We investigated the application of machine learning (ML) methods through symbolic regression (SR) to generate improved model fits that are not humanly possible within the time constrains of submission. We show utilities from ML algorithms can impact significantly the incremental cost-effectiveness ratio (ICER) and the quality adjusted life year (QALY) estimates.

### METHODS

We use data from a previously reported randomized (Erlotinib vs Best Supportive Care (BSC) trial (TOPICAL) in stage IIb-IV non-small cell lung cancer (NSCLC) patients (N=670). Health states were defined as progression free (PF), progressive disease (PD) and Death. EQ-5D-3L utilities were collected at baseline and monthly until disease progression.

#### Symbolic Regression

Symbolic regression (SR) attempts to fit an equation through a set of observed data points. In general, equations (model form) need to be identified first, prior to fit. Through SR, we can find the equation(s) that fit the sample data through an optimization criteria (e.g. R², Akikakes information criterion (AIC)). Genetic Programming (GP) is an implementation of evolutionary programming (Figure 1), where the problem-solving domain is modelled on a computer and the algorithm attempts to find a solution by the process of simulated evolution, employing the biological theory of genetics and the Darwinian principle of survival of the fittest. In this sense, GP is an ideal vehicle to implement SR. Data Modeler® using a Mathematica® platform utilizing all mathematical operators except step, hyperbolic and trigonometric functions were used.

### RESULTS

**Trial Demographics:** These have been published elsewhere; In general, N=350 vs 320 Erlotinib vs BSC; median age 77 yrs, 61% male; 35% stage IIIb; ECOG 0-1,2,3: 16%, 41%, 43% respectively.

| Health State | Observed | Linear Model | SR# |
|--------------|----------|--------------|-----|
| Alive and PF (n=33) | 0.745 (0.019) | 0.793 (0.027) | 0.744 (0.021) |
| PD (n=637) | 0.612 (0.0061) | 0.660 (0.011) | 0.611 (0.017) |
| AIC | 318 | 272 to 290 |
| R² | 4.2% | 6.5% - 9.2% |
| Incremental QALY | 0.035 (0.0163)* | 0.029 (0.033) |

#Mildly complex  to Most complex; models included covariates: ECOG, smoking status, age, gender ;*published in Khan et al, 2015

---

**REFERENCES**

1. Khan et al BMJ Open 2015 Jul 2;5(7)
2. SM Lee et al; Lancet Oncol, 2012 Nov;13(11): 1161-70
3. Hussein.S; Genetic Programming in Mathematica
   http://www.husseinsspace.com/research/publications/gpinmath.pdf