

# 1 Supplement

## 2 Supplement 1. Multiple imputation

3 **Table 1. Details of the multiple imputation process**

Software used:	IBM SPSS statistics version 25
Imputation method and keysettings:	Fully Conditional Specification Method Iterations; maximum iterations 10
Number of imputed data sets created	15
Variable include in the imputation procedure (imputed or used as predictors of missing data)	Age, sex, kidney failure, liver failure, heart disease, respiratory disease, Diabetes Mellitus, hiv-positive, down syndrome, antibiotic prescription, hospital admission, intensive care admission, death, body mass index, smoke status
Treatment of normally distributed variables	Linear regression
Treatment of binary/categorical variables	Logistic regression
Imputed values	1 <sup>st</sup> and 2 <sup>nd</sup> wave respectively: Body mass index missing values 84.3% (1434) and 61.9% (4274) (minimum values 17 and maximum 57) Smoke status missing values 84.3% (1434) and 60.6% (4182)

4

5 **Table 2. Effect multiple imputation on average BMI and percentage smokers**

	1th wave		2 <sup>nd</sup> wave	
	Before multiple imputation*	After multiple imputation*	Before multiple imputation*	After multiple imputation*
BMI > 29	42.2%	49.2%	43.6%	46.7%
Smoking status	23.8%	29.0%	16.4%	20.4%

6 BMI: Body Mass Index.

7 \* This percentage was calculated using the population number with known BMI or smoke status as  
8 denominator.

9