SUPPLEMENTAL APPENDICES

Understanding the Role of Genetic Heterogeneity in Smoking Interventions: Experimental Evidence from the Lung Health Study

Shubhashrita Basu   Jason Fletcher   Qiongshi Lu   Jiacheng Miao   Lauren L. Schmitz

APPENDIX A

List of Tables

List of Figures

APPENDIX B

Appendix A

Table A1. Mean Baseline Characteristics of the Lung Health Study (LHS) Sample by Genotyped Status

Variables	Not Genotyped	Genotyped	p-value of difference in means
Body mass index (BMI)	25.650	25.529	0.280
Cigarettes per day (CPD)	31.328	31.249	0.830
FEV_1/FVC	62.923	62.943	0.899
Age	48.306	48.537	0.236
Pack years	40.127	40.586	0.401
Years of education	13.598	13.614	0.844
Female	0.375	0.370	0.703
N	1,742	4,145	

Note: The table reports average respondent background characteristics for LHS participants stratified by genotyped status. All variables were measured at baseline before the treatment was administered. We additionally report the p-values of the differences in means for each variable.

Table A2. Impact of Treatment Status on the Probability of Being Genotyped in the Lung Health Study (LHS)

	Genotyped
Treated	0.001
	(0.013)
N	5,887

Note: The table reports the effect of treatment on the probability of being genotyped in the LHS. The total number of participants is 5,887. The regression controls for age fixed effects and sex. Robust standard errors are reported in parentheses.

Table A3. Incremental R^2 Values of Polygenic Indices and Demographic Characteristics

	Smoking cessation (all waves)	CPD (baseline)	BMI (baseline)	FEV$_1$/FVC (baseline)
PGIs				
SC PGI	0.19%			
Bootstrapped SE	(0.056)			
p-value	0.002			
SI PGI	0.03%			
Bootstrapped SE	(0.02)			
p-value	0.242			
CPD PGI		3.33%		
Bootstrapped SE		(0.55)		
p-value		0.000		
BMI PGI			8.82%	
Bootstrapped SE			(0.84)	
p-value			0.000	
FEV$_1$/FVC PGI				0.01%
Bootstrapped SE				(0.35)
p-value				0.000
Demographic characteristics				
Sex	0.07%	2.15%	8.39%	0.26%
Bootstrapped SE	(0.037)	(0.433)	(0.83)	(0.16)
p-value	0.053	0.000	0.000	0.097
Married	0.17%	0.03%	0.87%	0.01%
Bootstrapped SE	(0.056)	(0.062)	(0.31)	(0.05)
p-value	0.003	0.639	0.006	0.788

Note: The table reports the incremental R^2 values of PGIs and demographic characteristics for smoking cessation (SC), smoking initiation (SI), cigarettes per day (CPD), body mass index (BMI), and lung function, or forced expiratory volume in one second over forced vital capacity (FEV$_1$/FVC). For all the variables except smoking cessation, we use baseline measures (data on smoking cessation is not available before the treatment). Each estimate comes from a regression of the outcome variable on the PGI or demographic characteristic after residualizing out the first 20 genetic principal components (PCs) to account for population stratification. Bootstrapped standard errors with 1000 replications and corresponding p-values are reported in parentheses.

Table A4. Heterogeneous Treatment Effects of the LHS Intervention on Smoking Cessation by SC PGI

	Composite (1)	SIA (2)	SIP (3)
	Smoking Cessation		
Treated	0.229	0.232	0.227
	(0.011)	(0.014)	(0.014)
SC	-0.015	-0.013	-0.017
	(0.008)	(0.009)	(0.008)
Treat × SC PGI (std)	-0.009	-0.002	-0.014
	(0.011)	(0.014)	(0.013)
N	20,725	13,950	13,690

Note: The table reports the heterogeneous effects of the treatment on the probability of quitting smoking. Composite indicates both treatment arms combined. SIA: smoking intervention plus Atrovent inhaler group. SIP: smoking intervention plus placebo inhaler group. The PGI is standardized. All the regressions control for age fixed effects, wave fixed effects, sex, and the first 20 principal components (PCs) of the genetic data to account for population stratification. Standard errors clustered at the individual level are reported in the parentheses.

Table A5. Heterogeneous Treatment Effects of the LHS Intervention on Smoking Cessation by Baseline CPD and SI PGI

	Composite	SIA	SIP
Smoking Cessation			
Panel I			
Treated	0.229	0.230	0.227
	(0.011)	(0.014)	(0.014)
Baseline CPD (std)	-0.029	-0.030	-0.028
	(0.009)	(0.009)	(0.009)
Treat × Baseline CPD (std)	0.012	0.009	0.017
	(0.012)	(0.014)	(0.015)
Panel II			
Treated	0.231	0.233	0.228
	(0.011)	(0.014)	(0.014)
Baseline CPD (std)	-0.028	-0.029	-0.028
	(0.009)	(0.009)	(0.009)
Treat × Baseline CPD (std)	0.013	0.010	0.017
	(0.012)	(0.014)	(0.015)
SI PGI (std)	0.004	0.005	0.002
	(0.008)	(0.008)	(0.008)
Treat × SI PGI (std)	-0.025	-0.034	-0.013
	(0.011)	(0.013)	(0.013)
N	20,725	13,950	13,690

Note: Panel I of this table reports heterogeneous treatment effects of cigarettes smoked per day (CPD) at baseline on smoking cessation. These regressions control for age fixed effects, wave fixed effects, and sex. Panel II adds the smoking initiation (SI) PGI and its interaction. Regressions in Panel II also control for the first 20 principal components (PCs) of the genetic data to account for population stratification. Standard errors clustered at the individual level are reported in parentheses.

Table A6. Heterogeneous Treatment Effects of the LHS Intervention on Smoking Cessation by Smoking Initiation PGI and Annual Visit

	Annual Visit				
	1	2	3	4	5
Panel 1			Composite		
Treated	0.288	0.240	0.235	0.205	0.183
	(0.012)	(0.013)	(0.014)	(0.014)	(0.015)
SI PGI (std)	0.002	0.007	0.011	-0.003	0.002
	(0.007)	(0.009)	(0.010)	(0.011)	(0.012)
Treat × SI PGI (std)	-0.021	-0.020	-0.033	-0.014	-0.033
	(0.012)	(0.013)	(0.013)	(0.014)	(0.015)
N	4,145	4,145	4,145	4,145	4,145
Panel II			SIA		
Treated	0.296	0.243	0.232	0.204	0.187
	(0.015)	(0.016)	(0.016)	(0.017)	(0.017)
SI PGI (std)	0.003	0.007	0.011	-0.001	0.002
	(0.008)	(0.009)	(0.010)	(0.011)	(0.012)
Treat × SI PGI (std)	-0.031	-0.032	-0.038	-0.027	-0.043
	(0.015)	(0.016)	(0.016)	(0.017)	(0.017)
N	2,790	2,790	2,790	2,790	2,790
Panel III			SIP		
Treated	0.278	0.235	0.236	0.205	0.178
	(0.015)	(0.016)	(0.017)	(0.017)	(0.018)
SI PGI (std)	-0.001	0.004	0.008	-0.007	-0.000
	(0.008)	(0.009)	(0.010)	(0.011)	(0.012)
Treat × SI PGI (std)	-0.010	-0.004	-0.026	0.003	-0.020
	(0.015)	(0.016)	(0.016)	(0.017)	(0.018)
N	2,738	2,738	2,738	2,738	2,738

Note: This table reports the heterogeneous treatment effects of the SI PGI on smoking cessation stratified by annual visits post treatment. There were five follow-up visits recorded in the LHS. Panel I reports the effects for both treatment groups combined (SIA and SIP) relative to the control group. Panel II reports the effects for those participants who were assigned to the SIA group compared to the control group. Panel III reports the effects of the SIP participants in comparison with the control group. Each regression controls for age fixed effects, wave fixed effects, sex, and the first 20 principal components to account for population stratification.

Table A7. Heterogeneous Treatment Effects of the LHS Intervention on Cigarettes per Day by CPD PGI and Annual Visit

	Annual Visit				
	1	2	3	4	5
Panel 1			Composite		
Treated	-11.995	-9.850	-9.330	-7.727	-6.626
	(0.491)	(0.486)	(0.484)	(0.488)	(0.484)
CPD PGI (std)	1.814	1.544	1.872	1.673	1.665
	(0.437)	(0.429)	(0.427)	(0.436)	(0.441)
Treat × CPD PGI (std)	-0.961	-0.313	-0.799	-0.485	-0.560
	(0.514)	(0.512)	(0.506)	(0.513)	(0.515)
N	4,050	4,038	4,027	4,019	4,145
Panel II			SIA		
Treated	-12.319	-9.995	-9.657	-7.906	-6.935
	(0.565)	(0.559)	(0.550)	(0.557)	(0.550)
CPD PGI (std)	1.842	1.503	1.861	1.601	1.623
	(0.437)	(0.432)	(0.430)	(0.440)	(0.446)
Treat × CPD PGI (std)	-1.724	-0.824	-1.153	-0.805	-0.681
	(0.580)	(0.587)	(0.577)	(0.582)	(0.578)
N	2,727	2,719	2,708	2,700	2,790
Panel III			SIP		
Treated	-11.531	-9.589	-8.895	-7.522	-6.248
	(0.575)	(0.570)	(0.572)	(0.571)	(0.563)
CPD PGI (std)	1.886	1.637	1.937	1.761	1.741
	(0.438)	(0.434)	(0.431)	(0.440)	(0.444)
Treat × CPD PGI (std)	-0.174	0.184	-0.487	-0.182	-0.460
	(0.591)	(0.592)	(0.587)	(0.590)	(0.591)
N	2,675	2,671	2,657	2,663	2,738

Note: This table reports the heterogeneous treatment effect of the CPD PGI on CPD stratified by annual visits post treatment. There were five follow-up visits recorded in the LHS. Panel I reports effects for both treatment groups combined (SIA and SIP) relative to the control group. Panel II reports effects for participants who were assigned to the SIA group compared to the control group. Panel III reports effects for SIP participants compared to the control group. Each regression controls for age fixed effects, wave fixed effects, sex, and the first 20 principal components (PCs) of the genetic data to account for population stratification.
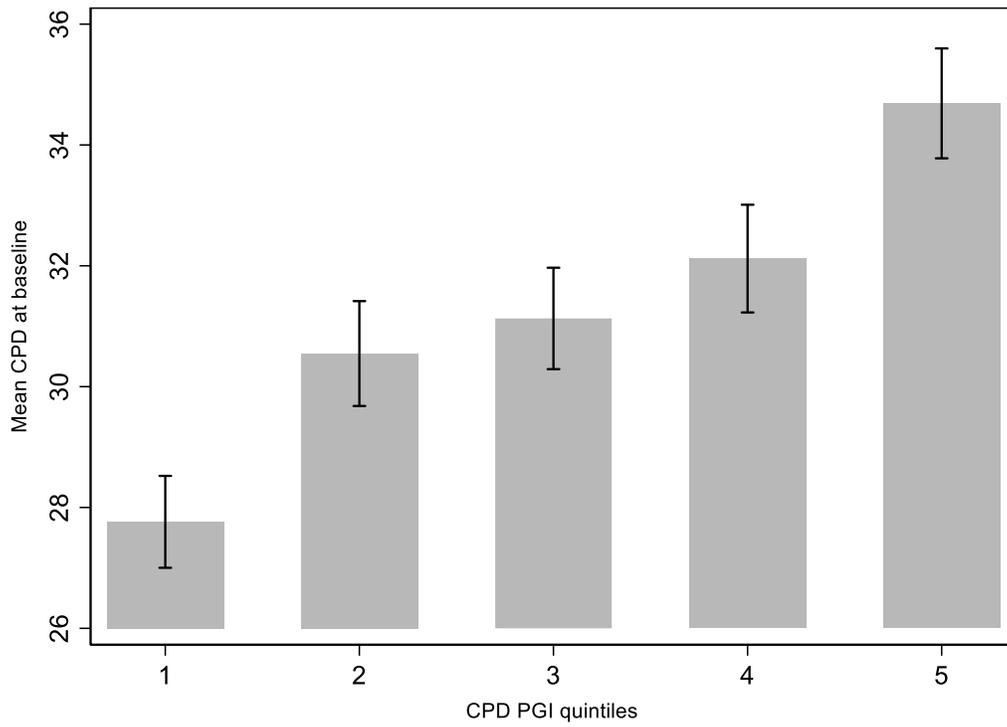
Figures

**Figure A1.** PGI Correlation Matrix in the LHS Sample
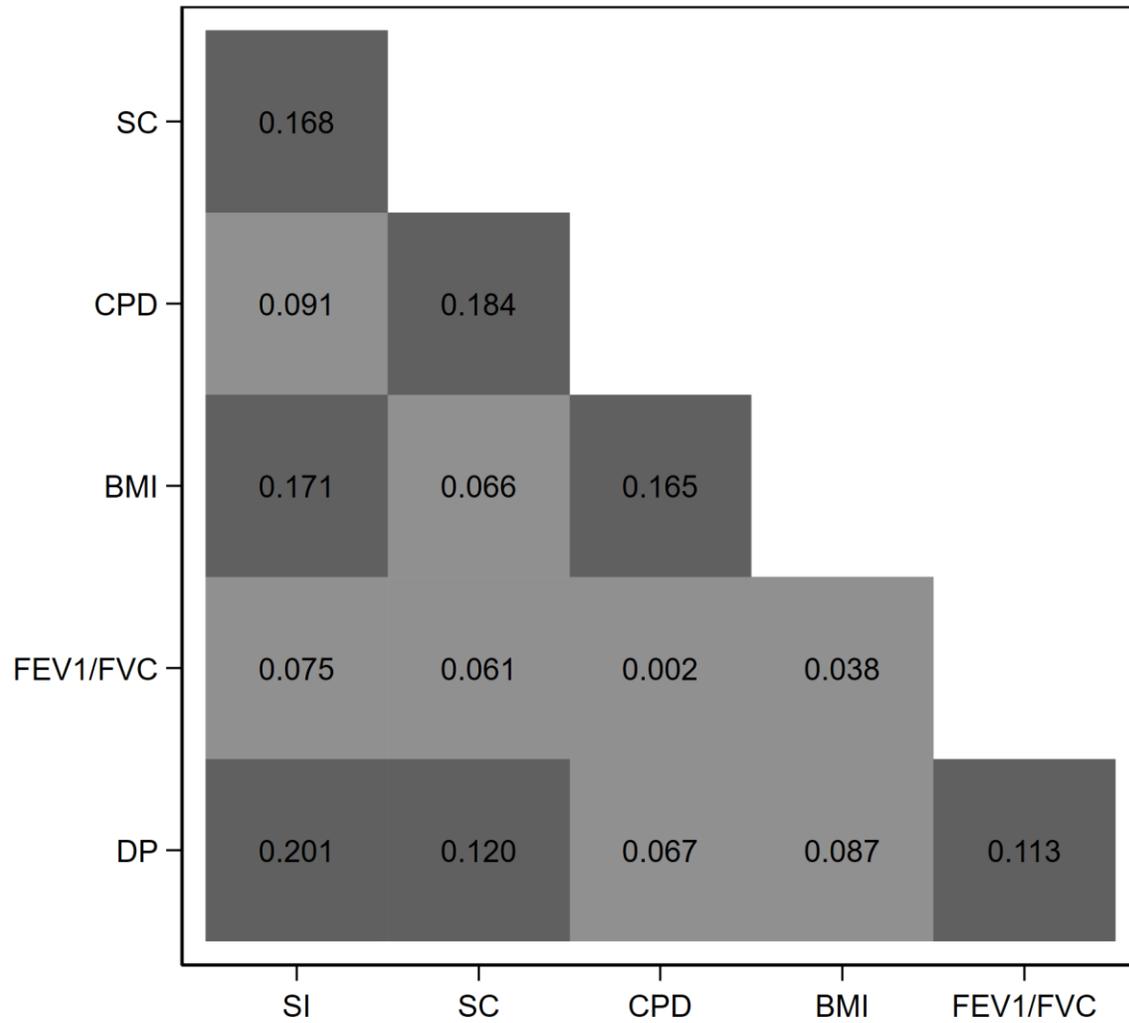
Note: The heatmap presents the correlation coefficients between the following PGIs: smoking initiation (SI), smoking cessation (SC), cigarettes per day (CPD), body mass index (BMI), lung function, or forced expiratory volume in one second over forced vital capacity (FEV$_1$/FVC), and depression (DP). Lighter cells indicate weaker correlations, while darker cells indicate stronger correlations.

**Figure A2.** Mean Cigarettes Per Day at Baseline by CPD PGI Quintile

Note: The figure plots the mean number of cigarettes per day (CPD) at baseline by CPD PGI quintiles. 95% confidence intervals are reported.

**Figure A3.** Mean Body Mass Index at Baseline by BMI PGI Quintile
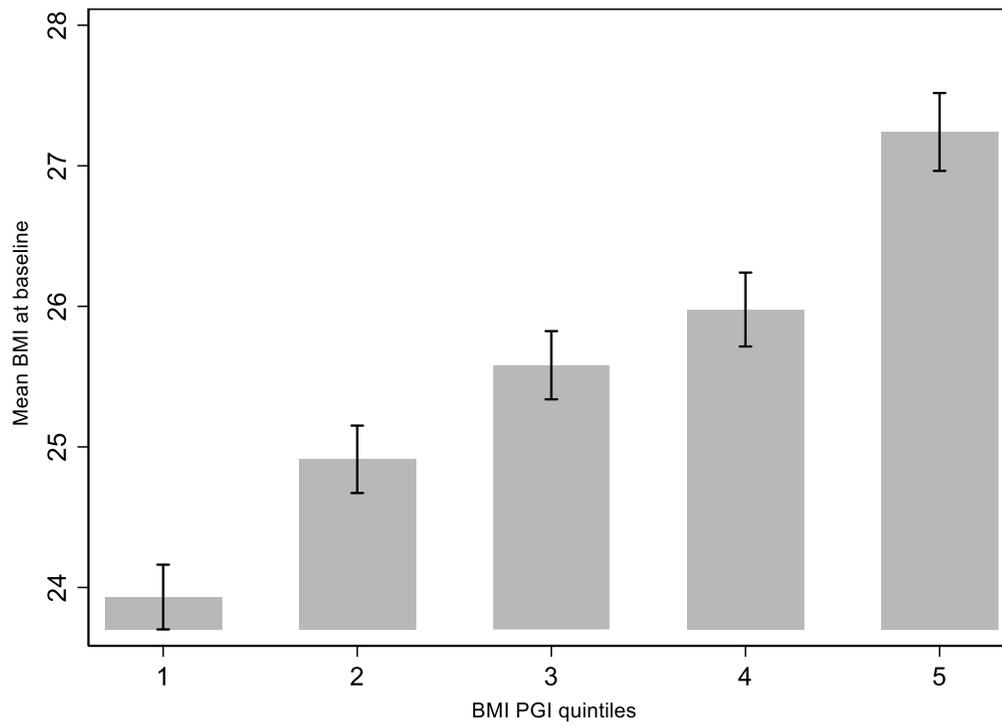
Note: This figure plots the mean body mass index (BMI) at baseline by BMI PGI quintiles. 95% confidence intervals are reported.

Figure A4. Mean FEV$_1$/FVC at Baseline by FEV$_1$/FVC PGI Quintile
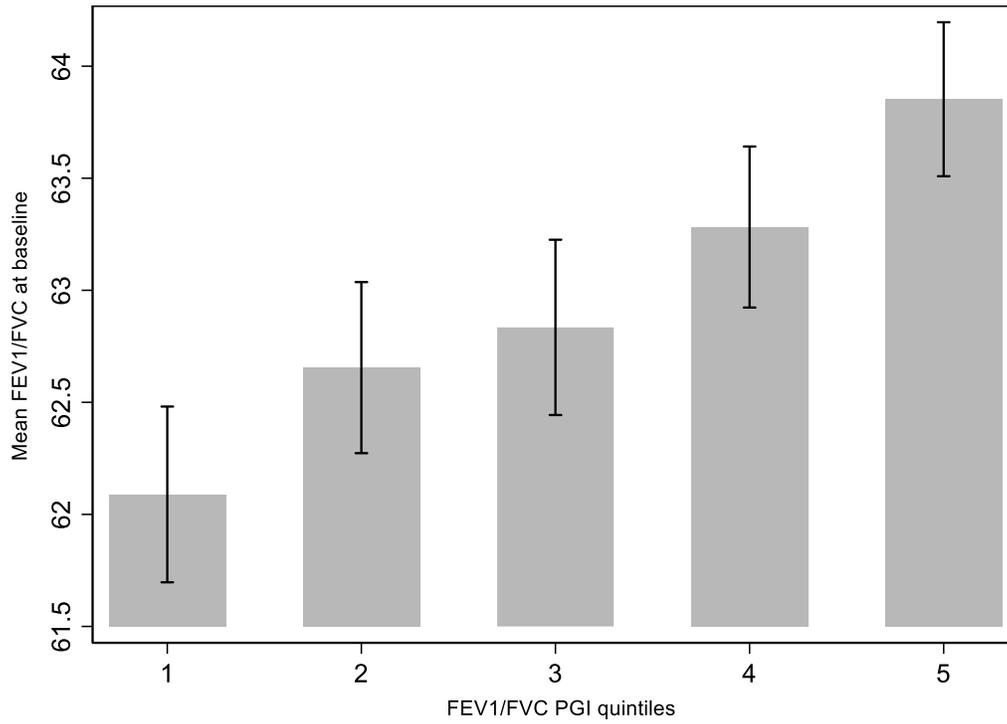
*Note:* The figure plots mean lung function, or forced expiratory volume in one second over forced vital capacity (FEV$_1$/FVC) at baseline by FEV$_1$/FVC PGI quintiles. 95% confidence intervals are reported.

Figure A5. Mean Depression Symptoms at Baseline by Depression PGI Quintile
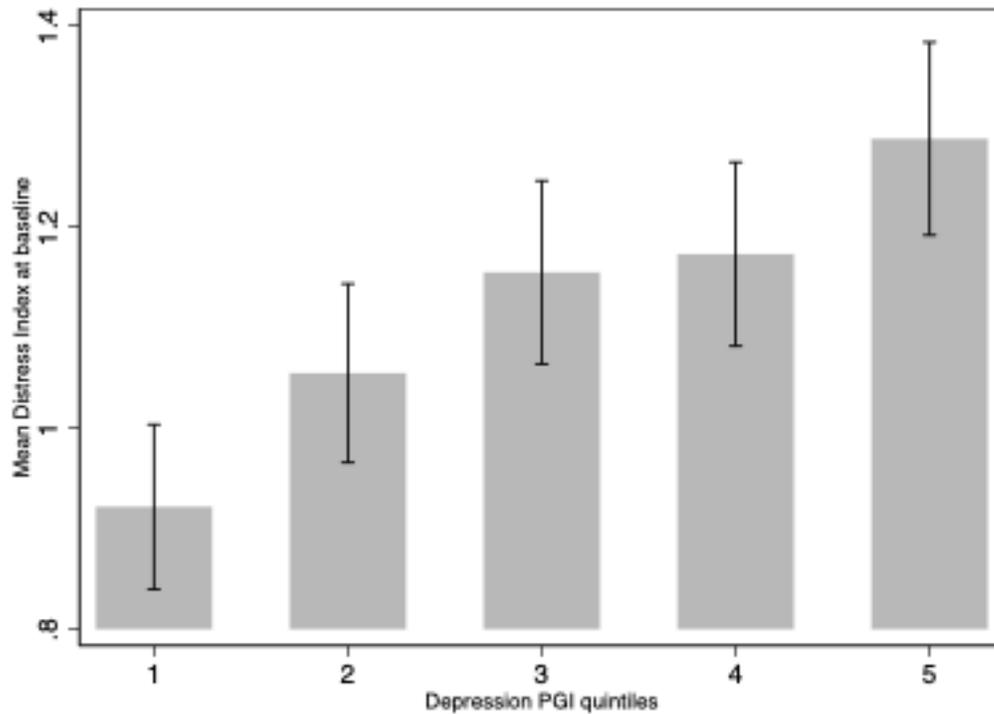

Note: This figure plots the average distress index at baseline by depression PGI quintile. The distress index is the sum of five symptoms related to depression: irritability, insomnia, mood changes, nervousness, and psychological illness. The LHS asks participants about these mental conditions via the following question: "Indicate the extent to which you have been troubled in the last four months by any of the following. Please indicate 'Severe,' 'Moderate,' 'Mild,' or 'Not at all.'" For each condition, we construct a dichotomous variable that takes the value of one if the respondent reports mild to severe symptoms and zero otherwise. We then form the distress index by summing across these mental health conditions at baseline. An increase in the distress index indicates worsening mental health. 95% confidence intervals are reported.

Figure A6. Per-wave Means of Smoking Cessation and CPD by Treatment Status
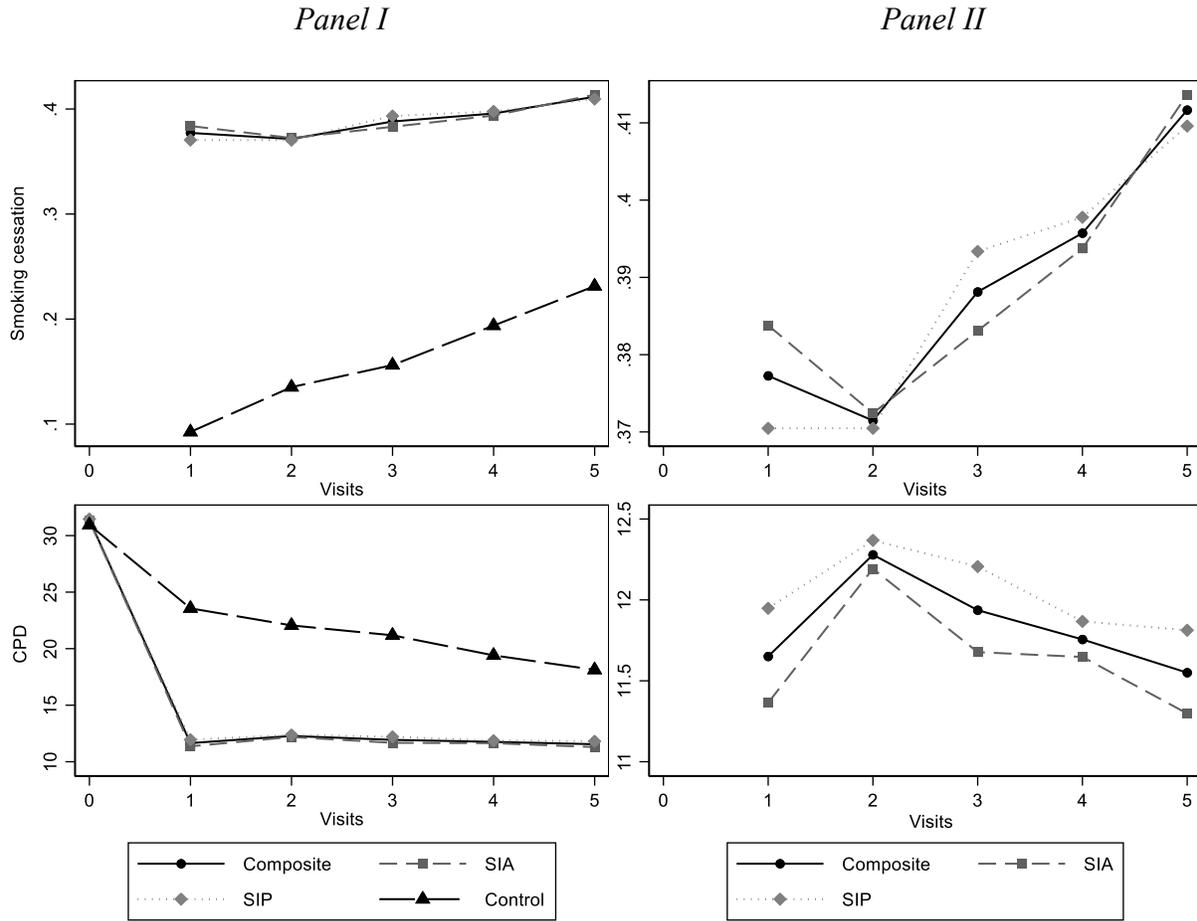
Panel I   *Panel II*

Note: Panel I shows the averages of smoking cessation and CPD by treatment status at baseline before treatment was administered and across all five annual follow-up visits. SIA: smoking intervention plus Atrovent inhaler (ipratropium bromide) group; SIP: smoking intervention plus placebo inhaler group. Averages are shown for the two treatment arms separately, for the composite treatment group (SIA and SIP), and the control group. Since we do not have information on smoking cessation at baseline, the first diagram under Panel I starts at the first visit. Panel II presents a zoomed-in version of the Panel I figures for the three treatment groups. Since baseline CPD was the same across groups, Panel II only plots means for the five follow-up visits.

Figure A7. Heterogeneous Treatment Effects of the LHS Intervention on Smoking Cessation by Baseline CPD and SI PGI
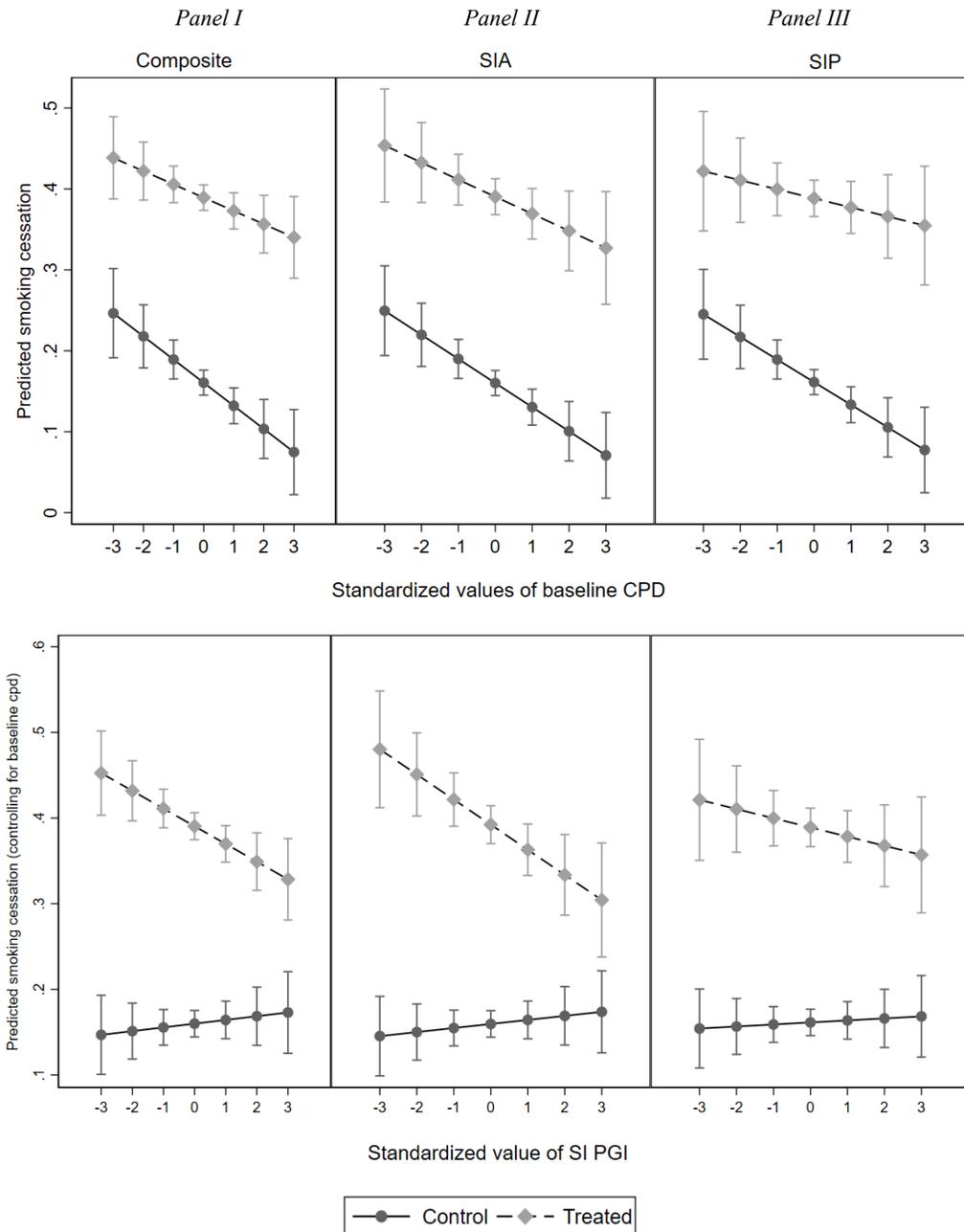
Note: Panel I corresponds to the composite treatment group ('1' if either SIA or SIP and '0' if control), Panel II corresponds to the SIA treatment group ('1' if SIA and '0' if control), and Panel III corresponds to the SIP treatment group ('1' if SIP and '0' if control). The top row presents the heterogeneous treatment effects on smoking cessation using variations in baseline CPD. These regressions control for age fixed effects, wave fixed effects, and sex. The lower row plots predicted treatment effects for smoking cessation by SI PGI that also control for baseline CPD and its interaction with the treatment. These regressions also control for the first 20 PCs of the genetic data. The grey diamonds represent the treated groups (composite/SIA/SIP) and the black circles represent the control group. Standard errors are clustered at the individual level and 95% confidence intervals are plotted with the estimates.

Appendix B

B1. List of Abbreviations

1. LHS: Lung Health Study

2. RCT: Randomized controlled trial

3. ATE: Average treatment effect

4. SIA: Smoking intervention plus inhaler with the active drug ipratropium bromide group

5. SIP: Smoking intervention plus placebo inhaler group

6. PGI: Polygenic index

7. PC: Principal component

8. SI: Smoking initiation

9. SC: Smoking cessation

10. CPD: Cigarettes smoked per day

11. FEV_1/FVC: Forced expiratory volume in one second over forced vital capacity

12. DP: Depression

13. GWAS: Genome-wide association study

14. SNP: Single nucleotide polymorphism

15. LD: Linkage disequilibrium

B2. Additional Information on the Lung Health Study Intervention

1. Exclusionary criteria

Individuals were excluded from the study if they had serious illnesses such as cancer, heart disease, stroke, or other significant medical conditions like high blood pressure, were using bronchodilators, beta-blockers, nitrates, or insulin, or if they consumed more than 25 alcoholic drinks per week or were binge drinkers. To reduce attrition, individuals were also excluded if they anticipated moving more than 75 miles from the clinical center or were unwilling to participate in the behavior intervention if randomized into the treatment group.

2. Intervention details

Each center recruited approximately 600 participants and assigned 400 individuals to the treatment group and 200 to the control group. Women comprised approximately 37 percent of enrolled subjects and 96 percent of the sample was white. The smoking cessation intervention program was designed to help individuals cope with their addiction and included several key features throughout the five-year trial (Connett et al. 1993; O'Hara et al. 1993):

- An initial message from a physician was delivered soon after randomization to give the participant information regarding their lung impairment, the adverse health effects of smoking, the importance of smoking cessation to reduce health risks, and a prescription for the inhaler and nicotine gum. Participants were encouraged to limit their nicotine gum use to the first 6 months of the trial and were provided with a reduction strategy if they had difficulty tapering their use.

- A meeting with an intervention specialist (held immediately after the physician meeting) explained the program and helped participants choose a quit date.

- An intensive 12-session group intervention program spread over 10 weeks provided cognitive and behavioral strategies for quitting and instructions for proper inhaler and nicotine gum use.

- Clinic visits every four months across all five years of follow-up.

- A maintenance program to minimize relapse and provide long-term support with problems such as weight gain and stress management.

- An extended intervention program for smokers who relapsed included options for individual counseling, physician visits, or additional group meetings.

References

Connett, J., J. Kusek, W. Bailey, P. O'Hara, and M. Wu. 1993. "Design of the Lung Health Study: A Randomized Clinical Trial of Early Intervention for Chronic Obstructive Pulmonary Disease." *Controlled Clinical Trials* 14: 3S-19S.

O'Hara, P., J. Grill, M. A. Rigdon, J. E. Connett, G. A. Lauger, and J. J. Johnston. 1993. "Design and Results of the Initial Intervention Program for the Lung Health Study." *Preventive Medicine* 22 (3): 304–315.

B3. Quality Control and Imputation of the Genotyped Data

We performed pre-imputation quality control (QC) on the genotype data using PLINK software (Purcell et al. 2007). We only kept autosomal non-biallelic SNPs with a minor allele frequency (MAF) > 0.01 and Hardy Weinberg equilibrium test p-value $\geq$ 1.0e-6. We used the UCSC liftOver tool to lift over the genome coordinates from hg18 to hg19. We phased and imputed the genotype data using the HRC reference panel version r1.1 2016 from the Michigan Imputation server (Das et al. 2016). After imputation, we removed the duplicated and strand ambiguous SNPs, SNPs with imputation quality<0.9, and SNPs with MAF<0.01. After QC, 12,030,369 SNPs remained. PGIs were then constructed using SNPs that overlapped between the LHS and GWAS samples. Number of SNPs used to construct each PGI: SI=1,044,939; SC=1,057,529; CPD=1,057,519; BMI=920,776; FEV_1/FVC=1,052,022; Depression=1,017,199.

References

Das, Sayantan, Lukas Forer, Sebastian Schönherr, Carlo Sidore, Adam E Locke, Alan Kwong, Scott I Vrieze, et al. 2016. "Next-Generation Genotype Imputation Service and Methods." *Nature Genetics* 48 (10): 1284–87.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A R Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I W De Bakker, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81 (3): 559–75.