

SUPPLEMENTARY MATERIALS

Supplementary 1 Sample Size Estimation and Investigation Process Sample Size Estimation

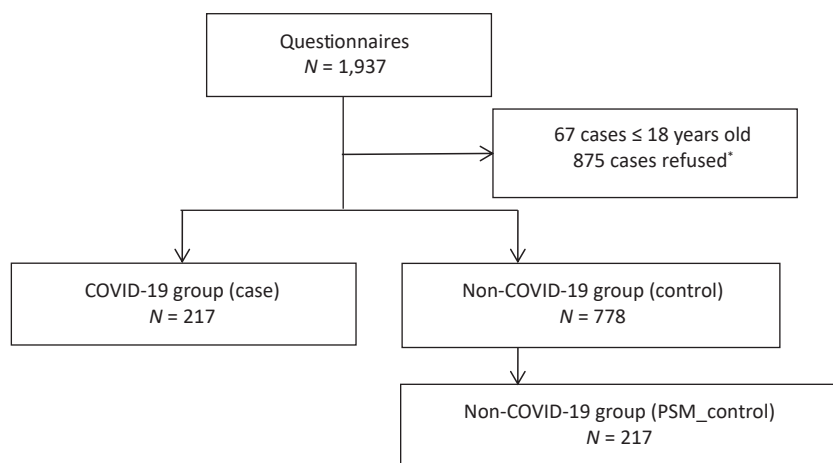
Our study assumed that a certain risk factor accounts for 20% in the case group and 30% in the control group, $\alpha = 0.05$, $\beta = 0.1$, and 10% non-response rate. A total of 1,085 individuals are required for the case-control study. The formula for calculating the sample size is as follows:

$$n = \left[Z_{1-\alpha/2} \sqrt{(1+1/r)P(1-P)} + Z_{\beta} \sqrt{P_1(1-P_1)/r + P_0(1-P_0)} \right]^2 / (P_1 - P_0)^2$$

where $\bar{p} = (P_1 + rP_0)/(1+r)$, P_1 is the proportion of risk factors in the case group, and P_0 is the proportion of risk factors in the control group, r is the matching ratio between the control and the case group, α is the probability of hypothesis testing type I error, β is the probability of type II error [Zhan SY. Epidemiology. 8th Ed. Beijing: People's Health Publishing House, 2017; 84–5. (In Chinese)].

Supplementary 2 Exploratory Factor Analysis

In the results of EFA, the KMO value was 0.615, and the Bartlett's spherical test yielded $\chi^2 = 931.821$ ($df = 36$), $P < 0.001$, suggesting that the index of nucleic acid sampling is suitable for factor analysis. There are three common factors with an initial eigenvalue > 1 in the correlation coefficient matrix, and the cumulative variance contribution rate is 56.843%. After removing the poorly-performed items “nucleic acid sampling frequency” and “type of mask when sampling”, the cumulative variance contribution rate rises to 64.189%. The three common factors can be interpreted as: (1) Environmental of sampling site, represented by sampling site location, outdoor/indoor of sampling site and the type of sample site; (2) Sampling practices of sampling personnel, represented by “the frequency of sampling from the right body part” and “sampling from which body part”; (3) Arrangements of sampling site, represented by queuing time and sampling site order, as shown in [Supplementary Table S1](#). The above three factors and nucleic acid sampling frequency were included in the multi-factor analysis.



Supplementary Figure S1. Flow chart of study participants. Participants with or without history of SARS-CoV-2 infection Beijing during November 2022 were considered. The left side of the diagram represents the flowchart for the COVID-19 group (community-found cases). The community-found cases were defined as cases, who could freely transmit the virus through social interactions before quarantine or medical treatment. The right side shows the inclusion flowchart for the non-COVID-19 group and matched through PSM, who living in the same community as the infected person. *There was no significant difference in age and gender between those refused and agreed to participate ($P > 0.05$).

Supplementary Table S1. Factor loading of nucleic acid sampling

Variables	Environmental of sampling site	Sampling practices of sampling personnel	Arrangement of sampling site
Queuing time	0.465	-0.071	0.690
Sampling site arrangement	0.406	-0.473	0.504
Sampling site location	0.559	0.282	-0.294
Outdoor/Indoor of Sampling site	-0.707	-0.310	0.097
Sampled from the right body part	-0.299	0.721	0.275
Sampled from which body part	-0.274	0.671	0.368
Type of sample site	0.777	0.318	-0.140