Appendix 1

Methods

Data collection

Data used in this study includes COVID-19 histories of individuals and a follow-up serum antibody test. As it is more available to conduct this research among healthcare workers, convenience sampling is a practical way to collect data. However, as participants are chosen based on ease of access, it may not represent the broader population accurately.

First, we initiated data collection by distributing questionnaires to public health workers in Xiamen healthcare system, including hospitals, CDCs, and primary healthcare organizations, and questionnaires were distributed to samples who were selected via convenient sampling. As we targeted individuals who had been infected by SARS-CoV-2 or had received vaccinations after December 1st, 2022. After collecting and reviewing the completed questionnaires, a total of 1,344 questionnaires were collected. Based on prior research indicating that NAbs for COVID-19 might persist for a maximum of 73 days (34), questionnaires that were incomplete or lacking essential infection-related information were excluded from the study. Finally, 1,191 of them being eligible to be included as the study's sample population.

Subsequently, we conducted serum antibody testing on the selected participants in four monthly intervals, specifically in February, March, April, and May. This longitudinal testing aimed to capture the variations in antibody titers over time.

MLR models

MLR models were established to analyze the relationship between individual baseline properties and COVID-19 antibody titers. Regression coefficients (β), 95% CIs, and standardized regression coefficients (β) are calculated and the calculation results are presented in *Figures S1-S6*.

GBTMs

The study constructs a GBTM, which can depict the characteristic dynamic changes of time-varying variables as the number of follow-up visits increases. It simultaneously divides the population into several latent class groups and establishes a latent growth class model within each category to describe the individual changes over time within the group. This model can not only reveal the relationships between different latent trajectories but also depict the fluctuations within the trajectory, thus providing a more realistic grouping of indicators and conducting predictive research (16).

The study uses the "traj" package in Stata 17.0 software for data analysis, first analyzing the dynamic changes in serum IgG, IgM, and NAb from baseline, the first to the third follow-up visits, and employing GBTM to identify latent clusters with similar trajectories. The study hypothesizes that the population serum may be divided into up to five main categories: gradual growth type, gradual decline type, unchanged type, growth and then decline type, decline and then growth type. The model is set with the highest order of 3, meaning the potential groupings of the population are between 1 and 5 groups, with orders 0–3, where the order reflects the speed of the trend changes. The model's effectiveness is evaluated using the BIC and the AIC, with the smallest absolute value of the indicators and closer to 0 indicating a better model fit. Additionally, a higher average posterior probability of group membership (AvePP) indicates a better model fit. To ensure the effectiveness of the grouping, each group composition in the model should account for at least 2–5% and must be consistent with medical knowledge.

After determining the GBTM groupings, χ^2 analysis is used to compare differences in serological change patterns by gender, age, and region, and analysis of variance is used to compare differences in serological change patterns by gender, age, and region. Multivariate linear regression and multivariate logistic regression are used to compare demographic differences between different serological change patterns. A P value of <0.05 is considered statistically significant. See the GBTM results of IgG, IgM and NAb in appendix available at https://cdn.amegroups.cn/static/public/jtd-23-1516-1.xlsx.

High-dimensional transmission dynamics model

$VEAFIR_pRV$ model was built under these assumptions:

In this model, we grouped the total population N into susceptible, V; exposed, E; asymptomatic infection, A; pre-symptomatic infection, F; symptomatic infection, I; removed/recovered that will not be reinfected, R_p; recovered and will possible be reinfected, R.

(I) First, for the total population, we have included various regions (i), age groups (j) and their vaccination status (k) in the compartment N when making the calculation.

(II) Transmission rate, which is in the dimension of [person·time^{-1}], contributes to the reduction of susceptible population (V) and the increase of exposed population (E). Usually, in an ODE model, it simplifies individual properties to the population averaged quantities, and

depicts the age-heterogeneity of transmission by the 'contact frequency matrix'. Therefore, we assumed that should age-specific contact frequency matrices for various regions are available, then for every individual in different age groups, the expected number of symptomatic infectious individuals he/she has contacted during a time interval could be calculated. Then the vaccine efficacy, which considered to be able to reduce transmissibility, was included in to calculation.

(III) In an ODE model, any infected individual is first categorized as exposed (E), then at time t, according to the natural history of COVID-19, there would be two results for exposed population E, they either become pre-symptomatic infection F or asymptomatic infections A. Assume that a proportion p of E is converted to A, and the proportion of E to I is (1 – p). It is generally believed that after a person is exposed to pathogens, the time interval of he/she gets invaded by the pathogen and able to emit it is called latent period. The rate of transformation from E to A is proportional to the amount of E with a scale factor of $p\omega E$ and ω is the latent period coefficient. And in the case of symptomatic infections, since there would be a time lag between the time of virus excretion and symptoms onset, we set ω' and ω'' as average incubation period from exposed (E) to pre-symptomatic infection (F) and average incubation period from pre-symptomatic infection (F) to infections (I), respectively.

(IV) After infection, individuals would be removed/recovered, however, some of the removed/recovered population could experience reinfection while others won't. Therefore, we set two endings for COVID-19 infections: one is the removed/recovered group that would not be infected again (R_p), and the other is removed/recovered group that would experience reinfection (R). Here we set the proportion of those become susceptible again after recovery as δ. At time t, the number of transfers to R and R_p is γI if the time interval between onset and diagnosis from a symptomatic infection I is γ; the number of transfers to R from A who is identified as asymptomatic infection is γ'I.

(V) As reinfection is often correlated with losing immunity, therefore, for R_p become susceptible, V depends on the immunity duration of the individual, here we introduce τ to be the average duration coefficient for immunity duration at time t.

Sensitivity analysis

This High-dimensional transmission dynamics model is an extension of basic SEIAR model, or we could consider it as an SEIAR model with multiple groups. Since the SEIAR models with multi-group are widely used by many studies (36,37), the sensitivity of other model parameters could be found in those references. In our model, the vector of VE, is multiplied on the group-wise contact matrix; which makes the sensitivity analysis analogous to those for the contact matrix.

Additional results

Dep. Variable:	IgG	No. Observations:	1191
Model:	GLM	Df Residuals:	1165
Model Family:	Gaussian	Df Model:	25
Link Function:	identity	Scale:	14.380
Method:	IRLS	Log-Likelihood:	-3264.3
Date:	Thu, 31 Aug 2023	Deviance:	16753.
Time:	08:59:39	Pearson chi2:	1.68e+04
No. Iterations:	3	Pseudo R-squ. (CS):	0.03206
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	15.7720	4.890	3.225	0.001	6.187	25.357
sex[T.Female]	-0.4605	0.377	-1.221	0.222	-1.200	0.279
resd[T.Yes]	-0.1378	0.270	-0.511	0.610	-0.667	0.391
cd[T.Yes]	0.4007	0.410	0.978	0.328	-0.403	1.204
loca[T.XiangAn district]	-1.7410	2.796	-0.623	0.533	-7.220	3.738
loca[T.TongAn district]	-1.6696	2.793	-0.598	0.550	-7.145	3.806
loca[T.HaiCang district]	-2.2075	2.802	-0.788	0.431	-7.698	3.283
loca[T.JiMei district]	-1.4597	2.783	-0.525	0.600	-6.914	3.995
loca[T.HuLi district]	-2.3232	2.787	-0.834	0.404	-7.785	3.139
loca[T.SiMing district]	-1.9249	2.772	-0.694	0.487	-7.358	3.509
vac[T.Basic immnue]	0.8088	1.071	0.755	0.450	-1.291	2.908
vac[T.Booster immnue]	0.4963	0.921	0.539	0.590	-1.309	2.302
phsms[T.very low]	1.0027	0.889	1.128	0.259	-0.740	2.745
phsms[T.low]	-0.2880	1.055	-0.273	0.785	-2.356	1.780
phsms[T.mid]	0.3893	0.878	0.443	0.657	-1.331	2.110
phsms[T.high]	0.8134	0.864	0.942	0.346	-0.880	2.506
phsms[T.very high]	-0.1828	0.985	-0.186	0.853	-2.114	1.748
PI[T.Yes]	-0.1824	0.671	-0.272	0.786	-1.497	1.133
symp[T.Yes]	0.0122	0.608	0.020	0.984	-1.180	1.204
hosp[T.Yes]	-1.4475	0.658	-2.201	0.028	-2.736	-0.159
fam[T.Yes]	0.8821	0.270	3.270	0.001	0.353	1.411
age	-0.0123	0.014	-0.898	0.369	-0.039	0.015
height	-0.0194	0.023	-0.840	0.401	-0.065	0.026
weight	0.0016	0.009	0.186	0.852	-0.015	0.018
smk	-0.0558	0.043	-1.309	0.190	-0.139	0.028
SI	-0.0016	0.013	-0.124	0.901	-0.026	0.023

Figure S1 Regression coefficients (β), 95% CIs for IgG with individual baseline properties. CI, confidence interval; IgG, immunoglobulin G.

```
Intercept                         12.23
sex[T.Female]                     -4.63
resd[T.Yes]                       -1.94
cd[T.Yes]                          3.71
loca[T.XiangAn district]          -2.36
loca[T.TongAn district]           -2.27
loca[T.HaiCang district]          -2.99
loca[T.JiMei district]            -1.99
loca[T.HuLi district]             -3.16
loca[T.SiMing district]           -2.63
vac[T.Basic immnue]                2.86
vac[T.Booster immnue]              2.04
phsms[T.very low]                  4.28
phsms[T.low]                      -1.03
phsms[T.mid]                       1.68
phsms[T.high]                      3.57
phsms[T.very high]                -0.70
PI[T.Yes]                         -1.03
symp[T.Yes]                        0.08
hosp[T.Yes]                       -8.35
fam[T.Yes]                        12.40
age                               -3.40
height                            -3.19
weight                             0.71
smk                               -4.96
SI                                -0.47
dtype: float64
```

Figure S2 Standardized regression coefficients (β) for IgG with individual baseline properties. IgG, immunoglobulin G.

Dep. Variable:	IgM	No. Observations:	1191
Model:	GLM	Df Residuals:	1165
Model Family:	Gaussian	Df Model:	25
Link Function:	identity	Scale:	45.255
Method:	IRLS	Log-Likelihood:	-3947.0
Date:	Thu, 31 Aug 2023	Deviance:	52722.
Time:	09:04:23	Pearson chi2:	5.27e+04
No. Iterations:	3	Pseudo R-squ. (CS):	0.008053
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-0.2712	8.675	-0.031	0.975	-17.275	16.732
sex[T.Female]	0.1216	0.669	0.182	0.856	-1.190	1.433
resd[T.Yes]	0.5441	0.479	1.137	0.256	-0.394	1.483
cd[T.Yes]	-0.0659	0.727	-0.091	0.928	-1.491	1.359
loca[T.XiangAn district]	-0.7255	4.959	-0.146	0.884	-10.445	8.994
loca[T.TongAn district]	-0.5136	4.956	-0.104	0.917	-10.226	9.199
loca[T.HaiCang district]	-0.3873	4.970	-0.078	0.938	-10.128	9.354
loca[T.JiMei district]	-0.2392	4.937	-0.048	0.961	-9.915	9.437
loca[T.HuLi district]	-0.7693	4.944	-0.156	0.876	-10.459	8.920
loca[T.SiMing district]	-0.2762	4.918	-0.056	0.955	-9.915	9.363
vac[T.Basic immnue]	0.3385	1.900	0.178	0.859	-3.386	4.063
vac[T.Booster immnue]	0.6315	1.634	0.387	0.699	-2.571	3.834
phsms[T.very low]	-0.3141	1.577	-0.199	0.842	-3.405	2.777
phsms[T.low]	1.2639	1.872	0.675	0.500	-2.405	4.932
phsms[T.mid]	-0.1162	1.557	-0.075	0.941	-3.169	2.936
phsms[T.high]	0.2525	1.532	0.165	0.869	-2.751	3.256
phsms[T.very high]	-0.2891	1.748	-0.165	0.869	-3.715	3.136
PI[T.Yes]	-0.6548	1.190	-0.550	0.582	-2.987	1.678
symp[T.Yes]	0.2038	1.079	0.189	0.850	-1.910	2.318
hosp[T.Yes]	-0.1212	1.166	-0.104	0.917	-2.407	2.165
fam[T.Yes]	0.3881	0.478	0.811	0.417	-0.550	1.326
age	0.0322	0.024	1.322	0.186	-0.016	0.080
height	0.0027	0.041	0.067	0.947	-0.078	0.083
weight	-0.0029	0.015	-0.193	0.847	-0.033	0.027
smk	-0.0143	0.076	-0.190	0.850	-0.163	0.134
SI	-0.0137	0.023	-0.607	0.544	-0.058	0.031

Figure S3 Regression coefficients (β), 95% CIs for IgM with individual baseline properties. CI, confidence interval; IgM, immunoglobulin M.

```
Intercept                      -0.21
sex[T.Female]                   1.22
resd[T.Yes]                     7.65
cd[T.Yes]                      -0.61
loca[T.XiangAn district]       -0.98
loca[T.TongAn district]        -0.70
loca[T.HaiCang district]       -0.52
loca[T.JiMei district]         -0.33
loca[T.HuLi district]          -1.05
loca[T.SiMing district]        -0.38
vac[T.Basic immnue]             1.20
vac[T.Booster immnue]           2.60
phsms[T.very low]              -1.34
phsms[T.low]                    4.54
phsms[T.mid]                   -0.50
phsms[T.high]                   1.11
phsms[T.very high]             -1.11
PI[T.Yes]                      -3.70
symp[T.Yes]                     1.27
hosp[T.Yes]                    -0.70
fam[T.Yes]                      5.46
age                             8.90
height                          0.45
weight                         -1.30
smk                            -1.27
SI                             -4.08
dtype: float64
```

Figure S4 Standardized regression coefficients (β) for IgM with individual baseline properties. IgM, immunoglobulin M.

Dep. Variable:	nAb	No. Observations:	1191
Model:	GLM	Df Residuals:	1165
Model Family:	Gaussian	Df Model:	25
Link Function:	identity	Scale:	2.9505e+06
Method:	IRLS	Log-Likelihood:	-10548.
Date:	Thu, 31 Aug 2023	Deviance:	3.4373e+09
Time:	09:06:53	Pearson chi2:	3.44e+09
No. Iterations:	3	Pseudo R-squ. (CS):	0.04587
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	6737.8837	2215.164	3.042	0.002	2396.243	1.11e+04
sex[T.Female]	-445.4320	170.808	-2.608	0.009	-780.210	-110.654
resd[T.Yes]	3.3422	122.250	0.027	0.978	-236.264	242.949
cd[T.Yes]	160.6590	185.645	0.865	0.387	-203.198	524.516
loca[T.XiangAn district]	-1883.5353	1266.276	-1.487	0.137	-4365.391	598.321
loca[T.TongAn district]	-1678.3478	1265.352	-1.326	0.185	-4158.392	801.697
loca[T.HaiCang district]	-2064.9435	1269.009	-1.627	0.104	-4552.156	422.268
loca[T.JiMei district]	-1823.9418	1260.520	-1.447	0.148	-4294.516	646.632
loca[T.HuLi district]	-1941.9523	1262.277	-1.538	0.124	-4415.971	532.066
loca[T.SiMing district]	-1826.6031	1255.714	-1.455	0.146	-4287.758	634.551
vac[T.Basic immnue]	1001.7013	485.227	2.064	0.039	50.674	1952.729
vac[T.Booster immnue]	768.6595	417.213	1.842	0.065	-49.062	1586.381
phsms[T.very low]	329.6995	402.675	0.819	0.413	-459.530	1118.929
phsms[T.low]	-301.3478	477.924	-0.631	0.528	-1238.063	635.367
phsms[T.mid]	79.2618	397.682	0.199	0.842	-700.180	858.704
phsms[T.high]	198.8071	391.255	0.508	0.611	-568.039	965.653
phsms[T.very high]	96.3143	446.271	0.216	0.829	-778.360	970.989
PI[T.Yes]	-3.9426	303.887	-0.013	0.990	-599.550	591.665
symp[T.Yes]	242.3555	275.415	0.880	0.379	-297.449	782.160
hosp[T.Yes]	-657.7967	297.845	-2.209	0.027	-1241.562	-74.032
fam[T.Yes]	559.6020	122.176	4.580	0.000	320.141	799.063
age	-3.0520	6.223	-0.490	0.624	-15.250	9.146
height	-16.3906	10.478	-1.564	0.118	-36.928	4.147
weight	-0.2215	3.892	-0.057	0.955	-7.849	7.406
smk	-54.6772	19.321	-2.830	0.005	-92.545	-16.809
SI	1.3280	5.759	0.231	0.818	-9.959	12.615

Figure S5 Regression coefficients (β), 95% CIs for NAb with individual baseline properties. CI, confidence interval; NAb, neutralizing antibody.

```
Intercept                      5,224.73
sex[T.Female]                 -4,479.39
resd[T.Yes]                       46.96
cd[T.Yes]                      1,486.51
loca[T.XiangAn district]      -2,555.00
loca[T.TongAn district]       -2,278.33
loca[T.HaiCang district]      -2,795.05
loca[T.JiMei district]        -2,485.46
loca[T.HuLi district]         -2,642.59
loca[T.SiMing district]       -2,498.62
vac[T.Basic immnue]            3,546.00
vac[T.Booster immnue]          3,164.63
phsms[T.very low]              1,406.40
phsms[T.low]                  -1,083.07
phsms[T.mid]                     342.35
phsms[T.high]                    872.81
phsms[T.very high]               370.71
PI[T.Yes]                        -22.29
symp[T.Yes]                    1,511.51
hosp[T.Yes]                   -3,793.57
fam[T.Yes]                     7,867.55
age                             -842.37
height                        -2,686.88
weight                           -97.74
smk                           -4,861.00
SI                               396.08
dtype: float64
```

Figure S6 Standardized regression coefficients (β) for NAb with individual baseline properties. NAb, neutralizing antibody.

References

36. Zhao Z, Chen Q, Zhao B, et al. Transmission pattern of shigellosis in Wuhan City, China: a modelling study. Epidemiology & Infection 2021;149:e249.

37. Zhao ZY, Zhu YZ, Xu JW, et al. A five-compartment model of age-specific transmissibility of SARS-CoV-2. Infect Dis Poverty 2020;9:117.