Supplementary Tables

Table S1: Prosocial nudge effects on medical students' intent-to-dropout by pre-COVID-19 intentions

Source of heterogeneity	(1) All	(2) Outcome: Post-COVID intent-to-dropout = 1 Pre-COVID-19 dropout=0	(3) Pre-COVID-19 dropout=1
Parent is a doctor (yes=1)			
Treatment	-0.005***	-0.005***	0.000
	(0.002)	(0.001)	(0.008)
Treatment × doctor parent	0.003	0.003	-0.016
	(0.006)	(0.005)	(0.029)
Teachers/Schoolmates on the frontline (yes=1)			
Treatment	-0.011***	-0.010***	-0.013
	(0.003)	(0.002)	(0.012)
Treatment × frontline	0.011***	0.009***	0.018
	(0.003)	(0.003)	(0.014)
Paid attention to COVID-19 (yes=1)			
Treatment	-0.006***	-0.007***	-0.004
	(0.003)	(0.002)	(0.010)
Treatment × attention	0.003	0.004	0.005
	(0.003)	(0.003)	(0.013)
Home internet access (yes=1)			
Treatment	-0.07**	-0.004*	-0.019*
	(0.003)	(0.002)	(0.010)
Treatment × internet	0.004	-0.000	0.029**
	(0.003)	(0.003)	(0.014)
College province near Wuhan (yes=1)			
Treatment	-0.006***	-0.004***	-0.011
	(0.002)	(0.001)	(0.009)
Treatment × near Wuhan	0.006*	-0.000	0.036**
	(0.004)	(0.003)	(0.018)
Hometown province confirmed cases > median (yes=1)			
Treatment	-0.007**	-0.005*	-0.009
	(0.003)	(0.002)	(0.012)
Treatment × Confirmed cases	0.003	0.000	0.011
	(0.003)	(0.003)	(0.015)

Notes: This table presents estimates on the overall and heterogeneous nudge effects on medical students' intent-to-dropout by pre-COVID-19 intentions. The overall treatment effect was obtained from Equation 4, while the heterogeneous treatment effects were obtained from Equation 5. Robust standard errors are clustered at the college-grade level. *** p<0.01, ** p<0.05, * p<0.1.

Table S2: Predicting medical student dropout: Model comparison

Model		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Linear		Logit		LASSO		Random Forest	
College Fixed Effect		No	Yes	No	Yes	No	Yes	No	Yes
Predicted									
dropout probability	Min	0.0158	-0.0687	0.0499	0.0152	0.0501	0.0194	0.0000	0.0000
	Mean	0.1380	0.1381	0.1381	0.1381	0.1381	0.1381	0.1382	0.1367
	Median	0.1346	0.1329	0.1310	0.1262	0.1308	0.1267	0.1274	0.0475
	Max	0.3222	0.4550	0.3927	0.5801	0.3895	0.5592	0.9700	1.0000
Classification accuracy									
Training (75% sample)		54.38%	56.53%	57.43%	60.05%	57.43%	57.43%	61.42%	81.79%
Validation (25% sample)		54.02%	55.91%	57.02%	59.50%	57.02%	57.02%	58.05%	63.14%

Notes: This table compares prediction results of four models that were used to predict medical student dropout behaviors. Linear model corresponds to columns 3 and 6 in Table 2. The other three models use the same covariates. The λ values in the LASSO models are 0.0001132 and 0.0004163, respectively. The top 5 important predictors in the Random Forest model (column 8) are undermatched in college admision, attention to COVID, home access to internet, study phase, teachers/schoolmates on the frontline.

Appendix: Survey

Survey on Online Learning of Undergraduate Clinical Medical Students in China

(Translated version for the items used in the paper)

Note: The raw datasets and original survey (in Chinese) can be accessed through Data User Agreement with the National Center for Health Professions Education Development of China (medudata.bjmu.edu.cn).

Section 1

1. Your present location (select the province and prefecture-level city):

2. Your enrolled university/college (select the university name):

3. Your Gender:

    A Male

    B Female

4. Your grade:

    A Grade 1

    B Grade 2

    C Grade 3

    D Grade 4

    E Grade 5

5. Your study phase:

    A General education

    B Basic medical education

    C Clinical medical education

    D Clerkship and rotation

6. Your enrolled program:

    A 5-year program

    B Rural 5-year program

    C 5+3-year program

    D 8-year program

7. Are you currently at home:

    A Yes

    B No, at the university

    C No, at another place

8. Your current residence location belongs to:

    A Rural area

    B Suburban area

    C Urban area

9. Do you have a doctor parent:

    A Yes, both parents are doctors

    B Yes, one parent is a doctor

    C No

10. How is your access to internet:

    A WIFI is in a good condition

    B WIFI is unstable

    C No WIFI, only mobile data are available

    D No WIFI, mobile data are unstable

11. Your current academic (GPA) ranking is:

A Top 25

B 25-50

C 50-75

D Bottom 25

12. Whether your college admission results are lower than expected:

A Yes

B No

13. Do you have teachers/schoolmates that you know well who have served at the frontline:

A Yes

B No

14. Please recall if you were planning to become a doctor after graduation before the COVID-19:

A I wanted to be a doctor

B I did not wanted to be a doctor

[Sections on student learning experiences during COVID]

Section 2

1. How serious do you think this COVID-19 outbreak is:

A Very serious

B Serious

C Not serious

2. Have you paid attention to the epidemic:

A Yes, I have paid great attention to it.

B I paid attention to it.

C I didn't pay attention.

[Nudge information randomly presented to treated students]

3. Do you plan to become a doctor after graduation:

    A Yes

    B No