Summary

Aim of the study
The primary objective of the Standardized Data Collection (SDC) for lung cancer patients is to develop and validate multi-factorial prediction models for different treatment outcomes. The long term aim is to build a Decision Support System based on validated prediction models that would facilitate individualized medicine.

Hypothesis
Our general hypothesis is that we will improve the performance of the prediction models (the Area Under the Curve (AUC) will be at least 0.05 higher) for survival and toxicity if we develop *multifactorial* models. The basic models of reference are based on demographical, clinical and treatment data. The improved multifactorial models will include additional clinical and/or imaging and/or genetic information and/or variables related to the quality of the treatment.

Study Design
This is a prospective cohort study.

Endpoints
Primary endpoint 1. Two-year survival rate.

Secondary endpoints 2. Change in dyspnea (early and late toxicity);
3. Change in cough (early and late toxicity);
4. Change in dysphagia (early and late toxicity);
5. Patient-rated generic quality of life (EuroQol).

Inclusion criteria
All patients planned for curatively intended primary or postoperative radiotherapy will be included. At the first visit, patients will be informed about the standardized data collection by the treating physician. The patient's written informed consent will be obtained.

Therapeutic regimens
Patients will be treated according to the institutional protocol. It is optional to collect a saliva sample for genetic profiling. Standard follow-up appointments to assess treatment toxicity will be scheduled at two time points: 2 weeks and 3 months after radiotherapy and it's optional to schedule follow-up appointments at three more points: 6, 12 and 24 months after radiotherapy (depending on the institute this is standard clinical practice or an extra visit). In addition, patients will fill in a short questionnaire, sent by mail, at 8 time points.

Sample size
It is expected that data of 864 patients can be analyzed, while additional genetic information will be available for 606 patients. The inclusion period will be fixed at 24 months. The power calculation is based on the minimum number of patients. It is assumed that 35% of the NSCLC patients will still be alive at two year time point. A sample of 404 deceased and 202 survivors achieves 89% power to detect a difference of 0.05 between the two models using a two-sided z-test at a significance level of 0.05.