Supplementary Materials to:

Periodontal and other oral bacteria and risk of lung cancer in the Atherosclerosis Risk in Communities (ARIC) Study

Baijun Zhou[1], Jiayun Lu[1], James D. Beck[2], Kevin L. Moss[2], Anna E. Prizment[3], Ryan T. Demmer[4], Kori A. Porosnicu Rodriguez[1,5], Corinne E. Joshu[1,6], Dominique S. Michaud[7], Elizabeth A. Platz[1,6]

[1]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

[2]Division of Comprehensive Oral Health/ Periodontology, Adams School of Dentistry, University of North Carolina, Chapel Hill, NC

[3]Division of Hematology, Oncology and Transplantation, University of Minnesota Medical School, and the University of Minnesota Masonic Cancer Center, Minneapolis, MN

[4]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN

[5]Johns Hopkins University School of Medicine, Baltimore, MD

[6]Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD

[7]Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA

Supplement Methods 1. Multiple Imputation

Missing values for smoking packyears (N=353) and family history of cancer (N=184) were imputed using the fully conditional specification (FCS) method. Variables used for imputation were age, field center, race (Black or White), ever use of hormone replacement therapy (HRT; females), sex, body mass index (BMI), cigarette smoking status (never, former, current), alcohol drinking status (never, former, current), diabetes status, calculated lifecourse socioeconomic status (SES), lung cancer status, and time of follow-up. The process of multiple imputation has three phases: imputation phase, analysis phase, and pooling phase. At the imputation phase, missing values for smoking packyears were imputed using the predictive mean matching method which randomly assigns a value from a set of observed values whose predicted values are closest to the predicted values for the missing values from the imputation model. Missing values for family history of cancer were imputed using the discriminant function method which is for the nominal response. The number of imputed datasets was 10, and the imputed values in each imputation were generated after 20 iterations. At the analysis phase, we ran Cox proportional hazards model using each of the 10 imputed datasets and obtained 10 sets of parameter estimates. At the pooling phase, the parameters from 10 analyzed datasets were combined for inference. The point estimate of the parameter estimate was the mean of 10 complete-data estimates.

$$\bar{\beta} = \frac{1}{m}\sum_{i=1}^{m}\widehat{\beta_i}$$

Where m is the number of imputed dataset; $\widehat{\beta_i}$ is the parameter estimate from the ith imputed dataset, $i = 1, 2, \dots, m$.

The variance of the parameter was the combination of the within-imputation variance and the across-imputation variance.

$$V = \frac{1}{m}\sum_{i=1}^{m}\widehat{s_i^2} + \left(1 + \frac{1}{m}\right)(m-1)\sum_{i=1}^{m}(\widehat{\beta_i} - \bar{\beta})^2$$

Where m is the number of imputed datasets; $\bar{\beta}$ is the pooled point parameter estimate; $\widehat{s_i}$ and $\widehat{\beta_i}$ are the variance estimate and point estimate from the ith imputed dataset, $i = 1, 2, \dots, m$.

Table S1. Distribution of pre-diagnostic concentrations of antibodies to 18 oral bacteria in lung cancer cases (N = 118) and non-cases (N = 4145) in ARIC [a]

Bacterium	% at or below LOD			% over 80th percentile [b] among those with concentrations above LOD	
	Case	Non-case	Overall	Case	Non-case
A. actinomycetemcomitans	16.95	20.05	19.96	26.53	19.79
A. viscosus	57.63	53.68	53.79	6.00	20.36
C. ochracea	55.93	54.07	54.12	19.23	20.01
C. rectus	61.86	61.98	61.98	11.11	20.24
E. corrodens	70.34	74.26	74.15	11.43	20.24
F. nucleatum	75.42	76.96	76.92	27.59	19.79
H. pylori	59.32	62.36	62.28	10.42	20.26
P. gingivalis	60.17	60.70	60.68	14.89	20.14
P. intermedia	35.59	39.47	39.36	19.74	20.01
P. micra	25.42	29.07	28.97	36.36	19.49
P. nigrescens	27.12	24.29	24.37	23.26	19.92
S. intermedius	55.93	56.02	56.02	23.08	19.91
S. noxia	70.34	75.37	75.23	14.29	20.18
S. oralis	47.46	54.52	54.33	24.19	19.84
S. sanguis	50.00	57.61	57.40	16.95	20.09
T. denticola	54.24	47.50	47.69	20.37	19.99
T. forsythensis	44.92	48.49	48.39	16.92	20.09
V. parvula	82.20	83.11	83.09	23.81	19.86
All bacteria	-	-	-	19.83	20.01
Red complex	-	-	-	16.67	20.09
Orange complex	-	-	-	24.77	19.86

[a] Lung cancer cases are participants who developed a first primary lung cancer during follow-up; All bacteria include 18 bacteria described in the study; Red complex bacteria include *P. gingivalis*, *T. denticola*, and *T. forsythensis*; Orange complex bacteria include *C. rectus*, *F. nucleatum*, *P. intermedia*, *P. micra*, and *P. nigrescens*; LOD = limit of detection (20 ng/mL).

[b] 80th percentile defined in the study population.

Table S2. Adjusted hazard ratios of non-small cell lung cancer incidence and lung cancer mortality by three levels of antibodies to 18 oral bacteria and to the sum of antibodies to total, red, and orange complex oral bacteria, by sex and race, in 4263 participants in ARIC [a]

Bacterium	Sex [b]				Race [c]			
	Male (N = 1918, 67 cases)		Female (N = 2345, 51 cases)		Black (N = 696, 19 cases)		White (N = 3567, 99 case)	
	HR (95% CI)	P_{trend} [d]	HR (95% CI)	P_{trend}	HR (95% CI)	P_{trend}	HR (95% CI)	P_{trend}
A. actinomycetemcomitans								
Group 1 (lowest)	1.00 (ref.)	**0.02**	1.00 (ref.)	0.86	1.00 (ref.)	0.75	1.00 (ref.)	0.07
Group 2	0.67 (0.31-1.47)		1.45 (0.67-3.13)		1.22 (0.21-7.02)		0.99 (0.56-1.77)	
Group 3 (highest)	1.41 (0.68-2.90)		1.15 (0.52-2.55)		1.32 (0.28-6.27)		1.45 (0.82-2.53)	
A. viscosus								
Group 1 (lowest)	1.00 (ref.)	0.83	1.00 (ref.)	0.69	1.00 (ref.)	0.26	1.00 (ref.)	0.90
Group 2	0.92 (0.50-1.68)		0.57 (0.26-1.24)		0.38 (0.08-1.78)		1.00 (0.61-1.64)	
Group 3 (highest)	1.06 (0.57-1.97)		0.84 (0.41-1.72)		0.42 (0.09-2.01)		1.03 (0.63-1.70)	
C. ochracea								
Group 1 (lowest)	1.00 (ref.)	0.61	1.00 (ref.)	0.53	1.00 (ref.)	0.63	1.00 (ref.)	0.78
Group 2	1.17 (0.63-2.18)		0.81 (0.39-1.71)		0.23 (0.03-1.85)		1.18 (0.72-1.93)	
Group 3 (highest)	1.16 (0.65-2.07)		0.80 (0.36-1.74)		1.20 (0.43-3.32)		1.07 (0.64-1.78)	
C. rectus								
Group 1 (lowest)	1.00 (ref.)	0.77	1.00 (ref.)	0.26	1.00 (ref.)	**0.04**	1.00 (ref.)	0.74
Group 2	1.52 (0.86-2.70)		1.06 (0.50-2.25)		0.52 (0.17-1.61)		1.48 (0.91-2.39)	
Group 3 (highest)	0.87 (0.43-1.75)		0.56 (0.22-1.46)		0.22 (0.05-1.02)		1.04 (0.59-1.86)	
E. corrodens								
Group 1 (lowest)	1.00 (ref.)	0.15	1.00 (ref.)	0.38	1.00 (ref.)	0.81	1.00 (ref.)	0.22
Group 2	1.39 (0.71-2.74)		0.76 (0.27-2.16)		0.27 (0.03-2.21)		1.38 (0.77-2.48)	
Group 3 (highest)	1.54 (0.83-2.86)		0.66 (0.23-1.87)		0.97 (0.30-3.17)		1.35 (0.78-2.35)	
F. nucleatum								
Group 1 (lowest)	1.00 (ref.)	0.31	1.00 (ref.)	0.31	1.00 (ref.)	0.75	1.00 (ref.)	0.31
Group 2	1.27 (0.61-2.62)		0.74 (0.26-2.09)		0.26 (0.03-2.05)		1.27 (0.68-2.35)	
Group 3 (highest)	1.38 (0.71-2.66)		0.57 (0.17-1.88)		1.03 (0.27-3.98)		1.31 (0.72-2.38)	
H. pylori								
Group 1 (lowest)	1.00 (ref.)	0.16	1.00 (ref.)	0.63	1.00 (ref.)	0.27	1.00 (ref.)	0.13

Group 2	0.88 (0.45-1.72)		0.55 (0.22-1.32)		0.39 (0.09-1.81)		0.81 (0.46-1.44)	
Group 3 (highest)	1.49 (0.83-2.68)		0.84 (0.40-1.79)		0.51 (0.17-1.51)		1.43 (0.89-2.30)	
P. gingivalis								
Group 1 (lowest)	1.00 (ref.)	0.79	1.00 (ref.)	0.27	1.00 (ref.)	0.13	1.00 (ref.)	0.90
Group 2	1.36 (0.75-2.44)		0.87 (0.43-1.76)		2.08 (0.67-6.43)		1.00 (0.60-1.64)	
Group 3 (highest)	0.95 (0.45-2.00)		0.59 (0.24-1.48)		0.57 (0.15-2.19)		0.96 (0.51-1.80)	
P. intermedia								
Group 1 (lowest)	1.00 (ref.)	0.40	1.00 (ref.)	0.70	1.00 (ref.)	0.60	1.00 (ref.)	0.16
Group 2	1.67 (0.90-3.11)		1.00 (0.50-1.99)		0.58 (0.16-2.03)		1.56 (0.96-2.53)	
Group 3 (highest)	1.50 (0.80-2.84)		1.14 (0.57-2.26)		0.62 (0.20-1.92)		1.54 (0.94-2.54)	
P. micra								
Group 1 (lowest)	1.00 (ref.)	0.16	1.00 (ref.)	0.50	1.00 (ref.)	0.56	1.00 (ref.)	0.41
Group 2	1.17 (0.54-2.51)		0.71 (0.36-1.42)		0.90 (0.28-2.91)		0.84 (0.47-1.48)	
Group 3 (highest)	1.56 (0.78-3.13)		0.73 (0.37-1.42)		0.71 (0.22-2.29)		1.10 (0.66-1.84)	
P. nigrescens								
Group 1 (lowest)	1.00 (ref.)	0.09	1.00 (ref.)	0.42	1.00 (ref.)	0.67	1.00 (ref.)	0.23
Group 2	0.90 (0.46-1.76)		1.06 (0.55-2.05)		1.25 (0.36-4.37)		1.04 (0.62-1.72)	
Group 3 (highest)	1.48 (0.78-2.79)		0.78 (0.37-1.66)		0.91 (0.24-3.35)		1.32 (0.79-2.20)	
S. intermedius								
Group 1 (lowest)	1.00 (ref.)	0.72	1.00 (ref.)	0.93	1.00 (ref.)	0.44	1.00 (ref.)	0.47
Group 2	0.95 (0.48-1.88)		1.11 (0.56-2.19)		1.42 (0.49-4.12)		0.94 (0.55-1.63)	
Group 3 (highest)	1.10 (0.62-1.97)		1.05 (0.50-2.18)		0.64 (0.17-2.43)		1.17 (0.73-1.88)	
S. noxia								
Group 1 (lowest)	1.00 (ref.)	0.81	1.00 (ref.)	0.36	1.00 (ref.)	0.35	1.00 (ref.)	0.54
Group 2	2.06 (1.16-3.66)		0.95 (0.37-2.44)		0.51 (0.11-2.41)		**1.96 (1.19-3.23)**	
Group 3 (highest)	0.92 (0.40-2.09)		0.62 (0.22-1.75)		1.71 (0.51-5.75)		0.77 (0.36-1.62)	
S. oralis								
Group 1 (lowest)	1.00 (ref.)	0.01	1.00 (ref.)	0.26	1.00 (ref.)	0.62	1.00 (ref.)	0.17
Group 2	1.21 (0.63-2.32)		0.91 (0.47-1.78)		1.06 (0.29-3.82)		1.06 (0.65-1.75)	
Group 3 (highest)	**2.19 (1.23-3.89)**		0.60 (0.25-1.46)		1.29 (0.45-3.72)		1.40 (0.86-2.27)	
S. sanguis								
Group 1 (lowest)	1.00 (ref.)	0.92	1.00 (ref.)	0.35	1.00 (ref.)	0.66	1.00 (ref.)	0.55
Group 2	1.68 (0.94-2.99)		0.84 (0.40-1.77)		1.80 (0.61-5.34)		1.21 (0.74-1.99)	
Group 3 (highest)	1.04 (0.55-1.96)		1.32 (0.68-2.57)		1.44 (0.40-5.25)		1.19 (0.73-1.93)	

T. denticola								
Group 1 (lowest)	1.00 (ref.)	0.39	1.00 (ref.)	0.03	1.00 (ref.)	0.37	1.00 (ref.)	0.81
Group 2	0.78 (0.41-1.46)		0.58 (0.28-1.18)		0.42 (0.12-1.50)		0.74 (0.44-1.23)	
Group 3 (highest)	1.25 (0.71-2.21)		0.41 (0.17-0.99)		0.55 (0.18-1.74)		0.94 (0.58-1.54)	
T. forsythensis								
Group 1 (lowest)	1.00 (ref.)	0.43	1.00 (ref.)	0.87	1.00 (ref.)	0.52	1.00 (ref.)	0.64
Group 2	**1.90 (1.08-3.32)**		0.92 (0.46-1.81)		0.70 (0.20-2.39)		**1.74 (1.11-2.74)**	
Group 3 (highest)	1.39 (0.73-2.65)		0.94 (0.45-1.95)		1.44 (0.44-4.66)		1.18 (0.69-2.00)	
V. parvula								
Group 1 (lowest)	1.00 (ref.)	0.85	1.00 (ref.)	0.34	1.00 (ref.)	0.81	1.00 (ref.)	0.86
Group 2 (highest)	1.06 (0.59-1.90)		0.63 (0.25-1.62)		0.85 (0.23-3.17)		1.05 (0.62-1.76)	
Red complex								
Group 1 (lowest)	1.00 (ref.)	0.84	1.00 (ref.)	0.15	1.00 (ref.)	0.14	1.00 (ref.)	0.84
Group 2	1.22 (0.67-2.21)		0.95 (0.50-1.80)		0.65 (0.20-2.14)		1.17 (0.73-1.87)	
Group 3 (highest)	1.01 (0.53-1.90)		0.58 (0.27-1.24)		0.38 (0.11-1.29)		0.99 (0.59-1.64)	
Orange complex								
Group 1 (lowest)	1.00 (ref.)	0.08	1.00 (ref.)	0.92	1.00 (ref.)	0.6	1.00 (ref.)	0.08
Group 2	1.55 (0.78-3.09)		1.17 (0.60-2.27)		0.89 (0.27-2.96)		1.54 (0.91-2.58)	
Group 3 (highest)	1.89 (0.98-3.66)		1.00 (0.49-2.04)		0.74 (0.23-2.35)		**1.68 (1.01-2.80)**	
P. intermedia + P. nigrescens								
Group 1 (lowest)	1.00 (ref.)	0.20	1.00 (ref.)	0.47	1.00 (ref.)	0.27	1.00 (ref.)	0.24
Group 2	1.37 (0.73-2.58)		1.07 (0.56-2.05)		1.26 (0.37-4.25)		1.34 (0.82-2.19)	
Group 3 (highest)	1.60 (0.85-2.99)		0.78 (0.37-1.65)		0.62 (0.17-2.23)		1.42 (0.86-2.33)	
All bacteria								
Group 1 (lowest)	1.00 (ref.)	0.33	1.00 (ref.)	0.44	1.00 (ref.)	0.12	1.00 (ref.)	0.32
Group 2	1.56 (0.81-3.00)		0.86 (0.44-1.69)		0.53 (0.16-1.72)		1.55 (0.94-2.57)	
Group 3 (highest)	1.53 (0.80-2.93)		0.75 (0.37-1.50)		0.38 (0.12-1.15)		1.41 (0.85-2.32)	

[a] For each bacterium, Group 1 was participants with an antibody concentration at or below the limit of detection (20 ng/mL); Group 2 and Group 3 were participants with an antibody concentration above the limit of detection, divided at the median concentration; if the proportion of at or below the limit of detection was larger than 80% for the antibody, only one group was set for those with an antibody level above the limit of detection; For the sum of bacteria (red complex, orange complex, *P. intermedia* + *P. nigrescens*, and all bacteria), participants were divided into tertiles, where Group 1 was the lowest tertile and Group 3 was the highest tertile; All bacteria include 18 bacteria described in the study; Red complex bacteria include *P. gingivalis*, *T. denticola*, and *T. forsythensis*; Orange complex bacteria include *C. rectus*, *F. nucleatum*, *P. intermedia*, *P. micra*, and *P. nigrescens*; HR = hazard ratio; CI = confidence interval; HRT = hormone replacement therapy; BMI = body mass index.

[b] Models by gender were adjusted for age, joint terms for field center and race (Black from Jackson; Black from Forsyth; White from Forsyth; White from Washington County [reference is White from Minneapolis]), cigarette smoking status, packyears smoked, alcohol drinking status, BMI, diagnosed diabetes status, undiagnosed diabetes status, at risk for diabetes status, family history of cancer, and socioeconomic status (SES was adjusted using a lifecourse SES variable); Models for females were additionally adjusted for HRT use.

[c] Models by race were adjusted for age, field center (Jackson for Black [reference is Forsyth], Forsyth and Washington County for White [reference is Minneapolis]), joint terms for sex and HRT use (female user, female nonuser [reference is men]), cigarette smoking status, packyears smoked, alcohol drinking status, BMI, diagnosed diabetes status, undiagnosed diabetes status, at risk for diabetes status, family history of cancer, and socioeconomic status (SES was adjusted using a lifecourse SES variable).

[d] P value for trend was from the Wald test of the coefficient for the ordinal variable of the 3 groups in a Cox proportional hazards regression.

Table S3. Adjusted hazard ratios of lung cancer incidence for antibodies to 18 bacteria and the sum of antibodies to total, red, and orange complex oral bacteria among ever-smokers (N = 2558) and among nondiabetics (N = 2166) in ARIC [a]

Bacterium	Ever-smokers [b] (N = 2558, 107 cases)	Nondiabetics [c] (N = 2166, 49 cases)
	HR (per IQR) (95% CI)	HR (per IQR) (95% CI)
A. actinomycetemcomitans	1.04 (0.93-1.17)	1.05 (0.87-1.27)
A. viscosus	0.93 (0.83-1.04)	0.87 (0.70-1.09)
C. ochracea	1.01 (0.88-1.15)	0.97 (0.77-1.23)
C. rectus	0.88 (0.74-1.04)	0.83 (0.63-1.09)
E. corrodens	0.99 (0.89-1.09)	0.93 (0.75-1.15)
F. nucleatum	1.03 (0.93-1.14)	1.02 (0.87-1.19)
H. pylori	0.97 (0.89-1.05)	0.89 (0.74-1.07)
P. gingivalis	0.94 (0.86-1.02)	0.95 (0.82-1.09)
P. intermedia	1.09 (0.99-1.19)	**1.19 (1.09-1.31)**
P. micra	1.06 (0.99-1.14)	0.98 (0.83-1.17)
P. nigrescens	1.05 (0.94-1.17)	**1.25 (1.12-1.40)**
S. intermedius	1.00 (0.95-1.04)	0.99 (0.92-1.06)
S. noxia	1.00 (0.98-1.01)	1.00 (0.98-1.02)
S. oralis	1.02 (0.98-1.06)	1.00 (0.93-1.07)
S. sanguis	1.01 (0.95-1.08)	0.99 (0.89-1.10)
T. denticola	0.94 (0.79-1.11)	0.93 (0.70-1.23)
T. forsythensis	1.01 (0.92-1.11)	1.04 (0.90-1.21)
V. parvula	0.94 (0.84-1.05)	0.89 (0.72-1.10)
Red complex	0.91 (0.78-1.06)	0.94 (0.76-1.18)
Orange complex	1.10 (0.97-1.26)	**1.24 (1.05-1.45)**
P. intermedia + *P. nigrescens*	1.08 (0.97-1.21)	**1.28 (1.14-1.44)**
All bacteria	1.02 (0.87-1.20)	1.05 (0.83-1.32)

[a] Antibodies to each bacterium and sum of antibody concentrations entered as a continuous variable in the model; All bacteria: 18 bacteria; Red complex: *P. gingivalis*, *T. denticola*, and *T. forsythensis*; Orange complex: *C. rectus*, *F. nucleatum*, *P. intermedia*, *P. micra*, and *P. nigrescens*; HR = hazard ratio; CI = confidence interval; IQR = interquartile range; HRT = hormone replacement therapy; BMI = body mass index.

[b] Ever-smokers include both current smokers and former smokers at Visit 4; Models were adjusted for age, joint terms for field center and race (Black from Jackson; Black from Forsyth; White from Forsyth; White from Washington County [reference is White from Minneapolis]), joint terms for sex and HRT use (female user, female nonuser [reference is men]), cigarette packyears smoked, alcohol drinking status, BMI, diagnosed diabetes status, undiagnosed diabetes status, at risk for diabetes status, family history of cancer, and socioeconomic status (SES was adjusted using a lifecourse SES variable).

[c] Nondiabetics include participants who had no diagnosed diabetes, no undiagnosed diabetes, and were not at risk for diabetes; Models were adjusted for age, joint terms for field center and race (Black from Jackson; Black from Forsyth; White from Forsyth; White from Washington County [reference is White from Minneapolis]), joint terms for sex and HRT use (female user, female nonuser [reference is men]), cigarette smoking status, packyears smoked, alcohol drinking status, BMI, family history of cancer, and socioeconomic status (SES was adjusted using a lifecourse SES variable).

Table S4. Adjusted hazard ratios of non-small cell lung cancer and lung cancer mortality for antibodies to 18 bacteria and the sum of antibodies to total, red, and orange complex oral bacteria, in 4263 participants in ARIC [a]

Bacterium	Non-small cell (N_{case} = 77)	Mortality (N_{case} = 83)
	HR (per IQR) (95% CI) [b]	HR (per IQR) (95% CI)
A. actinomycetemcomitans	1.00 (0.86-1.17)	1.03 (0.90-1.17)
A. viscosus	0.87 (0.73-1.03)	0.93 (0.82-1.07)
C. ochracea	0.99 (0.84-1.17)	0.98 (0.83-1.15)
C. rectus	0.88 (0.72-1.07)	0.81 (0.64-1.02)
E. corrodens	0.98 (0.86-1.11)	0.92 (0.78-1.09)
F. nucleatum	1.03 (0.91-1.15)	1.01 (0.88-1.15)
H. pylori	0.96 (0.88-1.06)	0.94 (0.84-1.05)
P. gingivalis	0.99 (0.93-1.06)	0.95 (0.87-1.05)
P. intermedia	**1.17 (1.08-1.26)**	1.10 (0.98-1.22)
P. micra	1.04 (0.95-1.14)	1.06 (0.97-1.16)
P. nigrescens	**1.15 (1.06-1.25)**	1.09 (0.95-1.24)
S. intermedius	1.00 (0.95-1.05)	0.96 (0.90-1.02)
S. noxia	1.01 (0.99-1.02)	0.98 (0.94-1.02)
S. oralis	1.03 (0.98-1.07)	0.94 (0.86-1.03)
S. sanguis	0.98 (0.90-1.07)	1.00 (0.93-1.08)
T. denticola	0.93 (0.76-1.14)	0.82 (0.65-1.04)
T. forsythensis	1.03 (0.95-1.13)	1.01 (0.89-1.15)
V. parvula	0.95 (0.83-1.09)	0.94 (0.83-1.08)
Red complex	1.00 (0.88-1.13)	0.91 (0.75-1.10)
Orange complex	**1.20 (1.06-1.36)**	1.12 (0.95-1.32)
P. intermedia + *P. nigrescens*	**1.20 (1.09-1.32)**	1.11 (0.97-1.28)
All bacteria	1.09 (0.92-1.30)	0.93 (0.75-1.16)

[a] Antibodies to each bacterium and sum of antibody concentrations entered as a continuous variable in the model; All bacteria: 18 bacteria; Red complex: *P. gingivalis*, *T. denticola*, and *T. forsythensis*; Orange complex: *C. rectus*, *F. nucleatum*, *P. intermedia*, *P. micra*, and *P. nigrescens*; HR = hazard ratio; CI = confidence interval; IQR = interquartile range; HRT = hormone replacement therapy; BMI = body mass index.

[b] All models were adjusted for age, joint terms for field center and race (Black from Jackson; Black from Forsyth; White from Forsyth; White from Washington County [reference is White from Minneapolis]), joint terms for sex and HRT use (female user, female nonuser [reference is men]), cigarette smoking status, packyears smoked, alcohol drinking status, BMI, diagnosed diabetes status, undiagnosed diabetes status, at risk for diabetes status, family history of cancer, and lifecourse socioeconomic status.

Table S5. Median (IQR) counts for pre-diagnostic DNA to 8 oral bacteria in lung cancer cases and non-cases in 1287 participants in ARIC [a]

Bacterium	Case (N=40)	Non-case (N=1247)	*P* Value [b]
A. actinomycetemcomitans	3844 (7189)	2341 (6905)	0.14
C. rectus	3902 (22024)	3270 (14347)	0.83
F. nucleatum	14182 (39003)	7263 (44652)	0.10
P. gingivalis	1688 (4455)	1272 (3774)	0.20
P. intermedia	6806 (22032)	5268 (20878)	0.40
P. nigrescens	6334 (16264)	4108 (17059)	0.82
T. denticola	6048 (11955)	3323 (11357)	0.75
T. forsythensis	1330 (10137)	1550 (5497)	0.90

[a] Lung cancer cases are participants who developed first primary lung cancer during the follow-up; IQR = interquartile range.
[b] P value was calculated with Wilcoxon rank-sum test.

Table S6. Median (IQR) concentration of pre-diagnostic antibodies to 8 oral bacteria (ng/mL) in participants with and without detected bacterial DNA in ARIC [a]

Bacterium	Not detected	Detected	*P* Value [b]
A. actinomycetemcomitans	76.84 (130.38)	79.65 (163.90)	0.50
C. rectus	13.69 (34.37)	14.61 (37.56)	0.19
F. nucleatum	7.12 (21.05)	7.42 (22.29)	0.75
P. gingivalis	12.66 (34.01)	13.56 (47.96)	0.19
P. intermedia	37.76 (93.73)	33.73 (82.23)	0.21
P. nigrescens	55.91 (119.83)	61.67 (146.71)	0.32
T. denticola	23.74 (29.46)	22.90 (29.29)	0.69
T. forsythensis	22.07 (43.87)	22.10 (40.34)	0.85

[a] IQR = interquartile range.

[b] P value was calculated with Wilcoxon rank-sum test.

Table S7. Adjusted hazard ratios of lung cancer by ratio of DNA-derived bacteria count to antibody concentration (both after log transformation) for 8 oral bacteria in 1287 participants in ARIC [a]

Bacterium	HR (95% CI) [b]
A. actinomycetemcomitans	
Group 1 (lowest)	0.94 (0.42-2.10)
Group 2	1.00 (ref.)
Group 3 (highest)	0.99 (0.45-2.14)
per IQR	1.22 (0.81-1.85)
C. rectus	
Group 1 (lowest)	1.21 (0.55-2.65)
Group 2	1.00 (ref.)
Group 3 (highest)	1.14 (0.51-2.57)
per IQR	1.21 (0.88-1.67)
F. nucleatum	
Group 1 (lowest)	0.69 (0.31-1.53)
Group 2	1.00 (ref.)
Group 3 (highest)	0.82 (0.38-1.77)
per IQR	1.32 (0.92-1.90)
P. gingivalis	
Group 1 (lowest)	0.47 (0.19-1.19)
Group 2	1.00 (ref.)
Group 3 (highest)	1.11 (0.54-2.27)
per IQR	**1.71 (1.16-2.50)**
P. intermedia	
Group 1 (lowest)	0.77 (0.33-1.82)
Group 2	1.00 (ref.)
Group 3 (highest)	1.28 (0.59-2.81)
per IQR	**1.93 (1.10-3.36)**
P. nigrescens	
Group 1 (lowest)	1.42 (0.65-3.14)
Group 2	1.00 (ref.)
Group 3 (highest)	1.29 (0.57-2.95)
per IQR	1.10 (0.64-1.91)
T. denticola	
Group 1 (lowest)	1.72 (0.76-3.90)
Group 2	1.00 (ref.)
Group 3 (highest)	1.12 (0.48-2.65)
per IQR	0.68 (0.37-1.27)
T. forsythensis	
Group 1 (lowest)	**3.31 (1.26-8.68)**
Group 2	1.00 (ref.)
Group 3 (highest)	**2.64 (1.01-6.95)**
per IQR	1.25 (0.75-2.10)

[a] For each ratio, participants were divided into tertiles, where Group 1 was the lowest tertile and Group 3 was the highest tertile; HR = hazard ratio; CI = confidence interval; IQR = interquartile range; HRT = hormone replacement therapy; BMI = body mass index.

[b] Model was adjusted for age, joint terms for field center and race (Black from Jackson; Black from Forsyth; White from Forsyth; White from Washington County [reference is White from Minneapolis]), joint terms for sex and HRT use (female user, female nonuser [reference is men]), cigarette smoking status, packyears smoked, alcohol drinking status, BMI, diagnosed diabetes status, undiagnosed diabetes status, at risk for diabetes status, family history of cancer, and lifecourse socioeconomic status.

Table S8. Adjusted hazard ratios of lung cancer incidence by joint categories of antibodies (>LOD, ≤LOD) and DNA-derived bacteria counts (detected, not detected) for 8 oral bacteria in 1287 participants in ARIC [a]

Bacterium	HR (95% CI) [b]
A. actinomycetemcomitans	
Group 1 (≤LOD/not detected)	1.00 (ref.)
Group 2 (≤LOD/detected)	2.11 (0.24-18.48)
Group 3 (>LOD/not detected)	0.86 (0.09-8.33)
Group 4 (>LOD/detected)	1.66 (0.20-13.52)
C. rectus	
Group 1 (≤LOD/not detected)	1.00 (ref.)
Group 2 (≤LOD/detected)	1.01 (0.37-2.76)
Group 3 (>LOD/not detected)	1.32 (0.40-4.32)
Group 4 (>LOD/detected)	0.92 (0.33-2.60)
F. nucleatum	
Group 1 (≤LOD/not detected)	1.00 (ref.)
Group 2 (≤LOD/detected)	2.69 (0.88-8.19)
Group 3 (>LOD/not detected)	1.88 (0.33-10.91)
Group 4 (>LOD/detected)	3.21 (1.00-10.29)
P. gingivalis	
Group 1 (≤LOD/not detected)	1.00 (ref.)
Group 2 (≤LOD/detected)	1.52 (0.62-3.73)
Group 3 (>LOD/not detected)	0.25 (0.03-2.09)
Group 4 (>LOD/detected)	0.90 (0.32-2.49)
P. intermedia	
Group 1 (≤LOD/not detected)	1.00 (ref.)
Group 2 (≤LOD/detected)	1.93 (0.53-7.06)
Group 3 (>LOD/not detected)	1.02 (0.25-4.07)
Group 4 (>LOD/detected)	1.25 (0.34-4.54)
P. nigrescens	
Group 1 (≤LOD/not detected)	1.00 (ref.)
Group 2 (≤LOD/detected)	1.28 (0.36-4.56)
Group 3 (>LOD/not detected)	0.73 (0.20-2.73)
Group 4 (>LOD/detected)	0.74 (0.24-2.29)
T. denticola	
Group 1 (≤LOD/not detected)	1.00 (ref.)
Group 2 (≤LOD/detected)	1.05 (0.33-3.35)
Group 3 (>LOD/not detected)	1.46 (0.43-4.96)
Group 4 (>LOD/detected)	0.59 (0.18-1.97)
T. forsythensis	
Group 1 (≤LOD/not detected)	1.00 (ref.)
Group 2 (≤LOD/detected)	0.87 (0.28-2.68)
Group 3 (>LOD/not detected)	2.04 (0.64-6.50)
Group 4 (>LOD/detected)	0.81 (0.27-2.38)

[a] For each bacterium, limit of detection (LOD) for antibody concentration is 20 ng/mL; HR = hazard ratio; CI = confidence interval; IQR = interquartile range; HRT = hormone replacement therapy; BMI = body mass index

[b] Model was adjusted for age, joint terms for field center and race (Black from Jackson; Black from Forsyth; White from Forsyth; White from Washington County [reference is White from Minneapolis]), joint terms for sex and HRT use (female user, female nonuser [reference is men]), cigarette smoking status, packyears smoked, alcohol drinking status, BMI, diagnosed diabetes status, undiagnosed diabetes status, at risk for diabetes status, family history of cancer, and lifecourse socioeconomic status.

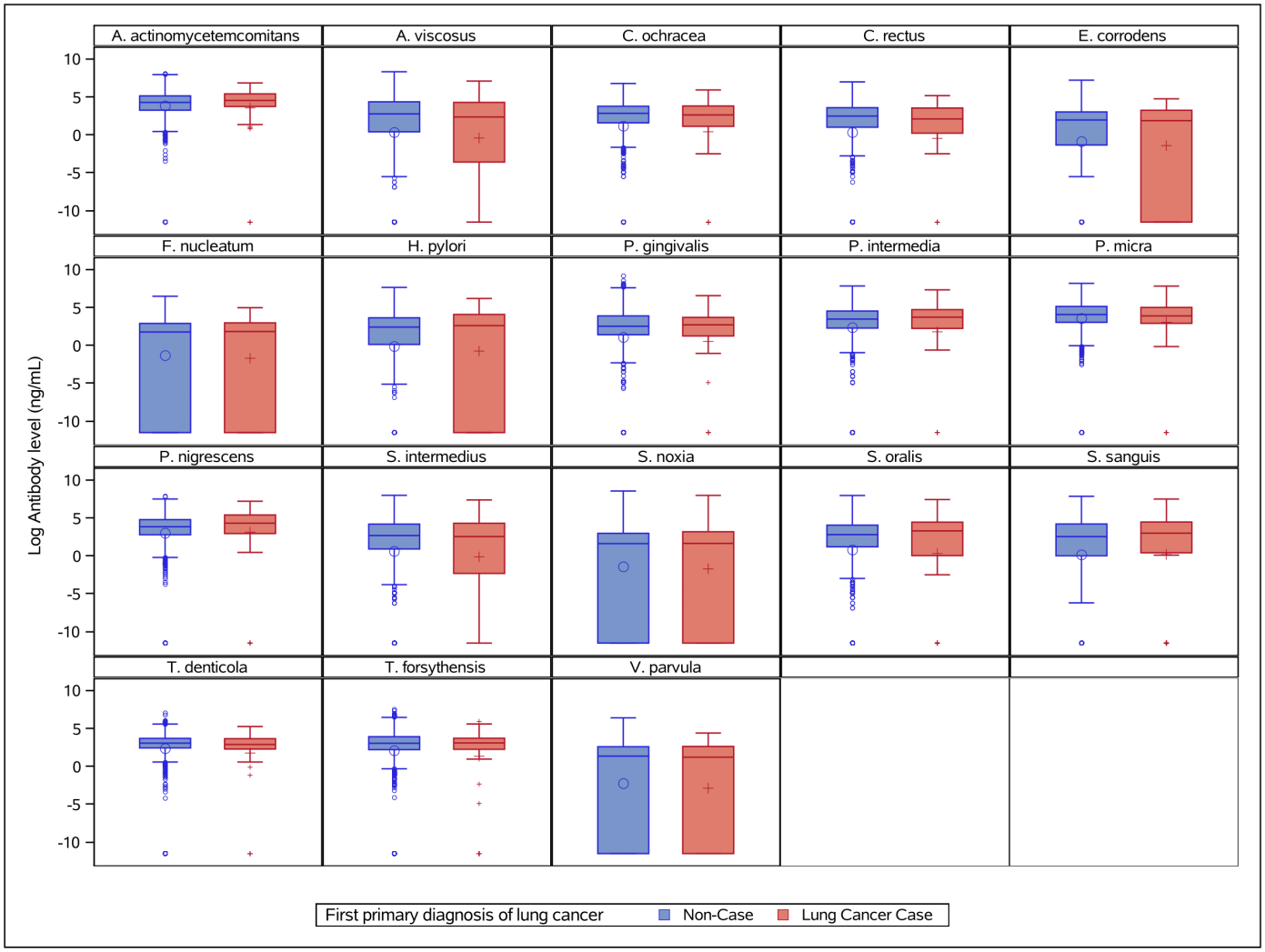
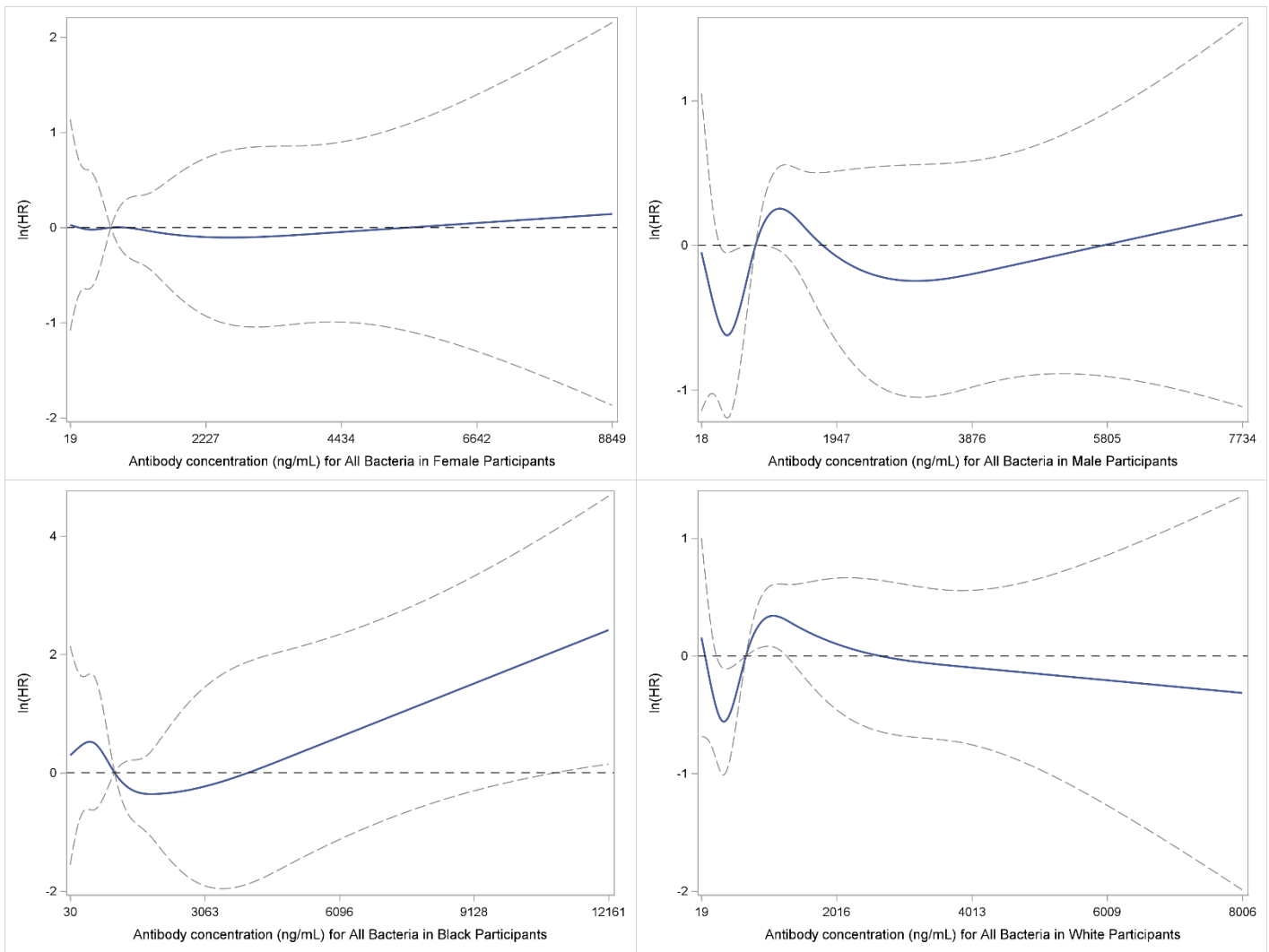
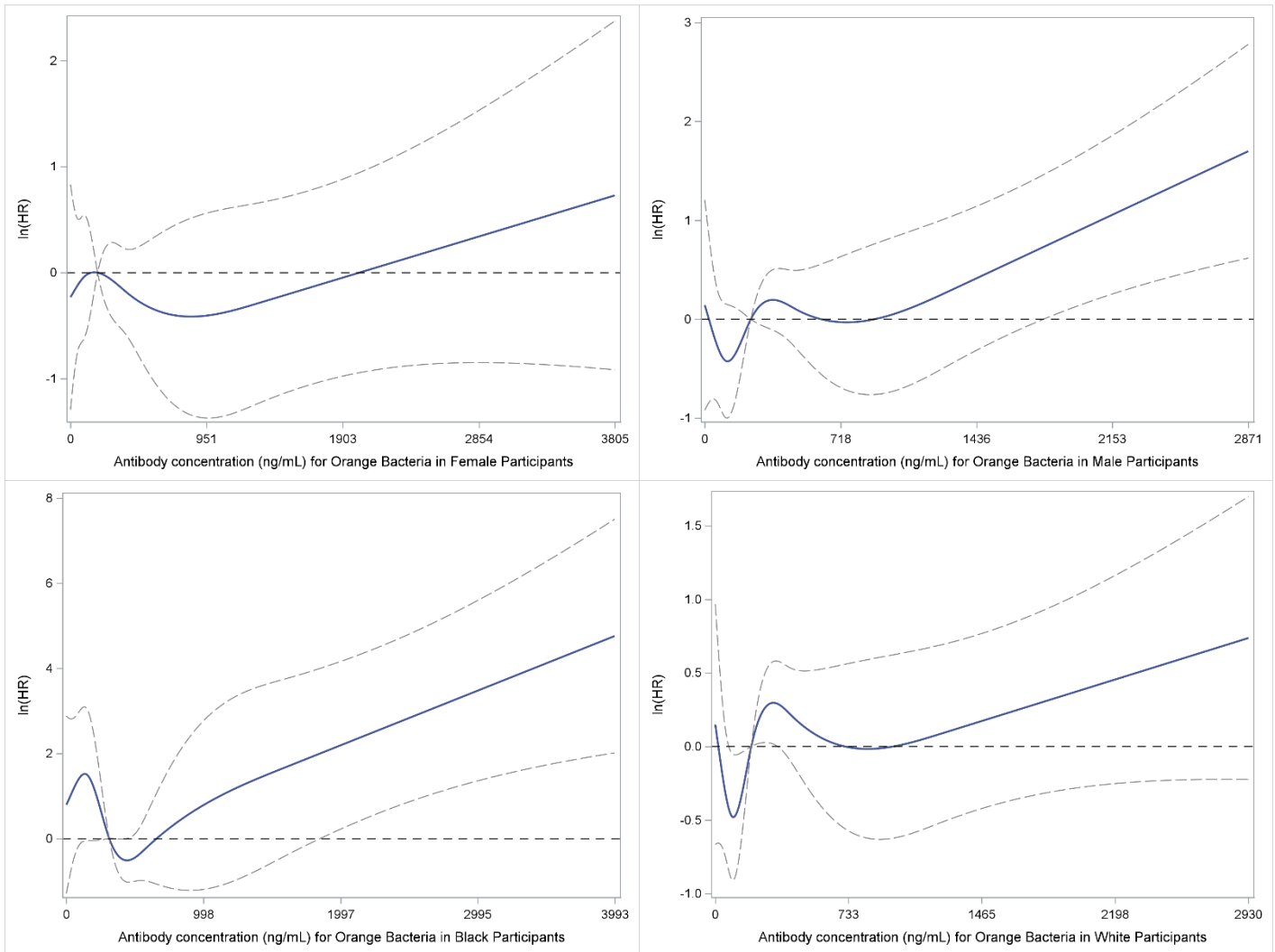


Figure S1. Box plots of the distribution of log-transformed concentration of pre-diagnostic antibodies to 18 oral bacteria (ng/mL) in 118 lung cancer cases and 4145 non-cases in ARIC.

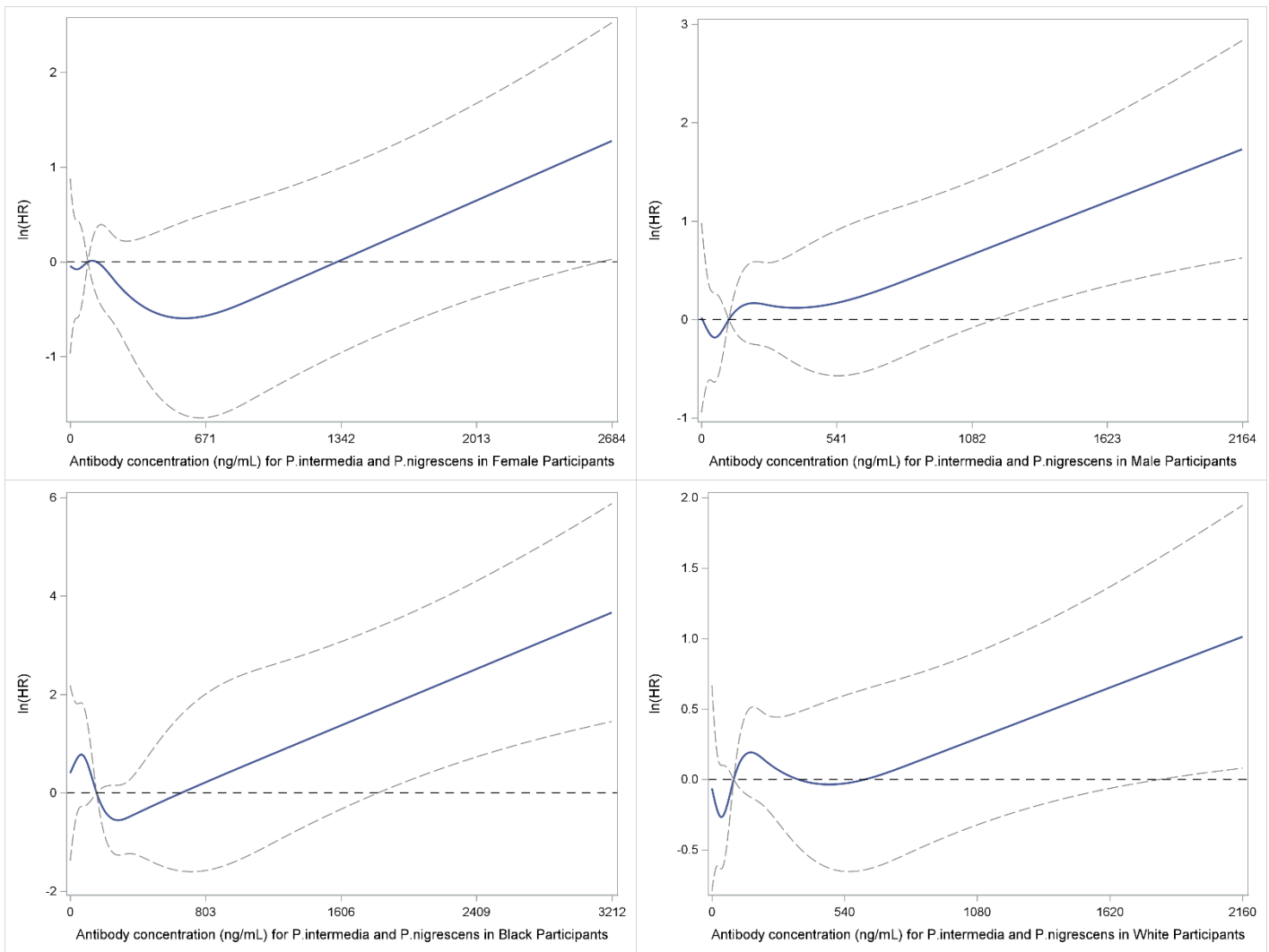
(A)


ln(HR)
Antibody concentration (ng/mL) for All Bacteria in Female Participants
Antibody concentration (ng/mL) for All Bacteria in Male Participants
Antibody concentration (ng/mL) for All Bacteria in Black Participants
Antibody concentration (ng/mL) for All Bacteria in White Participants

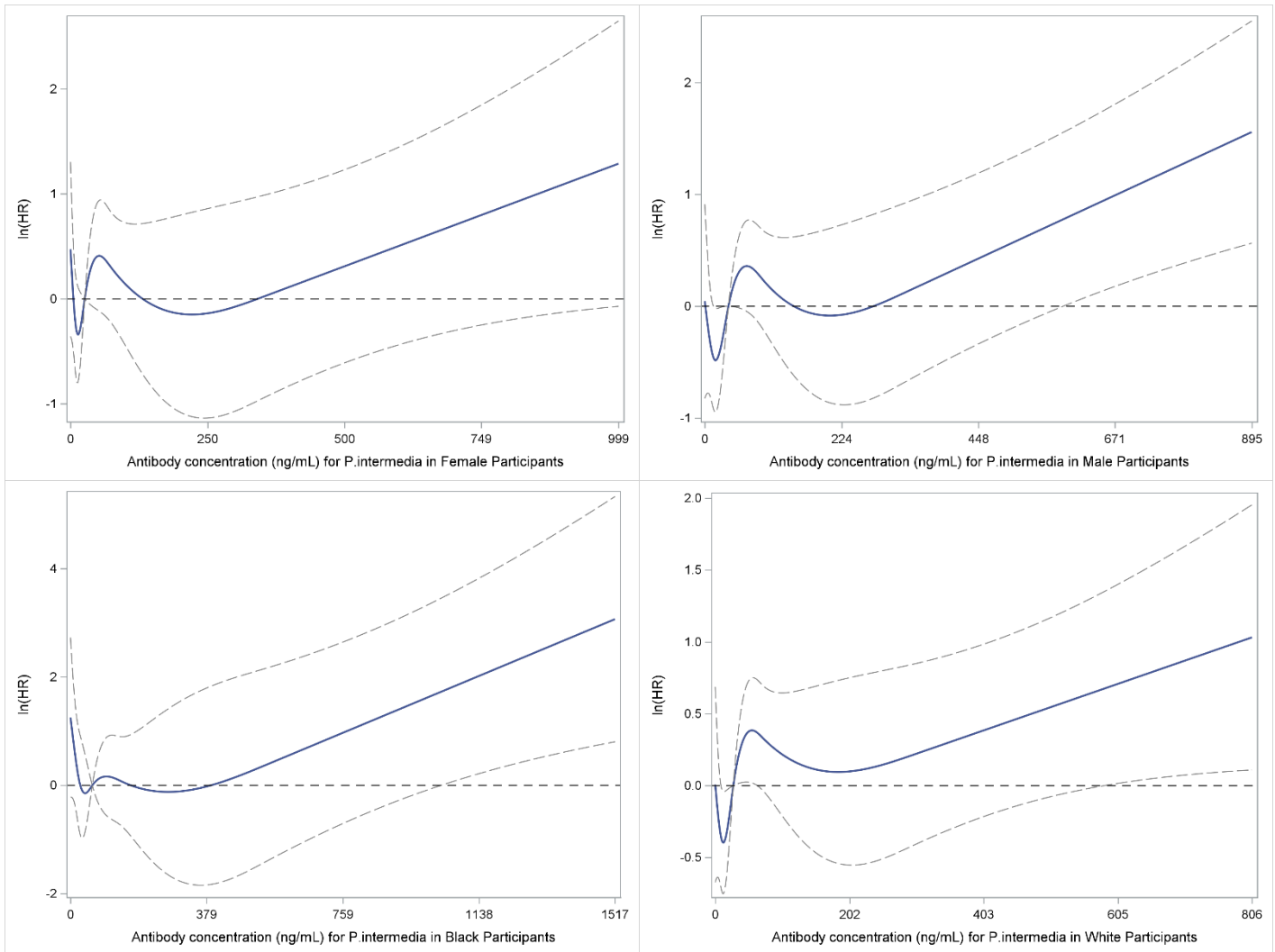
(B)


ln(HR)
Antibody concentration (ng/mL) for Orange Bacteria in Female Participants
Antibody concentration (ng/mL) for Orange Bacteria in Male Participants
Antibody concentration (ng/mL) for Orange Bacteria in Black Participants
Antibody concentration (ng/mL) for Orange Bacteria in White Participants

(C)


ln(HR)
Antibody concentration (ng/mL) for P.intermedia and P.nigrescens in Female Participants
Antibody concentration (ng/mL) for P.intermedia and P.nigrescens in Male Participants
Antibody concentration (ng/mL) for P.intermedia and P.nigrescens in Black Participants
Antibody concentration (ng/mL) for P.intermedia and P.nigrescens in White Participants

(D)


ln(HR)
Antibody concentration (ng/mL) for P.intermedia in Female Participants
Antibody concentration (ng/mL) for P.intermedia in Male Participants
Antibody concentration (ng/mL) for P.intermedia in Black Participants
Antibody concentration (ng/mL) for P.intermedia in White Participants

(E)

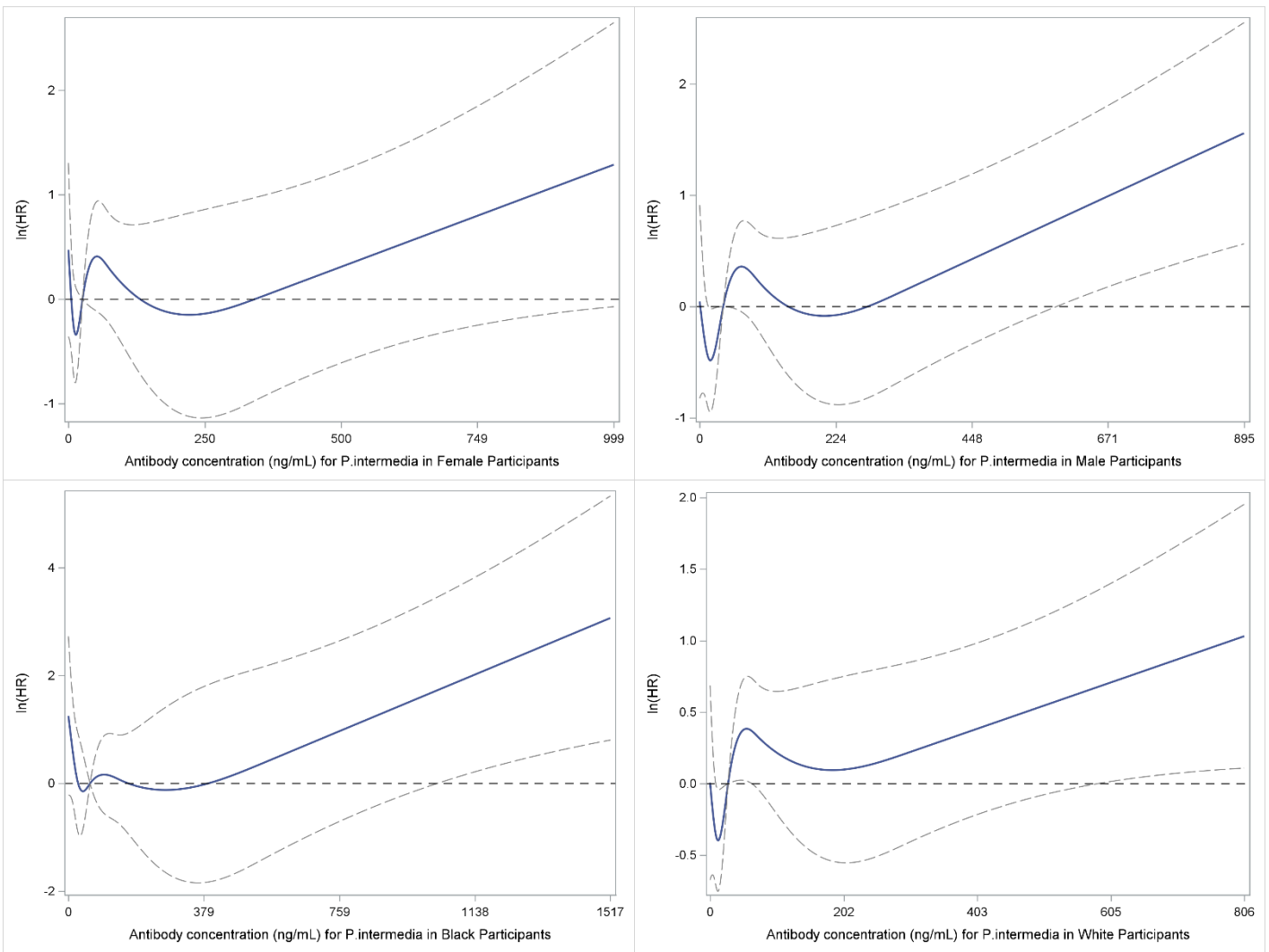


Figure S2. Hazard ratios and 95% CIs for the association between concentration of antibodies to 18 oral bacteria and lung cancer risk by sex and race, evaluated with restricted cubic splines. All models were fully adjusted; (A) Hazard ratios for the sum of concentrations of antibodies to all 18 oral bacteria; (B) Hazard ratios for the sum of concentrations of antibodies to orange complex bacteria; (C) Hazard ratios for the sum of concentrations of antibodies to *P. intermedia* and *P. nigrescens*; (D) Hazard ratios for concentration of antibodies to *P. intermedia*; (E) Hazard ratios for concentration of antibodies to *P. nigrescens*. Note that the y-axis (ln(HR)s) scale is optimized for each antibody association and subgroup; thus the scales differ. The purpose of these plots is to inform the linearity of the associations.