Supplementary material

Supplementary Tables

Table S1. Start dates of the data collection period, per study site.

Abbreviations: N, Number.

Table S3. Characteristics of study participants, according to receipt of BNT162b2 XBB.1.5-adapted vaccine and SARS-CoV-2 case classification, using various prior vaccination categories as reference groups.

Smoking history1

¹ Percentages exclude subjects with missing values.

² Obese classification was based on both BMI ≥30.0 kg/m2 and obesity as a categorical variable (yes/no).

³ The categories "0" and "1" among the number of chronic conditions are counted as separate categories of number of chronic conditions, based on the list of chronic conditions provided in Supplementary Table S2. Here, the cell counts among participants having "0–1" chronic conditions are provided as a composite.

⁴ The categories "Cancer" and "Immunodeficiency" among chronic conditions are identified as separate types of chronic conditions, as presented in Supplementary Table S2. Here, the cell counts among participants having "Immunodeficiency or cancer" are provided as a composite.

5 Pregnancy status counts and percentages were counted in female patients.

⁶246 patients out of the 310 (79.4%) who did not receive any COVID-19 vaccine in 2023-2024 season and had previously received ≥2 mRNA wildtype doses only, in this matched dataset, had received ≥3 mRNA wildtype doses only.

Supplementary Figures

Figure S1 Flow diagram of study population

¹ 14 patients infected with JN.1 outside of the JN.1 predominant period were included in the analysis.

Figure S2. Number of SARI patients enrolled according to SARS-CoV-2 case status over time A) overall and B) by country.

JN.1 variant predominance started on: 29 December 2023 in Belgium, 31 December 2023 in Germany, 1 January 2024 in Italy and 3 December 2023 in Spain. Vaccination campaigns began: 15 September 2023 in Belgium, 18 September 2023 in Germany, 25 September 2023 in Italy, 16 October 2023 in the Valencia region of Spain, and 26 September 2023 in the Catalonian region of Spain.

Figure S3. Vaccine effectiveness¹ against COVID-19 hospitalization in SARI patients who received at least one dose of BNT162b2 XBB.1.5-adapted **vaccine compared to patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season, by time since last dose, among adults ≥65 years.**

¹ Vaccine effectiveness estimates are adjusted for date of symptom onset, age, sex, and number of chronic conditions.

² 'Never vaccinated' subjects were excluded when calculating the median (IQR) of time since last vaccine dose in the unexposed group (patients who did not receive *any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season).*

Statistical Analyses Supplementary Material

A.1. General framework for the COVID-19 vaccine effectiveness analysis

To address potential intra-cluster correlation among subjects from the same study site, Generalized Estimating Equations (GEE) were employed for analyzing subject-level data. In this context, a study site is regarded as a distinct cluster. Heterogeneity between study sites may exist due to the differences in recruitment, local differences in the intensity of the epidemic and healthcare practices.

A.1.1. Analysis overview

VE against hospitalization due to laboratory-confirmed COVID-19 was estimated as:

$$
VE = (1 - OR) \times 100\%
$$

where OR denotes the odds ratio calculated as the ratio of the odds of vaccination among SARS-CoV-2 test-positive cases to the odds of vaccination among SARS-CoV-2 test-negative controls.

Let K denote the number of clusters (study sites), n_k represent the number of included patients from study site k ($k =$ 1, ..., K) (from now will be referred to as cluster). We denote Y_{ik} the variable representing the infection status of subject *i* from cluster k, $Y_{ik} = 1$ [if positive-case] and $Y_{ik} = 0$ [if negative-control]. X_1 is the variable representing the vaccination status of subject *i* from cluster k, $X_{1ik} = 1$ [if vaccinated] and $X_{1ik} = 0$ [if unvaccinated or other exposure definition depending on the analyses].

Assuming that Y_{ik} follow a binomial distribution with $\pi_{ik} = P(Y_{ik} = 1 | X_{1ik})$, the probability of being a COVID-19 positive case by laboratory-confirmed RT-PCR test or other equivalent tests (from now on, we will just refer to it as a COVID-19 positive case). Conditionally on the vaccination status and using a logit-function, we can relate:

$$
logit (\pi_{ik}) = log \left(\frac{\pi_{ik}}{1 - \pi_{ik}} \right) = \beta_0 + \beta_1 X_{1ik} [1]
$$

Hence, the odds of having COVID-19 in the vaccinated group (i.e. $X_1 = 1$) is $\pi_{\text{vaccinated}}/(1 - \pi_{\text{vaccinated}})$ $\exp(\beta_0 + \beta_1)$ while the odds of having COVID-19 in the reference group (either unvaccinated this season, or unvaccinated group depending on the objective, i.e., $X_1 = 0$) is $\pi_{reference}/(1 - \pi_{reference}) = \exp(\beta_0)$. Consequently, the odds ratio (OR) of having COVID-19 in the vaccinated group compared to the reference group is estimated as:

$$
OR = \frac{\pi_{\text{vaccinated}}/(1-\pi_{\text{vaccinated}})}{\pi_{\text{reference}}/(1-\pi_{\text{reference}})} = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1).
$$

Finally, the VE is calculated as follows:

$$
VE = (1 - \exp(\beta_1)) \times 100\%
$$

The GEE approach was used to obtain the population-averaged estimate (over study sites) of the parameter of interest (in this case, β_1), adjusting for other covariates when appropriate, depending on the analyses. We computed 95% Wald CIs, assuming normal sampling distributions of the estimates. For each VE estimate and the corresponding 95% CI, the median and IQR of time since last dose (measured in days) for the exposed subjects contributing to the estimate was calculated and reported.

A.1.2. Confounder adjustment

The estimated VE was adjusted for the symptom onset date, and other covariates [confounder-adjusted VE estimate].

The confounder-adjusted VE estimate was obtained from a logistic regression model that included the following covariates: vaccination status, symptom onset date, age, sex, and the number of chronic conditions. Date of symptom onset was included in the model to adjust for potential confounding due to changes in SARS-CoV-2 incidence as well as time-varying vaccination coverage. The effects of symptom onset date and patient age were modeled using cubic regression splines. The number of chronic conditions was modelled as a categorical variable, taking values 0 (no chronic condition present), 1 (only one chronic condition present), 2 (two chronic conditions present), or ≥ 3 (at least three chronic conditions present) and were based on the presence of the chronic conditions.

A.1.3. Generalized Estimating Equations (GEE) approach

A.1.3.1. Overview of the approach

Generalized Estimating Equations (GEE), initially proposed by [2,3] is a widely used approach for analyzing clustered data. GEE employ a marginal logistic regression model to address potential intra-cluster correlation, particularly in cases where outcomes of subjects from a study site may exhibit greater relatedness compared to subjects from other study sites. Treating a study site as a cluster, with the cluster size equal to the number of subjects from the corresponding study site included in the analysis, GEE models provide population-averaged results. Notably, GEE avoids the need to specify the full joint likelihood function while doing model fitting, a challenge in cases of discrete response variables, requiring only the specification of the first two marginal moments – the mean vector and the variance-covariance matrix.

In the GEE approach, the robust sandwich estimator, which relies on a working correlation matrix, yields a consistent and asymptotically normal estimator [3]. A working correlation matrix is a matrix presenting an assumed correlation structure between observations from the same cluster. The clusters are assumed to be independent of each other. If the assumed working correlation structure is close to the truth, some efficiency might be gained. However, the estimates from GEE are consistent even if the working correlation structure is mis-specified. Exchangeable, independent, and first-order autoregressive structures are examples of commonly used working correlation structures [2].

A1.3.2. Model fitting procedure

For all analyses, we assumed an independent working correlation among outcomes of subjects within the same cluster. Seaman et al and Kahan et al [4, 5] have shown that using an exchangeable working correlation structure might lead to biased estimates in case there are cluster based confounding (CBC) and/or informative cluster sizes (ICS) present, hence an independent working correlation structure is recommended. GEE models were fit using the PGEE package in R [6]. The PGEE package implements a penalized GEE procedure, as proposed by GEE procedure, as proposed by Wang et al. (2012) [7], designed for the analysis of correlated data with high-dimensional covariates. This method accommodates situations with the diverging number of parameters to be estimated. In our analyses, however, we slightly modified the written package to allow for a "ridge" penalty term. The inclusion of the ridge penalty term is useful to handle difficulty in model fitting when there is multicollinearity present [8]. Multicollinearity may be present when there is strong correlation among predictor variables which might lead to unstable and biased standard errors [9].

There are additional challenges in the process of model fitting. One common issue is complete separation which occurs when one of the cells in the 2x2 table [outcome by exposure] is 0 [10]. As a result, the estimated odds ratio is either 0 or infinity. Using logistic regression for odds ratio estimation often generates very extreme estimates (1e-8 or 1e8, for example, depending on specific data) corresponding to unrealistic VE estimates (approximately 100% or -1e-10). This problem may remain undetected when examining the 2x2 table of outcome and exposure in the absence of relevant covariate information, referred to as quasi-complete separation. It is quite probable that, for certain covariate values, the 2x2 table [outcome x exposure] contains a 0 cell, while the overall 2x2 table does not. Our analysis likely had sparse data due to relatively small sample sizes when incorporating covariates in the model. Consequently, if any entries in the 2x2 table of exposure and outcome have a 0 count, then the VE estimate was reported as NE (non-estimable) due to the presence of complete or quasi-complete separation and to avoid running a model fitting procedure that would yield nonsensical estimates for the VE coefficient.

GEE routines typically obtain initial coefficient estimates from corresponding generalized linear models (GLM) and update these parameters iteratively until convergence. However, in the presence of separation, utilizing a regular (i.e., unpenalized) GLM routine can result in infinite estimates of the odds ratio (on the log scale). To circumvent these issues, the initial coefficients are obtained from ridge regression (i.e., a GLM with a ridge penalty) using the glmnet package [11], and these coefficients are subsequently utilized in the PGEE call. The use of glmnet also facilitates obtaining an approximate estimate of the penalty parameter used in the penalized GEE model. This penalty parameter is selected by minimizing out-of-sample binomial deviance through five-fold cross-validation, employing the one-standard error rule. This procedure is achieved by using the glmnet package [12]. The initial coefficient associated with the penalty parameter determined by the one-standard error rule is chosen as the coefficient for the PGEE fitting procedure.

To ensure reproducibility in all analyses, the seed number 23 was employed by calling set.seed (23) before the glmnet function was called. The variables passed to the glmnet call were not necessary to be standardized (i.e., not rescaled

to have 0 mean and unit variance). This is because the function itself already implements data processing procedures to ensure that the included covariates (either categorical variables or B-spline bases of the continuous variables) are on the same scale. An intercept column should be added if it is not present in the model matrix. It is worth noting that only the spline terms are penalized in the glmnet package and the PGEE calls.

Finally, it is crucial to ensure that the data passed to glmnet and PGEE for model fitting are ordered by study site to guarantee accurate results produced by PGEE. Additionally, the scale parameter (scale.fix) for the PGEE model should be set to TRUE, which corresponds to a scale.value of 1 for binomial data [13]. To guarantee reliable VE estimates, the GEE analysis was only conducted if there were a minimum of 2 subjects in each cell of the 2x2 table (exposure x outcome) without considering stratification for other covariates. This condition leads to the requirement of a total sample size of at least 8 subjects. It is essential to highlight that, depending on the analysis with the inclusion of different sets of covariates, a sample size as modest as 8 could still pose challenges in model fitting or result in VE estimates with very wide confidence intervals.

Confounder adjusted VE estimate

For the confounder adjusted VE estimate, the effects of symptom onset date (X_2) , patient age in years (X_3) , patient gender assigned at birth (X_4) , and number of chronic conditions (X_5) were included in the GEE models. Formula [1] is extended to:

$$
logit (\pi_{ik}) = \beta_0 + \beta_1 X_{1ik} + s_2(X_{2ik}) + s_3(X_{3ik}) + \beta_4 X_{4ik} + \beta_5 X_{5ik}
$$
 [1a]

Here, X_2 and X_3 are treated as continuous variables. s_2 and s_3 are (cubic spline) smooth functions of the variables representing symptom onset date (X_2) and patient age (X_3) . The use of a smooth function allows capture of a flexible relationship between the odds of having COVID-19 and symptom onset date as well as patient age. For symptom onset date, we defined 2 knots corresponding to 2024-01-15 and 2024-02-15 for the cubic spline capturing the effect of symptom onset date. This assumption corresponds to a cubic spline term with 5 degrees of freedom. For the effect of age, we performed the analysis using 2 knots (at age 50 and 65 years) for its spline term, corresponding to 5 degrees of freedom. In any analysis with less than 10% of subjects younger than 50 years old, we performed the analysis using only 1 knot (at age 65 years) instead of the spline term of age, leading to a total number of degrees of freedom equal to 4.

A.1.4. Missing data

When a subject lacked data on outcome, exposure, or specific covariates (dependent on the model), that specific subject was excluded from the analysis. This method, referred to as complete case analysis (CCA), uses only subjects with complete information for all variables included in the analysis. When data is missing completely at random (MCAR), CCA yields unbiased estimates. Under the assumption that missingness happens completely at random (MCAR mechanism), the set of complete observations is considered a random sample from the population. Thus, CCA will not produce biased estimates [14].

Given that COVID-19 status was part of the primary data collection, it was anticipated that case status (test-positive or test-negative) should be available for nearly all subjects. On the other hand, data on exposure status and, in particular, potential confounders may be missing for a portion of subjects. These details are typically obtained from pre-existing medical records, vaccine registries, or other records, which existed before the SARI episode. Consequently, it is reasonable to assume that the MCAR mechanism holds. However, the CCA approach, which utilizes a subset of the whole sample when missingness occurs, can lead to loss of precision, especially when the number of subjects with missing data is relatively high. Consequently, if the CCA makes use of a dataset with more than 10% loss in sample size in comparison with the full dataset, techniques such as multiple imputation should be implemented to generate imputed datasets for subsequent analysis.

In the current study, there was no subject excluded because of missingness of test results. In total, 0·6% (17/2711) of SARI patients were excluded from analyses due to missing data on number of chronic conditions (n=12), sex $(n=4)$ and age $(n=1)$. Consequently, we did not perform any (multiple) imputation. Results reported in the main manuscript are considered to be reasonable with respect to the handling approach of missing data.

A.2 SAMPLE SIZE CALCULATIONS, TECHNICAL SPECIFICATIONS

A.2.1. Sample size calculation

The goal of the sample size calculations was to determine the minimum sample size necessary to ensure desirable properties of the VE estimates. An example of a desirable property is the width of the 95% CI. The 95% CI width is defined as the difference between the upper and lower values of the CI and can be viewed as a measure of the precision around the VE estimate.

Simulation-based methods to determine sample sizes were developed to closely mimic the actual study design and proposed analytical approaches at the cost of requiring additional parameter assumptions and computational burden.

A.2.2. Simulation-based approach

A.2.2.1. Data generation workflow

Notation

Before describing the data generation workflow, the following parameters which act as input for the model have to be defined:

 $VE_{x,overall}$: the overall VE of exposure x, the corresponding odds ratio is $OR_{x,overall} = 1 - \frac{VE_{x,overall}}{100}$.

 $c = P(unexposed|control)$: proportion of unexposed subjects among the controls

 $P_x = P(exposure x | exposed, control)$: brand share of exposure x among the exposed

 r : ratio of cases to control (that is, number of cases per one control)

General set-up

In each simulation run, a dataset is constructed by combining data generated for a number of individual sites. We will denote the total number of study sites as k and the total sample size as N. Additionally, it is assumed that each site contributes the same number of subjects $(\frac{N}{k})$. In order to allow for variability in the underlying vaccine effects across study sites, the VE can be different from site to site. In the next section, it is described how data for one site is generated given the study site-specific VEs for all exposures. The subsequent section describes how VE is varied across the study sites to introduce between-site variability.

A.2.3. Simulating data at the site level

For each site, $\frac{N}{k} \times \frac{r}{1+r}$ cases and $\frac{N}{k} \times \frac{1}{1+r}$ controls are simulated.

Vaccine exposure status for the controls is generated from a multinomial distribution with the probability of being unexposed equal to c and the probability of being exposed to brand x equal to $(1 - c)P_x$ (with $\sum_{x=1}^{\infty} P_x = 1$).

For each $\frac{N}{k} \times \frac{r}{1+r}$ of the cases, the vaccine exposure status is then generated from a multinomial distribution with the ^k ^{1+r} of being unexposed unexposed $P(unexposed | case) = \frac{1}{1 + \sum_{x} OR_x * \frac{(1-c)P_x}{c}}$ and the probability of being exposed to brand x equal to $P(exposure x \mid case) = OR_x * \frac{(1-c)P_x}{c} * P(unexposed \mid case)$, where

$$
\frac{P(exposure x \mid case)}{P(unexposed \mid case)} = OR_{x,site}
$$

$$
\frac{P(exposure x \mid control)}{P(unexposed \mid control)}
$$

A.2.4. Simulating study site-specific VE

A.2.4.1. Effect of vaccination

To incorporate the expected between-site heterogeneity, for each study site a site-specific odds ratio ($OR_{x, site}$) was generated from a log-normal distribution with a median of $1 - \frac{VE_{X,overall}}{100}$ and variance on the log scale of 0.05. The value of the variance parameter on the log scale was selected to be 0.05 as it introduced an amount of between-site heterogeneity and was in line with the heterogeneity seen in a previous database study [1]. Note that decreasing the value of this parameter leads to a decrease in the sample size requirements. The expected value of the VE over the sites is then equal to $100 \times \left(1 - \exp\left(\log\left(1 - \frac{V_{Ex,overall}}{100}\right) + \frac{0.05}{2}\right)\right)$.

A.2.4.2. Estimates and data obtained for each simulation

For each simulated dataset, an estimate of the VE and the corresponding 95% CI was obtained using the generalized estimating equations (GEE) method:

- The expected OR on the log-scale of the treatment effect was estimated using a logistic regression model with the disease status as the outcome and the exposure as a covariate. The estimates were obtained using the GEE method in which the sites were considered clusters and the variances were calculated using a robust sandwich estimator.
- The estimated log OR and the corresponding CI were then back-transformed to obtain an estimate of the mean overall VE and its 95% CI.
- The overall VE estimate and the width of the CI are stored for each simulation.

A.2.5. Number of simulations performed

For each combination of parameter settings, a total of 100 – 200 simulations are recommended. On an empirical basis, this number of simulation runs leads to stable Monte Carlo CIs while limiting the computational burden.

A.2.6. Summary measures of the simulation study

For each combination of study characteristics, the measure of interest is obtained from all simulations, which is:

The expected range of the 95% CI is defined as the mean range of the CI obtained from all simulations.

For each measure, 95% Monte Carlo CIs were constructed based on the respective Monte Carlo standard errors observed in the simulations.

A.2.7. Sample size calculations for the study reported in this manuscript

The sample size of the number of SARI patients in the autumn/winter 2023-2024 season required to obtain a VE estimate for the primary objective with an expected width of the 95% CI meeting a prespecified expectation relied on the following assumptions: An overall vaccination coverage of 40%, a proportion among vaccinated subjects vaccinated with at least one dose of BNT162b2 Omicron XBB.1.5-adapted vaccine in the autumn/winter 2023-2024 season of 90% (scenario 1) or 95% (scenario 2), an anticipated VE of 70% (scenario 1) or 40% (scenario 2) for BNT162b2 Omicron XBB.1.5-adapted vaccine [15], and a control-case ratio of 2:1 (scenario 1) or 4:1 (scenario 2). Sample size calculations were simulation-based, assuming data collection from 10 study sites and a GEE analysis. 10 study sites corresponds to the number of study sites that were anticipated to take part in the study (as of prior to study launch).

Targeting a VE estimate with a corresponding 95% CI width of ≤50% and given the assumptions specified above, a sample size of at least 410 SARI patients was calculated [\(Table 4\)](#page-18-2). Since the sample size calculation applied for crude VE estimates (not adjusting for any covariates), we conservatively inflated the obtained sample size by a factor of 1.2 to obtain the required sample sizes for adjusted estimates. Consequently, a size of at least $410 \times 1.2 = 492$ SARI patients (with $(410-82)$ x 1.2 = 394 controls, 82 x 1.2 = 98 COVID-19 cases) included in the analysis was targeted. This sample size allows estimation of adjusted VE with a 95% CI width ≤50%, assuming VE for BNT162b2 Omicron XBB.1.5-adapted vaccine of 70%, an overall vaccination coverage of 40%, and a BNT162b2 Omicron XBB.1.5 adapted vaccine proportion of 90%, with a control:case ratio of 4:1.

Table 1. The required number of SARI patients and COVID-19 cases to allow estimation of VE with an expected 95% CI width ≤50% based on a study including 10 sites, and overall vaccination coverage of 40%.

Abbreviations: CI, confidence interval; VE, vaccine effectiveness; SARI, severe acute respiratory infection

A.2.6. Adjusted sample size taking into account matching of cases and controls

Matching of COVID-19 cases and controls by study site and temporal proximity of symptom onset (2 weeks) was performed to mitigate potential selection bias and confounding by time trends. The matching process, however, was anticipated to result in a reduction of the sample size for analysis by approximately 40% based on observations from data for the autumn/winter 2022-2023 season (hospital admission between 02 October 2022 – 31 May 2023), given a matching ratio of 1 (positive) case to a maximum of 4 (negative) controls. Consequently, to achieve a sample size of 492 SARI patients for the estimation of VE for the primary objective, $492 \div (1 - 0.4) = 820$ SARI patients were required.

A.2.7. References

- 1. THOMPSON, M. G., STENEHJEM, E., GRANNIS, S., et al. (2021). Effectiveness of Covid-19 Vaccines in Ambulatory and Inpatient Care Settings. N Engl J Med, $385(15):1355-71$. DOI: Ambulatory and Inpatient Care Settings. N Engl J Med, 385(15):1355-71. DOI: [https://doi.org/10.1056/NEJMoa2110362.](https://doi.org/10.1056/NEJMoa2110362)
- 2. LIANG, K.-Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73(1):13-22. DOI[: https://doi.org/10.1093/biomet/73.1.13.](https://doi.org/10.1093/biomet/73.1.13)
- 3. ZEGER, S. L. & LIANG, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. Biometrics 4(1):121–130.
- 4. SEAMAN, S., PAVLOU, M. & COPAS, A. (2014). Review of methods for handling confounding by cluster and informative cluster size in clustered data. Stat Med, 33(30):5371-87. DOI: [https://doi.org/10.1002/sim.6277.](https://doi.org/10.1002/sim.6277)
- 5. KAHAN, B. C., LI, F., COPAS, A. J., et al. (2023). Estimands in cluster-randomized trials: choosing analyses that answer the right question. Int J Epidemiol, 52(1):107-18. DOI: 10.1093/ije/dyac131.
- 6. INAN, W. & WANG, L. (2017). PGEE: An R Package for Analysis of Longitudinal Data with High-Dimensional Covariates. The R Journal, 9(1):393-402. DOI: 10.32614/RJ-2017-030.
- 7. WANG, L., ZHOU, J. & QU, A. (2012). Penalized generalized estimating equations for high-dimensional longitudinal data analysis. Biometrics, 68(2):353-60. DOI: [https://doi.org/10.1111/j.1541-](https://doi.org/10.1111/j.1541-0420.2011.01678.x) [0420.2011.01678.x.](https://doi.org/10.1111/j.1541-0420.2011.01678.x)
- 8. HOERL, A. E. & KENNARD, R. W. (2000). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 42(1):80-6. DOI: 10.1080/00401706.2000.10485983.
- 9. VATCHEVA, K. P., LEE, M., MCCORMICK, J. B., et al. (2016). Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. Epidemiology (Sunnyvale), 6(2)DOI: 10.4172/2161-1165.1000227.
- 10. MANSOURNIA, M. A., GEROLDINGER, A., GREENLAND, S., et al. (2018). Separation in Logistic Regression: Causes, Consequences, and Control. Am J Epidemiol, 187(4):864-70. DOI: 10.1093/aje/kwx299.
- 11. FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw, 33(1):1-22. DOI.
- 12. TAY, J. K., NARASIMHAN, B. & HASTIE, T. (2023). Elastic Net Regularization Paths for All Generalized Linear Models. J Stat Softw, 106DOI: 10.18637/jss.v106.i01.
- 13. HARDIN, J. W. & HILBE, J. M. (2013). Generalized Estimating Equations (2nd ed.), New York, Chapman and Hall/CRC.
- 14. LEE, S.-Y. (2011). Handbook of latent variable and related models., Elsevier.
- 15. HOLM HANSEN, C., MOUSTSEN-HELMS, I. R., RASMUSSEN, M., et al. (2023). Effectiveness of the XBB.1.5 updated COVID-19 vaccine against hospitalisation: a nation-wide cohort study in Denmark, October 2023. Available from: https://ssrn.com/abstract=4627268 [Accessed: 01-12-2023].

A.3. Matching cases and controls

Up to 4 test-negative controls were exact matched to each case, based on

- Study site (and thus, Country)
- Temporal proximity of symptom onset (2 weeks)

Matching was conducted in R with an exact matching methodology. The number of controls matched to each case was examined. When more than 4 controls were available to match to 1 case, we selected controls based on distance between symptom onset dates between the controls and the case, with the rule that shorter distance had priority.

Matching was performed separately for each VE analysis. For example, for the analysis of the primary objective, only cases and controls who met the definitions of "vaccinated with at least one dose of BNT162b2 XBB vaccine" or "had not been vaccinated against COVID-19 during the 2023–2024 season" were matched. In the end, for the primary objective analyses, 8 test- positive cases were removed due to no matched controls found. There were 814 test-negative controls removed to reduce the case:control ratio from 1:6 to 1:4 at most.

A.4. Sensitivity analysis results

In this section, we provide results from various sensitivity analyses for all VE estimates reported in Figure 1 in the main manuscript.

A.4.1 Sensitivity analysis with regard to location and number of knots applied to symptom onset dates

Figure A.4.1a [Sensitivity analysis 1a] Vaccine effectiveness¹ against COVID-19 hospitalization in SARI patients who received at least one dose of **BNT162b2 XBB.1.5-adapted vaccine compared to patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season. The location of knots for symptom onset dates were placed at 10Jan2024, 31Jan2024, and 21Feb2024 (in increments of 3 weeks).**

¹Vaccine effectiveness estimates are adjusted for date of symptom onset, age, sex, and number of chronic conditions.

²*'Never vaccinated' subjects were excluded when calculating the median (IQR) of time since last vaccine dose in the unexposed group (patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season).*

³ The categories "Cancer" and "Immunodeficiency" among chronic conditions are counted as separate types of chronic conditions in all adjusted VE estimates. Here, the VE *estimate among participants having "Immunodeficiency or cancer" is provided as a composite.*

Figure A.4.1b [Sensitivity analysis 1b] Vaccine effectiveness¹ against COVID-19 hospitalization in SARI patients who received at least one dose of **BNT162b2 XBB.1.5-adapted vaccine compared to patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter** season. The location of knots for symptom onset dates were placed at 15Jan2024, 31Jan2024, and 15Feb2024.

¹ Vaccine effectiveness estimates are adjusted for date of symptom onset, age, sex, and number of chronic conditions.

² 'Never vaccinated' subjects were excluded when calculating the median (IQR) of time since last vaccine dose in the unexposed group (patients who did not receive *any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season).*

3 The categories "Cancer" and "Immunodeficiency" among chronic conditions are counted as separate types of chronic conditions in all adjusted VE estimates. Here, the VE estimate among participants having "Immunodeficiency or cancer" is provided as a composite.

A.4.2 Sensitivity analysis with regard to number and location of knots for age

Figure A.4.2a [Sensitivity analysis 2a] Vaccine effectiveness¹ against COVID-19 hospitalization in SARI patients who received at least one dose of **BNT162b2 XBB.1.5-adapted vaccine compared to patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season. The location of knots for age were placed at 60 and 75 years.**

¹Vaccine effectiveness estimates are adjusted for date of symptom onset, age, sex, and number of chronic conditions.

²*'Never vaccinated' subjects were excluded when calculating the median (IQR) of time since last vaccine dose in the unexposed group (patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season).*

³ The categories "Cancer" and "Immunodeficiency" among chronic conditions are counted as separate types of chronic conditions in all adjusted VE estimates. Here, the VE *estimate among participants having "Immunodeficiency or cancer" is provided as a composite.*

Figure A.4.2b [Sensitivity analysis 2b] Vaccine effectiveness¹ against COVID-19 hospitalization in SARI patients who received at least one dose of **BNT162b2 XBB.1.5-adapted vaccine compared to patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season. The location of knots for age were placed at 50, 65, and 80 years.**

¹Vaccine effectiveness estimates are adjusted for date of symptom onset, age, sex, and number of chronic conditions.

²*'Never vaccinated' subjects were excluded when calculating the median (IQR) of time since last vaccine dose in the unexposed group (patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season).*

³ The categories "Cancer" and "Immunodeficiency" among chronic conditions are counted as separate types of chronic conditions in all adjusted VE estimates. Here, the VE *estimate among participants having "Immunodeficiency or cancer" is provided as a composite.*

A.4.3 Sensitivity analysis with regard to VE estimates stratified by history of prior COVID-19 infection

Figure A.4.3 Vaccine effectiveness¹ against COVID-19 hospitalization in SARI patients who received at least one dose of BNT162b2 XBB.1.5-adapted **vaccine compared to patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season. Results for VE estimates stratified by history of prior COVID-19 infection are presented.**

¹ Vaccine effectiveness estimates are adjusted for date of symptom onset, age, sex, and number of chronic conditions.

² 'Never vaccinated' subjects were excluded when calculating the median (IQR) of time since last vaccine dose in the unexposed group (patients who did not receive *any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season).*

A.4.4 Sensitivity analysis with regard to VE estimates obtained from a matched analysis with matched units as cluster

Figure A.4.4 Vaccine effectiveness¹ against COVID-19 hospitalization in SARI patients who received at least one dose of BNT162b2 XBB.1.5-adapted **vaccine compared to patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season, adjusted for cluster covariate matching**

¹Vaccine effectiveness estimates are adjusted for date of symptom onset, age, sex, and number of chronic conditions.

²'Never vaccinated' subjects were excluded when calculating the median (IQR) of time since last vaccine dose in the unexposed group (patients who did not receive any dose of a *COVID-19 vaccine in the 2023–2024 autumn/winter season).*

³ The categories "Cancer" and "Immunodeficiency" among chronic conditions are counted as separate types of chronic conditions in all adjusted VE estimates. Here, the VE *estimate among participants having "Immunodeficiency or cancer" is provided as a composite.*

A.4.5 Sensitivity analysis with regard to VE estimates obtained from Spanish data only

Figure A.4.5 Vaccine effectiveness¹ against COVID-19 hospitalization in SARI patients who received at least one dose of BNT162b2 XBB.1.5-adapted **vaccine compared to patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season in Spain.**

¹Vaccine effectiveness estimates are adjusted for date of symptom onset, age, sex, and number of chronic conditions.

²^{*'Never vaccinated' subjects were excluded when calculating the median (IQR) of time since last vaccine dose in the unexposed group (patients who did not receive any dose of a} COVID-19 vaccine in the 2023–2024 autumn/winter season).*

³ The categories "Cancer" and "Immunodeficiency" among chronic conditions are counted as separate types of chronic conditions in all adjusted VE estimates. Here, the VE *estimate among participants having "Immunodeficiency or cancer" is provided as a composite.*

A.4.6 Sensitivity analysis with regard to VE estimates while accounting additionally for influenza vaccination

Figure A.4.6. Vaccine effectiveness¹ against COVID-19 hospitalization in SARI patients who received at least one dose of BNT162b2 XBB.1.5-adapted **vaccine compared to patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season, adjusted additionally for 2023-2024 influenza vaccination status.**

¹Vaccine effectiveness estimates are adjusted for 2023-2024 influenza vaccination status, date of symptom onset, age, sex, and number of chronic conditions.

²^{*'Never vaccinated' subjects were excluded when calculating the median (IQR) of time since last vaccine dose in the unexposed group (patients who did not receive any dose of a} COVID-19 vaccine in the 2023–2024 autumn/winter season).*

³ The categories "Cancer" and "Immunodeficiency" among chronic conditions are counted as separate types of chronic conditions in all adjusted VE estimates. Here, the VE *estimate among participants having "Immunodeficiency or cancer" is provided as a composite.*

A.4.7 Sensitivity analysis with regard to VE estimates where we separately adjusted for hypertension condition

Figure A.4.7. Vaccine effectiveness¹ against COVID-19 hospitalization in SARI patients who received at least one dose of BNT162b2 XBB.1.5-adapted **vaccine compared to patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season, adjusted separately for hypertension condition.**

¹ Vaccine effectiveness estimates are adjusted for date of symptom onset, age, sex, number of chronic conditions, and hypertension condition. For this specific analyses, we counted the number of chronic conditions based on the list specified in the manuscript without hypertension as this condition is added as an additional (binary) covariate into the model.

² 'Never vaccinated' subjects were excluded when calculating the median (IQR) of time since last vaccine dose in the unexposed group (patients who did not receive *any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season).*

³ The categories "Cancer" and "Immunodeficiency" among chronic conditions are counted as separate types of chronic conditions in all adjusted VE estimates. *Here, the VE estimate among participants having "Immunodeficiency or cancer" is provided as a composite.*

A.4.8 Sensitivity analysis with regard to VE estimates obtained using stricter criteria (symptom onset within 7 days)

Figure A.4.8.1 Vaccine effectiveness¹ against COVID-19 hospitalization in SARI patients who received at least one dose of BNT162b2 XBB.1.5-adapted **vaccine compared to patients who did not receive any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season, whose symptom onset date was within 7 days prior to and up to 72 hours after the hospital admission date.**

¹ Vaccine effectiveness estimates are adjusted for date of symptom onset, age, sex, and number of chronic conditions.

² 'Never vaccinated' subjects were excluded when calculating the median (IQR) of time since last vaccine dose in the unexposed group (patients who did not receive *any dose of a COVID-19 vaccine in the 2023–2024 autumn/winter season).*

3 The categories "Cancer" and "Immunodeficiency" among chronic conditions are counted as separate types of chronic conditions in all adjusted VE estimates. Here, the VE estimate among participants having "Immunodeficiency or cancer" is provided as a composite.