Supplementary material 1: Protection provided by covid-19 vaccines and previous infection against the Omicron variant

revised version

February 2, 2023

In this Supplementary material, we describe the data and the statistical methods we use in the paper.

Dataset

In total, our dataset contains 8,282,080 records of vaccinated and/or SARS-CoV-2 positive persons, out of which 10,937 were excluded for either inconsistency (e.g. invalid dates or wrong ordering of vaccine doses) or the lack of valid sex and/or age information. Additional 97,855 records were excluded as corresponding persons died before 7 December 2021, which is the start of our study period (34,145 were recorded to die from covid-19, 63,740 from other reasons; see Table S3).

As the source dataset consists only of those who were tested positive and/or were vaccinated, we added subjects who were neither tested positive nor vaccinated. In particular, we completed each sex-age category to the numbers reported by the Czech Statistical Office by December 31, 2020 – 10,701,777 inhabitants; consequently, our sample truly reflected the sex and age structure of the whole population, containing all the positive and/or vaccinated individuals. We neglected births and deaths of those added persons. Moreover, not all non-covid deaths were recorded as they are reported with a certain delay.

Methods: Cox regression with time-varying covariates

For the analyses of vaccine effectiveness and immunity protection, conducted separately for both the Omicron and Delta variants, we use the Cox proportional hazards model with time-varying covariates:

$$\lambda(t|\texttt{Immunity}_t, \texttt{AgeGr}, \texttt{Sex}) = \lambda_0(t) \exp\left\{\alpha\, \texttt{Immunity}_t + \beta\, \texttt{AgeGr} + \gamma\, \texttt{Sex}\right\}. \tag{1}$$

Here λ is the hazard function, λ_0 is the baseline hazard function, $\texttt{Immunity}_t$ is a vector of dummy variables, describing the immunity status of an individual at a time instant t, `AgeGr` is a vector of dummy variables determining the age category, and `Sex` is a dummy variable describing sex of an individual. In line with [**?**], we take time t as absolute, i.e. count it in days since the study start (7 December 2021 is considered day 0). A doubtless advantage of this approach is that all real forces affecting the hazard, like infection incidence, presence and intensity of counter-epidemic measures or weather, can be captured by the baseline hazard function on which the estimates of hazard rates, eventually used for calculation of vaccine effectiveness or immunity protection, do not depend.

For example, consider a female person aged 49, infected one year before the study start and vaccinated three weeks before the study start. At the beginning of the study period, the above covariates look like this:

$$\texttt{Immunity}_0 = (0, \ldots, 1, \ldots, 0),$$

where the position of the unit corresponds to the category `inf6+_full2-` (see below for a set of possible immunity classes),

$$\texttt{AgeGr} = (0, \ldots, 1, \ldots, 0),$$

where the position of the unit corresponds to the age category `45-49` (see below for all considered age classes) and

$$\texttt{Sex} = (0, 1),$$

where 1 corresponds to female.

We take the value `_noimmunity` corresponding to no immunization (no vaccination and no previous infection) as a reference for $\texttt{Immunity}_t$,[1] so the hazard function of a non-immunized person is

$$\lambda(t|\texttt{\_noimmunity}, \texttt{AgeGr}, \texttt{Sex}) = \lambda_0(t) \exp\left\{\beta\, \texttt{AgeGr} + \gamma\, \texttt{Sex}\right\}.$$

Consequently, the hazard ratio of the i-th immunity status is

$$HR_i = \frac{\lambda(t|i, \texttt{AgeGr}, \texttt{Sex})}{\lambda(t|\texttt{\_noimmunity}, \texttt{AgeGr}, \texttt{Sex})} = \exp\{\alpha_i\},$$

where α_i in an element of the vector parameter α corresponding to the immunity vector having the unit at position i.

Instead of model (1), we use, in some analyses (see Table S1), models

$$\lambda(t|\texttt{VaccStatus}_t, \texttt{InfPrior}_t, \texttt{AgeGr}, \texttt{Sex}) = \lambda_0(t) \exp\left\{\mu\, \texttt{VaccStatus}_t + \nu\, \texttt{InfPrior}_t + \beta\, \texttt{AgeGr} + \gamma\, \texttt{Sex}\right\} \tag{2}$$

or

$$\lambda(t|\texttt{InfPrior}_t, \texttt{AgeGr}, \texttt{Sex}) = \lambda_0(t) \exp\left\{\nu\, \texttt{InfPrior}_t + \beta\, \texttt{AgeGr} + \gamma\, \texttt{Sex}\right\}. \tag{3}$$

Here $\texttt{VaccStatus}_t$ is a dummy vector describing the vaccination status of an individual at a time instant t and $\texttt{InfPrior}_t$ is a vector of dummy variables, indicating temporal distance from the previous infection. Reference states corresponding to these new variables are `_novacc` (no vaccination) and `_noinf` (no previous infection).

Outcomes

The studied outcomes are the following:

Variable coding name	Verbal description
`VariantInf`	Infection by a selected virus variant
`VariantHosp`	Infection by the virus variant needing hospitalization
`VariantOxygen`	Infection by the virus variant neding oxygen therapy
`VariantICU`	Infection by the virus variant that needs ICU treatment

As needing hospitalization / oxygen therapy / ICU treatment we regard only those infections, after which the corresponding treatment happened within 15 days after a positive test.[2]

We also consider some additional outcomes for the purpose of censoring (see section on censoring below):

`Infected`	Infection by any virus variant
`DeadByOther`	Death without covid-19 positivity

[1] We take the age group $40-44$ and male sex as references for `AgeGr` and `Sex`, respectively.

[2] We set this limit for "fair treatment" of all the cases – if no limit had been set, then the hazard of earlier cases would be higher, which among other things would distort a comparison between the virus variants (most Delta cases happened earlier than the Omicron ones). On the other hand, as we are interested in a comparison of hazards rather than their absolute values, the fact that we "lose" some cases of hospitalization does not matter as we lose them "on both sides" of the comparison. Moreover, in rare cases of repeated hospitalizations, our input data lack information whether oxygen therapy took place, hence we exclude cases with repeated hospitalizations from our oxygen therapy analyses; here we also rely on "fairness" of this restriction for all the categories and the fact that this is the case for about one per-cent of hospitalizations.

The Covariates

$\texttt{Immunity}_t$

At the time t, the covariate takes value ...

`_noimmunity` if the subject was neither vaccinated nor underwent (confirmed) infection by t,

`X_alone` the condition X exclusively happened by time t,

`X_Y` meaning that conditions X and Y happened by time t.

The conditions `X` and `Y` could be:

`part` partially vaccinated (after 14 days following the first dose application) but not 14 days or more after full vaccination

`full2-` fully vaccinated no earlier than two months ago (from day 14 to day $14 + 61 - 1 = 74$ after the second dose application) but not 7 days or more after a booster dose

`full2+` fully vaccinated more than two months ago (day 75 or earlier after the second dose application) but not 7 days or more after a booster dose

`boost2-` booster vaccine effective no more than two months ago (from day 7 to day $7 + 61 - 1 = 67$ after the booster dose application)

`boost2+` booster vaccine effective more than two months ago (day 68 or earlier from the booster dose application)

`inf6-` infection underwent no more than six months (180 days) ago

`inf6+` infection underwent more than six months (180 days) ago.

$\texttt{VaccStatus}_t$

The $\texttt{VaccStatus}_t$ is either `_novacc` (not vaccintaed by t) or takes the form `V_S` where

`V` denotes the vaccine (BNT162b2 by `P`-fizer/BioNTech, mRNA-1273 by `M`-oderna, ChAdOx1 nCoV-1 by `A`-straZeneca, and Ad26.CoV2.S by `J`-ohnson & Johnson)

`S` denotes the state of vaccination (part = partial, full = full, boost = booster):

- **`S` = part:** 1 = first 61 days after the dose takes effect (14 days after application), 2 = later
- **`S` = full:** 1 = first 61 days after the dose takes effect (14 days after application), 2 = next 61 days, 3=later
- **`S` = boost:** 1 = first 61 days after the dose takes effect (7 days after application), 2 = later

$\texttt{InfPrior}_t$

This variable can takes the values:

`_noinf` if no previous infection happende by t

`inf1` time t is 60–180 days after the last confirmed infection (the re-infection cannot happen during the first 60 days by definition)

`inf2` 181–301 days after the last confirmed infection

`inf3` 302–422 days after the last confirmed infection

`inf4+` more than 422 days after the last confirmed infection

`AgeGr`

This covariate indicates the age of an individual in years at the time of entering the database and can take values $0-11$, $12-15$, $16-17$, $18-24$, $25-29$, $\ldots$, $75-79$, and $80+$.

`Sex`

`Sex` may be either `M` (male) or `Z` (female).

Study Period

Time is measured in days and we consider 7 December 2021 as time zero. The end of the study period is 13 February 2022 if `VariantInf` is the outcome, or 29 January 2022 for the remaining, hospitalizations-related model outcomes. The reason for making the latter shorter is the need to distinguish, for all infections, whether they led to hospitalization / oxygen therapy / ICU admission within 15 days.

Censoring

In all our analyses, we censor an individual (i.e. withdraw her/him from the study at the censoring time) if either

- The subject is infected by a virus variant other than that of interest or the variant has not been determined (the `Infected` outcome happens without the `VariantInf` outcome happening)
- The subject is reported to die (`DeadByOther` outcome happens)[3]
- The subject is reported to obtain a booster dose by AstraZeneca or Johnson & Johnson (there is only a few such records)

The Analyses Performed

Table S1 summarizes all the analyses we conducted by means of the Cox proportional hazards regression model and gives references to their sample inputs, i.e. excerpts from true input data. All those analyses were conducted separately for the Delta and the Omicron variants.

Discussion

A question naturally arises whether a competing-risks analysis, including both virus variants we consider, would not be more appropriate than the one we use, considering each virus variant separately. Although there is no doubt the former is more general and, under ideal circumstances, certainly more suitable, we chose the latter for the following reasons:

- The former approach is standard, having been used e.g. in the widely respected paper [**?**] which, to some extent, guarantees that it is correct to it use for a similar task.
- Given a competing-risks approach, the number of covariates in a single analysis would be much larger, which could create problems with estimation.
- Given that we are able to explain susceptibility of an individual by age, sex, immunity and the baseline hazard sufficiently, it is not unrealistic to assume that, conditionally on the covariates and time, the confirmed infection by Delta, the confirmed infection by Omicron and the infection with an undetermined variant are statistically (conditionally) independent (i.e. their probabilities depend only on the covariates and time), which would justify censoring by other variant and undetermined infections (see e.g. [**?**], p. 38).

[3] Deaths due to covid-19 always occur after infection, so any such outcome is censored before infection.

A question also is whether the fact that a variant has not been determined for all the infections does not distort our analyses. The truth is that, due to this fact, the hazards of infection by a particular virus variant are underestimated, equal to $\lambda_{\text{actual}} = \lambda_{\text{observed}}/p$, where p is the probability that an infection will be tested for its variant. However, HR's are independent of p as p cancels out provided that p does not depend on the immunity state. We view this independence as reasonable, because, in Czechia, the fact that the variant is determined during a PCR test depends principally on capabilities of a particular laboratory performing the test, but not on individual characteristics.

During our analyses it often happened that there were no outcomes for some `Immunity` states, especially for those combining the post-vaccination and the post-infection immunity. Alhough absence of these states does not contradict the use of Cox regression model (they can happen the more probably the closer λ is to zero), the estimation procedures in such cases return possibly confusing estimates of the corresponding hazard ratios: small non-zero values, possibly very close to zero but with confidence intervals not containing zero. Once this happens for an outcome of interest, we always report such results as "100 %, (no case)". If this happens for unreported or control variables, then we group then (into a category `Interactions` in the Oxygen and ICU analyses, for instance) or we exclude the input records containing them. As these records do not contribute to the likelihood whether included, grouped or omitted (having no events, they appear only in the denominator, contributing zero if omitted or only negligibly otherwise), the remaining estimates do not depend on the way we handle them, while omitting or grouping decreases dimensionality of the estimation, and hence minimizing risks of numerical and/or convergence errors.

To reduce dimensionality, we also group and/or omit records containing categories which are rare in the sense that there is a very low number of input records containing them; this can be done as we can assume that their HR's will be primarily explained by "their" coefficients, not influencing neither other dummies' coefficients nor the common ones due to their rarity. To be sure about this we conducted numerous sensitivity analyses, with the result that grouping, excluding or even leaving a small number of these categories in the analyses did not change the estimates of the HR's of interest. As an example, we present an alternative Omicron-Infection analysis (Supplementary Material 2, Sec. 20) where the rare categories, forming about 1 per cent of records, are grouped into a category named `other` instead of omitting; note that the results coincide up to less than 0.01, a majority even more sharply.

It is worth noting that our analyses comfortably satisfy recommended 10 events per variable (see the discussion in [**?**]). The only exception is the Omicron-ICU analysis, which is however very close to the requirement of 10 events per variable (note also that [**?**] suggest relaxing the recommendation to 5).

It is also necessary to stress that, from the mathematical point of view, we approximate waning patterns of individual immunity sources by piece-wise linear functions with constancy intervals coinciding with the intervals defining the corresponding dummy variables. The values of these functions are given by corresponding $1 - HR's$ (one means full protection, zero means no protection). Clearly this is an approximation while the actual patterns are most likely continuous; however, estimation of models including continuity (e.g. with piece-wise linear approximations of the waning functions) with millions records as input, is beyond the computational capabilities of available statistical packages we are aware of.

Example

Here we comment on the sample input for the Infection analysis, listed in Table S5. Next we explain, how input records are created for individual subjects. These examples also illustrate how counting of days within the study period and counting of days since receiving vaccination or containing infection relate one to another, with the former counting time within the study period and the latter used to calculate immunity state and its potential changes during the study period.

Subject 1. A male, aged 73, has not been infected until time zero (7 December 2021), while receiving his second vaccine dose on 27 May 2021. So, he found himself in the `full2-_alone` state since 10 June 2021 (i.e. 14 days after the second dose application). At time zero, which is 180 days after 10 June 2021, he thus found himself in the `full2+_alone` state. On 14 December 2021, he received a booster dose and hence

moved to the boost2-_alone state 7 days later (at time 14). Until the end of the study period (13 February 2022, corresponding to time 68), he had not been infected, so he remained in this state until time 68.

Subject 15. A male, aged 58, was infected on 21 January 2021 and received his second dose on 16 June 2021, so he found himself in the inf6-_full2- state since 30 June 2021 (i.e. 14 days after the second dose application). At time zero, now 320 days after infection and 160 days after the second dose application, he was in the inf6+_full2+ state. He received a booster dose on 5 January 2021, so he jumped to the inf6+_boost2- state 7 days later (at time 36). Until the end of the study period (13 February 2022, corresponding to time 68), he had not been infected, so he remained in this state until time 68.

Subject 417. A female, aged 25, has not been infected nor vaccinated until time zero (7 December 2021), so she found herself in the _noimmunity state at time zero. She received her first dose on 2 December 2021, so she moved to the part_alone state on 16 December 2021 (i.e. 14 days after the first dose application, corresponding to 9 days after time zero). Then she received the second dose on 22 December 2021 and jumped to the full2-_alone state on 5 January 2022 (i.e. 14 days after the second dose application). Until the end of the study period (13 February 2022, corresponding to time 68), she had not been infected nor boosted, so she remained in this state until time 68.

Subject 968. A female, aged 57, has not been infected until time zero (7 December 2021) and she received the second dose on 14 June 2021, so she found herself in the full2-_alone state since 28 June 2021 (i.e. 14 days after the second dose application). At time zero, which is 160 days after 28 June 2021, she thus found herself in the full2+_alone state. On 18 January 2022, she received a booster dose, so she jumped to the boost2-_alone state 7 days later (at time 49). Consequently, 50 days after time zero, she was infected by a variant of interest (which is considered to be Omicron for this example). Therefore, at the second record in Table S5 corresponding to this subject there is Infected and VariantInf equal to 1.

Subject 2757. A male, aged 86, has not been infected until time zero (7 December 2021) and he received a booster dose on 11 November 2021, so he found himself in the boost2-_alone state since 18 November 2021 (i.e. 7 days after the booster dose application). At time zero, which is 19 days after 18 November 2021, he was still in the boost2-_alone state. He remained in this state until his death, which is recorded 24 days after time zero.

Subject 22209. A male, aged 35, has not been infected nor vaccinated until time zero (7 December 2021), so he found himself in the _noimmunity state at time zero. He is recorded to become infected on 9 December 2021 (2 days after time zero) by a variant other than that of interest (or alternatively by an undetermined variant), which leads to his withdrawn from the study at the time of his infection (he is censored 2 days after time zero).

Subject 391175. A female, aged 28, has not been vaccinated until time zero (7 December 2021) and is recorded to become infected on 8 July 2021, so she found herself in the inf6-_alone state since 8 July 2021. At time zero, which is 152 days after 8 July 2021, she was still in this state. On 5 January 2022, she jumped to the inf6+_alone state, since she underwent her last infection more than 180 ago and has not been vaccinated until that date. Consequently, 52 days after time zero, she was infected by a variant of interest (which is considered to be Omicron for this example).

Methods: Logistic regression

In theses analyses, we use the logistic regression model

$$\frac{\mathbb{P}[\text{outcome}|\texttt{Variant},\texttt{Immunity},\texttt{AgeGr},\texttt{Sex}]}{1-\mathbb{P}[\text{outcome}|\texttt{Variant},\texttt{Immunity},\texttt{AgeGr},\texttt{Sex}]} = \exp\left\{\alpha+\beta\,\texttt{Variant}+\gamma\,\texttt{Immunity}+\delta\,\texttt{AgeGr}+\eta\texttt{Sex}\right\},$$

Table S1: Details on analyses by Cox regression.

Analysis	Outcome	Covariates	Model	Sample
Infections	VariantInf	Immunity, Sex, AgeGr	Eq. (1)	Tab. S5
Protection by infection	VariantInf	InfPrior, Sex, AgeGr	Eq. (3)	Tab. S6
Hospitalizations	VariantHosp	Immunity, Sex, AgeGr	Eq. (1)	Tab. S7
Oxygen therapy	VariantOxygen	Immunity, Sex, AgeGr	Eq. (1)	Tab. S7
ICU admission	VariantICU	Immunity, Sex, AgeGr	Eq. (1)	Tab. S7
Detailed infections	VariantInf	VaccStatus, InfPrior, Sex, AgeGr	Eq. (2)	Tab. S6
Detailed hospitalizations	VariantHosp	VaccStatus, InfPrior, Sex, AgeGr	Eq. (2)	Tab. S6

Table S2: Details on analyses by Logistic regression.

Analysis	Outcome	Covariates
Hospitalizations given infection	Hosp	Variant, Immunity, Sex, AgeGr
Oxygen given infection	Oxygen	Variant, Immunity, Sex, AgeGr
ICU given infection	ICU	Variant, Immunity, Sex, AgeGr
Oxygen given hospitalization	Oxygen	Variant, Immunity, Sex, AgeGr
ICU given hospitalization	ICU	Variant, Immunity, Sex, AgeGr

where `Variant` is a dummy indicating the Omicron variant and `Immunity` describes the immunity status at the time of the corresponding infection. Moreover, outcome is either `Hosp, Oxygen, ICU`, which are analogs of `VariantHosp, VariantOxygen, VariantICU`, respectively, now reflecting all infections for which the variant was discerned rather than just those with a selected variant. As these outcomes are rare for young individuals, the younger categories could suffer from a lack of events, so `AgeGr` takes only four possible values here: $0-24$, $25-39$, $40-64$, and $65+$ years. The odds ratio, corresponding to the Omicron variant, equals $OR = \exp\{\beta\}$. Table S2 summarizes all these analyses. The input data set is formed by all the individuals, infected (the first three analyses) or hospitalized (the remaining two analyses) between 7 December 2021 and 29 January 2022. Similarly to the Cox analyses, we exclude categories with zero events. Moreover, we exclude `Inf6+_alone` category in the Oxygen given hospitalization analysis, having a single occurrence and a single event. Sample input for all the logistic regression analyses can be found in Table S8.

Software implementation

The `coxph` function from `survival R` package was used for the Cox regression, the `glm` function from `stats R` package for the Logistic regression. The original software tool has been developed to transform the source data to the input of the functions. The source code can be retrieved from `https://github.com/bisop-repo/omicronprotection`.

Corrigendum to: Supplementary Materials of "Protection provided by covid-19 vaccines and previous infection against the Omicron variant"

In Supplementary material 2 (SM2), the Tables in Sections 7-10 displayed the numbers of hospitalization cases instead of oxygen therapy and ICU therapy cases, respectively. The tables have been corrected.

Moreover, some outputs of the Cox and Logistic regression analyses in SM2 contained confusing near-to-zero estimates of HRs and odds ratios corresponding to the categories with no events. Inclusion of these categories did not harm the estimates of the remaining HRs and ORs; however, this way of presentation

might be found confusing, especially in combination with an alleged error of the logistic regression R package producing nonsensical CIs for these categories. Newly, these categories are omitted in the outputs.

The authors regret these errors.

Further, partly in reaction to the Letter to Editor [**?**], some clarifications and other explanatory material has been added to Supplementary Material 1 (SM1) and to SM2:

- a Discussion section has been added to SM1,
- additional descriptions of the individual analyses including sample inputs were added to SM1,
- the legend in SM2 has been extended,
- the tables in Sections 3–4 and 11–14 in SM2 were changed for the their categories to match with the Cox outputs,
- graphs in SM2 have been graphically enlarged,
- some typos have been corrected.

Table S3: Descriptive statistics: population and epidemic characteristics in the Czech Republic; ISID = the Czech National Information System of Infectious Diseases.

	Dataset				Population by 07/12/2021			Events from 07/12/2021			
Age group	Number	Data errors	Dead by 07/12/2021 covid-19	other	From ISID	Added	Total	Infected	Hospitalized	Dead covid-19	Dead other
0-11	434936	23	6	17	434890	932215	1367105	167628	798	1	3
12-15	357588	14	0	11	357563	98925	456488	109893	175	1	3
16-17	170567	5	2	16	170544	24472	195016	51543	112	0	1
18-24	577030	69	6	94	576861	92631	669492	118759	433	0	13
25-29	466101	50	21	88	465942	154983	620925	87595	479	1	15
30-34	552204	28	48	156	551972	166959	718931	106904	634	5	30
35-39	587595	52	76	224	587243	166067	753310	116199	724	13	39
40-44	713128	48	151	453	712476	180845	893321	130422	799	19	84
45-49	778086	44	298	793	776951	105635	882586	126795	971	38	105
50-54	603655	23	455	1109	602068	89015	691083	81047	1029	67	167
55-59	581568	23	879	1792	578874	90859	669733	63633	1321	141	251
60-64	527736	16	1668	2892	523160	102305	625465	40469	1509	201	427
65-69	594362	21	3310	5374	585657	86761	672418	29413	2374	370	727
70-74	572325	19	5541	8917	557848	63329	621177	21452	2984	586	1160
75-79	418441	15	6670	10490	401266	15935	417201	14150	3165	691	1231
80+	445102	9	15014	31314	398765	48761	447526	14703	5983	1640	3304
Total	8380424	459	34145	63740	8282080	2419697	10701777	1280605	23490	3774	7560
Errors	No age	9929	No sex	481							

Table S4: Descriptive statistics: vaccine distribution among different age groups until 20 November 2021.

Age group	Completed vaccination					Booster doses				
	BNT162b2	mRNA-1273	ChAdOx1 nCoV-19	Ad26.CoV2.S	Total	BNT162b2	mRNA-1273	ChAdOx1 nCoV-19	Ad26.CoV2.S	Total
0-11	38353	55	3	11	38422	175	2	0	0	177
12-15	195722	4167	2	2	199893	16844	143	0	0	16987
16-17	120102	3052	33	341	123528	18058	689	0	0	18747
18-24	394263	26110	2348	44131	466852	105841	13953	2	12	119808
25-29	298948	25096	3547	36148	363739	93385	15294	4	10	108693
30-34	364532	28912	4877	37250	435571	131736	21530	3	10	153279
35-39	395946	31868	7156	37489	472459	165876	25716	10	27	191629
40-44	503173	40455	12059	41404	597091	249310	36474	17	33	285834
45-49	564756	44639	16643	43925	669963	321021	44898	17	40	365976
50-54	440085	37823	17806	34186	529900	274139	39573	19	41	313772
55-59	426529	37494	24522	31785	520330	293838	41891	21	40	335790
60-64	388409	36335	33399	28237	486380	307892	42194	23	47	350156
65-69	426983	45314	60188	27974	560459	381356	54634	39	97	436126
70-74	359947	58087	100680	20549	539263	379045	63896	59	128	443128
75-79	246125	46985	83302	12442	388854	276509	48422	53	106	325090
80+	264910	40559	67123	9655	382247	276295	42481	47	108	318931
Total	5428783	506951	433688	405529	6774951	3291320	491790	314	699	3784123

Table S5: Sample input to the Infection analysis in the Cox regression model, variant Omicron

Subject	T1	T2	Infected	DeadByOther	VariantInf	Immunity	AgeGr	Sex
1	0	14	0	0	0	full2+_alone	70-74	M
1	14	68	0	0	0	boost2-_alone	70-74	M
15	0	36	0	0	0	inf6+_full2+	55-59	M
15	36	68	0	0	0	inf6+_boost2-	55-59	M
417	0	9	0	0	0	_noimmunity	25-29	Z
417	9	29	0	0	0	part_alone	25-29	Z
417	29	68	0	0	0	full2-_alone	25-29	Z
968	0	49	0	0	0	full2+_alone	55-59	Z
968	49	50	1	0	1	boost2-_alone	55-59	Z
2757	0	24	0	1	0	boost2-_alone	80+	M
22209	0	2	1	0	0	_noimmunity	35-39	M
391175	0	29	0	0	0	inf6-_alone	25-29	Z
391175	29	52	1	0	1	inf6+ _alone	25-29	Z

Table S6: Sample input to the Protection by Infection and Detailed infection analyses in the Cox regression model, variant Omicron

Subject	T1	T2	Infected	DeadByOther	VariantInf	InfPrior	VaccStatus	AgeGr	Sex
1	0	14	0	0	0	_noinf	full3	70-74	M
1	14	68	0	0	0	_noinf	boost1	70-74	M
15	0	36	0	0	0	inf3	full3	55-59	M
15	36	68	0	0	0	inf3	boost1	55-59	M
417	0	9	0	0	0	_noinf	_novacc	25-29	Z
417	9	29	0	0	0	_noinf	part1	25-29	Z
417	29	68	0	0	0	_noinf	full1	25-29	Z
968	0	49	0	0	0	_noinf	full3	55-59	Z
968	49	50	1	0	1	_noinf	boost1	55-59	Z
2757	0	24	0	1	0	_noinf	boost1	80+	M
22209	0	2	1	0	0	_noinf	_novacc	35-39	M
391175	0	29	0	0	0	inf1	_novacc	25-29	Z
391175	29	52	1	0	1	inf2	_novacc	25-29	Z

Table S7: Sample input to Hospitalization, Oxygen and ICU analyses in the Cox regression model, variant Omicron

Subject	T1	T2	Variant Hosp	Variant Oxygen	Variant ICU	DeadBy Cov	Immunity	AgeGr	Sex
1	0	14	0	0	0	0	full2+_alone	70-74	M
1	14	53	0	0	0	0	boost2-_alone	70-74	M
3622	0	47	1	0	0	0	full2+_alone	18-24	Z
40485	0	33	0	0	0	0	full2-_alone	80+	Z
40485	33	43	1	1	0	0	full2+_alone	80+	Z
1657678	0	15	0	0	0	0	full2+_alone	80+	Z
1657678	15	46	1	0	1	0	boost2-_alone	80+	Z
1671783	0	42	1	1	1	0	full2+_alone	65-69	M

Table S8: Sample input to logistic regressions

Subject	Hosp	Oxygen	ICU	Variant	Immunity	AgeGr	Sex
686	0	0	0	Delta	full2+_alone	65	Z
3622	1	0	0	Omikron	full2+_alone	0-24	Z
4075	0	0	0	Delta	part_alone	25-39	M
40485	1	1	0	Omikron	full2+_alone	65	Z