Artificial Intelligence-Based Prognostication from Baseline Computed Tomography (CT) Scans in a Phase 3 Advanced Non-Small Cell Lung Cancer (aNSCLC) Trial

Omar F. Khan[1], Javier Montalt-Tordera[2], John Riskas[3], Vignesh Sivan[3], Oleksandra Samorodova[3], Jay Hennessy[3], Alan Mitchell[3], Sadegh Mohammadi[2], Emmanuelle di Tomaso[2], Topia Banerji[2], Felix Baldauf-Lenschen[3], Charles Glaus[2]

[1]Cumming School of Medicine, University of Calgary, Calgary, AB, Canada; [2]Bayer, various locations (Leverkusen, NRW, Germany; Cambridge, MA, USA; Whippany, NJ, USA; and Barcelona, Spain); [3]Altis Labs, Toronto, ON, Canada.

INTRODUCTION

- Clinical trials often utilize surrogate endpoints based on RECIST 1.1 criteria[1], which rely on one-dimensional measurements of a small sample of lesions. These measurements are known to have limited correlation with overall survival, the gold standard endpoint.
- CT imaging provides rich information regarding patient and disease status, but substantial fidelity is lost due to limitations in our ability to consistently interpret and extract data.
- Image-based prognostication (IPRO) is a fully-automated AI algorithm trained to predict mortality risk ('IPRO score') for aNSCLC patients.
- IPRO was trained and independently tested on a real-world dataset (poster #1362P). The real-world dataset contained:

8,276 aNSCLC Patients	25,995 Axial CT Series	17,412 Patient Years

- This study evaluates the performance and external validity of IPRO in an out-of-distribution dataset derived from NExUS, a phase III clinical trial in aNSCLC patients[2].

METHODS

- IPRO was retrospectively evaluated using data from NExUS (NCT00449033), a Bayer-sponsored phase III trial testing sorafenib and cisplatin versus placebo and cisplatin for first-line treatment of aNSCLC patients. As NExUS did not meet its primary endpoint, patients in both treatment arms were pooled for this analysis.
- IPRO scores were compared to baseline (BL) sum of longest diameters (SLD) per RECIST, a measurement of tumor burden that underpins several common surrogate endpoints. Patients with higher SLD are expected to be at higher risk of death than those with lower SLD.
- Harrell's c-index and standardized hazard ratios (HR) evaluated correlation between IPRO or SLD and survival. The Kaplan-Meier estimator was used to calculate the survival function and median overall survival (mOS) for covariate terciles.

IPRO predicted overall survival from baseline CT scans in a phase III aNSCLC clinical trial dataset, and shows potential for enhancing treatment effect quantification beyond RECIST-based SLD.

Figure 1. IPRO Model Structure. Patients with higher IPRO scores were anticipated to be at higher risk of death than those with lower IPRO scores. This figure depicts an example CT scan processed by IPRO.

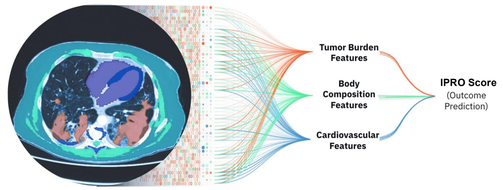
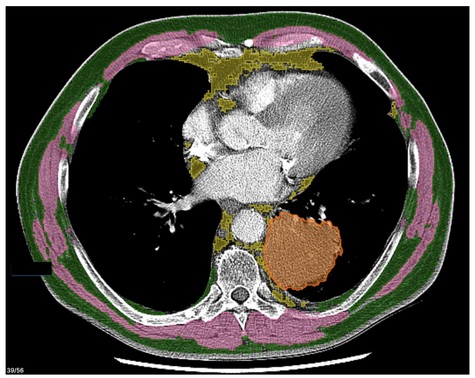


Figure 2. Example BL CT scan from NExUS with automated segmentations depicting a subset of lesions and tissue types. From each CT scan, over 5,000 features are extracted via segmentations yielding spatial imaging biomarkers (SIBs). These SIBs comprise the input to IPRO, thus enabling model explainability and correlative analyses.



Orange = Lesions
Yellow = Visceral Fat
Green = Subcutaneous Fat
Pink = Skeletal Muscle

Figure 3. Harrell's *c*-index (A) and standardized hazard ratio (B) for IPRO versus SLD, the sum of longest diameters underpinning many common surrogate endpoints.

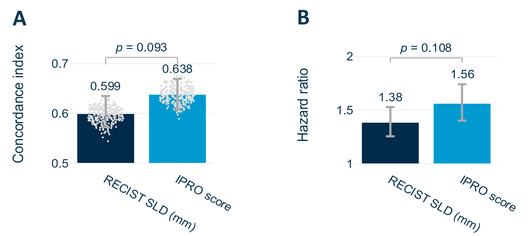
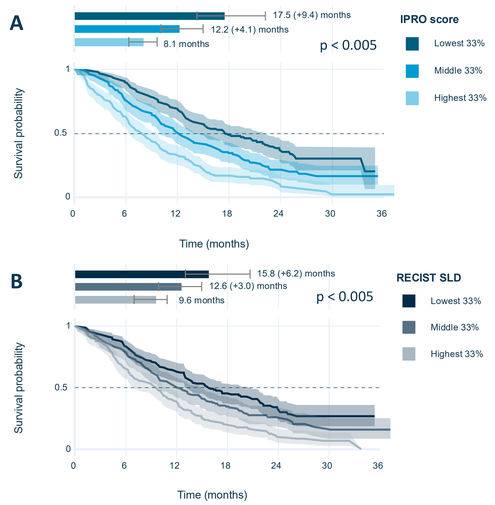


Figure 4. Survival function (curves) and median survival times (bar plots) by IPRO (A) and SLD (B) tercile. Bar plot annotations show the median survival time and the difference relative to the highest-risk tercile.



RESULTS

- 452 subjects were included in this study, with an average age of 58.8 years. 64.8% of subjects were male, and 88.3% had a stage IV malignancy (11.7% had stage IIIB). Median OS was 12.2 months (95% CI 10.8 – 13.3 months).
- IPRO significantly stratified OS across terciles ($p < 0.005$) (Figure 4A), with greater separation of curves than SLD (Figure 4B).
- There was a trend towards improved estimation of OS using IPRO compared to using SLD based on *c*-index (Figure 3A) and standardized HRs (Figure 3B).

DISCUSSION

- Our study results support the external validity of IPRO in predicting overall survival when applied to a novel aNSCLC dataset.
- There was a trend towards improved performance of IPRO relative to SLD, which underpins measurement of the most commonly used surrogate endpoints including overall response rate and progression-free survival.

FUTURE DIRECTIONS

- Ongoing work is assessing the ability of IPRO to detect OS treatment effect based on longitudinal imaging, and in other malignancies beyond aNSCLC.
- IPRO survival predictions can help improve risk stratification of patients in clinical trials, potentially increasing power to assess clinically relevant outcomes.

ACKNOWLEDGEMENTS

We thank the patients and their families, and all investigators involved in the original studies. We thank the Alberta Health Services Diagnostic Imaging and Cancer Research & Analytics teams for support with imaging data used to train IPRO, as well as the Digital Technology Supercluster for enabling this collaboration. We also thank Bayer's Radiology unit for support with imaging data from the NExUS trial.

References: 1. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1), *Eur J Cancer* 45:228-247. doi: 10.1016/j.ejca.2008.10.026. 2. Phase III, randomized, double-blind, placebo-controlled trial of gemcitabine/cisplatin alone or with sorafenib for the first-line treatment of advanced, nonsquamous non-small-cell lung cancer

Disclosures: OFK: Honoraria/Speaking Fees (Pfizer, AstraZeneca, Gilead, Knight Therapeutics, Novartis), Research Support (Altis Labs, Inc., Cogent Biosciences)
Presented at the European Society of Medical Oncology Congress 2024 in Barcelona, Spain (September 13-17, 2024)

Corresponding author contact: ofkhan@ucalgary.ca