# Methods Section Notes for Optellum-NCI collaboration

## **Task**

## (this section intended to be replaced by the paper’s Introduction)

We used an AI developed for discrimination of malignant and benign nodules, to see whether it could improve a clinical model for the prediction of a screening patient’s likelihood of developing cancer over the upcoming year. The existing model used a number of features extracted from the metadata of the US National Lung Screening Trial (NLST) to produce a likelihood score, given a screening output in year *y*, of developing cancer before the corresponding screening image taken at year *(y+1)*. These features included both patient clinical data (age, smoking history etc), and quantities extracted from a radiological read of the screening CT, such as maximal nodule size, the existence of any GGO nodules on the CT, and presence of emphysema within the patient.

## **Method**

## Data

A re-curated version of the NLST dataset CTs and metadata was used for this study. Each CT listed as containing at least one nodule was reviewed by a medical doctor or medical student, under expert supervision from University of Oxford Radiologists, and metadata records of the CT findings were reviewed and extended. In particular, the size, extent, location, margins and attenuation of each nodule were reviewed by the same small team of individuals, and the exact 3D location of each nodule was identified and recorded. Additional nodules not listed in the NLST metadata were also added as long as they were not fully calcified (since fully calcified nodules were not considered “positive findings” in the original NLST data). As well as reviewing all CTs on which one or more nodules was recorded, the team also reviewed all CTs of patients recorded as having developed lung cancer, and again fully reviewed and extended their mark-up and metadata. Patients who never had any reported nodules and also never had lung cancer were not considered or marked up.

## Nodule AI

An AI called the LCP-CNN (Lung Cancer Prediction Convolutional Neural Network) was trained on this augmented NLST set in an 8-fold cross-validated way. The AI training is outside the scope of this publication, but for reference, the training used class balancing, and also extensive pre-training on hundreds of thousands of non-NLST images in order to achieve strong performance both on the testing folds of the cross-validation, and also on several independent external datasets.

The AI was trained for the task of malignant vs benign classification, and produces a score where 0 indicates benignity, and 100 indicates malignancy.

## Features

The set of features extracted from the NLST metadata and the re-curated CT information represents the full space of information available to our new model.

[Table of each feature under consideration for the screening models. Detail which features are derived from where. LP or ND will use this list to produce the up-to-date release of Optellum’s nodule data and metadata for use in this project].

## New model

The new model we are fitting is an extension of <background description of Hilary’s work>. The particular form of the model being fitted is <equations>.

Two versions of the model were fitted using the feature set described above. In the first, the maximum LCP score for a given CT was *not* available as a feature, and in the second, it was available for selection. In both cases, feature selection was done using [Bayes Information Criterion or Lasso].

## Statistical Analysis

The two (or four?) models were then compared according to <TBD the necessary but sufficient statistical tests>, showing that in both cases the addition of the LCP-CNN gave a better fit to data.

## **Results**

We should defer writing up any results until we’ve reached consensus with everything above.