Supplementary material

MULTIPLE IMPUTATION VALIDITY

Summary statistics of original dataset, and the first and tenth multiple imputation (MI) datasets were compared (Supplementary Table 1). The imputed values were plausible and similar.

MULTIVARIABLE LOGISTIC REGRESSION MODEL DIAGNOSTICS

The median c statistic across the ten MI models was 0.753, indicating adequate discrimination. The median mean VIF was 1.48, with no exposure variable yielding a VIF greater than 5 in any imputed dataset, indicating no significant multicollinearity.

The Hosmer-Lemeshow (HL) statistic P values were non-significant using groupings of 5, 10 or 20, across all imputed datasets, indicating adequate calibration/goodness-of-fit (Supplementary Table 2). In the link test the _hat was significant, indicating that the included predictors were meanginful; and the _hat_squared was non-significant, indicating that the model was free from significant interactions (Supplementary Table 3).

### *Linearity of continuous variables*

Age (years) was the only continuous variable included in the model. Exploratory plots were inspected, and the squared and square-rooted terms were included in test models. In the squared age model, squared age was non-significant (OR: 1.00, 95%CI: 1.00-1.00; P = 0.430) and the inclusion of the squared term did not change the direction or statistical significance of any associations. In the square-rooted age model, square-rooted age was non-significant (OR: 2.15, 95%CI: 0.39-11.74; P = 0.376) and the inclusion of the square-rooted term did not change the direction or statistical significance of any associations. The plot of age against the predicted probability of PEI demonstrated an approximately linear relationship (Supplementary Figure 1).

Influential observations

Three outliers were identified from examination of the index plots, although none were leveraged (Supplementary Table 4). Re-running the model following the removal of each outlier, alone and in combination, did not change the direction or statistical significance of any associations and they were retained.

SENSITIVITY ANALYSES

Complete case analysis; MNAR Model 1 analysis
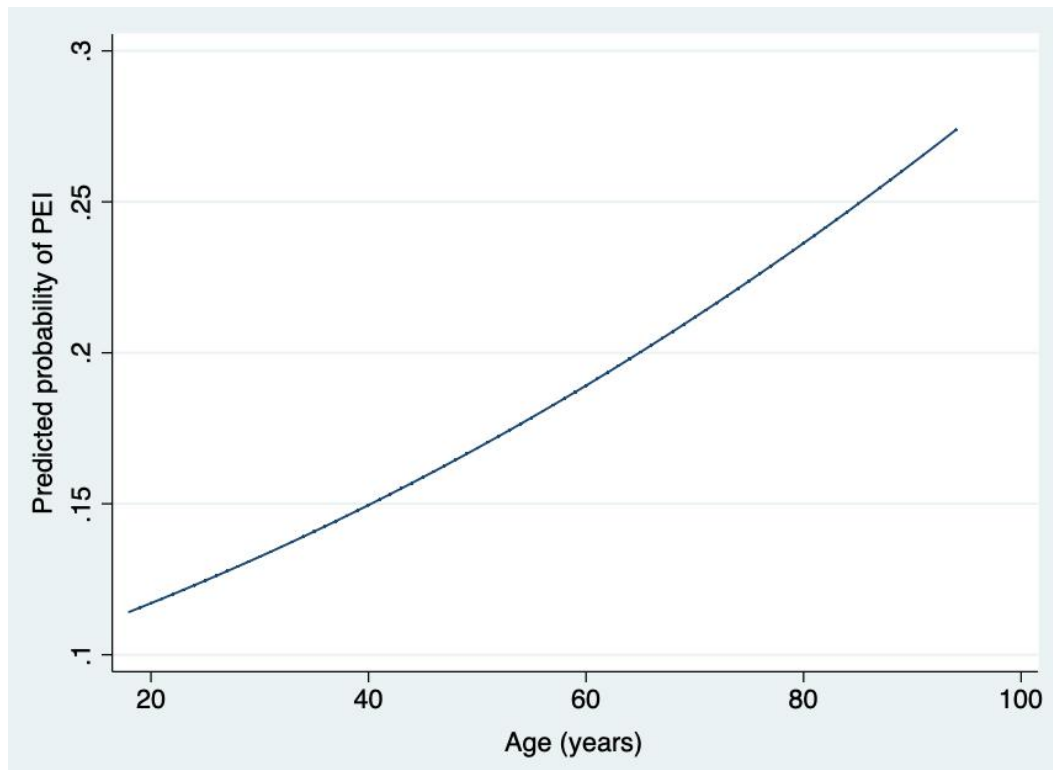
Supplementary Figure 1 Plot of age against the predicted probability of PEI, demonstrating an approximately linear relationship.

Supplementary Table 1 Multiple imputation dataset summary statistics

Variable with missing data	Mean	Min	Max
Alcohol excess			
Original dataset	0.378	0	2
MI dataset 1	0.336	0	2
MI dataset 10	0.324	0	2
Smoking history			
Original dataset	0.64	0	2
MI dataset 1	0.606	0	2
MI dataset 10	0.579	0	2
PPI therapy			
Original dataset	0.418	0	1
MI dataset 1	0.414	0	1
MI dataset 10	0.415	0	1

MI: Multiple imputation.

Supplementary Table 2 C statistic, mean variance inflation factor and Hosmer-Lemeshow goodness-of-fit test

Diagnostic test	Median (IQR) across the ten imputed datasets	Comments
C statistic	0.753 (0.005)	Adequate discrimination
Mean variance inflation factor (VIF)	1.48 (0.01)	Maximum VIF of 3.53 for any exposure variable across all imputed datasets
Hosmer-Lemeshow goodness-of-fit (HL) test		
HL statistic (5 groups)	2.015 (2.78)	
P value (5 groups)	0.630 (0.171)	Minimum P value of 0.108 across all imputed datasets
HL statistic (10 groups)	6.24 (4.27)	
P value (10 groups)	0.620 (0.449)	Minimum P value of 0.222 across all imputed datasets
HL statistic (20 groups)	16.11 (5.74)	
P value (20 groups)	0.585 (0.3819)	Minimum P value of 0.136 across all imputed datasets

HL: Hosmer-Lemeshow.

Supplementary Table 3 Link test

	Coefficient	*P* value	95%CI	Comments
_hat	0.834	< 0.001	0.56-1.10	Meaningful predictors included
_hatsq	-0.096	0.135	-0.22 to 0.03	No significant interactions detected

Supplementary Table 4 Influential observation diagnostics

Outlier	Pearson residual	Deviance residual	Leverage
Outlier 1	4.07	2.65	0.01
Outlier 2	4.07	2.65	0.01
Outlier 3	4.07	2.65	0.01

Supplementary Table 5 Complete case analysis

Variable	OR (95%CI)	*P* value
Demographics		
Age	1.01 (0.99, 1.02)	0.236
Sex		
Male	1.01 (0.63, 1.61)	0.969
Ethnicity		
Asian	2.61 (1.38, 4.91)	**0.003**
Black	1.12 (0.52, 2.39)	0.772
Other	1.85 (1.03, 3.31)	**0.039**
Comorbidities		
Type 2 diabetes	1.70 (0.97, 2.98)	0.066
Liver cirrhosis	1.14 (0.46, 2.86)	0.773
Chronic pancreatitis	8.91 (3.64, 21.84)	**< 0.001**
Pancreatic cancer	4.21 (0.91, 19.35)	0.065
Upper GI surgery	1.82 (0.74, 4.47)	0.189
CCF	1.05 (0.17, 6.64)	0.955
CKD	0.84 (0.22, 3.15)	0.797
PPI	2.01 (1.28, 3.15)	**0.002**
Alcohol excess		
Ex-excess	2.03 (0.84, 4.89)	0.115
Current excess	1.24 (0.62, 2.49)	0.545
Smoking history		
Ex-smoker	0.92 (0.49, 1.73)	0.802
Current smoker	1.41 (0.78, 2.55)	0.256

Ethnicity reference group is white ethnicity; Alcohol excess reference group is never excess; Smoking history reference group is never-smoker. GI: Gastrointestinal; CCF: Congestive heart failure; CKD: Chronic kidney disease; PPI: Proton pump inhibitor.

Supplementary Table 6 MNAR-Model 1

Variable	OR (95%CI)	*P* value
Demographics		
Age	1.00 (0.99, 1.02)	0.506
Sex		
Male	1.20 (0.83, 1.73)	0.343
Ethnicity		
Asian	2.14 (1.32, 3.47)	**0.002**
Black	1.21 (0.65, 2.25)	0.543
Other	1.41 (0.86, 2.30)	0.174
Comorbidities		
Type 2 diabetes	1.85 (1.19, 2.88)	**0.006**
Liver cirrhosis	0.92 (0.43, 1.96)	0.82
Chronic pancreatitis	7.76 (3.88, 15.50)	**< 0.001**
Pancreatic cancer	6.62 (1.70, 25.87)	**0.007**
Upper GI surgery	2.68 (1.37, 5.25)	**0.004**
CCF	1.53 (0.39, 6.03)	0.54
CKD	1.79 (0.79, 4.07)	0.166
PPI	1.72 (1.19, 2.47)	**0.004**
Alcohol excess		
Ex-excess	2.69 (1.27, 5.69)	**0.01**
Current excess	1.58 (0.88, 2.83)	0.122
Smoking history		
Ex-smoker	0.89 (0.52, 1.52)	0.662
Current smoker	1.24 (0.76, 2.02)	0.388

Ethnicity reference group is white ethnicity; Alcohol excess reference group is never excess; Smoking history reference group is never-smoker. GI: Gastrointestinal; CCF: Congestive heart failure; CKD: Chronic kidney disease; PPI: Proton pump inhibitor.

Supplementary Table 7 MNAR-Model 2

Variable	OR (95%CI)	*P* value
Demographics		
Age	1.00 (0.99, 1.01)	0.605
Sex		
Male	1.26 (0.88, 1.81)	0.211
Ethnicity		
Asian	1.98 (1.23, 3.18)	**0.005**
Black	1.22 (0.66, 2.26)	0.529
Other	1.39 (0.85, 2.27)	0.187
Comorbidities		
Type 2 diabetes	1.85 (1.20, 2.87)	**0.006**
Liver cirrhosis	1.05 (0.50, 2.21)	0.896
Chronic pancreatitis	8.44 (4.23, 16.85)	**< 0.001**
Pancreatic cancer	7.26 (1.84, 28.61)	**0.005**
Upper GI surgery	2.60 (1.39, 5.10)	**0.005**
CCF	1.55 (0.39, 6.10)	0.531
CKD	1.66 (0.73, 3.78)	0.227
PPI	1.78 (1.24, 2.55)	**0.002**
Alcohol excess		
Ex-excess	2.55 (1.18, 5.48)	**0.017**
Current excess	1.12 (0.71, 1.76)	0.64
Smoking history		
Ex-smoker	0.95 (0.53, 1.68)	0.853
Current smoker	1.14 (0.71, 1.85)	0.584

Ethnicity reference group is white ethnicity; Alcohol excess reference group is never excess; Smoking history reference group is never-smoker. GI: Gastrointestinal; CCF: Congestive heart failure; CKD: Chronic kidney disease; PPI: Proton pump inhibitor.