Supplementary Appendix

Supplement to: Goldberg Y, Mandel M, Bar-On YM, et al. Protection and waning of natural and hybrid immunity to SARS-CoV-2. N Engl J Med. DOI: 10.1056/NEJMoa2118946

This appendix has been provided by the authors to give readers additional information about the work.

Supplementary Appendix

Protection and waning of natural and hybrid COVID-19 immunity

Yair Goldberg[†1*] Ph.D., Micha Mandel[†2] Ph.D., Yinon M. Bar-On[3] M.Sc., Omri Bodenheimer[4] M.Sc., Laurence S. Freedman[5] Ph.D., Nachman Ash[4] M.D., Sharon Alroy-Preis[4] M.D., Amit Huppert[&5,6] Ph.D., Ron Milo[&3] Ph.D.

[1] Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Haifa, Israel

[2] Department of Statistics and Data Science, The Hebrew University of Jerusalem, Jerusalem, Israel

[3] Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot, Israel

[4] Israel Ministry of Health, Jerusalem, Israel

[5] The Bio-statistical and Bio-mathematical Unit, The Gertner Institute for Epidemiology & Health Policy Research, Sheba Medical Center, Ramat Gan, Israel

[6] The Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[†]Contributed equally.

[&] Contributed equally.

*corresponding author.

Yair Goldberg,
Faculty of Industrial Engineering and Management,
Technion - Israel Institute of Technology,
Haifa, Israel, 3200003
email: yairgo@technion.ac.il

Contents

Supplementary Methods 1 - Description of the Data

The analysis is based on the Israel Ministry of Health's database. Israel has experienced four pandemic waves, with the Delta (B.1.617.2) variant being the predominant variant during the fourth wave (the study period). During the third wave, Israel initiated a very rapid vaccination campaign, offering the BNT162b2 vaccine to all adult residents. The campaign opened on December 20, 2020, initially to people aged 60 years or older, and then gradually extended[1] until, on February 4, 2021, all individuals aged 16 or older were eligible to receive two doses of the vaccine. In March 2021, previously infected individuals were eligible to receive a single BNT162b2 dose after at least three months had elapsed from recovery from Covid-19. After the arrival of the Delta variant to Israel, a new Covid-19 wave began in mid-June 2021. Consequently, on July 30, 2021, the administration of a third (booster) dose was approved, first for people aged 60 years or older, and later for younger age groups.[2,3]

Israel has a centralized health system, where each resident belongs to one of four health maintenance organizations (HMOs). Polymerase Chain Reaction (PCR) tests for SARS-CoV-2 infections as well as vaccination against the virus are provided free of charge, and are directly reported to the Ministry of Health (MoH). The MoH established a centralized Covid-19 national database containing regularly updated information on all PCR tests and results, vaccination dates, and follow-up data on all infected individuals, including severity of disease and mortality. In this study, re-infection is defined as a positive PCR test in an individual who had a previous positive result on a sample taken at least 90 days earlier.[4] Severe disease is defined following the US NIH definition: a resting respiratory rate of more than 30 breaths per minute, an oxygen saturation of less than 94% while breathing ambient air, or a ratio of partial pressure of arterial oxygen to fraction of inspired oxygen of less than 300. The MoH database also includes basic demographic information, such as sex, age, place of residency, and population sector.

The MoH database is updated daily regarding death due to Covid-19, but only periodically regarding death for other reasons. Death data were last updated on August 6, 2021, thus people could contribute risk days after non-Covid related death. During the period of the study, the all-cause mortality for Israeli residents, including those not included in the study and those who died from Covid-19, were 363 death cases for individuals below age 40, 714 cases for individuals aged 40 to 60, and 7711 cases among individuals aged 60 or above. Out of these numbers, 1283 were Covid-19 related and appeared in the dataset. Thus, the number of days counted after death is small (about 0.25 million) relative to the total number of days at risk (about 300 million) in the cohorts, and most of them are in the 60+ age group.

Supplementary Methods 2 - Statistical modeling

Figure S6 describes the dynamics of individuals between immunity states. The five "blue" states are those studied in this paper. The transitions of interest are those to a new confirmed infection, and are described by cause-specific hazard functions or rate functions, denoted by h.

Let T_j , X_j and E_j denote the entrance calendar time to state j, the sojourn time in state j and the transition type from state j (for j being one of the 'blue' states in Figure S6); the transition type may be to a new immunity state depicted in the diagram, to a confirmed infection state or to a competing Exit state which combines all non-infection events that remove people from the study (e.g., vaccination with a different vaccine, traveling abroad, etc.). We assume that for each subject, (X_j, E_j) is independent of all preceding sojourn times conditionally on T_j, on exposure at time T_j, and on time-fixed covariates. In other words, the transition to a new state 'resets the clock' and the sojourn time and transition type from the new state are independent of all the history. More specifically, let t denote calendar time and let s denote the time since entering the last state, the cause specific hazard of infection in state j (as before, j is one of the 'blue' states depicted in Figure S6), is denoted by $h_j(t,s)$ and is assumed to take the following functional form:

$$h_j(t,s) = exp\{w_v 1(t \in W_v) + \sum_r \quad a_r 1(g(t) \in R_r) + \sum_k \quad b_{jk} 1(s \in I_{jk}) + \sum_l \quad c_l x_l\}.$$

Here w_v is the parameter of the calendar week W_v at time t. a_r is the coefficient of R_r, the interval between the rth and $(r+1)$st deciles of the exposure measure $g(t)$ (see the Statistical Analysis section). The parameters b_{jk} define a piecewise constant function for the effect of the sojourn time in a state, that is, the two-month sub-cohorts that are indicated by I_{jk}. And c_l is the coefficient of the lth time-fixed covariate x_l. For ease of notation, we omit the index of the individual from the formula. We also assume that the cause-specific hazards of leaving the state and exiting the study for reasons other than infection are indexed by parameters different from those in $h_j(t,s)$, and thus, contain no direct information on the parameter of interest. We denote such hazards by $r_j(t,s)$.

The conditional independence assumption and the piecewise constant structure of the effect of sojourn times implies that the likelihood of each individual can be partitioned into parts corresponding to the different states. Specifically, in our setting where time is measured in days, the contribution to the likelihood is a product of terms such as:

- $h_j(t,s)exp\{-h_j(t,s) - r_j(t,s)\}$ if a person gets an infection on date t after being in state j for s days.
- $r_j(t,s)exp\{-h_j(t,s) - r_j(t,s)\}$ if a person left state j from reason other than infection on date t after being in state j for s days.
- $exp\{-h_j(t,s) - r_j(t,s)\}$ if neither of the two aforementioned events occur on date t after being s days in state j.

A simple inspection of the product over all individuals and all dates reveals that it is factorized into terms containing the parameters of interest (w_t, a_r, b_{jk}, c_l) and a term containing the other unknown parameters ($r_j(t,s)$). As the partition is by days, it can easily accommodate left truncation as well. Moreover, the part containing the parameters of interest has the form of a Poisson likelihood where events on different dates are considered independent.

A simple introduction to the use of piecewise constant hazards and its connection to the Poisson model is given in the lecture notes of Germán Rodríguez[5]; for a more formal treatment see the works of Laird and Olivier[6] and Efron.[7]

Supplementary Methods 3 - Calculation of adjusted rates

The adjusted rates were calculated as follows. After fitting the Poisson model, the expected number of cases for sub-cohort k was calculated by assuming all individuals belonging to that sub-cohort. These estimates were calculated by applying the results of the Poisson model to all individuals with their coefficient of sub-cohort being replaced with that of sub-cohort k: $exp(\gamma_k + \beta \times covariates)$, where exp is the exponential function, γ_k is the coefficient of sub-cohort k, and β is the row vector of coefficients for all other covariates (see Table S7). These estimates (expressed per 100,000 person days) are reported as the adjusted rates.

Supplementary Analysis 1 - Handling Missing Data

As this is a national-level dataset, the dataset is incomplete and some variables have missing values. Table S5 summarizes the characteristics of individuals with various missing information. For 6677 (0.1%) individuals, the sex variable is missing. As can be seen in Table S5, the infection rate in this group (2.8%) is similar to the infection rate in the population (2.9%). A total of 115,929 (1.9%) individuals had missing data on area of residence. This variable is used to calculate the daily exposure risk and determine the population sector. Table S5 shows that the group with missing area of residence was younger than the rest of the population and most had received a booster dose or were recovered from Covid-19. The table also shows that this group had a lower infection rate (0.2%) compared to the rate in the population (2.9%). In the main analysis, an average exposure risk was imputed to individuals with missing data on residency. As a sensitivity analysis, we also performed a multiple imputation analysis by imputing the city code using weights according to the number of residents in the different cities, and repeating this process 10 times. We further examined the results when these individuals were omitted from the analysis. The comparison between the results of these methods appears in Table S6. Evidently, mean imputation and multiple imputation give identical results to one decimal accuracy, and removing individuals with missing data also has a negligible effect on the results.

The data contain information on 84,128 individuals who received only a single vaccine dose. This group was not included in the analysis, as it comprised a small fraction of the population (1.4%) who did not comply with any official vaccination protocol. Table S5 presents the data for individuals in this group. They were somewhat younger compared to the whole population, and a higher proportion of them belonged to the Arab sector; the rate of documented infection was higher than in the population (3.7% vs 2.9%).

Supplementary Analysis 2 - Sensitivity analysis for misclassification

As mild and asymptomatic infections are not always detected, individuals included in all the cohorts might have actually unknowingly been previously infected with Covid-19. This can introduce bias into the estimate of rates and rate ratios. The following is a simple approach to quantifying the bias that could arise from the misclassification of individuals unknowingly previously infected with Covid-19 into one of the cohorts of previously uninfected individuals. This bias pertains to the following cohorts in this study:

1. Previously uninfected vaccinated with 2 doses.
2. Previously uninfected vaccinated with 3 doses.
3. Previously infected and recovered.
4. Previously infected and recovered and then vaccinated with 1 dose.
5. Vaccinated with 1 or 2 doses and then infected and recovered.

In addition, there is the cohort of previously uninfected and unvaccinated persons, who are not analyzed in this paper. To study the sensitivity of the results to misclassification, we need to distinguish between the true immune status of individuals and their observed status, that is, the cohort to which they are classified. We used the following notation and assumption:

True values

Cohort	Subscript	Number of person-days in sample	Rate of new infection
Uninfected and unvaccinated	0	n_0	r_0
Uninfected and vaccinated with 2 doses	1	n_1	r_1
Uninfected and vaccinated with 3 doses	2	n_2	r_2
Recovered and unvaccinated	3	n_3	r_3
Recovered then vaccinated with 1 dose	4	n_4	r_4
Vaccinated with 1 or 2 doses then recovered	5	n_5	r_5

Of course, we do not observe these numbers or rates of infection because a proportion p of the Recovered person-days are misclassified as unvaccinated person-days. As a result of the misclassification, we observe the following values:

Observed values

Cohort	Subscript	Number of person days in sample	Rate of new infection
Uninfected and unvaccinated	0	n_0*	r_0*
Uninfected and vaccinated with 2 doses	1	n_1*	r_1*
Uninfected and vaccinated with 3 doses	2	n_2*	r_2*
Recovered and unvaccinated	3	n_3*	r_3*
Recovered then vaccinated with 1 dose	4	n_4*	r_4*
Vaccinated with 1 or 2 doses then recovered	5	n_5*	r_5*

We make the following assumptions for the relation between the true and recorded values:

1. For every person-day recorded in groups 3 and 4 (those who were infected before receiving any vaccine), there were another p/(1-p) person-days that were not reported as infected. (This is equivalent to assuming that of the truly recovered individuals, a proportion p were not reported.) We assume that these (n_3*+ n_4*) p/(1-p) person-days are distributed among groups 0-2 (the uninfected groups) at a ratio of n_0: n_1: n_2 , namely, in the same ratio as the true sizes of these groups.

2. A different assumption needs to be made for the misclassification of recovered person-days from group 5, because they were vaccinated before they got infected. We assume that n_5* p/(1-p) of them are misclassified into either group 1 or group 2 (the groups of uninfected vaccinated persons) and their ratio in these groups is n_1: n_2.

With these assumptions, the relations between recorded and true numbers of person-days are:

$n_0^* = n_0 + [(n_3^* + n_4^*)\, p / (1-p)]\, [n_0 / (n_0 + n_1 + n_2)]$

$n_1^* = n_1 + [(n_3^* + n_4^*)\, p / (1-p)]\, [n_1 / (n_0 + n_1 + n_2)] + [n_5^*\, p / (1-p)]\, [n_1 / (n_1 + n_2)]$

$n_2^* = n_2 + [(n_3^* + n_4^*)\, p / (1-p)]\, [n_2 / (n_0 + n_1 + n_2)] + [n_5^*\, p / (1-p)]\, [n_2 / (n_1 + n_2)]$

$n_3^* = (1-p)\, n_3$

$n_4^* = (1-p)\, n_4$

$n_5^* = (1-p)\, n_5$

Knowing the values of n_0^* to n_5^* and p, one can solve these equations to obtain the values of n_0 to n_5.

We make the following assumptions regarding the observed rates of new infection in the different cohorts.

1. There is no difference in the rates of new infection among the correctly classified and misclassified recovered individuals. That is, $r_3^* = r_3$, $r_4^* = r_4$, and $r_5^* = r_5$.

2. The (n3+n4) p/(1-p) person-days that pertain to individuals who were infected (but not reported) before receiving any vaccine and to those who are later vaccinated with 2 doses (group 1) or 3 doses (group 2) have a lower infection rate than their uninfected counterparts. We assume that the relative risk factor is 0.1, which is a little lower than the factors of 0.10 to 0.18 based on infection rates in sub-cohorts of group 4 compared to those of group 1 (see Table 2). (Assuming this lower relative risk factor of 0.1 may, therefore, modestly exaggerate our estimate of the bias caused by the misclassification of infected individuals.)

3. The n5 p/(1-p) person-days that pertain to individuals who were infected after vaccination (but not reported) and who are included in the group vaccinated with 2 doses (group 1) have an infection rate equal to r_5. Those who received a third dose (group 2) have an infection rate equal to $0.1r_2$, using the same argument as in assumption 2 above.

With these assumptions, we can write the equations for computing the rates of infection in groups 1 and 2. To do this, we need some extra notation. We denote the proportions of person-days in observed group 1 as follows: those who are correctly classified as q_0; those who are misclassified from groups 3 and 4 as q_{34}; and those who are misclassified from group 5 as q_5. Their sum equals 1. It so happens that under our assumptions, the same proportions apply to the person-days observed in group 2. With this notation:

(1) $r_1^* = q_0\, r_1 + q_{34}\, (0.1\, r_1) + q_5\, r_5$

(2) $r_2^* = q_0\, r_2 + (q_{34} + q_5)\, (0.1\, r_2)$

The proportions q can be calculated easily from the solutions of n_1 to n_5, and knowing the values of r_1* and r_2*, together with assuming that $r_5 = r_5$*, these equations can be solved for r_1 and r_2. Note that with this approach, the rates of infection need not be crude, but could be those adjusted for confounding. In our calculations below, we use the adjusted rates.

Bias in the estimated rates of infection

Under the above assumptions, it is clear that:

(i) the estimated rates of infection among cohorts 3-5 are unbiased; and

(ii) the estimated rates of infection among cohorts 1 and 2 are biased, with the bias depending on the misclassification proportion p.

The following table presents the recorded number of person-days in the various cohorts and the calculated true numbers under our assumptions, i.e., the solution to equations (1) and (2) above. The calculations are shown for two values of the proportion of infected individuals who are not reported: p=0.5 and p=0.7.

Cohort	Recorded person-days	"True" person-days p=0.5	"True" person-days p=0.7
0*	61,497,760	56,271,112	49,202,683
1	184,214,308	167,304,073	144,826,398
2	80,428,946	73,045,847	63,231,975
3	17,880,995	35,761,990	59,603,317
4	9,670,155	19,340,310	32,233,850
5	1,968,832**	3,937,664	6,562,773

* This cohort was not used in the main analysis.
** Including those vaccinated with two doses then recovered.

From these numbers, the following proportions of person-days for the correctly classified infected before vaccination but not reported, and infected after vaccination but not reported cohorts were calculated. They apply both to cohort 1 and cohort 2.

For p=0.5. Correctly classified: 0.9082; infected before vaccination: 0.0844; infected after vaccination: 0.0074.

For p=0.7 Correctly classified: 0.7862; infected before vaccination: 0.1965; infected after vaccination: 0.0174

The observed and recalculated crude rates of infection in Cohorts 1 and 2 are presented in the following table. Rates are per 100,000 person-days exposure.

Cohort	Recorded rate of infection	Recalculated rate of infection p=0.5	Ratio of recalculated to recorded p=0.5	Recalculated rate of infection p=0.7	Ratio of recalculated to recorded p=0.7
1	76.3	83.1	1.09	94.3	1.23
2	7.3	8.0	1.09	9.0	1.23

From Table 1, the crude rate of infection per 100,000 person-days for recovered individuals that were then vaccinated is 8.5. Thus, the crude relative risk of recovered and vaccinated to previously uninfected and vaccinated is 8.5/76.3 = 0.11 based on the observed misclassified data, compared to 8.5/83.1 = 0.10 after the adjustment for 50% non-reporting of infection, and 8.5/94.3 = 0.09 after the adjustment for 70% non-reporting of infection.

Under the above assumptions, the rates of the recovered cohorts, with or without vaccination, are not affected by the non-reporting of infections. Neither does the ratio of rates of infection in previously uninfected individuals receiving 3 doses compared to rates in those receiving two doses change due to non-reporting of infections, since the rates of both groups are inflated by the same factor after adjustment. Similarly, we would not expect our estimates of rates of waning within the different non-recovered cohorts to be greatly affected by non-reporting of infections, since very similar inflation factors resulting from the adjustment would be expected to apply to the sub-cohorts within a given cohort. The conclusions from the analysis are dependent on the assumptions made in this simple model; different assumptions may be expected to show a somewhat different magnitude of the effects. However, the bias toward underestimation of the infection rate among vaccinated uninfected individuals is likely to be real on the assumption that those who recovered from Covid-19 and were misclassified to the vaccinated cohorts were more protected from reinfection than their uninfected counterparts.

Table S1: ***Demographic and clinical characteristics of the different Recovered unvaccinated sub-cohorts.The table presents the proportion of person-days at risk and the number of events that were used in the analysis; study period: August 1, 2021, to September 30, 2021.***

Group	Recovered unvaccinated 4-6 months ago *Person-days at risk = 1,264,699*			Recovered unvaccinated 6-8 months ago *Person-days at risk = 7,210,192*			Recovered unvaccinated 8-10 months ago *Person-days at risk = 3,243,005*			Recovered unvaccinated 10-12 months ago *Person-days at risk = 3,262,444*			Recovered unvaccinated 12+ months ago *Person-days at risk = 2,900,655*		
	% person days at risk	*# Infect-ions*	*# severe Covid-19*	*% person days at risk*	*# Infect-ions*	*# severe Covid-19*	*% person days at risk*	*# Infect-ions*	*# severe Covid-19*	*% person days at risk*	*# Infect-ions*	*# severe Covid-19*	*% person days at risk*	*# Infect-ions*	*# severe Covid-19*
Female	*57.8%*	*69*	*0*	*55.6%*	*681*	*2*	*55.3%*	*455*	*2*	*47.9%*	*599*	*1*	*48.1%*	*566*	*4*
Male	*42.2%*	*52*	*2*	*44.4%*	*543*	*7*	*44.7%*	*292*	*2*	*52.1%*	*598*	*4*	*51.9%*	*536*	*1*
Age 16-39	*67.1%*	*104*	*0*	*68.2%*	*989*	*3*	*64.0%*	*551*	*3*	*66.3%*	*915*	*0*	*66.1%*	*802*	*0*
Age 40-59	*24.4%*	*13*	*0*	*24.4%*	*208*	*3*	*27.0%*	*180*	*1*	*24.7%*	*249*	*3*	*25.1%*	*265*	*3*
Age 60+	*8.6%*	*4*	*2*	*7.3%*	*27*	*3*	*9.0%*	*16*	*0*	*9.1%*	*33*	*2*	*8.8%*	*35*	*2*
General Jewish	*59.3%*	*94*	*2*	*55.4%*	*789*	*5*	*45.6%*	*415*	*1*	*48.3%*	*679*	*2*	*43.7%*	*545*	*2*
Arab	*11.0%*	*16*	*0*	*25.8%*	*335*	*2*	*26.2%*	*195*	*1*	*35.2%*	*466*	*1*	*37.3%*	*451*	*0*
Ultra-Orthodox	*29.7%*	*11*	*0*	*18.7%*	*100*	*2*	*28.2%*	*137*	*2*	*16.6%*	*52*	*2*	*19.0%*	*106*	*3*

Table S2: *Demographic and clinical characteristics of the different Vaccinated two and three doses sub-cohorts.The table presents the proportion of person-days at risk and number of events that were used in the analysis; study period: August 1, 2021, to September 30, 2021.*

Group	Vaccinated two doses 0-2 months ago *Person-days at risk = 4,889,023*			Vaccinated two doses 2-4 months ago *Person-days at risk = 9,247,416*			Vaccinated two doses 4-6 months ago *Person-days at risk = 98,156,329*			Vaccinated two doses 6-8 months ago *Person-days at risk = 71,921,540*			Vaccinated three doses *Person-days at risk = 80,428,946*		
	% person days at risk	# Infect-ions	# severe Covid -19	% person days at risk	# Infect-ions	# severe Covid -19	% person days at risk	# Infect-ions	# severe Covid-19	% person days at risk	# Infect-ions	# severe Covid -19	% person days at risk	# Infect-ions	# severe Covid-19
Female	*51.1%*	*689*	*9*	*52.4%*	*2,405*	*24*	*50.9%*	*39,876*	*158*	*51.4%*	*34,505*	*388*	*51.0%*	*2,835*	*70*
Male	*48.9%*	*460*	*12*	*47.6%*	*1,497*	*19*	*49.1%*	*30,209*	*212*	*48.6%*	*30,837*	*550*	*49.0%*	*3,010*	*108*
Age 16-39	*79.5%*	*913*	*0*	*68.1%*	*2,901*	*1*	*63.8%*	*48,182*	*16*	*44.0%*	*32,475*	*27*	*16.4%*	*1,156*	*1*
Age 40-59	*13.7%*	*138*	*5*	*22.0%*	*735*	*10*	*28.5%*	*18,287*	*80*	*36.0%*	*23,665*	*169*	*30.1%*	*2,042*	*13*
Age 60+	*6.8%*	*98*	*16*	*9.9%*	*266*	*32*	*7.8%*	*3,616*	*274*	*20.0%*	*9,202*	*742*	*53.4%*	*2,647*	*164*
General Jewish	*75.6%*	*887*	*18*	*70.5%*	*3,149*	*32*	*71.8%*	*56,494*	*267*	*77.7%*	*50,644*	*754*	*89.2%*	*4,957*	*155*
Arab	*6.7%*	*104*	*2*	*6.6%*	*363*	*1*	*5.8%*	*6,104*	*24*	*5.1%*	*5,757*	*47*	*4.1%*	*417*	*12*
Ultra-Orthodox	*17.8%*	*158*	*1*	*22.9%*	*390*	*10*	*22.4%*	*7,487*	*79*	*17.2%*	*8,941*	*137*	*6.7%*	*471*	*11*

Table S3: Demographic and clinical characteristics of the different Recovered then vaccinated one dose sub-cohorts.The table presents the proportion of person-days at risk and number of events that were used in the analysis; study period: August 1, 2021, to September 30, 2021.

	Recovered then Vaccinated one dose 0-2 months ago *Person-days at risk = 2,321,324*			Recovered then Vaccinated one dose 2-4 months ago *Person-days at risk = 2,064,746*			Recovered then Vaccinated one dose 4-6 months ago *Person-days at risk = 3,979,206*			Recovered then Vaccinated one dose 6-8 months ago *Person-days at risk = 1,304,879*		
Group	*% person days at risk*	*# Infect-ions*	*# severe Covid -19*	*% person days at risk*	*# Infect-ions*	*# severe Covid -19*	*% person days at risk*	*# nfect-ions*	*# severe Covid -19*	*% person days at risk*	*# Infect-ions*	*# severe Covid -19*
Female	*53.3%*	*61*	*0*	*54.5%*	*52*	*1*	*50.8%*	*245*	*7*	*47.0%*	*83*	*0*
Male	*46.7%*	*44*	*1*	*45.5%*	*42*	*1*	*49.2%*	*219*	*0*	*53.0%*	*76*	*3*
Age 16-39	*65.7%*	*79*	*0*	*59.8%*	*69*	*0*	*52.2%*	*295*	*0*	*53.0%*	*100*	*1*
Age 40-59	*25.1%*	*19*	*1*	*28.8%*	*15*	*0*	*31.1%*	*126*	*3*	*30.4%*	*49*	*1*
Age 60+	*9.1%*	*7*	*0*	*11.5%*	*10*	*2*	*16.7%*	*43*	*4*	*16.6%*	*10*	*1*
General Jewish	*54.6%*	*62*	*0*	*60.4%*	*68*	*1*	*50.7%*	*299*	*5*	*55.5%*	*110*	*2*
Arab	*25.8%*	*23*	*0*	*20.4%*	*18*	*1*	*21.0%*	*103*	*1*	*21.3%*	*26*	*0*
Ultra-Orthodox	*19.7%*	*20*	*1*	*19.2%*	*8*	*0*	*28.2%*	*62*	*1*	*23.2%*	*23*	*1*

Table S4: Demographic and clinical characteristics of the different Vaccinated one dose then recovered sub-cohorts.The table presents the proportion of person-days at risk and number of events that were used in the analysis; study period: August 1, 2021, to September 30, 2021.

	Vaccinated one dose then Recovered 4-6 months ago Person-days at risk = 364,111			Vaccinated one dose then Recovered 6-8 months ago Person-days at risk = 1,011,858		
Group	*% person days at risk*	*# Infect- ions*	*# severe Covid-19*	*% person days at risk*	*# Infect- ions*	*# severe Covid-19*
Female	*55.1%*	*29*	*0*	*51.2%*	*88*	*1*
Male	*44.9%*	*7*	*0*	*48.8%*	*105*	*0*
Age 16-39	*61.7%*	*28*	*0*	*49.3%*	*128*	*0*
Age 40-59	*29.5%*	*8*	*0*	*34.4%*	*54*	*0*
Age 60+	*8.9%*	*0*	*0*	*16.3%*	*11*	*1*
General Jewish	*58.7%*	*27*	*0*	*61.6%*	*141*	*1*
Arab	*8.7%*	*4*	*0*	*17.7%*	*34*	*0*
Ultra-Orthodox	*32.7%*	*5*	*0*	*20.7%*	*18*	*0*

Table S5: Comparison of characteristics between individuals used in the final analysis and groups that were excluded (missing sex and 1 vaccine dose) or had missing residency data; study period: August 1, 2021, to September 30, 2021.

	Missing Sex	*1 Vaccine Dose*	*Missing Residency*	*Complete Data*	*Overall*
	(N=6677)	*(N=84,128)*	*(N=115,929)*	*(N=5,855,572)*	*(N=5,946,377)*
Age					
16-39	*3706 (55.5%)*	*53,926 (64.1%)*	*66,505 (57.4%)*	*2,626,552 (44.9%)*	*2,684,184 (45.1%)*
40-59	*1927 (28.9%)*	*22,004 (26.2%)*	*42,881 (37.0%)*	*1,787,378 (30.5%)*	*1,811,309 (30.5%)*
60+	*1044 (15.6%)*	*8198 (9.7%)*	*6543 (5.6%)*	*1,441,642 (24.6%)*	*1,450,884 (24.4%)*
Gender					
Unknown	*6677 (100%)*	*0 (0%)*	*0 (0%)*	*0 (0%)*	*6677 (0.1%)*
female	*0 (0%)*	*39,579 (47.0%)*	*52,346 (45.2%)*	*2,997,668 (51.2%)*	*3,037,247 (51.1%)*
male	*0 (0%)*	*44,549 (53.0%)*	*63,583 (54.8%)*	*2,857,904 (48.8%)*	*2,902,453 (48.8%)*
Sector					
General Jewish	*5546 (83.1%)*	*48,345 (57.5%)*	*115,929 (100%)*	*4,525,848 (77.3%)*	*4,579,739 (77.0%)*
ultra-Orthodox Jewish	*555 (8.3%)*	*8580 (10.2%)*	*0 (0%)*	*411,515 (7.0%)*	*420,650 (7.1%)*
Arabs	*576 (8.6%)*	*27,203 (32.3%)*	*0 (0%)*	*918,209 (15.7%)*	*945,988 (15.9%)*
Covid-19 Outcome					
No Documented Infection	*6489 (97.2%)*	*80,913 (96.2%)*	*115,729 (99.8%)*	*5,683,644 (97.1%)*	*5,771,046 (97.1%)*
Documented Infection	*188 (2.8%)*	*3154 (3.7%)*	*198 (0.2%)*	*170,013 (2.9%)*	*173,355 (2.9%)*
Severe Disease	*0 (0%)*	*61 (0.1%)*	*2 (0.0%)*	*1915 (0.0%)*	*1976 (0.0%)*
Vaccination Status					
2 Doses	*2741 (41.1%)*	*0 (0%)*	*76 (0.1%)*	*2,082,562 (35.6%)*	*2,085,303 (35.1%)*
3 Doses	*3014 (45.1%)*	*0 (0%)*	*81312 (70.1%)*	*3,155,246 (53.9%)*	*3,158,260 (53.1%)*
Recovered	*609 (9.1%)*	*0 (0%)*	*32327 (27.9%)*	*330,552 (5.6%)*	*331,161 (5.6%)*
Recovered then Vaccinated	*214 (3.2%)*	*0 (0%)*	*683 (0.6%)*	*216,773 (3.7%)*	*216,987 (3.6%)*
Vaccinated then Recovered	*99 (1.5%)*	*0 (0%)*	*1531 (1.3%)*	*70,439 (1.2%)*	*70,538 (1.2%)*
1 Dose	*0 (0%)*	*84,128 (100%)*	*0 (0%)*	*0 (0%)*	*84,128 (1.4%)*

Table S6. Comparison of Poisson regression analyses with different methods for handling missing residency data. For each sub-cohort, the table shows the rate ratio of confirmed infections between individuals with a fresh second vaccine dose (up to two months) who were not previously infected relative to each of the other sub-cohorts.

Cohort	Mean Imputation	Multiple Imputation	Dropping Missing
Recovered Unvaccinated 4-6	2.0 [1.7, 2.4]	2.0 [1.7, 2.4]	2.1 [1.7, 2.5]
Recovered Unvaccinated 6-8	1.5 [1.4, 1.6]	1.5 [1.4, 1.6]	1.6 [1.4, 1.7]
Recovered Unvaccinated 8-10	1.0 [0.9, 1.1]	1.0 [0.9, 1.1]	1.1 [1.0, 1.2]
Recovered Unvaccinated 10-12	0.7 [0.7, 0.8]	0.7 [0.7, 0.8]	0.8 [0.7, 0.8]
Recovered Unvaccinated 12+	0.7 [0.6, 0.8]	0.7 [0.6, 0.8]	0.7 [0.7, 0.8]
3 Doses 0-2	2.6 [2.4, 2.7]	2.6 [2.4, 2.7]	2.7 [2.5, 2.9]
2 Doses 0-2	Reference	Reference	Reference
2 Doses 2-4	0.5 [0.4, 0.5]	0.5 [0.4, 0.5]	0.5 [0.4, 0.5]
2 Doses 4-6	0.3 [0.3, 0.3]	0.3 [0.3, 0.3]	0.3 [0.3, 0.3]
2 Doses 6-8	0.2 [0.2, 0.3]	0.2 [0.2, 0.3]	0.2 [0.2, 0.3]
Recovered then 1 Dose 0-2	5.7 [4.6, 6.9]	5.6 [4.6, 6.9]	5.9 [4.8, 7.2]
Recovered then 1 Dose 2-4	5.0 [4.0, 6.1]	5.0 [4.0, 6.1]	5.2 [4.2, 6.4]
Recovered then 1 Dose 4-6	2.0 [1.8, 2.3]	2.0 [1.8, 2.3]	2.1 [1.9, 2.4]
Recovered then 1 Dose 6-8	1.8 [1.5, 2.2]	1.8 [1.5, 2.2]	1.9 [1.6, 2.3]
1 Dose then Recovered 4-6	2.0 [1.4, 2.8]	2.0 [1.4, 2.8]	2.1 [1.5, 2.9]
1 Dose then Recovered 6-8	1.3 [1.1, 1.5]	1.3 [1.1, 1.5]	1.4 [1.2, 1.6]

Table S7. Coefficients of Poisson regression for confirmed infection.

term	estimate	std.error
(Intercept)	-9.814	0.035
Age group: 40-59	-0.099	0.006
Age group: 60+	-0.396	0.009
Sex: male	-0.151	0.005
Sector: Ultra-Orthodox Jewish	0.357	0.009
Sector: Arab	-0.324	0.008
Exposure risk: 2nd decile (2.14,3.19]	0.740	0.021
Exposure risk: 3rd decile (3.19,4.20]	0.966	0.020
Exposure risk: 4th decile (4.20,5.16]	1.080	0.020
Exposure risk: 5th decile (5.16,6.18]	1.237	0.020
Exposure risk: 6th decile (6.18,7.32]	1.404	0.019
Exposure risk: 7th decile (7.32,8.63]	1.502	0.019
Exposure risk: 8th decile (8.63,10.3]	1.619	0.019
Exposure risk: 9th decile (10.30,12.60]	1.823	0.019
Exposure risk: 10th decile (12.60,7.15e+03]	2.100	0.019
Week 2	0.169	0.012
Week 3	0.201	0.012
Week 4	0.252	0.012
Week 5	0.264	0.013
Week 6	0.164	0.013

Week 7	0.045	0.014
Week 8	0.064	0.014
Week 9	0.070	0.017
Cohort: Recovered unvaccinated 12+	0.358	0.042
Cohort: Recovered unvaccinated 10-12	0.299	0.041
Cohort: Recovered unvaccinated 8-10	-0.028	0.047
Cohort: Recovered unvaccinated 6-8	-0.414	0.041
Cohort: Recovered unvaccinated 4-6	-0.702	0.096
Cohort: Vaccinated three doses 0-2	-0.944	0.033
Cohort: Vaccinated two doses 6-8	1.436	0.030
Cohort: Vaccinated two doses 4-6	1.189	0.030
Cohort: Vaccinated two doses 2-4	0.759	0.034
Cohort: Recovered then vaccinated 1 dose 6-8	-0.601	0.085
Cohort: Recovered then vaccinated 1 dose 4-6	-0.716	0.055
Cohort: Recovered then vaccinated 1 dose 2-4	-1.604	0.107
Cohort: Recovered then vaccinated 1 dose 0-2	-1.733	0.102
Cohort: Vaccinated 1 dose then recovered 6-8	-0.268	0.078
Cohort: Vaccinated 1 dose then recovered 4-6	-0.688	0.169

Table S8: Summary of the results regarding confirmed infections of the Poisson regression analysis for all sub-cohorts by age groups. For each group, the table shows the estimated covariate-adjusted confirmed infection rate per 100,000 person-days at risk. 95% confidence intervals without adjustment for multiplicity are given in square brackets.

Cohort	**16-39**	**40-59**	**60+**
Recovered unvaccinated 4-6	*14.2 [11.6, 17.4]*	*5.0 [3.0, 8.7]*	*4.2 [1.6, 10.5]*
Recovered unvaccinated 6-8	*17.4 [16.4, 18.5]*	*10.7 [9.4, 12.3]*	*4.8 [3.3, 6.9]*
Recovered unvaccinated 8-10	*24.6 [22.5, 26.9]*	*20.1 [17.5, 23.3]*	*5.5 [3.4, 9.3]*
Recovered unvaccinated 10-12	*34.5 [32.4, 36.8]*	*26.0 [23, 29.5]*	*9.7 [6.8, 13.8]*
Recovered unvaccinated 12+	*34.9 [32.5, 37.3]*	*31.5 [28, 35.4]*	*12.4 [8.9, 17.2]*
Vaccinated three doses 0-2	*10.7 [10.1, 11.4]*	*8.9 [8.5, 9.3]*	*5.6 [5.4, 5.8]*
Vaccinated two doses 0-2	*22.9 [21.5, 24.5]*	*20 [17.0, 23.6]*	*26.5 [21.5, 32.1]*
Vaccinated two doses 2-4	*53.0 [51.3, 55.1]*	*40.9 [38.2, 44.1]*	*29.3 [26.2, 33]*
Vaccinated two doses 4-6	*77.6 [76.9, 78.3]*	*71.5 [70.4, 72.5]*	*50.1 [48.5, 51.8]*
Vaccinated two doses 6-8	*98.0 [96.9, 99.1]*	*89.7 [88.6, 90.8]*	*72.1 [70.8, 73.7]*
Recovered then vaccinated one dose 0-2	*4.5 [3.6, 5.5]*	*2.9 [1.8, 4.5]*	*3.0 [1.4, 6.1]*
Recovered then Vaccinated one dose 2-4	*5.4 [4.3, 6.8]*	*2.5 [1.5, 4.2]*	*4.3 [2.3, 8]*
Recovered then Vaccinated one dose 4-6	*12.6 [11.2, 14.1]*	*9.6 [8.0, 11.4]*	*6.0 [4.5, 7.9]*
Recovered then Vaccinated one dose 6-8	*13.8 [11.5, 16.6]*	*12.3 [9.3, 16.3]*	*4.7 [2.5, 8.9]*
Vaccinated one dose then Recovered 4-6	*14.0 [9.7, 20.1]*	*8.6 [4.3, 16.7]*	*No events*
Vaccinated one dose then Recovered 6-8	*22.2 [18.6, 26.4]*	*13.4 [10.2, 17.2]*	*9.7 [7.0, 13.2]*

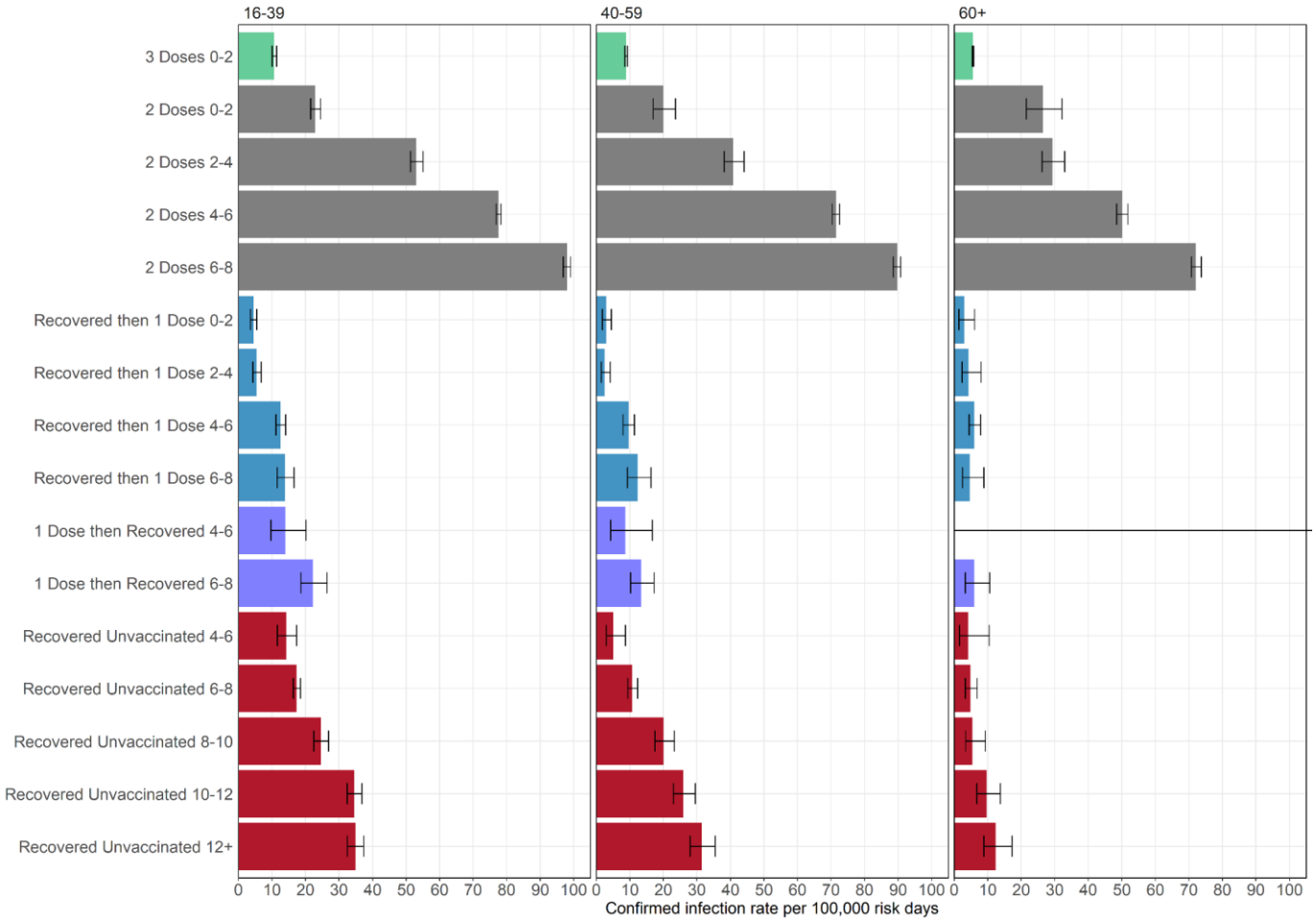


Figure S1: Estimated covariate-adjusted rates of confirmed infections per 100,000 at-risk days obtained from the Poisson regression analysis for the study period August 1, 2021, to September 30, 2021, stratified by age and sub-cohorts. Confidence intervals are not adjusted for multiplicity.

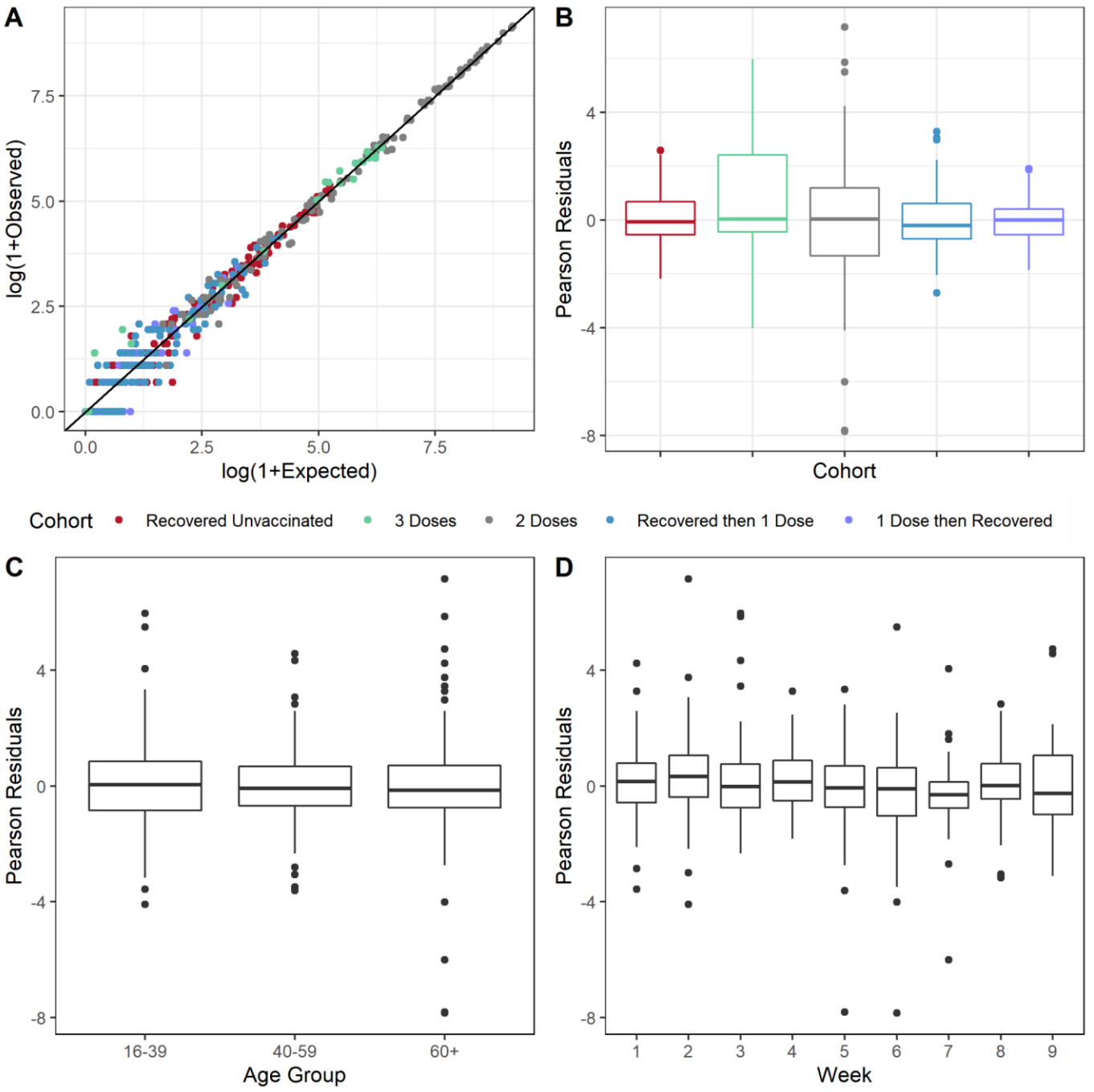


Figure S2: Residual Analysis. Pearson residuals for groups defined by combinations of sub-cohort, age group and week are calculated as $(Observed - Expected)/\sqrt{Expected}$*, where* $Observed$ *is the actual number of confirmed infections in the group and* $Expected$ *is the predicted number calculated by the fitted model.*

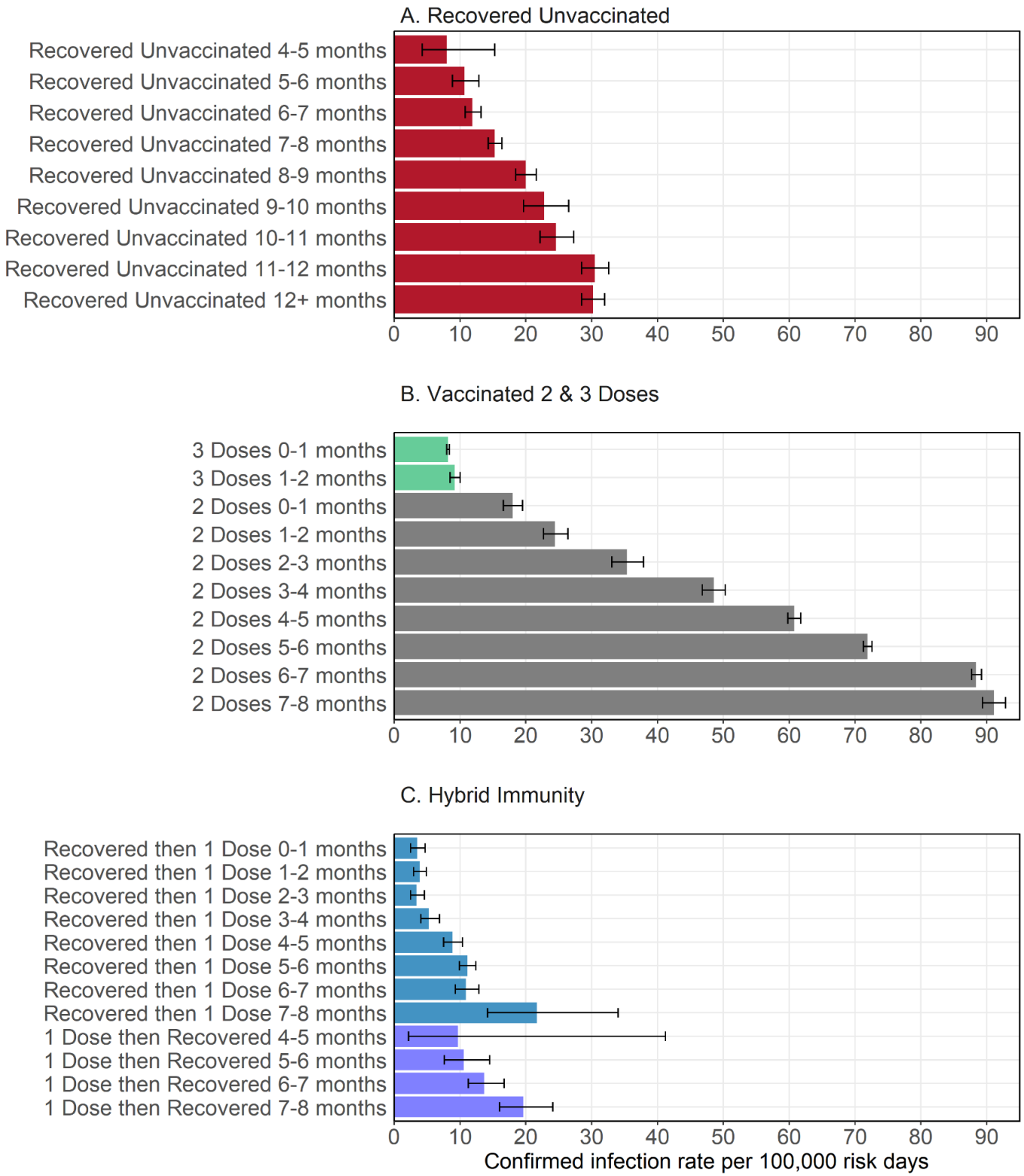


Figure S3: Estimated covariate-adjusted rates of confirmed infections per 100,000 at-risk days obtained from the Poisson regression analysis for the study period August 1, 2021, to September 30, 2021 using one-month sub-cohorts. Confidence intervals are not adjusted for multiplicity.

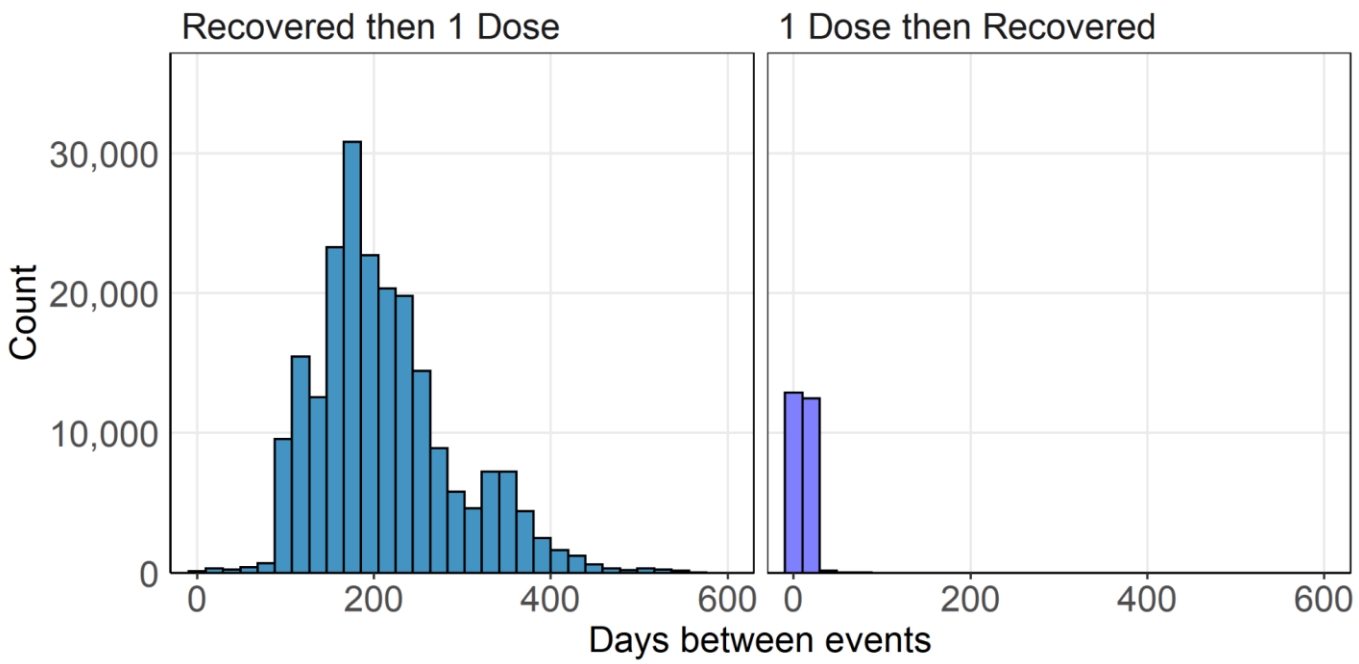


Figure S4: The distribution of time between infection and vaccination (left) in the Recovered then Vaccinated cohort, and between vaccination and infection (right) in the Vaccinated then Recovered cohort. The latter is shorter as individuals become doubly vaccinated starting January 2021.

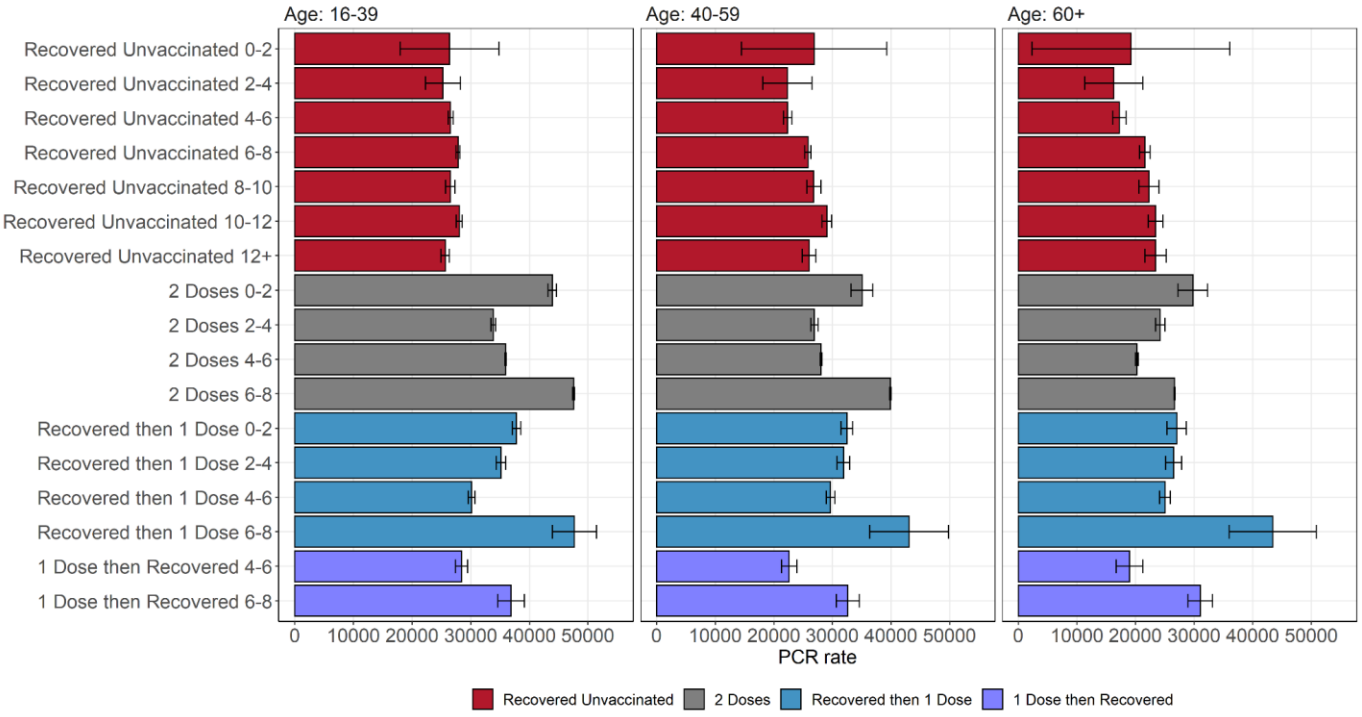
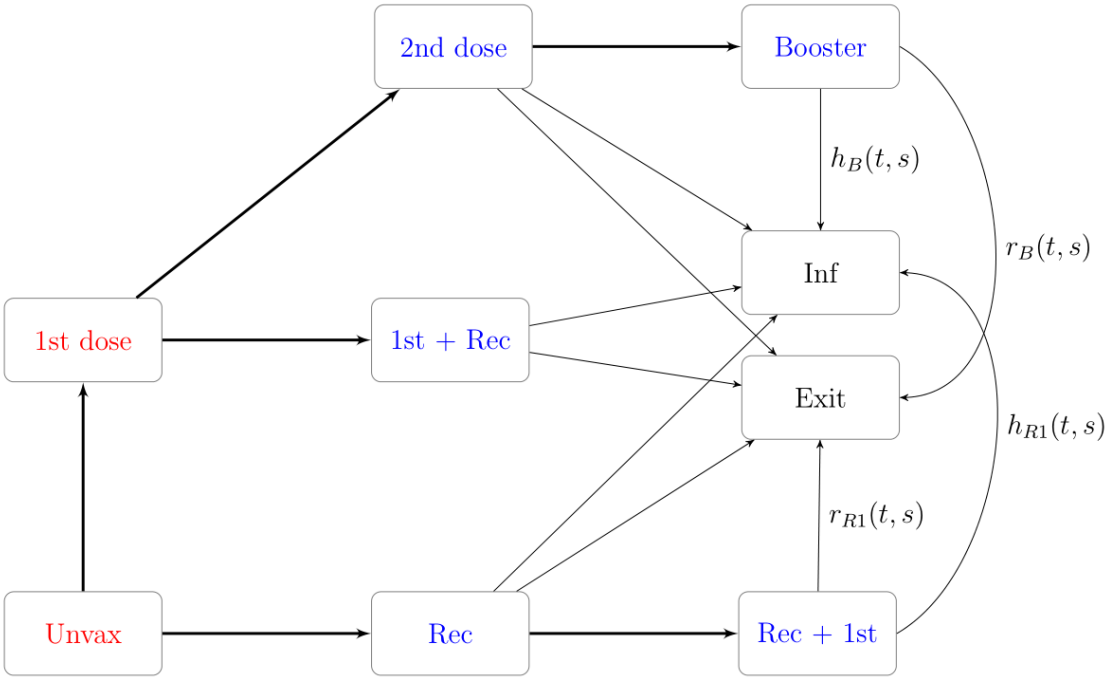


Figure S5: PCR testing rates by sub-cohort and age group. Bars indicate the number of individuals, per 100,000, who performed at least one test during the study period. Individuals were associated with their sub-cohort at the beginning of the study. The 95% confidence intervals are not adjusted for multiplicity.



Figure

Figure S6. Dynamic of individuals between immunity states. Unvax stands for unvaccinated, Rec stands for Recovered, Inf stands for confirmed infection during the study period, and Exit stands for all other events leading to an exit from the study. The current study focuses on the cause-specific hazard of a new infection from states appearing in blue in the diagram (all states excluding Unvax and 1st dose, appearing in red). Individuals can leave each blue state to any other blue state that is joined to it in the diagram. They can also leave each blue state and enter the infected state or the Exit state. Events that trigger an exit from the study include vaccinated with a vaccine other than BNT162b2, vaccinated with an additional dose not appearing in the diagram, or traveling abroad. The cause-specific hazards that are of interest are labeled for a selected sample of the pathways (Booster and Rec + 1st to Inf and Exit). They depend on calendar time t and on the sojourn time in a state s. The hazards for Infection are denoted by h. and for Exit by r.

References

1. Shilo S, Rossman H, Segal E. Signals of hope: gauging the impact of a rapid national vaccination campaign. Nat Rev Immunol 2021;21(4):198–9.
2. Bar-On YM, Goldberg Y, Mandel M, et al. Protection Across Age Groups of BNT162b2 Vaccine Booster against Covid-19 [Internet]. 2021 [cited 2021 Oct 24]. Available from: https://www.medrxiv.org/content/10.1101/2021.10.07.21264626v1
3. Bar-On YM, Goldberg Y, Mandel M, et al. Protection of BNT162b2 Vaccine Booster against Covid-19 in Israel. N Engl J Med 2021;385(15):1393–400.
4. SARS-CoV-2 variants of concern and variants under investigation in England - Technical briefing 19 [Internet]. Public Health England; 2021. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1005517/Technical_Briefing_19.pdf
5. Rodríguez G. Lecture Notes on Generalized Linear Models [Internet]. 2007. Available from: URL: https://data.princeton.edu/wws509/notes/
6. Laird N, Olivier D. Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques. J Am Stat Assoc 1981;76(374):231–40.
7. Efron B. The two-way proportional hazards model. J R Stat Soc Ser B Stat Methodol 2002;64(4):899–909.