THE LANCET
Infectious Diseases

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Singanayagam A, Hakki S, Dunning J, et al. Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: an observational, longitudinal, cohort study. *Lancet Infect Dis* 2021; published online Oct 28. https://doi.org/10.1016/S1473-3099(21)00648-4.

Supplementary Material

Supplementary Tables:

Supplementary Figure:

Supplementary Methods:

Supplementary Tables

Table S1: Demographics of PCR-positive participants infected with delta, alpha and pre-alpha strains. Of the 163 PCR-positive participants, 19 were index cases, 125 were household contacts and 19 were non-household contacts. Chi-squared tests were performed to determine differences for each characteristic between pre-alpha-, alpha- and delta- infected cases. *BMI missing for participants under the age of 18.

(1) Table includes the data of 1 alpha index case and 18 delta index cases. If these index cases are removed, leaving only contacts in the comparison, the P-values for the demographic characteristics are as follows: sex, p = 0.64, age, p = 0.01, BMI, p = 0.74, ethnicity, p = 0.63, co-morbidities, p = 0.71, smoking status, p = 0.11, type of contact, p = 0.20, vaccination status, p < 0.01.

Characteristics(1)		**Total (n = 163)**	**Pre-alpha (n = 50)**	**Alpha (n = 42)**	**Delta (n = 71)**	***P*-value**
Sex	Female (%)	89 (55)	29 (58)	23 (55)	37 (52)	0.80
	Male (%)	74 (45)	21 (42)	19 (45)	34 (48)	
Age	Median (IQR)	36 (26 - 50)	39 (29 - 51)	35 (27-49)	34 (18 - 49)	-
	<18 years (%)	24 (15)	3 (6)	3 (7)	18 (25)	0.01
	18-49 years (%)	97 (60)	31 (62)	30 (71)	36 (51)	
	50-64 years (%)	37 (23)	15 (30)	9 (21)	13 (18)	
	≥ 65 years (%)	5 (3)	1 (2)	0 (0)	4 (6)	
BMI*	Underweight (%)	2 (1)	0 (0)	1 (2)	1 (1)	0.89
	Normal (%)	58 (36)	19 (38)	15 (36)	24 (34)	
	Overweight (%)	38 (23)	14 (28)	10 (24)	14 (207)	
	Obese (%)	27 (17)	9 (18)	9 (21)	9 (13)	
	Morbidly obese (%)	7 (4)	3 (6)	3 (7)	1 (1)	
	Unknown (%)	31 (19)	5 (10)	4 (10)	22 (31)	-
Ethnicity	White (%)	133 (82)	43 (86)	35 (83)	55 (78)	0.517
	Non-white (%)	20 (12)	5 (10)	4 (10)	11 (16)	
	Unknown (%)	10 (6)	2 (4)	3 (7)	5 (7)	-
Comorbidities	Yes (%)	55 (34)	15 (30)	15 (36)	25 (35)	0.80
	No (%)	108 (66)	35 (70)	27 (64)	46 (65)	
Smoking status	Current (%)	14 (9)	3 (6)	7 (17)	4 (6)	0.08
	Former (%)	18 (11)	9 (18)	2 (5)	7 (10)	
	Never (%)	126 (77)	36 (72)	33 (79)	57 (80)	
	Unknown (%)	5 (3)	2 (4)	0 (0)	3 (4)	-
Type of contact	Household (%)	125 (76.5)	43 (86)	29 (69)	53 (74.7)	0.20
	Non-household (%)	19 (12)	7 (14)	12 (29)	0 (0)	
	Index (%)	19 (12)	0 (0)	1 (2)	18 (25)	
Vaccination status	Fully vaccinated (%)	39 (24)	0 (0)	1 (2)	38 (54)	p < 0.01
	Unvaccinated (%)	113 (69)	50 (100)	41 (98)	22 (31)	
	Partially vaccinated (%)	11 (7)	0 (0)	0 (0)	11 (16)	

Table S2: Demographic characteristics of delta-infected unvaccinated, partially-vaccinated and fully-vaccinated participants. Chi-squared tests were performed to determine differences for each characteristic between fully vaccinated, partially vaccinated and unvaccinated infected cases. *BMI missing for participants under the age of 18.

(1) Table includes the data of 18 index cases (7 fully-vaccinated,. 3 partially-vaccinated and 8 unvaccinated). If these index cases are removed, the P-values for the demographic characteristics for contacts only (n=53) are as follows: sex, p = 0.02, age, p < 0.01, ethnicity, p = 0.59, BMI, p = 0.66, co-morbidities, p = 0.64, symptomatic status p = 0.94.

Characteristics(1)		**Total (n=71)**	**Fully vaccinated (n=38)**	**Partially vaccinated (n=10)**	**Unvaccinated (n=23)**	***P*-value**
Sex	Female (%)	37 (52)	24 (67)	6 (60)	7 (30)	0.04
	Male (%)	34 (48)	14 (33)	4 (40)	16 (70)	
	Unknown	0 (0)	0 (0)	0 (0)	0 (0)	-
Age	Median (IQR)	34 (18-49)	49 (41– 55)	34 (31 – 39)	13 (11 – 17)	-
	<18 years (%)	18 (25)	0 (0)	0 (0)	18 (78)	p < 0.01
	18-49 years (%)	36 (51)	22 (58)	9 (90)	5 (22)	
	50-64 years (%)	13 (18)	12 (32)	1 (10)	0 (0)	
	≥65 years (%)	4 (6)	4 (11)	0 (0)	0 (0)	
	Unknown (%)	0 (0)	0 (0)	0 (0)	0 (0)	-
Ethnicity	White (%)	48 (68)	26 (68)	8 (80)	14 (61)	0.71
	Non-white (%)	18 (25)	9 (24)	2 (20)	7 (30)	
	Unknown (%)	5 (7)	3 (8)	0 (0)	2 (9)	-
BMI*	Underweight (%)	1 (1)	1 (3)	0 (0)	0 (0)	0.69
	Normal (%)	23 (32)	14 (37)	5 (50)	4 (17)	
	Overweight (%)	14 (20)	11 (29)	3 (30)	0 (0)	
	Obese (%)	9 (13)	7 (18)	2 (20)	0 (0)	
	Morbidly obese (%)	1 (1)	1 (3)	0 (0)	0 (0)	
	Unknown (%)	23 (32)	4 (11)	0 (0)	19 (83)	-
Comorbidities	Yes (%)	23 (32)	15 (39)	3 (30)	5 (22)	0.36
	No (%)	46 (65)	22 (58)	7 (70)	17 (74)	
	Unknown (%)	2 (3)	1 (3)	0 (0)	1 (4)	-
Symptomatic	Symptomatic (%)	36 (51)	17 (45)	6 (60)	13 (57)	0.90
	Asymptomatic (%)	23 (32)	12 (32)	4 (40)	7 (30)	
	Unknown (%)	12 (17)	9 (25)	0 (0)	3 (13)	-

Table S3 - Demographics of PCR-positive and PCR-negative delta variant contacts. Chi-squared tests were performed to determine differences for each characteristic between PCR-positive and PCR-negative cases, except time between second vaccination and recruitment for which Mann Whitney test was performed . *BMI missing for participants under the age of 18. One PCR-negative contact was excluded as their vaccination status was unknown.

Characteristics		Total (n=231)	PCR-positive (n=53)	PCR-negative (n=178)	*P*-value
Sex	Female (%)	127 (55)	26 (49)	101 (57)	0.30
	Male (%)	101 (44)	27 (51)	74 (42)	
	Unknown (%)	3 (1)	0 (0)	3 (2)	-
Age	Median (IQR)	41 (28 - 49)	41 (19-49)	41 (29 - 49)	-
	<18 years (%)	30 (13)	12 (23)	18 (10)	0.05
	18-49 years (%)	145 (63)	28 (53)	117 (66)	
	50-64 years (%)	43 (19)	9 (17)	34 (19)	
	≥65 years (%)	10 (4)	4 (8)	6 (3)	
	Unknown (%)	3 (1)	0 (0)	3 (2)	-
BMI*	Underweight, <18.5 (%)	6 (3)	1 (2)	5 (3)	0.83
	Normal, 18.5-25 (%)	102 (44)	20 (38)	82 (46)	
	Overweight, 25-30 (%)	58 (25)	12 (23)	46 (26)	
	Obese, 30-40 (%)	25 (11)	7 (13)	18 (10)	
	Morbidly obese, >40 (%)	2 (1)	0 (0)	2 (1)	
	Unknown* (%)	38 (17)	13 (25)	25 (14)	-
Ethnicity	White (%)	196 (85)	42 (80)	154 (87)	0.48
	Non-white (%)	27 (12)	8 (15)	19 (11)	
	Unknown (%)	8 (4)	3 (6)	5 (3)	-
Comorbidities	Yes (%)	77 (33)	20 (38)	57 (32)	0.54
	No (%)	154 (67)	33 (62)	121 (68)	
Smoking status	Current (%)	19 (8)	3 (6)	16 (9)	0.76
	Former (%)	25 (11)	6 (11)	19 (11)	
	Never (%)	183 (79)	42 (79)	141 (79)	
	Unknown (%)	4 (2)	2 (4)	2 (1)	-
Type of contact	Household contact (%)	205 (89)	53 (100)	152 (85)	1
	Non-household contact (%)	26 (11)	0 (0)	26 (15)	
Vaccination status	Fully Vaccinated (%)	138 (60)	30 (57)	108 (61)	0.07
	Partially vaccinated (%)	47 (20)	7 (13)	40 (23)	
	Non-vaccinated (%)	46 (20)	16 (30)	30 (17)	
Fully vaccinated contacts: Time between 2nd vaccination and enrolment	Median days (IQR)	74 (35-105)	101 (74-120)	64 (32-97)	p < 0.01

Table S4 – Model selection using Leave One Out Cross Validation (LOO-CV). Five model variants were examined. The most predictive model for both the ORF1ab and E gene Ct value fits was the most complex: fitting parameters to the four variant/vaccination-status groups (pre-alpha, alpha, delta-unvaccinated and delta-vaccinated) together with three correlation coefficients describing the within-group correlation between peak titre, growth rate and decline rate. ELPD = expected log pointwise predictive density [1]. Other terms and approach defined in ref [1]. Analysis used the R package loo [2]. See Supplementary Methods for details of priors.

Dataset	Model description	Number of parameters fitted	Mean ELPD difference	Standard Error of ELPD difference	LOO estimate of ELPD	Standard error in LOO estimate of ELPD	Mean estimate of effective number of parameters	Standard error in estimate of effective number of parameters	Proportion of observations with Pareto k diagnostic values >=0.7
ORF1ab gene Ct values (1951 data points)	*4 groups, correlation, informative prior on growth, peak VL independent of group but dependent on age*	*552 (4 per participant + 20)*	*0*	*0*	*-4110.71*	*57.844*	*482.192*	*15.259*	*0.90%*
	4 groups, correlation, informative prior on growth	554 (4 per participant + 22)	-2.948	1.627	-4113.66	57.834	483.339	15.314	0.80%
	1 group, correlation, uninformative prior on growth	545 (4 per participant + 13)	-8.037	2.51	-4118.75	57.745	487.798	15.335	0.80%
	1 group, correlation, informative prior on growth	545 (4 per participant + 13)	-9	2.536	-4119.71	57.753	489.067	15.318	0.80%
	4 groups, no correlation, informative prior on growth	551 (4 per participant + 19)	-13.267	5.631	-4123.98	57.819	486.531	15.824	0.60%
	1 group, no correlation, uninformative prior on growth	542 (4 per participant + 10)	-17.378	6.05	-4128.09	57.608	487.366	15.746	1.00%
E gene Ct values (1934 data points)	*4 groups, correlation, informative prior on growth, peak VL independent of group but dependent on age*	*552 (4 per participant + 20)*	*0*	*0*	*-4255.60*	*54.578*	*474.364*	*15.072*	*0.90%*
	4 groups, correlation, informative prior on growth	554 (4 per participant + 22)	-0.674	1.463	-4256.27	54.674	476.651	15.212	0.9%
	1 group, correlation, informative prior on growth	545 (4 per participant + 13)	-6.901	2.99	-4262.50	54.538	478.208	15.445	0.8%
	4 groups, no correlation, informative prior on growth	551 (4 per participant + 19)	-8.143	4.467	-4263.74	54.538	473.719	15.377	1.4%
	1 group, correlation, uninformative prior on growth	545 (4 per participant + 13)	-8.544	3.332	-4264.14	54.415	479.578	15.487	1.0%
	1 group, no correlation, uninformative prior on growth	542 (4 per participant + 10)	-9.817	5.663	-4265.41	54.369	465.844	15.333	1.3%

Table S5 – Posterior estimates of key summary statistics for the most predictive model fitted to ORF1ab gene Ct value data, describing peak VL, growth rate of VL and decline rate of VL. Group-level population averages and within-sample averages shown (see Supplementary Methods for definitions); while mean and median estimates are very similar for both types of statistic, the within-sample estimates have narrower credible intervals than the group-level estimates (akin to the difference between a population variance and a sample variance). Note that for the most predictive model, mean peak VL is not fitted as a group-specific parameter, but as a parameter affecting all groups.

Statistic type	Statistic	Mean	Median	2.5% percentile	97.5% percentile
Population (group-level)	Peak log10 VL/ml in 50-year-olds: all groups	8.14	8.14	7.95	8.32
	VL growth rate/day: pre-alpha	7.45	6.87	4.11	14.14
	VL growth rate/day: alpha	7.2	6.65	4.05	13.69
	VL growth rate/day: delta-unvaccinated	6.46	5.94	3.39	12.59
	VL growth rate/day: delta-vaccinated	6.2	5.71	3.49	11.9
	VL decline rate/day: pre-alpha	1.59	1.58	1.34	1.87
	VL decline rate/day: alpha	1.89	1.88	1.55	2.32
	VL decline rate/day: delta-unvaccinated	1.82	1.8	1.37	2.39
	VL decline rate/day: delta-vaccinated	2.19	2.17	1.74	2.72
Within-sample	Peak log10 VL/ml in 50-year-olds: pre-alpha	8.1	8.09	7.9	8.29
	Peak log10 VL/ml in 50-year-olds: alpha	8.15	8.15	7.94	8.37
	Peak log10 VL/ml in 50-year-olds: delta-unvaccinated	8.09	8.09	7.74	8.42
	Peak log10 VL/ml in 50-year-olds: delta-vaccinated	8.19	8.19	7.99	8.41
	VL growth rate/day: pre-alpha	7.35	6.68	4.11	14.52
	VL growth rate/day: alpha	6.23	5.55	3.49	13.03
	VL growth rate/day: delta-unvaccinated	4.88	4.16	2.19	11.78
	VL growth rate/day: delta-vaccinated	4.96	4.43	3.01	10.19
	VL decline rate/day: pre-alpha	1.54	1.54	1.43	1.68
	VL decline rate/day: alpha	1.89	1.87	1.64	2.21
	VL decline rate/day: delta-unvaccinated	1.83	1.81	1.54	2.2
	VL decline rate/day: delta-vaccinated	2.19	2.18	1.88	2.57

Table S6. As Table S5, but for the most predictive model fitted to E gene Ct value data.

Statistic type	Statistic	Mean	Median	2.5% percentile	97.5% percentile
Population (group-level)	Peak log10 VL/ml in 50-year-olds: all groups	8.21	8.21	8.01	8.4
	VL growth rate/day: pre-alpha	6.26	5.9	3.8	10.87
	VL growth rate/day: alpha	5.95	5.59	3.61	10.4
	VL growth rate/day: delta-unvaccinated	5.36	5.04	3.06	9.57
	VL growth rate/day: delta-vaccinated	5.24	4.93	3.2	9.15
	VL decline rate/day: pre-alpha	1.46	1.45	1.24	1.72
	VL decline rate/day: alpha	1.62	1.61	1.34	1.97
	VL decline rate/day: delta-unvaccinated	1.69	1.67	1.28	2.21
	VL decline rate/day: delta-vaccinated	2.07	2.05	1.65	2.57
Within-sample	Peak log10 VL/ml in 50-year-olds: pre-alpha	8.19	8.18	7.98	8.39
	Peak log10 VL/ml in 50-year-olds: alpha	8.24	8.24	8.02	8.46
	Peak log10 VL/ml in 50-year-olds: delta-unvaccinated	8.16	8.16	7.82	8.5
	Peak log10 VL/ml in 50-year-olds: delta-vaccinated	8.27	8.27	8.06	8.48
	VL growth rate/day: pre-alpha	6.85	6.38	4.03	12.47
	VL growth rate/day: alpha	5.27	4.86	3.27	9.69
	VL growth rate/day: delta-unvaccinated	4.67	4.15	2.36	10.01
	VL growth rate/day: delta-vaccinated	4.69	4.33	3.07	8.47
	VL decline rate/day: pre-alpha	1.43	1.43	1.32	1.57
	VL decline rate/day: alpha	1.58	1.57	1.41	1.82
	VL decline rate/day: delta-unvaccinated	1.66	1.65	1.42	1.97
	VL decline rate/day: delta-vaccinated	2.05	2.05	1.76	2.4

Table S7– Group-level and global posterior parameter estimates for the most predictive model fitted to ORF1ab gene Ct value data. See Supplementary Methods for parameter definitions.

Parameter	Mean	Median	2.5% percentile	97.5% percentile	Effective sample size
μ_1	18.733	18.734	18.298	19.166	13298
ν_a	0.546	0.547	-0.037	1.12	16381
$\mu_{2,1}$	0.985	0.985	0.661	1.314	15071
$\mu_{2,2}$	0.951	0.951	0.636	1.267	13873
$\mu_{2,3}$	0.836	0.836	0.463	1.203	14868
$\mu_{2,4}$	0.801	0.802	0.473	1.125	13224
$\mu_{3,1}$	0.307	0.307	0.151	0.463	7754
$\mu_{3,2}$	0.484	0.483	0.301	0.674	7982
$\mu_{3,3}$	0.441	0.44	0.163	0.715	3573
$\mu_{3,4}$	0.626	0.626	0.403	0.846	9860
δ_1	1.607	1.602	1.239	2.013	8978
δ_2	1.379	1.367	1.028	1.798	6616
δ_3	0.544	0.543	0.458	0.643	9005
$c_{1,2}$	0.417	0.426	0.139	0.645	7725
$c_{1,3}$	0.073	0.075	-0.222	0.358	9517
$c_{2,3}$	-0.444	-0.453	-0.665	-0.179	8929
σ_v	2.507	2.505	2.339	2.683	11134
p	0.109	0.108	0.088	0.131	12703
x_0	-0.557	-0.521	-3.259	1.904	7259
σ_0	9.257	9.211	7.694	11.089	11874

Table S8 – As Table S7, but for the most predictive model fitted to E gene Ct value data. See Supplementary Methods for parameter definitions.

Parameter	Mean	Median	2.5% percentile	97.5% percentile	Effective sample size
μ_1	18.901	18.904	18.438	19.351	11751
ν_a	0.499	0.498	-0.112	1.099	15282
$\mu_{2,1}$	1.033	1.033	0.701	1.351	15349
$\mu_{2,2}$	0.981	0.981	0.671	1.283	13963
$\mu_{2,3}$	0.87	0.872	0.507	1.238	13990
$\mu_{2,4}$	0.855	0.855	0.532	1.171	11656
$\mu_{3,1}$	0.234	0.233	0.079	0.391	5960
$\mu_{3,2}$	0.342	0.343	0.16	0.526	7775
$\mu_{3,3}$	0.378	0.379	0.108	0.646	12238
$\mu_{3,4}$	0.581	0.582	0.36	0.799	8347
δ_1	1.634	1.629	1.221	2.074	6691
δ_2	1.223	1.21	0.896	1.62	6537
δ_3	0.523	0.521	0.441	0.617	9865
$c_{1,2}$	0.333	0.342	0.004	0.597	7157
$c_{1,3}$	0.143	0.146	-0.171	0.438	7339
$c_{2,3}$	-0.391	-0.397	-0.645	-0.093	7114
σ_v	2.686	2.685	2.512	2.868	10357
p	0.095	0.095	0.075	0.118	12808
x_0	-1.981	-1.931	-5.509	1.259	3925
σ_0	9.732	9.68	7.889	11.853	15109

Table S9. Posterior probabilities for the most predictive model ORF1ab gene Ct value data. Values above the diagonal show the posterior probability that the within-sample mean value of the parameter (peak VL, growth rate or decline rate) is larger for the group specified in the respective column title than that referenced in the row title. Values below the diagonal give the posterior probability that the population (group-level) mean value of the parameter is larger for the group specified in the row title than that referenced in the column title. Note that group-level comparisons of Peak VL are not available for the most predictive model, for which only group-specific VL growth and decline rates are fitted. Within-sample posterior probabilities are generally (but not always) more certain (closer to 0 or 1) than population values. The posterior probability that that one group has a parameter estimate less than another group is just 1 minus the posterior probability that that the former group has a parameter estimate greater than the latter group. delta-U = delta-unvaccinated, delta-V = delta-vaccinated. Probabilities are derived from 20,000 posterior samples and have sampling errors of <0.01.

Peak VL	*pre-alpha*	*alpha*	*delta-U*	*delta-V*
pre-alpha		0.7	0.48	0.8
alpha			0.34	0.62
delta-U				0.73
delta-V				

VL growth rate	*pre-alpha*	*alpha*	*delta-U*	*delta-V*
pre-alpha		0.31	0.16	0.14
alpha	0.44		0.27	0.28
delta-U	0.27	0.32		0.57
delta-V	0.21	0.25	0.44	

VL decline rate	*pre-alpha*	*alpha*	*delta-U*	*delta-V*
pre-alpha		1	0.96	1
alpha	0.93		0.38	0.93
delta-U	0.79	0.4		0.94
delta-V	0.99	0.84	0.85	

Table S10. As table S9 but for the most predictive model fitted to E gene Ct value data.

Peak VL	*pre-alpha*	*alpha*	*delta-U*	*delta-V*
pre-alpha		0.68	0.44	0.75
alpha			0.31	0.58
delta-U				0.73
delta-V				

VL growth rate	*pre-alpha*	*alpha*	*delta-U*	*delta-V*
pre-alpha		0.21	0.16	0.13
alpha	0.41		0.34	0.36
delta-U	0.26	0.32		0.55
delta-V	0.22	0.29	0.48	

VL decline rate	*pre-alpha*	*alpha*	*delta-U*	*delta-V*
pre-alpha		0.92	0.95	1
alpha	0.82		0.67	0.99
delta-U	0.83	0.59		0.97
delta-V	1	0.95	0.88	

Supplementary Figure

Figure S1. E gene VL trajectories from 14 days before peak to 28 days after for n=117 participants infected with pre-alpha (red), alpha (green) or delta (unvaccinated – blue, fully vaccinated = purple) variants. Black points = measured values, curves = model posterior median estimate, grey = 95% credible region.

Supplementary Methods

1. Upper respiratory tract sampling

In both ATACCC1 and ATACCC2, following detailed instructions and demonstration by a research nurse, participants collected self-performed combined nose (anterior nares) and throat swabs (a single swab used for throat then nose) at home for up to 14-20 consecutive days. If throat sampling was not tolerated, only the nose was swabbed. Samples were collected same-day and delivered to PHE for SARS-CoV-2 PCR testing.

2. SARS-CoV-2 RT-qPCR

For each swab sample, UTM was aliquoted into lysis buffer containing the exogenously added internal control (IC), prior to purification of nucleic acid. Following automated extraction of viral RNA from the sample, real-time PCR was performed. The PHE triplex assay uses TaqPath™ 1-Step Multiplex Master Mix to amplify the targets over 40 cycles and specifically detects SARS-CoV-2 in the ORF1Ab assay target[3], and Sarbecoviruses including SARS CoV-2 in the E gene target[4]. The assays were optimized and validated locally on the ABI QuantiStudio 7 Flex instrument, using the Invitrogen TaqPath Multiplex Master Mix. Two sets of primers and probes are used to detect SARS-CoV-2, with a third set used to amplify an 80 base-pair sequence of the coat-protein gene of soil-borne cereal mosaic virus that acts as a control for exogenous extraction and RT-PCR. Cycle threshold (Ct) values were determined using QuantStudio software, with the threshold set individually for each channel according to the exponential growth curves and above background fluorescence of negative controls. All assay results were analysed using pre-determined threshold values and SARS-CoV-2 was reported as detected if either ORF1ab or E gene is detected at Ct < 35, or if both targets are detected at Ct >35 and <40. SARS-CoV-2 infection status was assigned to a participant if SARS-CoV-2 RNA was detected by PCR in two or more consecutive daily samples. Samples received up to 12 October 2020 were tested using a duplex version of the PHE assay (ORF1Ab and internal control only), as this was the assay used by the reference laboratory prior to switching to the triplex assay.

3. Conversion of Ct values to viral genome copies

The national reference laboratory performed separate work (unpublished) using quantitative *in vitro* transcripts to generate estimates of the relationship between Ct values and viral RNA copy number for the PCR tests conducted by Public Health England. The resulting equation for determining RNA copies per reaction from an Orf1ab gene Ct value is exp((37.933-Ct)/1.418) and for E gene Ct values is exp((37.564-Ct)/1.394). To calculate RNA copies/ml, the RNA copies/reaction were multiplied by the relevant dilution factor. RNA was eluted from 150µl sample, eluted in 100µl and 5µl RNA used per reaction = (1000/150) x (100/5) = 133.3333; therefore, the RNA copies/reaction values were multiplied by 133.3333 to determine RNA copies/ml.

4. Whole genome sequencing and lineage assignments

Samples with a positive RT-qPCR result were submitted for WGS to either PHE or Imperial College to assign lineages.

Samples with the highest viral load were chosen. For WGS performed at PHE for alpha and pre-alpha samples, viral amplicons were generated using the ARTIC amplicon generation method and sequenced on the Illumina sequencing platform (HiSeq or NextSeq) using the Illumina Nextera library preparatory kits. Raw sequences were trimmed and aligned against a SARS-CoV-2 reference genome (NC_045512.2). Consensus sequences were generated using a PHE bioinformatics pipeline, and samples with >80% genome coverage were included in the analysis. For WGS performed at Imperial College for alpha and delta variants, automated RNA extraction was performed using a CyBio FeliX (Analytik Jena) and innuPREP Virus TS RNA Kit 2.0 (Analytik Jena) according to the manufacturer's instructions, with a sample volume of 200 µl, without carrier RNA and with an elution volume of 50 µl. RT-qPCR was repeated using an in-house protocol.[5] cDNA synthesis was then performed using the LunaScript RT SuperMix Kit (NEB) according to the manufacturer's instructions with a total reaction volume of 20 µl and extracted sample volume of 5 µl. Libraries were generated using the EasySeq™ RT-PCR SARS CoV-2 (novel coronavirus) Whole Genome Sequencing kit v3 (Nimagen) according to the manufacturer's instructions. Samples were then pooled and purified with AMPure XP (Beckman Coulter) magnetic beads. Suitable quality of libraries was confirmed using a Tapestation (Agilent) and concentrations were measured using the Qubit 1x dsDNA High Sensitivity Assay Kit (Thermofisher Scientific) and Qubit 4 Fluorometer (ThermoFisher Scientific). Pooled libraries were then diluted down to 55 pM. The final pool was then run on an iSeq 100 (Illumina) with a total of 322 cycles (151 bp paired reads and 10 bp indices). Generated fastq files were processed using the EasySeq variant pipeline (v0.8.1)[5] which is a Nextflow[6] pipeline that uses

fastp[7], BWA MEM[8], SAMtools[9], BCFtools[9], LoFreq[10], mosdepth[11], BEDtools[12], SnpEff[13] and MultiQC[14] to QC, trim and assemble the reads (using reference sequence NC_045512.2) and then generate a consensus sequence and variant report before assigning a PANGO lineage[6] using pangolin (v3.1.11, lineages version 2021-09-17)[7]. Genomically-confirmed alpha variant status was reported if all lineage defining non-synonymous changes[8] were called as alternate base. Genomically-probable alpha variant was reported if at least five lineage defining non-synonymous changes were called as alternate base and all other positions either N or mixed base. In the analyses presented here, both genomically-probable and genomically-confirmed cases were included. pre-alpha status was assigned to cases where alpha infection, and infections caused by other variants of concern or variants under investigation had been excluded.

5. Modelling viral kinetics

Cases were included in the modelling analysis if they matched one or more of the following criteria

- Follow-up at least up to and including day 12 (where day 1 is day of enrolment)
- First viral load measurement was undetectable (i.e. an incident case)
- Last two viral load measurements were undetectable (i.e. fully resolved decline)

These choices ensured there were sufficient data points to allow individual-specific kinetic parameters to be estimated, while maximising inclusion of incident cases (and therefore power to infer VL growth rate). All partially vaccinated cases were excluded, together with one fully vaccinated alpha case (given there is no power to estimate group parameters from a single case).

To model viral kinetics, we used a simple phenomenological model of viral titre[15] during disease pathogenesis:

$$v(\tau) = v_{max}\frac{(a+b)}{be^{-a(\tau-\tau_{max})} + ae^{b(\tau-\tau_{max})}} \qquad (1)$$

This function shows exponential growth at rate a for $\tau \ll \tau_{max}$, exponential decline at rate b for $\tau \gg \tau_{max}$, and has a maximum of v_{max} at time $\tau = \tau_{max}$. It reproduces dynamics similar to explicit within-host models[16] and has the benefit of being a continuous rather than a piecewise linear (on the $\ln v$ scale) function[17].

PCR cycle threshold (CT) values were converted to estimates of viral copies/ml of sample using the following calibration equation:

$$\ln v = \kappa + \frac{x}{\alpha} \qquad (2)$$

Here $x = 40 - CT$, where CT is the CT value measured (always £40) and 40 is the limit of detection for the PCR test used. For the ORF1ab target, calibration experiments gave estimates of κ =3.435 and α =1.418. For the E gene target, κ =3.145 and α =1.394. These are just transformed versions of the equations given in Supplementary Methods section 1 above.

Test accuracy and model misspecification was modelled with a mixture model by assuming there was a probability p of a test giving an observation x drawn from a normal error distribution with mean x_0 and standard deviation σ_0, and probability $1-p$ of it being drawn from the true distribution, assumed to be normally distributed with mean $\alpha(\log v - \kappa)$ (from equations 1 and 2), standard deviation σ_v (representing the accuracy of Ct measurements). Hence the overall log-likelihood of observing x when $x > 0$ is

$$l(x) = \log[p n(x|x_0, \sigma_0) + (1-p) n(x|\alpha(\log v - \kappa), \sigma_v)] \qquad (3)$$

where $n()$ represents the probability density function (pdf) of the normal distribution. We use $n(\mu, \sigma)$ to represent the distribution with mean μ and standard deviation σ, and $n(x|\mu, \sigma)$ to represent the probability density of that distribution at point x.

When $x = 0$ (the limit of detection) the observation of viral load is censored (the true value could be any value below that limit of detection). Hence the log-likelihood is

$$l(x=0) = \log[p\mathcal{N}(0|x_0, \sigma_0) + (1-p)\mathcal{N}(0|m(\log v - k), \sigma_v)] \qquad (4)$$

where $\mathcal{N}()$ represents the cumulative density function of the normal distribution.

The error distribution $n(x|x_0, \sigma_0)$ represents both false negative and false positive errors, with the proportion of each being governed by the value of x_0, $x_0 = 0$ giving a 1:1 ratio of each. As importantly, it also allows for a degree of misspecification of the quasi-mechanistic model given in equation (1) by accommodating viral kinetic profiles that are poorly represented by simple fixed rate exponential growth and then decline of viral titre.

Our statistical framework differs from previous work[17] in inferring the parameter p from the data, in using a more flexible (and inferred) functional form for the error distribution, and in representing the likelihood of Ct values above the limit of detection with the correct cumulative distribution function.

We used a Bayesian hierarchical model to fit this model to the entire dataset of sequential Ct values measured for all study participants. Given only a minority of participants were incident (i.e. initially undetectable virus) cases in our dataset, our ability to infer viral growth rates was limited. We therefore supplemented the observed Ct values with exposure data for the 19 participants who were non-household contacts of index cases and had a unique date of exposure identified; the Ct data for these participants were supplemented by a pseudo-absence data point (i.e. undetectable virus) on the date of exposure.

Parameters v_{max}, a, b and τ_{max} are fitted on a participant-specific basis. Since the first three of these parameters are positive, we fit the log of each parameter. Letting $i = 1..3$ index these parameters and $j = 1..N$ index participants, we define a parameter matrix $\theta_{i,j}$ such that

$$\begin{aligned} v_{max,j} &= \exp(\theta_{1,j}) \\ a_j &= \exp(\theta_{2,j}) \\ b_j &= \exp(\theta_{3,j}) \end{aligned} \qquad (5)$$

The hierarchical structure in the inference model is represented by grouping participants on the basis of the infecting variant and their vaccination status, such that the group of participants j is $k(j)$. A single group model was fitted, which implicitly assumes that viral kinetic parameters vary by subject but not by variant or vaccination status, and a four group model was also explored, where $k = 1, 2, 3, 4$ represents pre-alpha, alpha, unvaccinated (UV) delta and fully-vaccinated (FV) delta respectively. All but one of the infections observed in vaccinated individuals were caused by delta. We excluded the one alpha infection in a vaccinated individual from the analysis, given a single subject gives insufficient power to estimate group-level parameters.

We used a non-centred representation for the hierarchical model and defined

$$\theta_{i,j} = \mu_{i,k(j)} + \delta_i z_{i,j} \qquad (7)$$

Here $z_{i,j}$ are fitted participant-specific parameters with multivariate normal prior

$$\boldsymbol{z_i} \sim \boldsymbol{n}(0, \boldsymbol{C}) \qquad (8)$$

Here the vector $\boldsymbol{z_i}$ has three dimensions and elements $\{z_{i,1}, z_{i,2}, z_{i,3}\}$, $\boldsymbol{n}(0, \boldsymbol{C})$ is a three-dimensional multivariate normal distribution with zero mean and a standard deviation of one for each component, and $\boldsymbol{C}$ is a correlation matrix to be estimated. We use "LKJ" priors with a Cholesky factor representation[18] when estimating $\boldsymbol{C}$, with prior $\boldsymbol{C} \sim lkj\_corr\_cholesky(1)$**.** Estimating $\boldsymbol{C}$ allows us to examine whether there is within-group correlation between peak viral titre, viral growth rather, and viral decline rate; e.g. whether individuals with faster growth rates also have higher than average peak viral titres.

The $\mu_{i,k}$ are the group mean values of $\theta_{i,\cdot}$, and δ_i (³0, by definition) are the standard deviations of individual $\theta_{i,j}$ values within each group. The δ_i were assumed to be the same for all groups since numbers of participants in some groups were insufficient to allow independent group-specific $\delta_{i,k}$ to be reliably estimated.

When fitting the single group model ($k = 1$ for all j) for the group mean parameters we assumed the following weakly informative priors (noting $i = 1$ corresponds to peak viral load/ml on a natural log scale, while $i = 2$ and 3 correspond to the natural log of viral growth rate and decline rate per day, respectively) for all groups, informed by previously reported viral kinetic profiles[19–21]:

$$\mu_{1,k} \sim n(15, 15) \quad \mu_{2,k} \sim n(1, 1.4) \quad \mu_{3,k} \sim n(0.5, 1.4) \qquad (9)$$

When fitting the four-group model, the limited numbers of subjects in some groups meant that Markov Chain Monte-Carlo (MCMC) mixing was poor with the prior on $\mu_{2,k}$ given above (since only a minority of participants have viral growth observed in the measured Ct profiles), so we used a more informative prior based on the one-group model posterior estimates for the growth rate parameter:

$$\mu_{2,k} \sim n(0.9, 0.2) \qquad (10)$$

For model comparison purposes, the single group model was also refitted with this prior for growth rate.

For both the one and four-group models, the group level standard deviations δ_i were given normal priors truncated below at 0 (represented by $n_{\geq 0}$) of

$$\delta_1 \sim n_{\geq 0}(0, 10) \quad \delta_2 \sim n_{\geq 0}(0, 1) \quad \delta_3 \sim n_{\geq 0}(0, 1) \qquad (11)$$

Time τ was defined on a participant-specific basis such that $\tau = 0$ was the day of the lowest Ct value (highest x) measured. However, the peak viral load may not have been observed, particularly since for most participants, we only observed the decline phase of their viral load trajectories. We therefore fitted τ_{max} for each participant, with prior $n(0, 4)$ and time being measured in days.

The error probability p was fitted on a log scale such that $(-\log p) \sim n_{\geq 0}(5, 2)$ (*i.e.* giving a relatively uninformative truncated [above at 1] lognormal prior for p with mean 0.04 and standard deviation 0.12) while x_0 was given a prior of $n(0, 1)$ (giving a 1:1 ratio of false positives and negatives). Both σ_0 and σ_v were given the relatively uninformative prior of $n_{\geq 1}(3, 3)$, where the minimum lower bound of 1 prevented MCMC divergence issues associated with exploring very low values of these measurement precision related parameters.

Initial model selection was made using leave-one-out cross-validation[1] to compare models with a one-group or four group (pre-alpha, alpha, delta-UV, delta-FV) hierarchical model with or without fitted correlation coefficients between individual-level parameters determining peak VL and VL growth and decline rates, and with informative or non-informative priors on VL growth rates. This selected the four-group model with fitted correlation coefficients (Table S4). However, resulting participant-specific estimates of peak VL (but not growth and decline rates) showed a marked and significant correlation with age (the correlation coefficient being highest with log(age)) in exploratory analysis. This motivated examination of models where mean peak VL could vary with age. Specifically, we examined models where peak VL was proportional to the log of age:

$$v_{max,j} = \exp\left(\theta_{1,j} + \nu_a \log\left(\frac{a_j}{50}\right)\right) \qquad (12)$$

Here a_j is the age of participant j, and ν_a is a slope parameter which was given prior $n(0, 0.5)$. We also examined four group models where $\theta_{2,j}$ and $\theta_{3,j}$ but not $\theta_{1,j}$ varied by group (*i.e.* $\mu_{1,k} = \mu_1$). We note that form of equation 12 implies that $\mu_{1,k}$ gives the population mean estimate of log peak VL for 50-year-olds.

The most predictive model allowed $\mu_{2,k}$ and $\mu_{3,k}$ to vary across the four groups, with μ_1 common to all groups but with peak VL varying as given in equation 12 (Table S4). Models with variation in peak VL by group but not age had more fitted parameters and were marginally less predictive, as judged by the expected log pointwise predictive density (ELPD). Models where peak VL varied with both group and age were less predictive and showed poorer convergence. This suggests some confounding between VOC-variant group and age, driven by UV delta cases, all but four of which were under 18 years old. These under-18 cases were the youngest cases in our dataset, bar one pre-alpha child case.

We computed group-level population means of log peak viral titre, viral growth rate and viral decline rate:

$$\ln(V_{max,k}) = \mu_{1,k}$$

$$A_k = \exp\left(\mu_{2,k} + {\delta_2}^2/2\right) \qquad (13)$$

$$B_k = \exp\left(\mu_{3,k} + {\delta_3}^2/2\right)$$

Note the equations for A_k and B_k reflect the lognormal distribution for $\exp\left(\theta_{i,j}\right)$ implied by equations 7 and 8.

We also compute within-sample means of log peak viral titre, viral growth rate and viral decline rate across the n_k participants within each group k:

$$\overline{\ln(v_{max,k})} = \sum_{j|k(j)=k} \theta_{1,j}/n_k$$

$$\overline{a_k} = \sum_{j|k(j)=k} \exp(\theta_{2,j})/n_k \qquad (14)$$

$$\overline{b_k} = \sum_{j|k(j)=k} \exp(\theta_{3,j})/n_k$$

The variables defined in equation 13 are able to be computed for the four variant/vaccination status groups even for the single group model which does not allow for between-group variation in the $\mu_{i,k}$ parameters.

We computed posterior probabilities that the variables defined in equations 13 and 14 were larger for one group than another.

Modelling was conducted in R version 4.0.3[22] using the package RStan[23] to fit the models using Hamiltonian MCMC methods. Leave-one-out cross validation[1] with moment-matching using the R package loo[2] was used for model comparison. For each model fit, 10 MCMC chains of 7,000 iterations were undertaken. The first 3,000 iterations of each chain were used for equilibration, and chains were thinned by half, giving a total of 20,000 posterior samples for each model. Standard Stan diagnostics were used to confirm convergence, mixing and adequate effective sample sizes.

6. Computer programs

Modelling was conducted in R version 4.1.0[22] using the package RStan[23] to fit the models using Hamiltonian MCMC methods. Leave-one-out cross validation[1] using the R package loo[2] was used for model comparison. Demographic data was summarized using R version 4.1.0[22]. The 'stats' (v4.1.0)[22] and 'epikit' (v0.1.2)[24] packages were used to conduct Pearson's Chi-squared tests and calculate secondary attack rates, respectively.

Supplementary References

1 Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput 2017; 27: 1413–32.

2 Vehtari A, G.J., Magnusson M, Yao Y, Bürkner P, Paananen T GA. loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models:R package version 2.4.1. 2020. 2020. https://mc-stan.org/loo/.

3 China CDC. Primers and probes for detection 2019-nCoV. 2020. http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.htm (accessed March 30, 2021).

4 Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance 2020; 25. DOI:10.2807/1560-7917.ES.2020.25.3.2000045.

5 Wolters F, Coolen JPM, Tostmann A, et al. Novel SARS-CoV-2 Whole-genome sequencing technique using Reverse Complement PCR enables easy, fast and accurate outbreak analysis in hospital and community settings. bioRxiv 2020; : 2020.10.29.360578.

6 Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol 2017; 35: 316–9.

7 Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018; 34: i884–90.

8 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013; published online March 16. http://arxiv.org/abs/1303.3997 (accessed Sept 23, 2021).

9 Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. Gigascience 2021; 10. DOI:10.1093/gigascience/giab008.

10 Wilm A, Aw PPK, Bertrand D, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res 2012; 40: 11189–201.

11 Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 2018; 34: 867–8.

12 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010; 26: 841–2.

13 Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) 2012; 6: 80–92.

14 Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016; 32: 3047–8.

15 Ferguson NM, Kien DT, Clapham H, et al. Modeling the impact on virus transmission of Wolbachia-mediated blocking of dengue virus infection of Aedes aegypti. Sci Transl Med 2015; 7: 279ra37.

16 Ke R, Zitzmann C, Ribeiro RM, Perelson AS. Kinetics of SARS-CoV-2 infection in the human upper and lower respiratory tracts and their relationship with infectiousness. Preprint 2020. https://www.medrxiv.org/content/10.1101/2020.09.25.20201772v1.full.pdf.

17 Kissler SM, Fauver JR, Mack C, et al. Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies. PLoS Biol 2021; 19: e3001333.

18 Stan Development Team. Stan function reference: Correlation Matrix Distributions. https://mc-stan.org/docs/2_27/functions-reference/correlation-matrix-distributions.html.

19 He X, Lau EH, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. medRxiv 2020; : 2020.03.15.20036707.

20 To KK-W, Tsang OT-Y, Leung W-S, et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. Lancet Infect Dis 2020; published online March. DOI:10.1016/S1473-3099(20)30196-1.

21 Wölfel R, Corman VM, Guggemos W, et al. Virological assessment of hospitalized patients with COVID-2019. Nature 2020; 581: 465–9.

22 R Core Team. R: A Language and Environment for Statistical Computing. 2020. https://www.r-project.org/.

23 Stan Development Team. RStan: the R interface to Stan, R package version 2.21.2. 2020. http://mc-stan.org/.

24 Schumacher ZNK and D. epikit: Miscellaneous Tools for the ‘R4Epis’ Project. 2020. https://cran.r-project.org/package=epikit.