

### Supplementary Table S1

**Complete case analysis: Table 4 Regression coefficients, OR (95% CI) and SE for covariates in the final model for the probability of lung cancer in pulmonary nodules.**

Covariates	$\beta$ -coefficient	Standard error	OR (95%CI)	<i>P</i> -values
Intercept	-2.4131	1.4291	-	0.091
Age (years)	-0.0248	0.0188	0.976(0.940-1.012)	0.186
Gender (female)	0.5363	0.1619	1.710(1.245-2.348)	<0.0001
Asthma	-0.9266	0.2019	0.396(0.267-0.588)	<0.0001
Bronchitis	1.7979	0.1981	6.037(4.059-8.901)	<0.0001
Asbestos exposure	0.7507	0.1697	2.119(1.519-2.954)	<0.0001
Previous malignancy	0.5021	0.1983	1.652(1.120-2.437)	0.011
Family history of cancer				
Early onset *	2.0818	0.2019	8.019(5.398-11.912)	<0.0001
Late onset*	1.5308	0.1964	4.622(3.145-6.792)	<0.0001
Smoking duration (years)	0.0573	0.0104	1.059(1.038-1.081)	<0.0001
Forced vital capacity (Litres)	-1.1718	0.1080	0.310(0.251-0.383)	<0.0001
Nodule type (solid as reference)				
Nonsolid	1.5866	0.3670	4.887(2.380-10.033)	<0.0001
Part solid	0.4941	0.2870	1.639(0.934-2.876)	0.085
Nodule location				
Upper vs. middle or lower lobe	-0.2373	0.1604	0.792(0.579-1.085)	0.147
Nodular volume (mm <sup>3</sup> )	0.000863	0.000117	1.001(1.000-1.001)	<0.0001

Table S1 is complete case analyses – which is a replicate of the odds ratios and beta coefficients of the data with missing data, which was found to be very similar to that with multiple imputations.