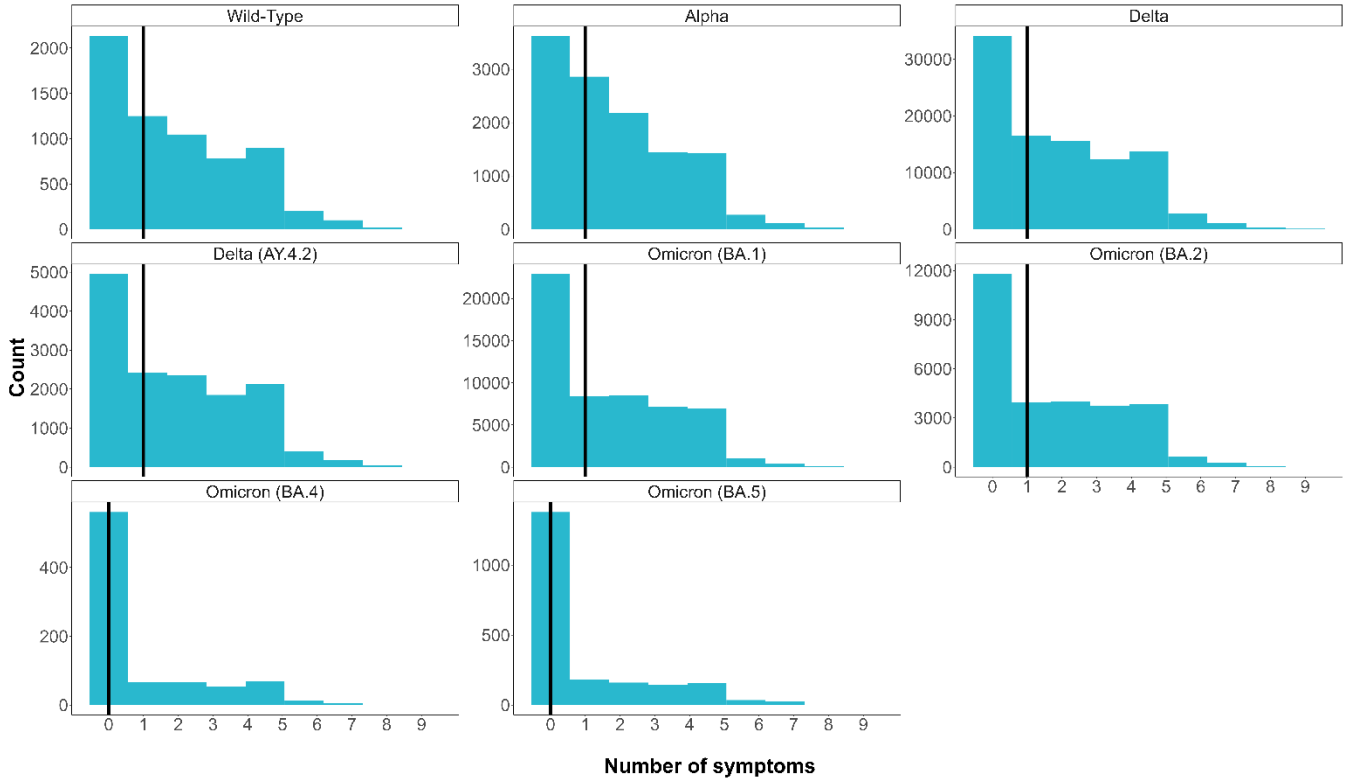
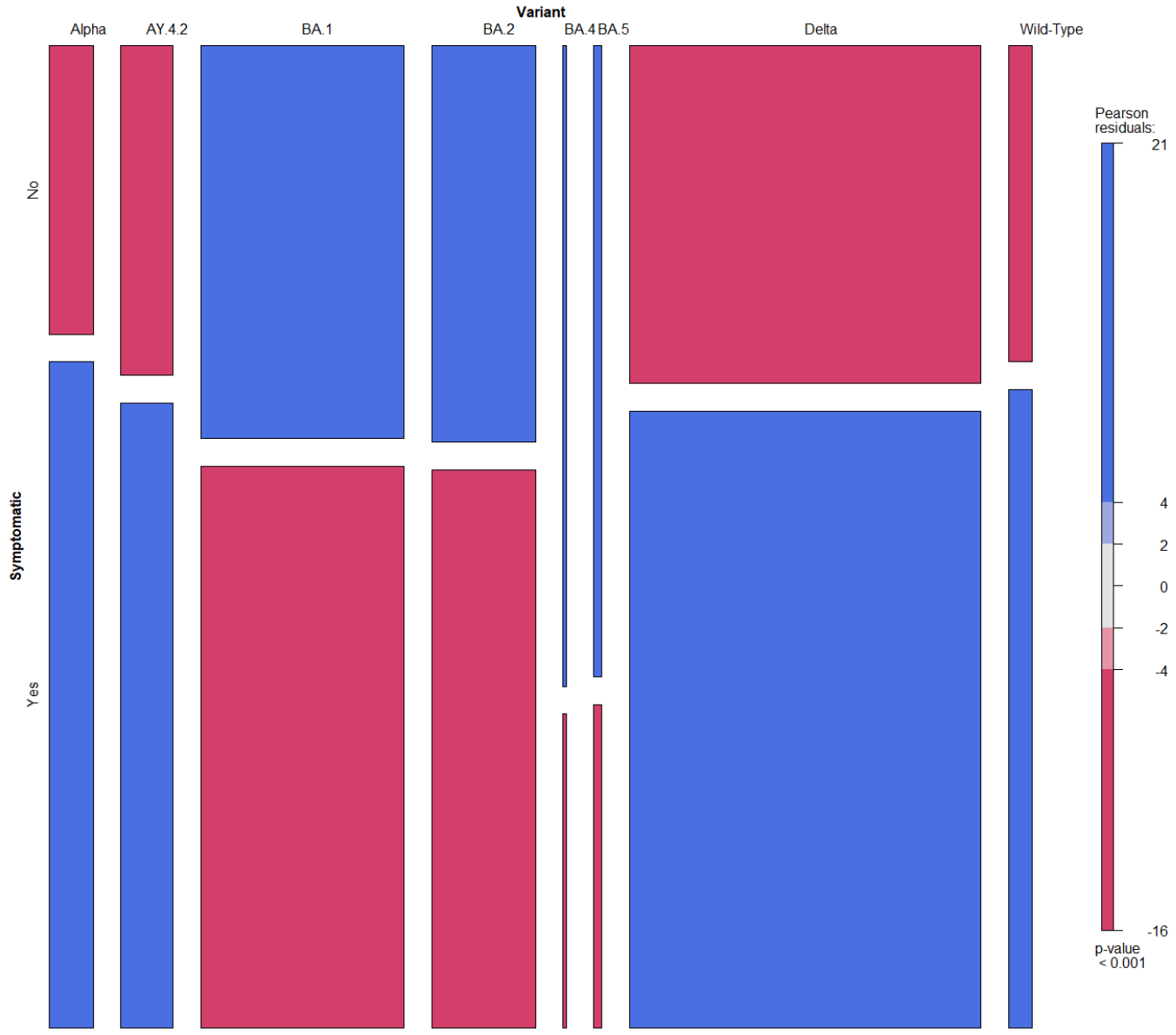
Supporting Information

Additional overall symptomatic status descriptive results



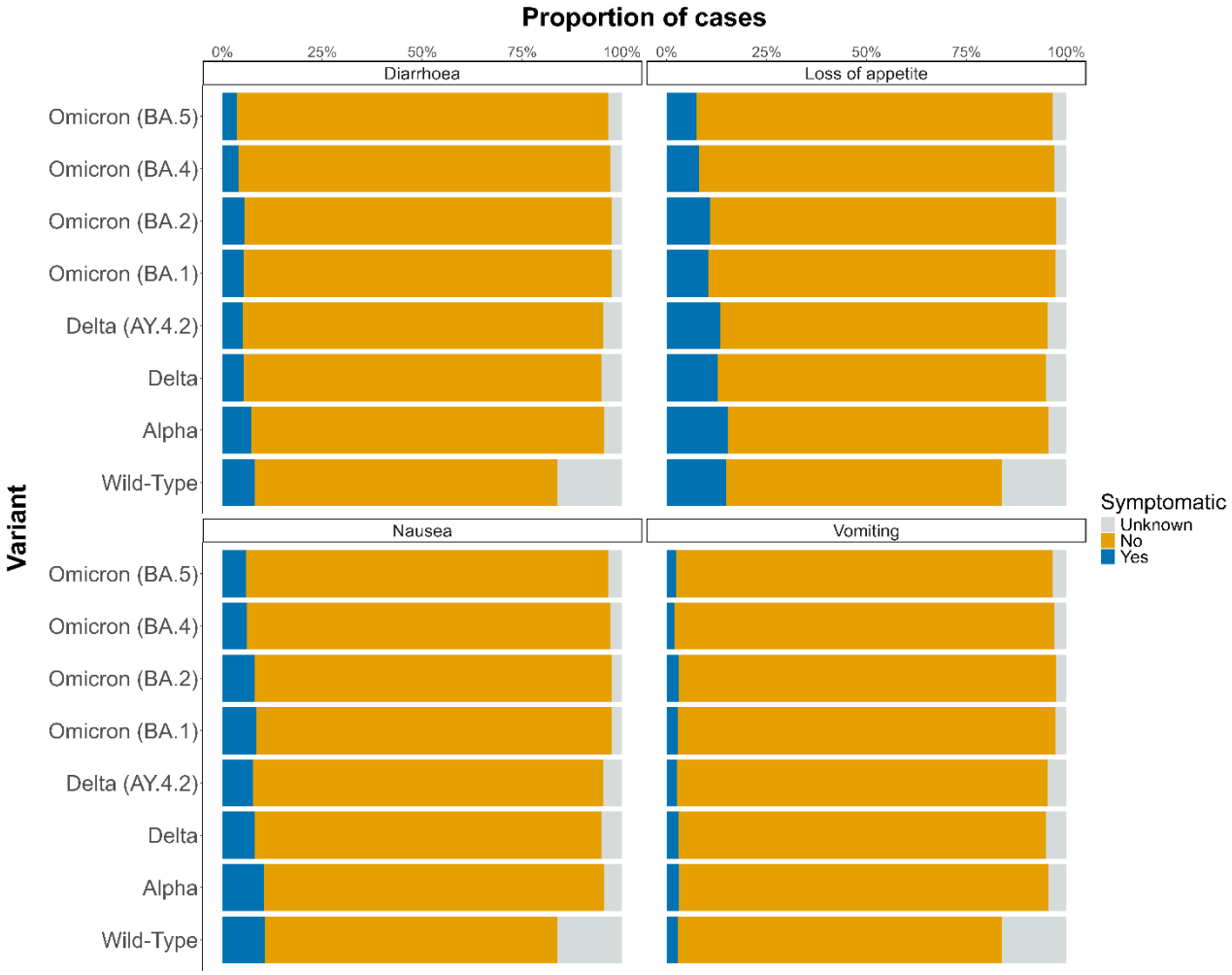
Supporting Information Figure 1. Histogram of distribution of total number of symptoms reported per case faceted by SARS-CoV-2 variant. In each panel, the vertical black line indicates the median number of symptoms reported for a given variant. Abbreviations: SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2.



Supporting Information Figure 2. Mosaic plot of symptomatic status vs. SARS-CoV-2 variant. Abbreviations: SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2. The area of the rectangular tiles represents the conditional relative frequency for a cell in the contingency table of variant vs. symptomatic status. Each tile is coloured to show the deviation from the expected frequency (residual) of cases for a given symptomatic status for a given variant, from a Pearson X^2. Red tiles indicate significant negative residuals, where the frequency of cases which are symptomatic/asymptomatic for a given variant is less than expected, while blue tiles indicate significant positive residuals where the frequency of cases which are symptomatic/asymptomatic for a given variant is greater than expected. The intensity of the colour represents the magnitude of the residual. The vertical length of the bars is proportional to the number of observations in symptomatic status within each level of

variant. Cases with an unknown symptomatic status were excluded from the above visualisation.

Additional symptom-specific descriptive results



Supporting Information Figure 3. Proportion of cases in which less commonly reported symptom of interest was reported by SAR-CoV-2 variant. Less common symptoms included diarrhoea, nausea, vomiting and loss of appetite. Abbreviations: SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2

Additional symptom-specific modelling results

Diarrhoea

The odds ratios derived from the fitted diarrhoea status model are reported in Supporting Table 1. A likelihood ratio test confirmed that including variant as a model predictor yielded an improved model fit ($\chi 2$ (7) = 157.14, $p \leq 0.001$).

Supporting Table 1. Odds ratios, 95% confidence intervals and p-values for the fitted diarrhoea status model. 12,388/214,796 cases included in the model reported diarrhoea as a symptom. AIC = 93,180.

Characteristic	OR[1]	95% CI[1]	p-value
Variant			
Wild-Type	—	—	
Alpha	0.71	0.63, 0.80	**<0.001**
Delta	0.62	0.55, 0.69	**<0.001**
Delta (AY.4.2)	0.58	0.51, 0.66	**<0.001**
Omicron (BA.1)	0.54	0.48, 0.61	**<0.001**
Omicron (BA.2)	0.55	0.48, 0.62	**<0.001**
Omicron (BA.4)	0.38	0.26, 0.54	**<0.001**
Omicron (BA.5)	0.35	0.27, 0.45	**<0.001**
Vaccination status			
Unvaccinated	—	—	
Vaccinated	0.81	0.77, 0.86	**<0.001**
Unknown vaccination status	0.84	0.77, 0.92	**<0.001**
Age group			
<5	—	—	
5-14	0.33	0.28, 0.38	**<0.001**
15-29	0.50	0.43, 0.58	**<0.001**
30-59	0.73	0.64, 0.85	**<0.001**
60+	0.57	0.49, 0.66	**<0.001**

Characteristic	OR[1]	95% CI[1]	p-value
Sex			
Female	—	—	
Male	0.86	0.83, 0.89	**<0.001**
Keyworker status			
No	—	—	
Yes	1.49	1.43, 1.56	**<0.001**
Reinfection			
No	—	—	
Yes	0.86	0.71, 1.04	0.13

[1]OR = Odds Ratio, CI = Confidence Interval

Vomiting

The odds ratios derived from the fitted vomiting status model are reported in Supporting Table 2. Likelihood ratio test confirmed that including variant as a model predictor yielded an improved model fit ($\chi 2$ (7) = 28.47, $p \leq 0.001$).

Supporting Table 2. Odds ratios, 95% confidence intervals and p-values for the fitted vomiting status model. 6,388/214,796 cases included in the model reported vomiting as a symptom. AIC = 56,583.

Characteristic	OR[1]	95% CI[1]	p-value
Variant			
Wild-Type	—	—	
Alpha	0.86	0.71, 1.04	0.12
Delta	0.91	0.77, 1.09	0.3
Delta (AY.4.2)	0.79	0.65, 0.96	**0.018**
Omicron (BA.1)	0.92	0.77, 1.10	0.3
Omicron (BA.2)	1.06	0.89, 1.28	0.5
Omicron (BA.4)	0.77	0.45, 1.25	0.3
Omicron (BA.5)	0.91	0.65, 1.25	0.6
Vaccination status			

Characteristic	OR[1]	95% CI[1]	p-value
Unvaccinated	—	—	
Vaccinated	0.62	0.57, 0.66	**<0.001**
Unknown vaccination status	0.79	0.70, 0.89	**<0.001**
Age group			
<5	—	—	
5-14	0.53	0.45, 0.64	**<0.001**
15-29	0.47	0.40, 0.56	**<0.001**
30-59	0.41	0.34, 0.48	**<0.001**
60+	0.32	0.27, 0.38	**<0.001**
Sex			
Female	—	—	
Male	0.71	0.68, 0.75	**<0.001**
Keyworker status			
No	—	—	
Yes	1.21	1.13, 1.30	**<0.001**
Reinfection			
No	—	—	
Yes	0.65	0.49, 0.85	**0.002**

[1]OR = Odds Ratio, CI = Confidence Interval

Nausea

The odds ratios derived from the fitted nausea status model are reported in Supporting Table 3. Likelihood ratio test confirmed that including variant as a model predictor yielded an improved model fit ($\chi 2$ (7) = 121.77, $p \leq 0.001$).

Supporting Table 3. Odds ratios, 95% confidence intervals and p-values for the fitted nausea status model. 18,737/214,796 cases included in the model reported nausea as a symptom. AIC = 124,269.

Characteristic	OR[1]	95% CI[1]	p-value
Variant			
Wild-Type	—	—	
Alpha	0.77	0.69, 0.85	**<0.001**
Delta	0.63	0.57, 0.70	**<0.001**
Delta (AY.4.2)	0.59	0.53, 0.67	**<0.001**
Omicron (BA.1)	0.63	0.57, 0.70	**<0.001**
Omicron (BA.2)	0.61	0.55, 0.68	**<0.001**
Omicron (BA.4)	0.49	0.36, 0.66	**<0.001**
Omicron (BA.5)	0.49	0.39, 0.59	**<0.001**
Vaccination status			
Unvaccinated	—	—	
Vaccinated	0.77	0.74, 0.81	**<0.001**
Unknown vaccination status	0.80	0.74, 0.86	**<0.001**
Age group			
<5	—	—	
5-14	4.75	3.67, 6.25	**<0.001**
15-29	5.49	4.26, 7.22	**<0.001**
30-59	5.31	4.12, 6.98	**<0.001**
60+	3.09	2.39, 4.08	**<0.001**
Sex			
Female	—	—	
Male	0.61	0.59, 0.63	**<0.001**
Keyworker status			
No	—	—	
Yes	1.46	1.41, 1.52	**<0.001**
Reinfection			

Characteristic	OR[1]	95% CI[1]	p-value
No	—	—	
Yes	0.74	0.63, 0.86	**<0.001**

[1]OR = Odds Ratio, CI = Confidence Interval

Appetite loss

The odds ratios derived from the fitted loss of appetite status model are reported in Supporting Table 4. Likelihood ratio test confirmed that including variant as a model predictor yielded an improved model fit ($\chi 2$ (7) = 497.42, $p \leq 0.001$).

Supporting Table 4. Odds ratios, 95% confidence intervals and p-values for the fitted loss of appetite status model. 27,168/214,796 cases included in the model reported loss of appetite as symptom. AIC = 160,813.

Characteristic	OR[1]	95% CI[1]	p-value
Variant			
Wild-Type	—	—	
Alpha	0.82	0.75, 0.90	**<0.001**
Delta	0.78	0.72, 0.85	**<0.001**
Delta (AY.4.2)	0.83	0.76, 0.92	**<0.001**
Omicron (BA.1)	0.59	0.54, 0.64	**<0.001**
Omicron (BA.2)	0.62	0.56, 0.67	**<0.001**
Omicron (BA.4)	0.45	0.35, 0.58	**<0.001**
Omicron (BA.5)	0.42	0.35, 0.50	**<0.001**
Vaccination status			
Unvaccinated	—	—	
Vaccinated	0.74	0.71, 0.76	**<0.001**
Unknown vaccination status	0.83	0.78, 0.89	**<0.001**
Age group			
<5	—	—	

Characteristic	OR[1]	95% CI[1]	p-value
5-14	0.75	0.67, 0.84	**<0.001**
15-29	0.95	0.85, 1.06	0.3
30-59	0.97	0.87, 1.08	0.6
60+	0.82	0.73, 0.92	**<0.001**
Sex			
Female	—	—	
Male	0.75	0.73, 0.77	**<0.001**
Keyworker status			
No	—	—	
Yes	1.51	1.46, 1.56	**<0.001**
Reinfection			
No	—	—	
Yes	0.70	0.60, 0.81	**<0.001**

[1]OR = Odds Ratio, CI = Confidence Interval

Additional overall symptomatic status modelling results

Interaction model

A complementary binary logistic regression analysis was used to model the relationship between overall symptom status (symptomatic versus asymptomatic) and variant adjusted for: cases' age group, sex, vaccination status, keyworker status, reinfection status and the interaction between vaccination status and variant. Results are reported as adjusted odds ratios with 95% confidence intervals and AIC. Cases whose age, sex, keyworker status or symptom status were unknown were excluded from the regression analyses. The same approach was also used to explore associations between specific symptoms and variant, adjusted for vaccination status, age group, sex, keyworker status, reinfection status and the vaccination status-variant interaction. Data for Alpha, Delta, Delta (AY.4.2), Omicron (BA.1), Omicron (BA.2), Omicron (BA.4) and Omicron (BA.5) cases were included in the models as some cases associated with each of these variants had received vaccination. Alpha was

therefore the reference variant for the analyses reported below, which are therefore not directly comparable with the models reported in the main text but are reported here for completeness. The glm() function in the stats package (v4.1.3, part of base R) was again used for regression modelling.

There was a significant reduction in the odds of being symptomatic across all variants, compared to those infected with Alpha, whilst holding the other model variables fixed, most notably for Omicron (BA.4) and Omicron (BA.5) (smallest OR = 0.12, smallest 95% CI = [0.07, 0.19], $p < 0.001$).

Supporting Table 5. Odds ratios, 95% confidence intervals and p-values for the fitted overall symptomatic status model. 129,751/208,404 cases included in the model reported being symptomatic. AIC =262,529.

Characteristic	OR[1]	95% CI[1]	p-value
Variant			
Alpha	—	—	
Delta	0.57	0.54, 0.60	**<0.001**
Delta (AY.4.2)	0.54	0.50, 0.59	**<0.001**
Omicron (BA.1)	0.52	0.48, 0.55	**<0.001**
Omicron (BA.2)	0.51	0.47, 0.55	**<0.001**
Omicron (BA.4)	0.11	0.06, 0.20	**<0.001**
Omicron (BA.5)	0.13	0.08, 0.20	**<0.001**
Vaccination status			
Unvaccinated	—	—	
Vaccinated	0.68	0.61, 0.76	**<0.001**
Unknown vaccination status	0.76	0.69, 0.85	**<0.001**
Age group			
<5	—	—	
5-14	0.85	0.78, 0.93	**<0.001**
15-29	0.77	0.71, 0.84	**<0.001**
30-59	0.70	0.65, 0.76	**<0.001**

Characteristic	OR[1]	95% CI[1]	p-value
60+	0.43	0.40, 0.47	**<0.001**
Sex			
Female	—	—	
Male	0.90	0.88, 0.91	**<0.001**
Keyworker status			
No	—	—	
Yes	3.30	3.21, 3.40	**<0.001**
Reinfection			
No	—	—	
Yes	0.80	0.73, 0.87	**<0.001**
Variant / vaccination status interaction			
Delta * Vaccinated	2.09	1.87, 2.34	**<0.001**
Delta (AY.4.2) * Vaccinated	2.30	2.02, 2.62	**<0.001**
Omicron (BA.1) * Vaccinated	1.38	1.22, 1.54	**<0.001**
Omicron (BA.2) * Vaccinated	1.42	1.25, 1.61	**<0.001**
Omicron (BA.4) * Vaccinated	2.39	1.30, 4.62	**0.007**
Omicron (BA.5) * Vaccinated	2.13	1.33, 3.52	**0.002**
Delta * Unknown vaccination status	1.27	1.12, 1.44	**<0.001**
Delta (AY.4.2) * Unknown vaccination status	1.14	0.92, 1.41	0.2
Omicron (BA.1) * Unknown vaccination status	1.01	0.88, 1.16	0.9
Omicron (BA.2) * Unknown vaccination status	1.14	0.97, 1.35	0.12
Omicron (BA.4) * Unknown vaccination status	0.80	0.26, 2.30	0.7
Omicron (BA.5) * Unknown vaccination status	1.44	0.78, 2.69	0.2

[1]OR = Odds Ratio, CI = Confidence Interval

Interrupted time sensitivity model

The purpose of this sensitivity analysis was to determine if the effect of variant on the odds of being symptomatic across the entire study period (over which testing policies were highly variable) that is reported within the main manuscript, remained when restricting the analysis to a period of more stable community testing. Binary logistic regression was again used to model the relationship between overall symptom status (symptomatic versus asymptomatic) and variant adjusted for: cases' age group, sex, vaccination status, keyworker status and reinfection status. Cases whose age, sex, keyworker status or symptom status were unknown were again excluded from the regression analyses. The results are reported as adjusted odds ratios with 95% confidence intervals and AIC. The study data was filtered for cases whose PCR tests were conducted between 31st December 2020 and 31st December 2021, a period of relatively stable community testing. Alpha, Delta, Delta (AY.4.2) and Omicron (BA.1) were included in the model. Omicron (BA.2) was excluded due to low case numbers. The glm() function in the stats package (v4.1.3, part of base R) was again used for regression model fitting.

Whilst not directly comparable to the model reported in the main text due to different reference variants etc, the pattern of a significant reduction of odds, particularly so with Omicron (BA.1), remained when the analysis was repeated with the restricted time period. Indeed, the odds of being symptomatic were significantly lower for all variants compared to Alpha, when holding all other model variables fixed. Most notably, for Omicron (BA.1) (OR = 0.43, 95% CI = [0.41, 0.46], $p < 0.001$).

Supporting Table 6. Odds ratios, 95% confidence intervals and p-values for the fitted overall symptomatic status model between 31st December 2020 and 31st December 2021. 86,046/133,495 cases included in the model reported being symptomatic. AIC = 168,313.

Characteristic	OR[1]	95% CI[1]	p-value
Variant			
Alpha	—	—	
Delta	0.74	0.71, 0.78	**<0.001**
Delta (AY.4.2)	0.77	0.72, 0.81	**<0.001**
Omicron (BA.1)	0.43	0.41, 0.46	**<0.001**

Characteristic	OR[1]	95% CI[1]	p-value
Vaccination status			
Unvaccinated	—	—	
Vaccinated	1.22	1.18, 1.26	**<0.001**
Unknown vaccination status	0.84	0.79, 0.90	**<0.001**
Age group			
<5	—	—	
5-14	0.74	0.67, 0.82	**<0.001**
15-29	0.87	0.78, 0.96	**0.005**
30-59	0.76	0.68, 0.84	**<0.001**
60+	0.52	0.47, 0.58	**<0.001**
Sex			
Female	—	—	
Male	0.90	0.88, 0.92	**<0.001**
Keyworker status			
No	—	—	
Yes	2.90	2.78, 3.02	**<0.001**
Reinfection			
No	—	—	
Yes	0.49	0.40, 0.60	**<0.001**

[1]OR = Odds Ratio, CI = Confidence Interval

Multiple imputations by chained equations model

As noted in the main text, some observations were excluded from our analyses due to missing data, including in our outcome variable. Multiple imputation is one flexible approach to imputing missing data. Whilst we note that some authors argue that following imputation, those cases with imputed outcomes have nothing further to contribute to subsequent analyses as random variation in the imputed response only adds noise to estimates (e.g.,

von Hippel, 2007), we performed a missing data sensitivity analysis to explore any potential impact of missing outcome as well as predictor data on the findings from our overall symptomatic status model. Missing data in both the outcome variable, symptomatic status, and the predictor variables, keyworker status (both of these variables are drawn from the TTP contact tracing dataset so missingness co-occurs across these variables), sex and age, were handled using multiple imputation by chained equations (MICE) implemented in the mice package in R (v 3.16.0) (van Buuren & Groothuis-Oudshoorn, 2011). Five imputed datasets (m = 5) were generated using logistic regression for binary variables, and multinomial logistic regression for categorical variables with more than two levels (i.e., age group). Imputations were based on all variables included in the subsequent analysis model, under the assumption that data were missing at random (MAR), a weaker assumption than the missing completely at random (MCAR) assumption underlying a complete case analysis. As per the approach described in the main text, each imputed dataset was then analysed using a binary logistic regression model estimated using the glm() function in the stats package (v4.1.3, part of base R). Overall symptomatic status was regressed on variant, age group, sex, vaccination status, keyworker status and reinfection status. The results were pooled across imputations using Rubin's rules, and are reported below as adjusted odds ratios with 95% confidence intervals and the model AIC (the range across imputations). All eight variants were included in the model. The results of this sensitivity analysis were almost identical to the model reported in the main manuscript text, with only very minor changes in the magnitudes of odds of being symptomatic. The odds of being symptomatic were significantly lower for all variants, with the exception of Alpha, compared to Wild-Type, when holding all other model variables fixed, most notably for Omicron (BA.4) and Omicron (BA.5) (smallest OR = 0.16, 95% CI = [0.13, 0.19], $p < 0.001$).

Supporting Table 7. Odds ratios, 95% confidence intervals and p-values for the fitted overall symptomatic status model where multiple imputation of chained equations has been utilised to impute missing data for age, sex, symptom status and keyworker status. AIC range: 284,622.7 – 284,859.8.

Characteristic	OR[1]	95% CI[1]	p-value
Variant			
Wild-Type	—	—	
Alpha	0.95	0.88, 1.03	0.2
Delta	0.68	0.63, 0.73	**<0.001**
Delta (AY.4.2)	0.69	0.64, 0.74	**<0.001**

Characteristic	OR[1]	95% CI[1]	p-value
Omicron (BA.1)	0.46	0.43, 0.49	**<0.001**
Omicron (BA.2)	0.46	0.43, 0.49	**<0.001**
Omicron (BA.4)	0.16	0.13, 0.19	**<0.001**
Omicron (BA.5)	0.17	0.15, 0.19	**<0.001**
Vaccination status			
Unvaccinated	—	—	
Vaccinated	1.21	1.17, 1.24	**<0.001**
Unknown vaccination status	0.87	0.83, 0.91	**<0.001**
Age group			
<5	—	—	
5-14	0.81	0.75, 0.88	**<0.001**
15-29	0.75	0.69, 0.82	**<0.001**
30-59	0.70	0.65, 0.76	**<0.001**
60+	0.41	0.38, 0.45	**<0.001**
Sex			
Female	—	—	
Male	0.90	0.88, 0.92	**<0.001**
Keyworker status			
No	—	—	
Yes	3.22	3.13, 3.31	**<0.001**
Reinfection			
No	—	—	
Yes	0.82	0.75, 0.90	**<0.001**

[1]OR = Odds Ratio, CI = Confidence Interval

Mixed effects model

A complementary mixed-effects logistic regression model was used to model the relationship between overall symptom status (symptomatic versus asymptomatic) and variant adjusted for: cases' age group, sex, vaccination status, keyworker status and reinfection status (all included as fixed effects), with a random intercept for individual-specific reference codes, thereby accounting for nonindependence of observations in the dataset due to reinfection. The mixed effects model was estimated using the glmer function from the lme4 package in R (v1.1-31) (Bates et al., 2015). Estimation was performed using the Laplace approximation (`nAGQ = 0`) with the "bobyqa" optimizer to improve convergence. Cases whose age, sex, symptomatic or keyworker status were unknown were excluded. Results are reported below as adjusted odds ratios (conditional on the random intercept) with 95% confidence intervals and AIC. The standard deviation (SD), variance and Intraclass-correlation coefficient (ICC) are also reported for the random intercept component of this model.

Consistent with the results of the corresponding fixed effects-only model reported in the main manuscript text, the odds of being symptomatic were significantly lower for all variants - other than Alpha - compared to Wild-Type, when holding all other model variables fixed and accounting for reinfection in specific individuals, most notably for Omicron (BA.4) and Omicron (BA.5) (smallest OR = 0.16, 95% CI = [0.14, 0.20], $p < 0.001$). Crucially, however, the AIC for this mixed effect model slightly exceeded that of the equivalent fixed effects-only analysis reported in the main text (270,487.4 versus 270,486), such that the latter is preferred. Indeed, additional mixed effects models were also estimated for the nine specific symptoms and in all cases, the observed AIC exceeded that of the corresponding fixed effects-only models. The results of the overall symptomatic mixed effects model are nevertheless reported below by way of example and for reader interest.

Supporting Table 8. Conditional odds ratios, 95% confidence intervals and p-values for the fitted overall symptomatic status model with a random intercept for individual-specific reference codes. 134,031/214,796 cases included in the model reported being symptomatic. AIC = 270,487.4. SD (intercept) = 0.089; variance = 0.008; ICC = 0.002.

Characteristic	OR[1]	95% CI[1]	p-value
Variant			
Wild-Type	—	—	
Alpha	0.98	0.91, 1.06	0.6
Delta	0.70	0.66, 0.75	**<0.001**

Characteristic	OR[1]	95% CI[1]	p-value
Delta (AY.4.2)	0.71	0.66, 0.77	**<0.001**
Omicron (BA.1)	0.47	0.44, 0.51	**<0.001**
Omicron (BA.2)	0.48	0.44, 0.51	**<0.001**
Omicron (BA.4)	0.16	0.14, 0.20	**<0.001**
Omicron (BA.5)	0.17	0.15, 0.19	**<0.001**
Vaccination status			
Unvaccinated	—	—	
Vaccinated	1.21	1.17, 1.24	**<0.001**
Unknown vaccination status	0.91	0.87, 0.95	**<0.001**
Age group			
<5	—	—	
5-14	0.84	0.78, 0.92	**<0.001**
15-29	0.78	0.72, 0.85	**<0.001**
30-59	0.73	0.67, 0.79	**<0.001**
60+	0.43	0.39, 0.46	**<0.001**
Sex			
Female	—	—	
Male	0.90	0.88, 0.92	**<0.001**
Keyworker status			
No	—	—	
Yes	3.21	3.12, 3.30	**<0.001**
Reinfection			
No	—	—	
Yes	0.82	0.76, 0.90	**<0.001**

[1]OR = Odds Ratio, CI = Confidence Interval

References

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/JSS.V067.I01

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/JSS.V045.I03