

Appendix 1: Supplementary results

Figure S1. Percent bias difference between treatment and control groups, before and after propensity score matching. This illustrates the imbalance in the observed covariates before and after matching.

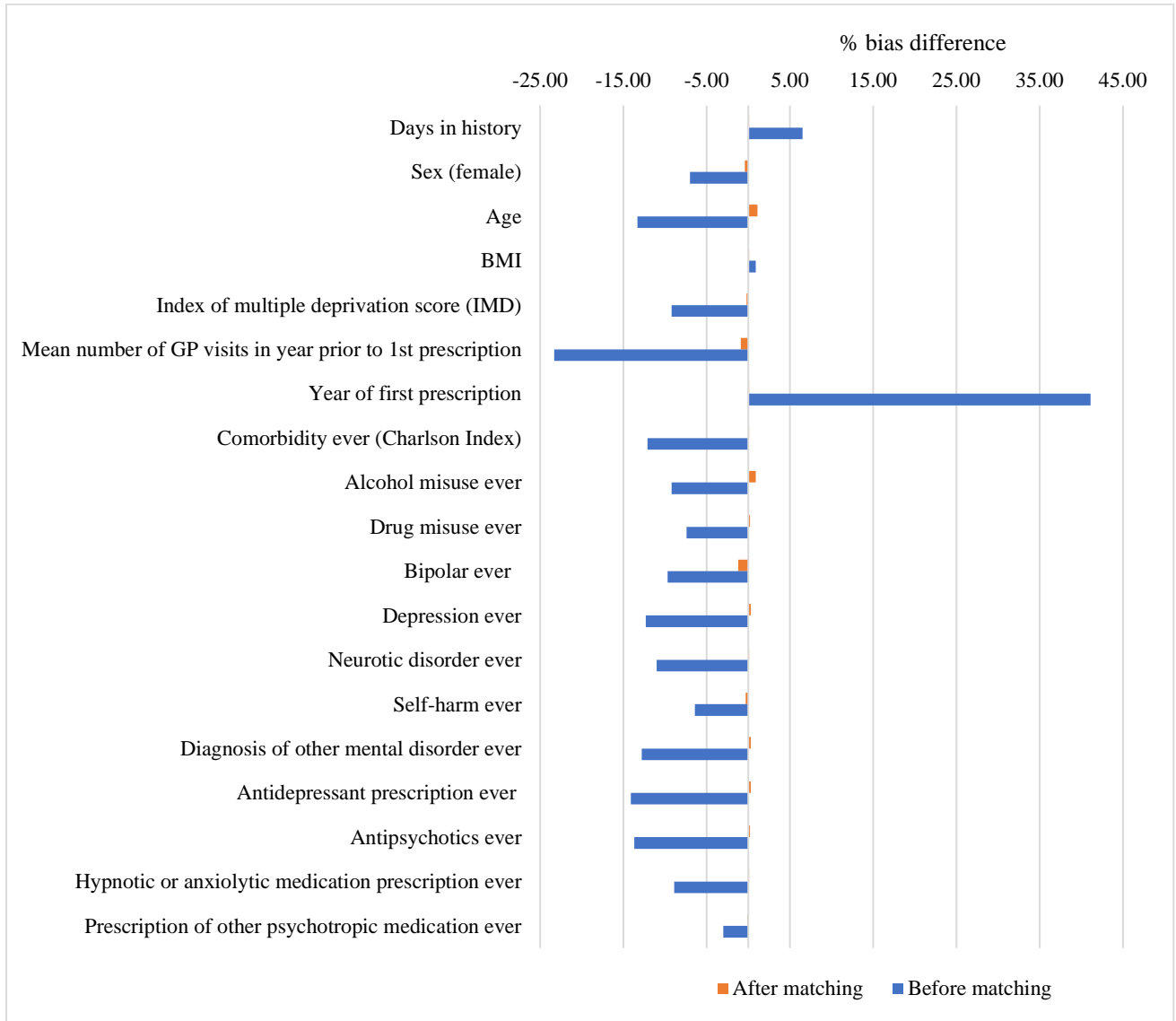


Figure S2. Kernel density estimation of groups' propensity scores before and after matching

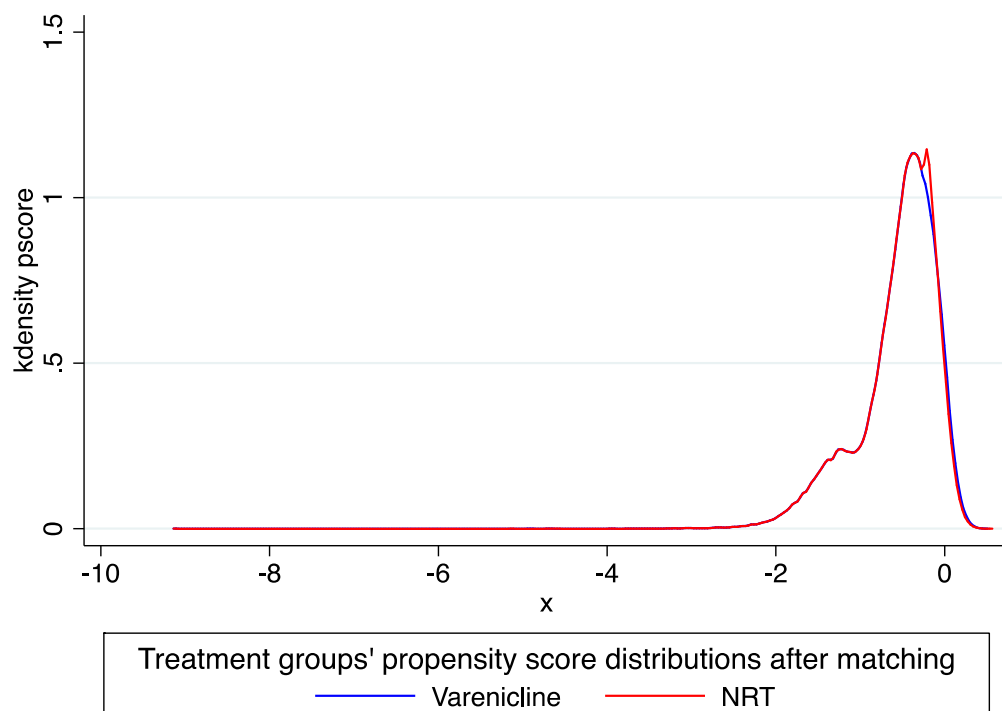
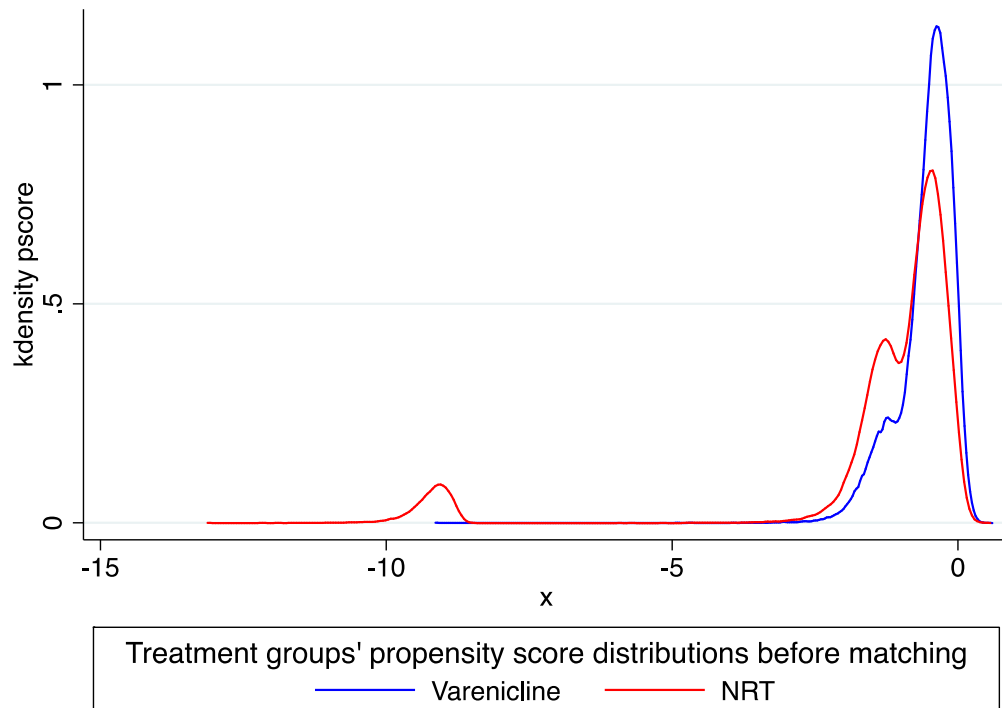


Figure S3. Bias terms of linear regression ■ and instrumental variable methods ▲: Binary and continuous covariates. These plots provide an indication of the bias caused by omitting a single observed covariate from an analysis.

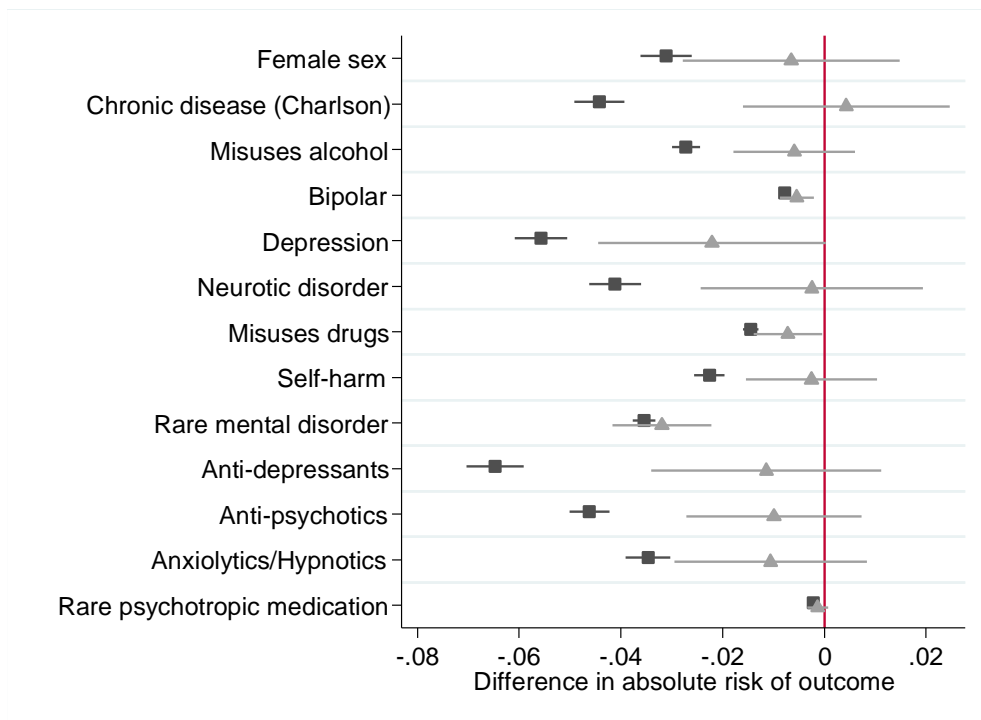
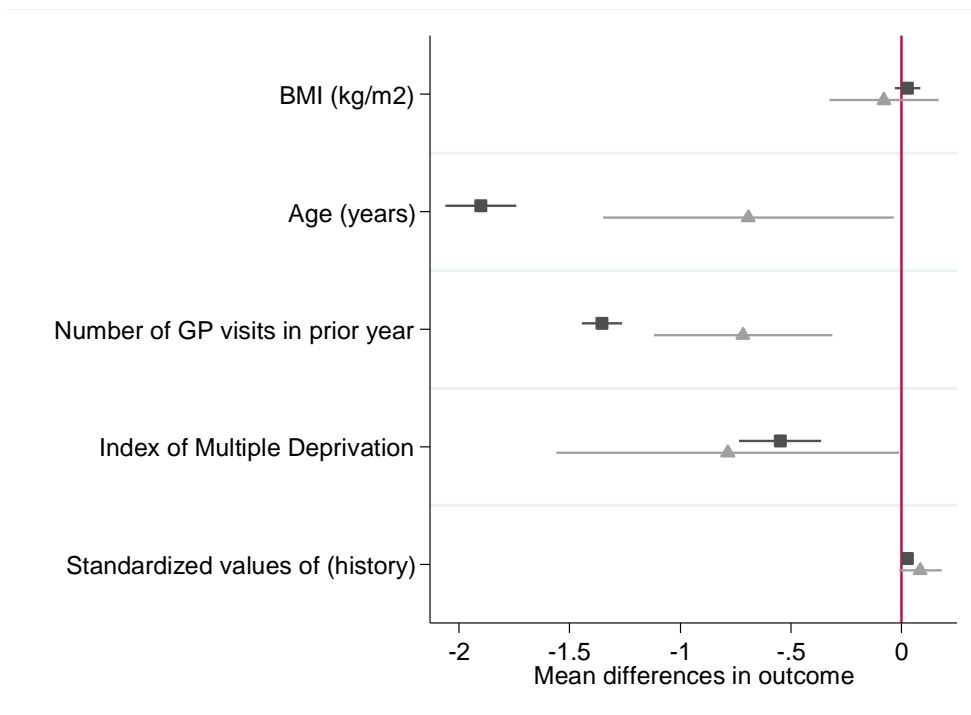
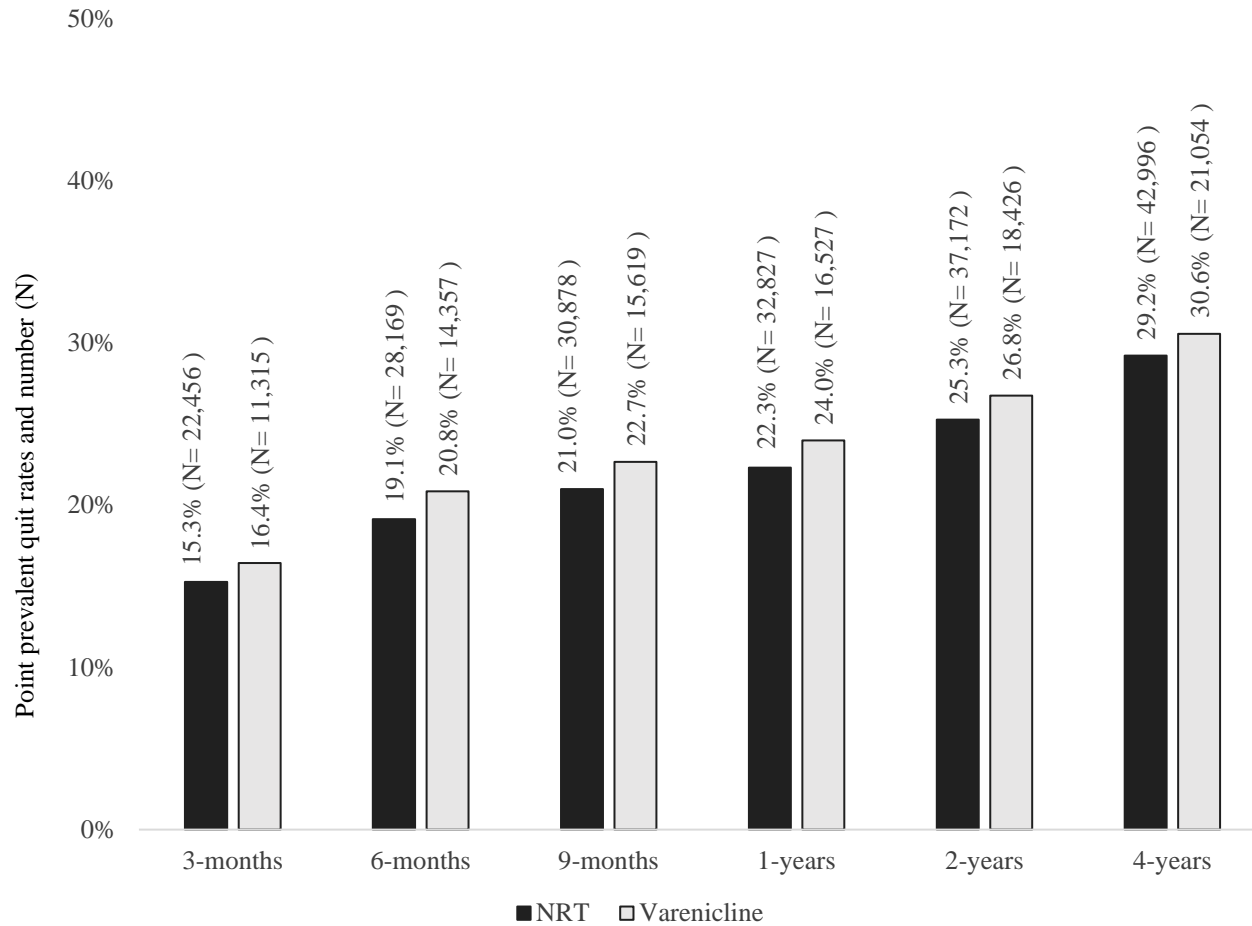


Figure S4. Point prevalence quit rates by instrumental variable condition at 3, 6 and 9-months and 1, 2 and 4-years after exposure, N=216,022*



*4114 patients were excluded from the instrumental variable analysis as they were the first individuals to consult with each GP, thus for these individuals we do not have data about the GP previous prescribing behaviour to enable generation of the instrument.

Table S1. Propensity score matched models: Odds-ratios and 95% confidence intervals for the association between prescription of varenicline versus NRT and smoking cessation at 3, 6 and 9-months and 1, 2 and 4-years after exposure, N=141,218*

Odds-ratio (95% confidence interval) ‡					
3-months	6-months	9-months	1-year	2-years	4-years
1.42 (1.37 to 1.48)	1.45 (1.40 to 1.51)	1.40 (1.35 to 1.45)	1.35 (1.30 to 1.39)	1.27 (1.23 to 1.30)	1.19 (1.16 to 1.22)

‡Model adjusted for propensity score. *Missing BMI and IMD values were imputed using multiple imputation (16).

Table S2. Conventional and instrumental variable linear regression models: Risk difference per 100 patients treated and 95% confidence intervals for the association between varenicline versus NRT and smoking cessation at 3, 6 and 9-months and 1, 2 and 4-years after exposure, N=216,022

Model	3-months	6-months	9-months	1-year	2-years	4-years
	Risk difference (95% confidence interval)					
Linear regression model‡	5.09 (4.61 to 5.58)	6.41 (5.91 to 6.91)	6.15 (5.65 to 6.64)	6.15 (5.65 to 6.64)	5.03 (4.57 to 5.50)	4.33 (3.86 to 4.80)
Instrumental variable linear regression model‡‡	4.13 (2.24 to 6.01)	6.51 (4.52 to 8.500)	6.42 (4.40 to 8.43)	5.97 (3.94 to 8.00)	4.76 (2.77 to 6.74)	4.06 (2.09 to 6.03)
Partial F-statistic*	12466.37	12466.37	12466.37	12466.37	12466.37	12466.37
Hausman test	0.46 P=0.50	0.52 P=0.47	0.86 P=0.35	0.93 P=0.33	0.12 P=0.73	0.11 P=0.74

‡Conventional linear regression model adjusted for age, sex and year of 1st prescription. ‡‡ Instrumental variable linear regression model adjusted only for year of 1st prescription. *This table presents partial F statistics (i.e. the test of the association of the instrument and the prescription) and the Hausman test of endogenous the exposure. 4114 patients were excluded from the instrumental variable analysis as they were the first individuals to consult with each GP, thus for these individuals we do not have data about the GP previous prescribing behavior to enable generation of the instrument.

Effectiveness of varenicline stratified by neighbourhood deprivation

Table S3. Effectiveness of varenicline at 3, 6 and 9-months, and 1, 2 and 4-years after first prescription in the least deprived areas (IMD scores 1 to 10). Effect estimates and 95% confidence intervals presented for each analytic technique.

Analysis technique	3-months	6-months	9-months	1-year	2-years	4-years
Effect estimate (95% confidence interval)						
Logistic regression model ¹	1.45 (1.37 to 1.54) p<0.0001	1.50 (1.42 to 1.57) p<0.0001	1.46 (1.39 to 1.54) p<0.0001	1.38 (1.31 to 1.45) p<0.0001	1.29 (1.23 to 1.35) p<0.0001	1.19 (1.14 to 1.24) p<0.0001
Propensity score matched logistic regression model ²	1.41 (1.33 to 1.50) p<0.0001	1.43 (1.35 to 1.51) p<0.0001	1.23 (1.17 to 1.30) p<0.0001	1.18 (1.08 to 1.28) p=0.0003	1.07 (0.99 to 1.16) p=0.1063	1.00 (0.92 to 1.08) p=0.9841
Instrumental variable analysis ³	1.49 (-2.05 to 5.04) p=0.4097	3.84 (0.05 to 7.64) p=0.0473	4.46 (0.64 to 8.28) p=0.0222	4.13 (0.32 to 7.93) p=0.0334	2.79 (-1.07 to 6.65) p=0.1567	0.59 (-3.30 to 4.48) p=0.7661

1 N= 52,534; data reported are partial adjusted odds-ratios, models were adjusted for age, sex and year of prescription. 2 N= 31,407; data reported are odds-ratios, models were adjusted for propensity score. 3 N= 51,436; data reported are risk difference per 100 patients treated; models were adjusted for year of prescription. ‡Missing IMD values were not imputed, and patients with missing IMD data were excluded from analyses to ensure comparability of results across samples.

Table S4. Effectiveness of varenicline at 3, 6 and 9-months, and 1, 2 and 4-years after first prescription in the most deprived areas (IMD scores 11 to 20). Effect estimates and 95% confidence intervals presented for each analytic technique.‡

Analysis technique	3-months	6-months	9-months	1-year	2-years	4-years
Effect estimate (95% confidence interval)						
Logistic regression model ¹	1.38 (1.31 to 1.46) p<0.0001	1.43 (1.36 to 1.50) p<0.0001	1.37 (1.31 to 1.44) p<0.0001	1.33 (1.27 to 1.39) p<0.0001	1.28 (1.23 to 1.34) p<0.0001	1.22 (1.17 to 1.26) p<0.0001
Propensity score matched logistic regression model ²	1.32 (1.23 to 1.41) p<0.0001	1.35 (1.27 to 1.42) p<0.0001	1.23 (1.17 to 1.29) p<0.0001	1.45 (1.35 to 1.54) p<0.0001	1.37 (1.29 to 1.46) p<0.0001	1.32 (1.24 to 1.40) p<0.0001
Instrumental variable analysis ³	0.66 (-2.44 to 3.75) p=0.6785	3.09 (-0.11 to 6.30) p=0.0584	2.07 (-1.20 to 5.33) p=0.2144	2.15 (-1.21 to 5.52) p=0.2098	2.85 (-0.40 to 6.11) p=0.0856	2.54 (-0.64 to 5.72) p=0.1171

¹ N= 72,247; data reported are partial adjusted odds-ratios, models were adjusted for age, sex and year of prescription. ² N= 40,243; data reported are odds-ratios, models were adjusted for propensity score. ³ N= 71,041; data reported are risk difference per 100 patients treated; models were adjusted for year of prescription. ‡Missing IMD values were not imputed, and patients with missing IMD data were excluded from analyses to ensure comparability of results across samples.

Table S5. Number (N) and percent (%) of patients missing outcome data by treatment at all follow-ups

	3-months	6-months	9-months	1-year	2-years	4-years
NRT	69.4%	56.0%	45.9%	37.2%	19.8%	10.8%
	N=103,743/ 149,526	N=83,803/ 149,526	N=68,665/ 149,526	N=55,696/ 149,526	N=29,608/ 149,526	N=16,151/ 149,526
Varenicline	65.6%	53.6%	44.9%	37.4%	21.4%	12.7%
	46,312/ 70,610	37,819/ 70,610	31,732/ 70,610	26,400/ 70,610	15,129/ 70,610	8,984/ 70,610

Table S6. Comparison of baseline characteristics between the whole sample and patients missing 2-year outcome data

Characteristic	NRT		Varenicline	
	Patients with missing outcome data (N=29,608)	Whole sample (N= 149,526)	Patients with missing outcome data (N= 15,129)	Whole sample (N= 70,610)
Age at time of first prescription ¹	40.7 (14.2)	46.4 (15.5)	40.5 (11.9)	44.5 (13.2)
Sex (female)	43.9% (12,988)	53.7% (80,348)	39.6% (5,995)	50.2% (35,466)
Index of multiple deprivation score (IMD)* ²	12	12	12	12
Mean number of GP visits 1-year prior to first prescription ¹	5.5 (5.7)	7.9 (7.4)	4.5 (4.6)	6.3 (6.1)
BMI* ¹	26.4 (2.8)	26.4 (6.4)	26.5 (2.7)	26.5 (5.9)
Year of first prescription ²	2009	2009	2010	2010
Days of history ¹	3,060.3 (1894.2)	3158.7 (1892.1)	3,210.6 (2011.3)	3283.9 (1976.6)
Comorbidity ever (Charlson Index) (17, 18)	21.8% (6,465)	37.6% (56,274)	19.7% (2,981)	31.9% (22,523)
Alcohol misuse	7.0% (2,076)	8.3% (12,422)	4.9% (742)	6.0 (4,199)
Drug misuse ever	3.2% (948)	3.1% (4,595)	1.94% (294)	1.9% (1,357)
Bipolar ever	<1% (109)	1% (1,464)	<1% (14)	<1% (160)
Depression ever	28.4% (8,398)	35.0% (52,233)	23.5% (3,554)	29.2% (20,615)
Neurotic disorder ever	19.4% (5,737)	24.7% (36,921)	15.5% (2,340)	20.1% (14,189)
Self-harm ever	9.1% (2,700)	10.6% (15,903)	7.5% (1,137)	8.7% (6,169)
Other mental disorder ever	6.0% (1,780)	6.9% (10,343)	3.8% (574)	4.0% (2,832)
Antidepressant prescription ever	41.5% (12,282)	50.1% (74,921)	35.1% (5,304)	43.1% (30,435)
Antipsychotic prescription ever	13.7% (4,066)	20.0% (29,873)	10.2% (1,547)	14.8% (10,459)
Hypnotics/anxiolytics prescription ever	16.2% (4,784)	21.1% (31,513)	13.5% (2,036)	17.6% (12,415)
Other psychotropic medication	<1% (117)	<1% (473)	<1% (32)	<1% (120)

*Missing data: BMI data was missing for 14.2% (N= 31,169); IMD data was missing for 43.3% (N= 95,355). Missing BMI and IMD values were imputed using multiple imputation (16). 1 Data presented are mean and standard deviation. 2 Data presented are median.

Table S7. Multivariable logistic regression models: Comparison of estimates derived from the main analysis and the sensitivity analysis. Fully adjusted odds-ratios and 95% confidence intervals for the association between varenicline versus NRT and smoking cessation at 3, 6 and 9-months and 1, 2 and 4-years after exposure, N=220,136

	3-months	6-months	9-months	1-year	2-years	4-years
	Odds-ratios (95% confidence interval) ‡‡					
Main analysis (missing outcome data=smoking) (19)	1.42 (1.38 to 1.47)	1.46 (1.42 to 1.50)	1.40 (1.36 to 1.44)	1.34 (1.31 to 1.38)	1.26 (1.23 to 1.29)	1.19 (1.16 to 1.21)
Sensitivity (missing outcome data=multiply imputed)	1.37 (1.33 to 1.40)	1.40 (1.36 to 1.44)	1.34 (1.31 to 1.38)	1.29 (1.26 to 1.33)	1.23 (1.20 to 1.26)	1.17 (1.14 to 1.19)

‡‡ Data reported are odds-ratios and models were fully adjusted for all baseline covariates. Missing BMI and IMD values were imputed using multiple imputation (16).

Table S8. Estimated linear regression and instrumental variable bias components

Covariate	N	Difference per 100 patients treated (95% Confidence intervals)		Test for heterogeneity
		Ordinary least squares	Instrumental variables	
Male sex	124,397	3.56 (2.91 to 4.21)	0.93 (-1.72 to 3.57)	0.041
Median age (SD)	124,397	-224.59 (-246.21 to -202.96)	-152.99 (-234.67 to -71.31)	0.065
Body mass index (SD)	107,582	1.43 (-6.05 to 8.90)	-6.99 (-37.60 to 23.61)	0.573
Alcohol misuse	124,397	-2.60 (-2.92 to -2.29)	-1.04 (-2.31 to 0.23)	0.011
Drug misuse	124,397	-1.46 (-1.65 to -1.27)	-1.01 (-1.80 to -0.21)	0.248
Least deprived fifth of patients	124,273	-2.81 (-4.18 to -1.44)	-1.39 (-7.05 to 4.27)	0.531
Most deprived fifth of patients	124,273	2.13 (1.32 to 2.94)	4.28 (0.94 to 7.61)	0.118
Median number of GP visits in year before treatment (SD)	124,397	-150.71 (-162.89 to -138.52)	-131.89 (-183.24 to -80.54)	0.401
Previous use of				
Hypnotics/Anxiolytic	124,397	-3.41 (-3.95 to -2.86)	-1.50 (-3.73 to 0.74)	0.070
Antipsychotic	124,397	-4.88 (-5.41 to -4.35)	-2.68 (-4.79 to -0.58)	0.026
Anti-depressant	124,397	-6.26 (-6.98 to -5.54)	-1.83 (-4.61 to 0.96)	0.001
Statins	124,397	-2.75 (-3.21 to -2.29)	-0.81 (-2.71 to 1.09)	0.032
Anti-hypertensive	124,397	-2.31 (-2.77 to -1.84)	-1.44 (-3.36 to 0.48)	0.350
Diabetic medication	124,397	-2.46 (-2.80 to -2.13)	-0.30 (-1.76 to 1.16)	0.002
Previous diagnosis of				
Self-harm	124,397	-2.15 (-2.53 to -1.76)	-1.17 (-2.79 to 0.45)	0.214
Myocardial infarction	124,397	-1.12 (-1.30 to -0.94)	0.18 (-0.60 to 0.97)	0.001
Chronic obstructive pulmonary disease	124,397	-0.81 (-1.12 to -0.49)	-0.60 (-1.91 to 0.71)	0.744
Chronic disease (Charlson index)	124,397	-4.93 (-5.59 to -4.27)	-0.71 (-3.30 to 1.89)	0.001

Notes: Bias components estimated via GMM. The null hypothesis of the heterogeneity test is that there is no difference between the linear regression and the linear regression bias terms.

1 **Table S9.** Adjusted relative outcome rate among patients treated with varenicline or nicotine replacement
 2 therapy using propensity score methods. Follow-up at 3, 6, 9, 12, 24, 48 months.

		Number of events	Number of patients	Odds-ratio (95% Confidence interval)
Mortality:				
All cause	3	192	80362	0.41 (0.30 to 0.57)
	6	362	77343	0.41 (0.32 to 0.52)
	9	508	74102	0.46 (0.38 to 0.56)
	12	684	70566	0.49 (0.42 to 0.57)
	24	1172	56184	0.68 (0.60 to 0.77)
	48	1389	27871	0.72 (0.64 to 0.81)
Cardiovascular disease	3	52	80222	0.44 (0.24 to 0.81)
	6	111	77092	0.48 (0.32 to 0.72)
	9	152	73746	0.58 (0.42 to 0.81)
	12	193	70075	0.62 (0.46 to 0.83)
	24	351	55363	0.75 (0.61 to 0.94)
	48	406	26889	0.75 (0.61 to 0.91)
Respiratory disease	3	74	80244	0.27 (0.15 to 0.49)
	6	136	77117	0.35 (0.23 to 0.52)
	9	181	73775	0.40 (0.29 to 0.56)
	12	245	70127	0.42 (0.31 to 0.55)
	24	422	55434	0.61 (0.50 to 0.75)
	48	539	27023	0.66 (0.55 to 0.79)
Hospital admission for				
All causes	3	6150	80959	0.70 (0.66 to 0.73)
	6	10365	78654	0.70 (0.67 to 0.73)
	9	13796	76199	0.71 (0.68 to 0.74)
	12	16315	73379	0.76 (0.73 to 0.79)
	24	22287	61413	0.86 (0.83 to 0.89)
	48	19543	34141	0.94 (0.89 to 0.98)
Cardiovascular disease	3	1336	80421	0.68 (0.61 to 0.76)
	6	2277	77498	0.71 (0.65 to 0.77)
	9	2973	74393	0.74 (0.69 to 0.80)
	12	3628	70983	0.77 (0.72 to 0.83)
	24	5337	57038	0.89 (0.84 to 0.94)
	48	5388	29058	0.98 (0.92 to 1.04)
Respiratory disease	3	1437	80375	0.64 (0.57 to 0.71)
	6	2305	77417	0.69 (0.63 to 0.75)
	9	3014	74280	0.74 (0.68 to 0.79)
	12	3583	70804	0.76 (0.71 to 0.81)
	24	5080	56716	0.86 (0.81 to 0.91)
	48	4975	28636	0.95 (0.89 to 1.02)

Primary care diagnosis of:

		Number of events	Number of patients	Odds-ratio (95% Confidence interval)
Myocardial infarction	3	70	78786	0.40 (0.24 to 0.68)
	6	110	75660	0.67 (0.45 to 0.98)
	9	146	72335	0.70 (0.50 to 0.97)
	12	177	68706	0.81 (0.60 to 1.09)
	24	280	54118	0.91 (0.72 to 1.16)
	48	317	26134	1.03 (0.82 to 1.30)
Chronic obstructive pulmonary disease	3	674	75336	0.67 (0.57 to 0.79)
	6	868	72390	0.71 (0.62 to 0.82)
	9	1070	69269	0.77 (0.68 to 0.88)
	12	1237	65816	0.80 (0.71 to 0.90)
	24	1682	51966	0.88 (0.80 to 0.97)
	48	1652	25266	0.98 (0.89 to 1.09)

3

4 **Table S10.** Adjusted Relative Outcome Frequency Among Patients Treated With Varenicline or Nicotine Replacement Therapy Using Propensity
5 score Methods. Follow-up at 3, 6, 9, 12, 24, 48 Months. Non-imputed data. Reproduced without changes from Davies et al. 2018.²¹
6 <https://onlinelibrary.wiley.com/doi/abs/10.1111/add.14146>

7

Outcome	Follow-up length	Number of patients N	Percentage difference (95% confidence intervals) Fully adjusted
Number of GP visits	3	80,185	11.10 (6.63 to 15.76)
	6	77,005	2.86 (-0.74 to 6.59)
	9	73,617	-0.18 (-3.35 to 3.08)
	12	69,909	-1.37 (-4.29 to 1.63)
	24	55,032	-1.98 (-4.53 to 0.63)
	48	26,462	-1.15 (-4.04 to 1.82)
Number of hospitalizations	3	81,840	-7.20 (-8.23 to -6.17)
	6	80,493	-11.05 (-12.36 to -9.73)
	9	78,935	-13.52 (-14.99 to -12.03)
	12	76,990	-12.89 (-14.49 to -11.26)
	24	67,202	-9.97 (-11.97 to -7.93)
	48	39,405	-5.30 (-8.21 to -2.29)
Number of hospitalizations for respiratory disease	3	80,421	-1.83 (-2.36 to -1.30)
	6	77,498	-2.88 (-3.59 to -2.17)
	9	74,393	-3.32 (-4.15 to -2.49)
	12	70,983	-3.62 (-4.57 to -2.66)
	24	57,038	-2.93 (-4.38 to -1.46)
	48	29,058	-1.11 (-3.98 to 1.84)
Number of hospitalizations for cardiovascular disease	3	80,375	-2.27 (-2.82 to -1.71)
	6	77,417	-3.13 (-3.84 to -2.43)
	9	74,280	-3.40 (-4.22 to -2.56)
	12	70,804	-3.85 (-4.79 to -2.90)
	24	56,716	-3.57 (-4.94 to -2.18)
	48	28,636	-2.36 (-5.10 to 0.47)

8

Table S11. Means and bias of baseline covariates before and after propensity score matching. Reproduced without changes from Davies et al. 2018.²¹
<https://onlinelibrary.wiley.com/doi/abs/10.1111/add.14146>

		Mean		% bias	% Bias reduction
		Varenicline	NRT		
History	Unmatched	3298	3143	7.9	
	Matched	3298	3303	-0.2	97
Male	Unmatched	0.50	0.46	7.8	
	Matched	0.50	0.50	1.1	86.3
Age (years)	Unmatched	44.4	46.7	-15.6	
	Matched	44.4	44.5	-0.4	97.3
BMI (kg/m2)	Unmatched	26.5	26.4	0.9	
	Matched	26.5	26.5	0.2	80.9
Missing BMI	Unmatched	0.14	0.13	2.2	
	Matched	0.14	0.14	-0.1	96.1
Prescription year:	2007 Unmatched	0.11	0.23	-34.4	
	Matched	0.11	0.11	-1.8	94.7
2008	Unmatched	0.14	0.16	-4	
	Matched	0.14	0.14	1.4	65.2
2009	Unmatched	0.16	0.15	2.6	
	Matched	0.16	0.16	0.7	70.7
2010	Unmatched	0.18	0.13	14.2	
	Matched	0.18	0.19	-1.2	91.3
2011	Unmatched	0.16	0.11	15.9	
	Matched	0.16	0.17	-0.6	96.3
2012	Unmatched	0.13	0.08	14.6	
	Matched	0.13	0.12	0.7	95.4
2013	Unmatched	0.09	0.06	14.3	
	Matched	0.09	0.09	1.3	90.7
2014	Unmatched	0.02	0.01	5.3	
	Matched	0.02	0.02	0	99.6
Misuses alcohol	Unmatched	0.05	0.07	-9.6	
	Matched	0.05	0.05	0.3	96.5
Misuses drugs	Unmatched	0.02	0.03	-7.5	
	Matched	0.02	0.02	0	99.4
Index of Multiple Deprivation	Unmatched	11.2	11.7	-9.2	
	Matched	11.2	11.3	-1.8	80.3
Missing IMD	Unmatched	0.001	0.001	-0.1	
	Matched	0.001	0.001	0.2	-60.6
Hypnotics	Unmatched	0.17	0.20	-8.9	
	Matched	0.17	0.17	-0.4	95.4
Antipsychotics	Unmatched	0.14	0.20	-14.9	

		Mean		%bias	% Bias reduction
		Varenicline	NRT		
Antidepressants	Matched	0.14	0.14	-0.2	98.7
	Unmatched	0.43	0.50	-14.2	
Statins	Matched	0.43	0.44	-1.6	88.5
	Unmatched	0.15	0.19	-12.1	
Antihypertensives	Matched	0.15	0.15	-1.2	90.3
	Unmatched	0.17	0.21	-10.3	
Diabetic medications	Matched	0.17	0.17	-0.2	97.9
	Unmatched	0.06	0.09	-10	
Previously diagnosed with: Self-harm	Matched	0.06	0.07	-0.7	92.5
	Unmatched	0.09	0.10	-6	
Myocardial infarction	Matched	0.09	0.09	-0.7	88
	Unmatched	0.02	0.03	-9.1	
Chronic obstructive pulmonary disease	Matched	0.02	0.02	-0.4	96
	Unmatched	0.06	0.08	-7.9	
Chronic disease (Charlson index)	Matched	0.06	0.06	-0.6	92
	Unmatched	0.32	0.38	-13.6	
	Matched	0.32	0.32	-0.6	95.5