of ≥2 (i.e., containing ≥1 nodule)		Non-contrast chest CT scans from a referral-based clinic for high-risk screen-detected or incidental PNs	
Overall (n = 7573 LDCTs from 4331 patients)		Overall (n = 1768)	
Age	64 (59-69) years	Age	68.1 years (IQR 61.2-74.6)
Sex		Sex	69.1 years (IQR 60.7-76.6)
Male	2292 (53%)	Male	752 (43%)
Female	2039 (47%)	Female	1016 (57%)
Race		Race	66.8 years (IQR 61.8-71.3)
White	5758 (76%)	White	1546 (87.4%)
Black	219 (3%)	Black	58 (3.3%)
Asian	142 (2%)	Asian	69 (3.9%)
Other	191 (3%)	Other	9 (0.5%)
Missing	1264 (17%)	Missing	4 (0.4%)
Smoking status		Smoking status	
Current	2355 (54%)	Current	472 (27%)
Former	1939 (45%)	Former	832 (47%)
None	20 (0.5%)	Never	464 (26%)
Missing	17 (0.1%)	Missing	11 (2%)
Patients who developed lung cancer	169 (4%)	Patients who developed lung cancer	168 (29%)

Scans obtained between 2015 and 2021
Median follow up: 1.94 years (IQR 0.92-2.96)

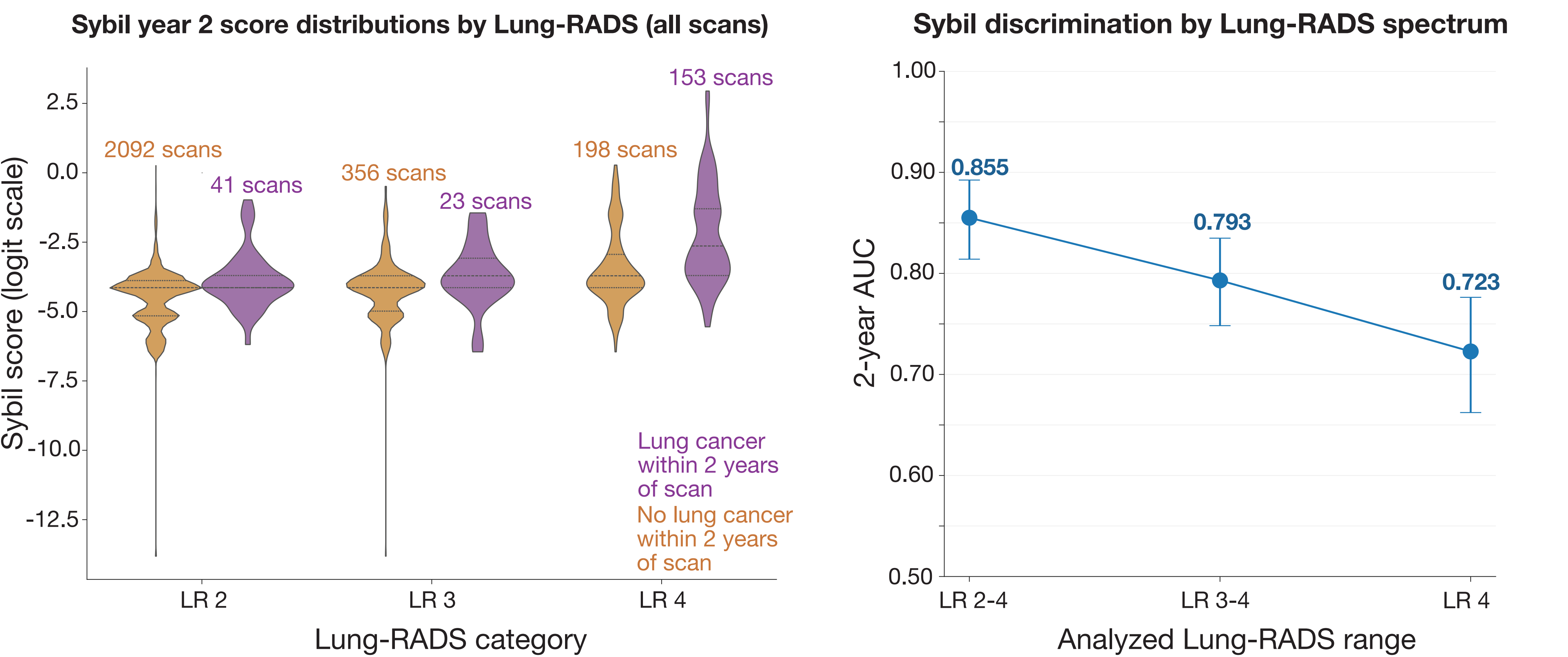
Screen-Detected Nodule Cohort

Sybil's discrimination was strong in a screening cohort containing a wide range of nodules (Lung-RADS 2-4).



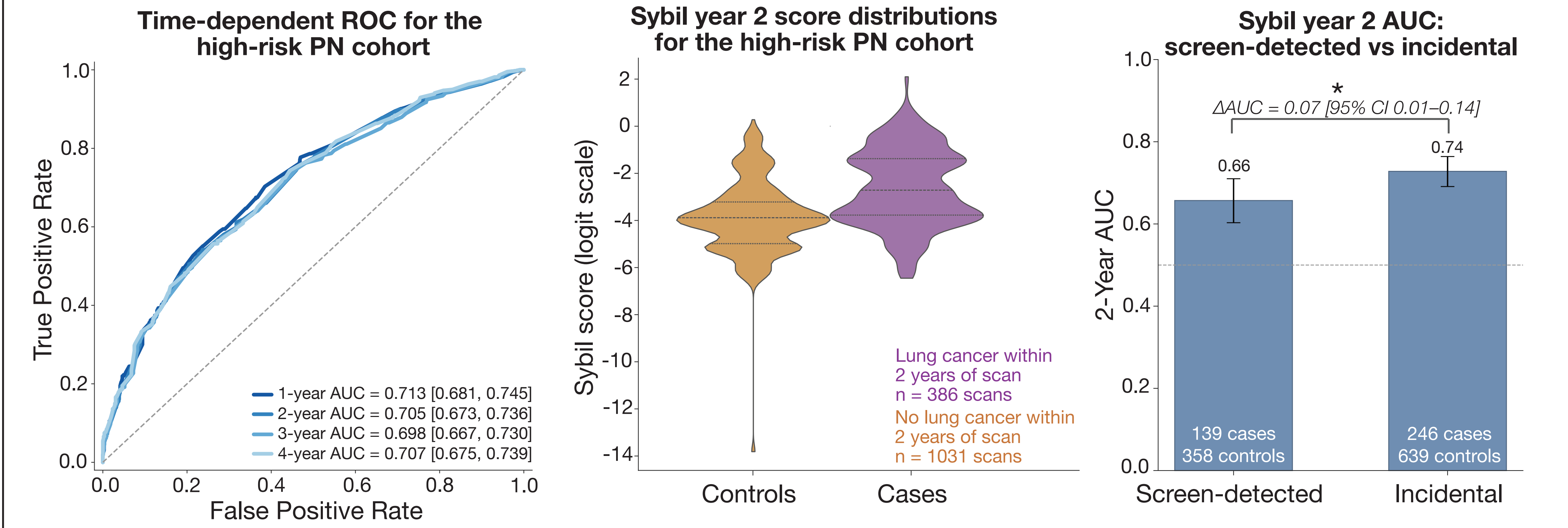
Sybil's automated risk scores increased with Lung-RADS category, demonstrating a modest correlation (controls: $p = 0.203$ [95% CI 0.167, 0.240]; cases: $p = 0.442$ [95% CI 0.342, 0.535]) with a validated radiologist-assigned risk category despite requiring no radiology assessment. This suggests Sybil and Lung-RADS capture partially overlapping risk information.

Within the screening cohort, 2-year discrimination declined as the analyzed Lung-RADS range narrowed, consistent with a spectrum effect. When restricted to Lung-RADS 4, Sybil retained moderate risk stratification (AUC 0.723).

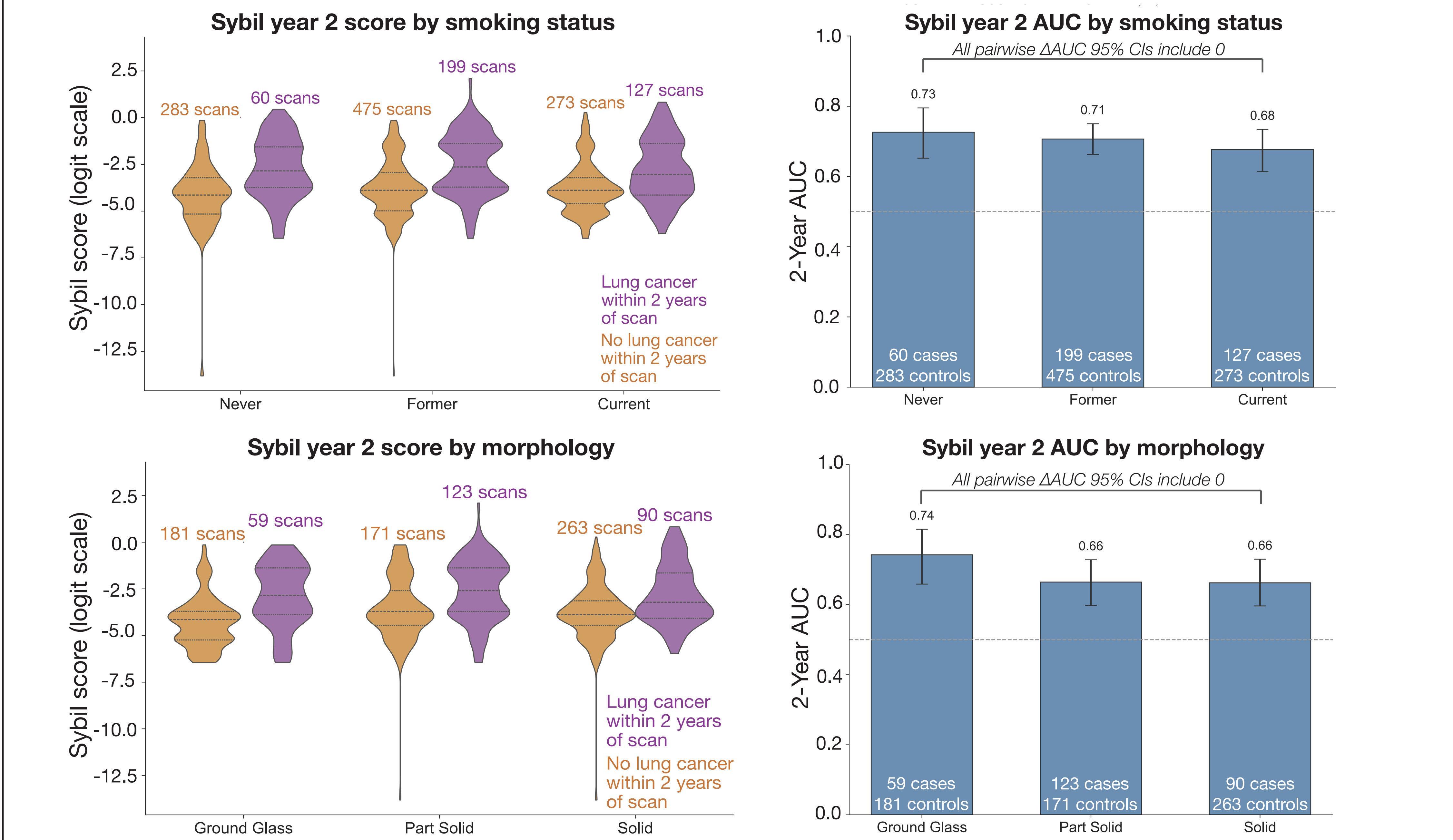


High-Risk Pulmonary Nodule Cohort

Sybil retained discrimination in a more homogeneous population of high-risk PNs, but performance was attenuated compared to the Lung-RADS 2-4 cohort, likely due in part to a spectrum effect. Discrimination was modestly better in incidental than in screen-detected PNs.



Discrimination and absolute Sybil scores were comparable across tobacco use histories and PN morphologies, despite having been trained on scans belonging to individuals with significant tobacco use (NLST).



Conclusions and Future Directions

- Sybil's discrimination of cancer risk was strong in a cohort containing a wide range of PNs (Lung-RADS 2-4) and remained good--albeit attenuated--in a population restricted to only high-of high risk PNs referred to multi-disciplinary specialty clinic.
- Sybil performance decreased as the analyzed nodule risk range narrowed, consistent with a spectrum effect and suggesting that Sybil and Lung-RADS capture partially overlapping risk information. This may partly explain the lower AUC observed in the referral-based high-risk PN cohort; spectrum restriction may also influence performance when Sybil is applied solely to a sub-cohort of scans from a larger population chosen for lack of visible nodules (Lee et al., Radiology 2025).
- Sybil's discrimination was modestly higher for incidental versus screen-detected nodules and broadly comparable across smoking histories and nodule morphologies.
- Sybil provides scan-level risk estimates, notnodule-level risk estimates. Future work of the Sybil Implementation Consortium will explore the value of adding a nodule-specific risk score.