

International Journal of Population Data Science

Journal Website: www.ijpds.org



The development and validation of a risk score : LUnC Cancer Improved Decisions (LUCID)

Ofran Almossawi¹, Luke Stephenton², Ruth Keogh³, Karla Diaz-Ordaz², Zhe Wang⁴, Andrew Challenger⁴, David Dodwell⁴, Martin Forster³, Kenneth Man², and Li Wei²

¹Great Ormond Street Hospital, London, United Kingdom; University College London, London, United Kingdom

²University College London, London, United Kingdom

³University College London Hospital NHS Foundation Trust, London, United Kingdom

⁴University of Oxford, Oxford, United Kingdom

Less than 20% of patients diagnosed with advanced lung cancer will survive beyond five years and half of these will suffer a serious adverse event (SAE) caused by systemic anti-cancer therapy (SACT) resulting in a hospital attendance. We applied counterfactual prediction models, grounded in causal inference, to estimate outcomes for hypothetical treatment scenarios. This approach was made possible using data linkage to gather necessary information on possible predictors and confounders. We aimed to develop and internally validate a risk prediction model for hospital admissions following the first dose of SACT in patients with advanced NSCLC. Adult patients diagnosed between 2016 and 2019 were included. Four treatment regimens were considered: Regimen A (double agent chemotherapy), Regimen B (chemotherapy + immunotherapy), Regimen C (immunotherapy), and Regi-

men D (single-agent chemotherapy). The outcome was any unplanned hospital admission within 30 days of SACT initiation. The cohort was developed by linking Hospital Episode Statistics, Cancer Registry, and SACT datasets. Model development involved multivariable logistic regression and counterfactual prediction methods. 20,076 patients were included, of whom 36% experienced the outcome. Logistic regression and penalised models both showed a modest AUC of 0.61. An ensemble machine learning approach did not outperform logistic regression. Counterfactual analysis showed that regimens A and C were associated with the lowest risk of hospital admission, while regimen D was linked to the highest risk. The LUCID score was made possible using data linkage, and provides a promising tool for predicting hospital admissions following SACT in advanced NSCLC patients.

