**Title: All-cause Mortality in Relation to COVID-19 Vaccination: Analysis of the UK ONS Public Data**

**Abstract:**

The widespread availably of COVID-19 vaccines has presented a unique opportunity to study real-world impacts on population-wide mortality. Although clinical trials demonstrated high vaccine efficacy against COVID-19, there are still questions about the vaccines’ effectiveness in reducing all‐cause mortality, especially in the context of the ever‐changing pandemic. Using publicly available data from the UK Office for National Statistics (ONS), this study examines whether vaccination status correlates with overall mortality rate in England from April 2021 to May 2023. We employ both rate ratio (RR) calculation, stratified by vaccine dose and timing, and Negative Binomial regression to account for overdispersion and evaluate the net association between vaccination and mortality. While monthly RRs occasionally exceed 1.0 (suggesting higher mortality in certain dose-age combinations), the aggregated Negative Binomial model indicates a consistent protective effect across all age groups, with vaccinated individuals generally experiencing lower all-cause mortality. We discuss potential sources of bias, including underestimation of the unvaccinated population, health-vaccine bias, indication bias, and the harvesting effect, that could explain seemingly contradictory findings. Despite data limitation, our results underscore the complexity of interpreting observational vaccine data and the apparent overall benefit of COVID-19 vaccination in reducing all-cause mortality.

**1. Introduction**

With the outbreak of the COVID-19 pandemic, global public health has faced unprecedented challenges. As of October 2021, over 240 million confirmed COVID-19 cases and more than 4.8 million related deaths have been reported worldwide (WHO, 2021b). Several measures were adopted by governments of the globe (travel restrictions, lockdowns, mask mandates and finally mass vaccination campaigns) to reduce the number of deaths. These include the development of several vaccines, including mRNA-based formulations (Pfizer-BioNTech’s BNT162b2 and Moderna’s mRNA-1273), adenovirus-vectored vaccines (Oxford-AstraZeneca’s ChAdOx1 nCoV-19), and others, represented landmark in pandemic response efforts.

Although clinical trials have shown that these vaccines are very efficient in preventing the symptoms of COVID 19 and reducing mortality, there is ongoing debate concerning their broader influence on all-cause mortality (Baden et al., 2021; Skowronski & De Serres, 2021). In the early days of pandemic, numerous observational studies focused specifically on COVID-19-related deaths. However, measuring the effect of vaccination on all-cause mortality may yield additional insights, reflecting whether vaccines confer “spillover” benefits or, conversely, carry unforeseen risks that could influence non–COVID-19 deaths.

Excess mortality, which is the number of deaths above the seasonal or historical average, has been noted to rise in multiple countries (Schöley et al., 2019; Yao et al., 2023). England and Wales witnessed a surge in excess mortality that began in March 2022 and lasted most of the year 2022 (Hussain, 2022). From April 2021 to May 2023, excess deaths in these two countries amounted to approximately 130,000 above the five-year average (Office for National Statistics, 2023).

The United Kingdom maintains one of the most detailed public health data collection systems worldwide, with the Office for National Statistics (ONS) regularly publishing mortality statistics stratified by age group and COVID-19 vaccination status. With more than 75% of the eligible population vaccinated by the end of 2021 (Our World in Data, 2024), England and Wales offer a robust data environment to explore whether vaccination truly influences overall mortality. Studying all-cause mortality rather than solely COVID-19 mortality can shed light on broader health implications—both positive and negative arising from large-scale immunization efforts.

This study utilizes ONS data on all-cause mortality to examine whether vaccination status and timing are associated with overall death rates in England between April 2021 and May 2023. The analysis includes the computation of rate ratios (RRs) for various vaccination categories (e.g., first dose ≥ 21 days, second dose ≥ 21 days) relative to unvaccinated individuals, stratified by age group and month. Additionally, Poisson and Negative Binomial regression models are used to account for overdispersion in mortality counts and

evaluate the net effect of vaccination across

different age brackets. Potential biases—such as

the underestimation of unvaccinated populations, healthy-vaccine bias, indication bias, and the harvesting effect—are also discussed to provide context for the findings.

By combining descriptive and inferential approaches, this study aims to offer a greater understanding of how vaccine-induced effects vary over time, across doses, and among age groups, while also addressing limitations in the ONS dataset, such as missing comorbidity data and incomplete temporal coverage.

**2. Methods**

**2.1 Data Collection and Preprocessing**

All data were obtained from the UK Office for National Statistics (ONS) platform, including total mortality by vaccination status from April 2021 to May 2023. These data are publicly available and can be freely analysed and published, provided that the source is properly acknowledged.

From the ONS Excel file, we used “Table 2,” which stratifies all-cause deaths by age and vaccination status, thereby allowing for the calculation of relative risks (RR). We excluded any deaths classified as COVID-19 because, for many months—especially early in the pandemic and in younger age brackets—the number of COVID-19 deaths was less than three, preventing reliable RR calculations. The spreadsheet defines seven age groups (18–39, 40–49, 50–59, 60–69, 70–79, 80–89, 90+). Each group is further broken down by vaccination status: Unvaccinated, First dose less than 21 days, First dose at least 21 days, Second dose less than 21 days, Second dose at least 21 days, Third dose or booster less than 21 days, Third dose or booster at least 21 days.

Where the death count was listed as “<3,” we recoded these values to 0. This step enabled us to visualize mortality trends for each age group over time despite early low vaccination rates. We then created stacked graphs to examine the monthly distributions of mortality by vaccination status for each age group (see Supplementary Materials, Tables S1–S7 and Figures S1–S7).

**2.2 Statistical analysis**

**2.2.1 Computing Relative Risk (RR)**

We first computed the relative risk (RR) for all-cause mortality comparing each vaccinated group with the unvaccinated, based on the age-standardized rates available from the ONS (Office for National Statistics, 2023). The RR is: Their 95% confidence intervals were calculated according to the following formula (Soliani, 2008; Norman & Streiner, 2015):

(Equation 1)

Confidence intervals at 95% were calculated according to the following formula (Soliani, 2008; Norman & Streiner, 2015):

(Equation 2)

where “ln(RR)” is the natural log of the RR and “SEln(RR)” is the standard error of the natural log of the RR.

To obtain SEln(RR) or each vaccination status by age group, we applied (Alessandria et al., 2024):

(Equation 3)

where, “V.Pop.” is the vaccinated population, “Un.Pop.” is the unvaccinated population, and “Exp.Stand.D.” is the expected standardized death – calculated by multiplying the standardized rate by the actual population, divided by 100,000 (Equation 3)

(Alessandria et al., 2024).

(Equation 4)

Because RR is a ratio of two standardized rates, expected standardized deaths adjust for differences in population composition

We then calculated the p-value following Altman and Bland (2011):

)(Equation 5)

where z = and SE is from Equation 1

**2.2.2 Poisson Regression**

Next, we assessed whether vaccination status (vaccinated vs. unvaccinated) was associated with mortality using a Poisson regression model. Poisson regression is suited to count data, such as total deaths, assuming the

mean equals the variance (Ogallo et al., 2022). We began by filtering for deaths attributed to “All causes,” then recoded vaccination status

as a binary variable: “Vaccinated” (any dose) vs. “Unvaccinated” (no dose). The Poisson model takes the form:

(Equation 6)

where:

is the expected number of deaths in group i

is the intercept, representing the baseline log-mortality rate for the unvaccinated group,

quantifies the log-rate difference between vaccinated and unvaccinated individuals, and

ln(PersonYears) is included as an offset to account for varying exposure time across groups, ensuring the model estimates mortality rates rather than raw events counts.

The exponentiated coefficient exp () yields the rate ratio (RR), which quantifies the relative mortality risk for vaccinated compared to unvaccinated individuals. RR below 1 suggests a protective effect of vaccination. In the contrast, RR larger than 1 indicates higher mortality risk in the vaccinated group.

**2.2.3 Addressing Overdispersion**

In many real-world mortality datasets, variance often exceeds the mean, leading to *overdispersion* (Ibarra-Espinosa et al., 2022). We identified this issue by observing that the variance in death counts was substantially larger than the mean across all age groups (for instance, a mean of 238.83 vs. a variance of 16,564.34 for ages 18–39). Consequently, Poisson regression alone was unsuitable for these data.

To address this, we employed Negative Binomial regression models, which extend the Poisson model by introducing a dispersing parameter ( that accounts for the extra variability. This adjustment allows the variance to exceed the mean, with the relationship (Xie, 2020):

(Equation 7).

where represents the expected count of deaths for group i, and is the estimated dispersion parameter. The Negative Binomial maintains the same log‐linear structure as Poisson regression, such that exponentiating the coefficient of the “Vaccinated (vs. Unvaccinated)” variable again gives the RR:

(Equation 8)

As before, an RR < 1 implies a protective effect of vaccination, whereas an RR > 1 would indicate higher mortality among the vaccinated group.

**2.3 Data Visualization**

To illustrate the findings, we utilized several visualization techniques. First, whisker and box plots presented the relative risk (RR) of all-cause mortality in vaccinated individuals compared to varied vaccination statuses. A vertical reference line at RR = 1 indicated no difference in risk. In addition, we plotted logarithmic mortality rates by vaccination status to illustrate the log-linear assumptions underlying the Negative Binomial model and to show how these assumptions manifest across different age groups. Collectively, these visual tools revealed a consistent pattern of reduced mortality risk

(RR < 1) associated with vaccination in most age groups, with RRs approaching zero indicating a stronger protective effect.

All data processing was performed in RStudio (version 2024.12.0).

**3. Results:**

**3.1 About ONS dataset**

We used the latest version of the UK Office

for National Statistics (ONS) dataset, which draws on the 2021 Census and links deidentified census records to National Health Service (NHS) numbers. Individuals lacking

an NHS number or those with multiple entries are excluded. Consequently, the dataset

covers 51,786,812 people—over 90% of England’s population on Census Day 2021 (Office for National Statistics, 2022). An additional 103,142 individuals were excluded due to erroneous or inconsistent vaccination data, bringing the total excluded population to approximately 4.7 million.

**3.2 Rate Ratios (RR)**

Whisker and box plots of all‐cause mortality rate ratios (RR) by vaccination status are displayed in Supplementary Materials Section Figures S1-S7. These plots compare each vaccinated group (e.g., first dose ≥21 days, second dose ≥21 days, etc.) to the unvaccinated population across different age brackets. The main findings are:

**First dose (21days)**

For individuals who received their first dose at least 21 days prior, the boxplots indicate elevated rate ratios (RRs) compared to unvaccinated individuals in most age groups. In the 18–39 group, the RR shows significant variation, with a median value slightly above 1.0 but with a range extending up to 4–5, suggesting elevated mortality risk during certain periods. For older age groups (e.g., 60–69 and 70–79), the median RR increases, with upper whiskers and outliers indicating occasional periods of very high relative risk (e.g., >4). These findings suggest that early post-vaccination periods might include elevated risks, likely driven by specific underlying factors.

**Second Dose (21days)**

The RR values for the second dose (≥21 days) show a more balanced trend compared to the first dose. In the 40–49 and 50–59 groups, the boxplots demonstrate that the median RR is closer to 1.0, indicating a reduction in relative risk compared to the first dose. However, for older groups, such as 70–79 and 80–89, the RRs still exhibit variability, with the median slightly exceeding 1.0 in some cases. The overall trend suggests a decline in risk over time, although certain periods exhibit elevated risks.

**Third Dose (21days)**

For the third dose (≥21 days), the RRs generally display a protective effect in most age groups. The median RRs often fall below 1.0, particularly in younger groups such as 40–49 and 50–59. In older groups (e.g., 70–79 and 80–89), the median values hover around 1.0, with a few outliers exceeding 2.0, reflecting variability. The range of RRs narrows for most age groups compared to earlier doses, indicating more consistent trends.

**Fourth Dose (21days)**

The fourth dose (≥21 days) demonstrates a strong protective effect in nearly all age groups. The median RR values are below 1.0 across the board, with tight interquartile ranges and minimal outliers. Notably, in the 70–79 and 80–89 age groups, the fourth dose shows the most consistent protective trend, with RRs concentrated well below 1.0. The younger groups, such as 40–49, also exhibit a similar protective trend, with no significant spikes or outliers.

**Other Vaccination Categories (<21 days across doses)**

For individuals in the "less than 21 days" category across doses, the boxplots show significant variability in RRs, often exceeding 1.0. This trend is most pronounced in the first and second doses, where the upper whiskers extend beyond 2.0 or

3.0 in some age groups. For subsequent doses, such as the third and fourth, the RRs stabilize and tend to cluster closer to or below 1.0.

**3.3: Negative Binomial Regression**

A table of numbers and a number of numbers

Description automatically generated with medium confidence

Figure 8: Negative Binomial Regression analysis by age groups

Figure 8 presents a Negative Binomial regression by age group, where the dependent variable is the count of death in

each group. In this model, the coefficients for the “Vaccinated (Vs. Unvaccinated)” variable reflects the log-count difference between vaccinated and unvaccinated individuals; hence, a negative coefficient indicates a protective effect of vaccination on all-cause mortality. Exponentiating each coefficient provides a rate ratio (RR), or the ratio of the estimated death rate for the vaccinated population compared to the unvaccinated population. A negative coefficient below 0 translates to an RR below 1, indicating lower mortality risk among those vaccinated.

In detail, all coefficients for “Vaccinated (Vs. Unvaccinated)” are negative and statistically significant (p-value <0.01). For instance, in the 18-39 age group, the coefficient is -0.157(p<0.05), corresponding to rate ratio of approximately . Interpreted substantively, this suggests that vaccinated individuals in this age group face about a 15% lower death rate compared to those who are unvaccinated. In the 40-49 group, the coefficient is -0.438 (p<0.01), translating to a rate ratio of about 0.65, or a roughly 35% reduction in death rates among the vaccinated population. The 50-59 and 60-69 age groups both have coefficients around -0.612 ~-0.622 (p<0.01), indicating an even stronger protective effect of vaccination, where mortality rates are lowered by approximately 46% compared to unvaccinated individuals.

Among the older groups, the coefficients remain negative and significant, but the magnitude is smaller. For 70-79-year-olds, the coefficient of -0.537 (p<0.01) corresponds to a rate ratio near 0.58, implying that vaccination reduces death rates by around 42%. In the 80-89 bracket, the coefficient is -0.292 (p<0.01), giving a rate ratio around 0.75, or a 25% reduction in mortality. Finally, even in the 90+ age group, the coefficient remains negative at -0.093, associated with an RR of about 0.91. While this indicates a comparatively smaller protective effect (around 9% lower mortality), it still shows that vaccination confers a statistically significant reduction in risk among the oldest individuals.

In summary, while the RRs (section 3.2) suggest that some dose-age combinations experience increased risks over specific intervals, the overall Negative Binomial regression (combining all vaccination statuses) indicates that, on balance, vaccinated individuals in every age group face a significantly lower rate of all-cause mortality than their unvaccinated counterparts.

**4. Discussion**

In this study, we investigated all‐cause mortality trends in the UK ONS dataset, stratified by vaccination status. The boxplot and Negative Binomial regression analyses collectively reveal a nuanced relationship between vaccination status and all-cause mortality. The boxplots illustrate the variability in rate ratios (RRs) across vaccination statuses and age groups, showing elevated RRs for the first and second dose (≥21 days) and gradual stabilization with subsequent doses. Protective effects become more evident with the third and fourth doses, particularly in middle-aged and older groups.

Furthermore, Figure 9 (Negative Binomial regression results) illustrates that when considering the vaccinated population in aggregate (i.e., combining all doses), all‐cause mortality is consistently lower than in the unvaccinated group across every age bracket. This protective effect appears strongest among middle‐aged adults (40–79 years), likely

due to moderate baseline mortality and fewer comorbidities than those in the 90+ years bracket. Although the effect size in the oldest age groups is smaller—often because of multiple chronic conditions and frailty—the coefficients remain negative and statistically significant, indicating that vaccination status is firmly and consistently linked with reduced all‐cause mortality in this dataset.

A graph with black dots and white text

Description automatically generated

Figure 9: Vaccination effect on mortality by age group in negative binomial models

Our findings are partially consistent with studies conducted in Italian provinces (Acuti et al., 2022; Rosso et al., 2023), which also identified higher all-cause mortality risks among individuals with one or two doses compared to the unvaccinated. Additionally, those studies reported large protective effects for three-dose recipients, beyond what could be attributed solely to COVID-19 mortality. However, these findings were later attributed to significant biases, notably Immortal Time Bias (ITB) (Berrino et al., 2023), which occurs when the survival period is misclassified, preventing participants from experiencing the event of interest (Vail et al., 2021). Correcting for ITB in those studies rendered the apparent strong protection from third-dose vaccination non-significant. Similarly, Alessandria et al. (2024) reported higher all-cause death hazard ratios for individuals with one or two doses and no significant protection among those with three or more doses in the same dataset, along with a modest but statistically significant reduction in life expectancy among populations receiving two or three doses.

Our own data also indicate that certain age groups receiving one or two doses showed higher mortality risk variability, while the Negative Binomial regression underscored the aggregate protective effect of vaccination. Balancing these seemingly contradictory results remains challenging, given the limited information on individual health status and other potential factors, we propose several hypotheses in subsequent sections.

**4.1 Underestimation of the unvaccinated population**

One possible explanation for the contradictory patterns—such as instances where the vaccinated group appears to have a higher all-cause mortality risk in certain months or age groups—is that the UK Office for National Statistics (ONS) dataset might underestimate the true size of the unvaccinated population. As outlined in Section 3.1, approximately 4.7 million individuals are excluded because they lacked a National Health Service (NHS) number or had other data inconsistencies. Moreover, those without an NHS number or General Practitioner (GP) registration typically cannot receive a COVID-19 vaccine. These exclusions disproportionately affect unvaccinated people; when such individuals die, their deaths are still certified and captured by the ONS with unvaccinated population status, since the death registration process requires vaccination status. Consequently, a systemic bias could arise throughout the study period.

Section 3.2 illustrates this potential underestimation. The elevated RRs for one and two doses across all age groups could be significantly influenced by the larger-than-actual unvaccinated population in the denominator. If a considerable number of unvaccinated individuals are excluded from the dataset, the true RR value may be lower than recorded—skewing comparisons and making the “observed” mortality in the unvaccinated group appear higher than it is

in reality. Consequently, vaccinated groups might seem to have worse outcomes in certain months when, in fact, the high RR could reflect an artificially inflated unvaccinated mortality denominator. The discrepancy between the boxplot RRs and the regression results underscores the impact of underestimating the unvaccinated population. The regression analysis provides a more accurate estimate of the protective effect by mitigating some of these biases through its modelling approach.

**4.2 The Healthy-vaccine Bias**

A second key factor suggested by Sections 3.2 and 3.3is the healthy‐vaccine bias, which individuals who receive vaccines often exhibit healthier behaviours or characteristics than those who do not (Furst et al., 2024). As shown in Section 3.2, people who are 21 days past their second or third dose typically display rate ratios (RR) below 1.0, indicating a protective effect of vaccination. Furthermore, voluntary adherence to preventive measures (e.g. vaccination) has been shown to reduce mortality by nearly half, and in some cases with reduction of 2.5–3 or more times (Wright et al., 2021; Pinsky et al., 2007). However, in the short term (the first 21 days after dose 1), RRs often hover at or above 1.0, potentially because individuals with serious illnesses defer vaccination, temporarily inflating mortality rates among those who get vaccinated without delay (Furst et al., 2024). Over time, once these very frail individuals remain unvaccinated, the protective effect among the relatively healthier vaccinated population may become more apparent.

Meanwhile, the Negative Binomial analysis reveals consistently negative coefficients for “Vaccinated vs. Unvaccinated” across all age groups, determining a statistically significant reduction in all‐cause mortality. This protective effect may partly relate to healthier behaviours—such as better diets, more exercise, and lower tobacco or alcohol use—among the vaccinated (Eiser, 2021), However, these factors are not captured in detail in the ONS dataset, highlighting the role of unmeasured health-related lifestyle factors in amplifying the observed benefits of vaccination. In addition, Immortal Time Bias could further inflate survival times for certain vaccinated groups, if the period before actual vaccination is misclassified as “vaccinated” time ([Hernán](https://pubmed.ncbi.nlm.nih.gov/?term=Hern%C3%A1n+MA&cauthor_id=39494894), 2025).

A related phenomenon is indication bias. Early vaccination campaigns prioritized clinically vulnerable individuals, including older adults with multiple health conditions, who often received doses earlier (Apampodi et al., 2024). This prioritization may have temporarily elevated short-term RR values among older recipients, as their pre-existing health risks were disproportionately represented in the early vaccinated cohort. Over time, as healthier older adults joined the unvaccinated pool, RR estimates declined. This trend is particularly evident in Section 3.2, where individuals aged 70–79 and 80–89 initially exhibit RRs of 4-5 after their first dose, which later fall below 2. Such shifts underscore how cohort composition can yield seemingly contradictory findings about vaccine efficacy; the vaccine may appear less effective initially when early recipients represented a frailer subgroup. As the frail fraction’s influence diminishes, the protective effect of vaccination becomes clearer.

The healthy-vaccine bias also explains variations in the protective effect observed across age groups in the Negative Binomial analysis. Middle-aged adults (40–79 years) derive a more pronounced benefit from vaccination due to their relatively healthier baseline characteristics and fewer comorbidities (Bonanad et al., 2020). Conversely, the oldest age group (90+ years), characterized by higher frailty and chronic conditions, exhibits a weakened protective effect. For the 18–39 age group, the protective effect is less apparent, largely because of their already low baseline mortality rates (Bhopal et al., 2021). Even substantial relative reductions in mortality result in smaller absolute differences in this cohort. Additionally, younger adults may engage in behaviours, such as higher social interaction or different occupational exposures, that hinder the observable impact of vaccination on mortality.

In summary, the healthy-vaccine bias significantly enhances the apparent benefits of vaccination in reducing mortality, particularly among age groups with better baseline health profiles. However, the interplay of unmeasured health behaviours and cohort dynamics complicates efforts to isolate the vaccine’s independent protective effect.

**4.3 The harvesting effect**

A further hypothesis is the harvesting effect, sometimes referred to as “mortality displacement.” This concept describes a short‐term shift in mortality or severe health outcomes, whereby deaths or illnesses occur slightly earlier than they otherwise would, yet the overall long‐term mortality rate remains relatively unchanged (Walkowiak et al., 2023). In the context of COVID‐19 vaccination, the first dose often shows a pronounced initial mortality peak—likely because priority was given to the most clinically fragile individuals, many of whom died sooner due to their underlying conditions.

The early loss of these frail individuals means that those who proceed to receive the second dose are generally healthier. Indeed, because vaccination initially targeted older age groups, by April 2021, many older adults had already completed their first dose and started on their second. This phenomenon can only observe in younger groups (18–39 and 40–49), where we can determine a distinct initial mortality peak associated with the first dose—perhaps already in decline—followed by a smaller peak for the second dose, starting from lower values and reaching a reduced maximum compared to that of the first dose.

With respect to the third dose, the initial mortality peak is essentially absent across all age cohorts. One possible explanation is that many of the most vulnerable individuals died before reaching the third‐dose stage, thereby reducing the frailest segment of the population eligible for subsequent doses (Riou et al., 2023). In this way, the harvesting effect may help account for the noticeably low-rate ratios (RR) that appear early on for the third dose, and average protective effect shown in the negative binomial analysis, illustrating how selective survival in a population can shape vaccination‐mortality patterns over time.

**5. Limitation**

A significant limitation of this study is the absence of data for the first three months of 2021 in the most recent ONS datasets. This gap makes it impossible to establish baseline mortality rates prior to the initiation of the COVID-19 vaccination program. Access to a pre-vaccination reference period would have been invaluable for understanding normal mortality trends and accurately assessing the subsequent impact of vaccination.

Moreover, the ONS dataset collection concludes in May 2023, which restricts the analysis of longer-term mortality trends. As a result, the study cannot evaluate whether changes in mortality patterns occurred after May 2023 due to shifts in COVID-19 prevalence, the implementation of new health policies, or the effects of subsequent vaccination campaigns. This temporal limitation prevents the capture of evolving mortality trends that may have emerged beyond the dataset's coverage.

In summary, the lack of pre-vaccination mortality data and the restricted timeframe of the dataset constrain the study’s ability to provide a comprehensive analysis of mortality patterns, particularly in the context of long-term changes and health policy developments after May 2023.

**6. Conclusion**

This study leveraged publicly available data from the UK Office for National Statistics (ONS) to investigate the relationship between COVID-19 vaccination status and all-cause mortality. The analysis revealed complex patterns across vaccination doses and age groups. In the short term, individuals recently vaccinated with the first dose exhibited elevated mortality rates compared to the unvaccinated reference group, particularly among older age cohorts. These initial spikes were likely driven by confounding by indication, where vaccination campaigns prioritized clinically vulnerable individuals, and harvesting effects, where frail individuals with pre-existing conditions succumbed earlier. Over time, as individuals progressed to their second and third doses, these mortality rates decreased substantially, with rate ratios (RRs) often falling below 1.0, indicating a robust protective effect of vaccination.

The Negative Binomial regression further validated this trend, showing a statistically significant reduction in all-cause mortality across all age groups for fully vaccinated individuals. Younger cohorts (18–39 years) demonstrated smaller absolute reductions due to their already low baseline mortality, while middle-aged groups (40–79 years) experienced the most pronounced benefits. In contrast, among the oldest cohorts (80+ years), the protective effect was more attenuated but still present, reflecting the higher frailty and prevalence of chronic conditions in this population.

Interpreting these findings requires acknowledging several confounding factors. The healthy-vaccine bias, where vaccinated individuals tend to have healthier baseline behaviours and characteristics, may enhance the apparent protective effect. Similarly, confounding by indication inflates early mortality rates among the most vulnerable individuals vaccinated first. The harvesting effect also contributes to short-term spikes in mortality by excluding frail individuals from subsequent vaccination phases. Together, these factors illustrate the inherent challenges of assessing vaccine efficacy using real-world observational data and complicate causal interpretations. Overall, this analysis demonstrates the complex interplay between vaccination, cohort dynamics, and mortality patterns. While vaccination is clearly associated with reduced all-cause mortality, understanding its long-term effects requires more granular data, including individual-level comorbidity and behavioural factors, as well as extended observation periods. Future research should aim to address these gaps by incorporating detailed health metrics and examining mortality trends over longer timeframes to better elucidate the enduring impacts of COVID-19 vaccination.

**Reference:**

Acuti Martellucci, C., Flacco, M. E., Soldato, G., Di Martino, G., Carota, R., Caponetti, A., & Manzoli, L. (2022). Effectiveness of COVID-19 Vaccines in the General Population of an Italian Region before and during the Omicron Wave. Vaccines, 10(5), 662. <https://doi.org/10.3390/vaccines10050662>

Agampodi, S., Tadesse, B. T., Sahastrabuddhe, S., Excler, J. L., & Kim, J. H. (2024). Biases in COVID-19 vaccine effectiveness studies using cohort design. Frontiers in medicine, 11, 1474045. https://doi.org/10.3389/fmed.2024.1474045

Altman, D. G., & Bland, J. M. (2011). How to obtain the P value from a confidence interval. *BMJ (Clinical research ed.)*, *343*, d2304. <https://doi.org/10.1136/bmj.d2304>

Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S. A., Rouphael, N., Creech, C. B., McGettigan, J., Khetan, S., Segall, N., Solis, J., Brosz, A., Fierro, C., Schwartz, H., Neuzil, K., Corey, L., Gilbert, P., … COVE Study Group (2021). Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. The New England journal of medicine, 384(5), 403–416. <https://doi.org/10.1056/NEJMoa2035389>

Berrino, F., Donzelli, A., Bellavite, P., & Malatesta, G. (2023). COVID-19 vaccination and all-cause and non-COVID-19 mortality. A revaluation of a study carried out in an Italian Province. Vaccinazione COVID-19 e mortalità per tutte le cause e non COVID-19. Rivalutazione di uno studio condotto in una provincia italiana. *Epidemiologia e prevenzione*, *47*(6), 374–378. <https://doi.org/10.19191/EP23.6.A643.075>

Bhopal, S. S., Bagaria, J., Olabi, B., & Bhopal, R. (2021). Children and young people remain at low risk of COVID-19 mortality. The Lancet. Child & adolescent health, 5(5), e12–e13. <https://doi.org/10.1016/S2352-4642(21)00066-3>

Bonanad, C., García-Blas, S., Tarazona-Santabalbina, F., Sanchis, J., Bertomeu-González, V., Fácila, L., Ariza, A., Núñez, J., & Cordero, A. (2020). The Effect of Age on Mortality in Patients With COVID-19: A Meta-Analysis With 611,583 Subjects. Journal of the American Medical Directors Association, 21(7), 915–918. https://doi.org/10.1016/j.jamda.2020.05.045

Eiser A. R. (2021). Could Dietary Factors Reduce COVID-19 Mortality Rates? Moderating the Inflammatory State. Journal of alternative and complementary medicine (New York, N.Y.), 27(2), 176–178. <https://doi.org/10.1089/acm.2020.0441>

Fürst, T., Bazalová, A., Fryčák, T., & Janošek, J. (2024). Does the healthy vaccinee bias rule them all? Association of COVID-19 vaccination status and all-cause mortality from an analysis of data from 2.2 million individual health records. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*, *142*, 106976. <https://doi.org/10.1016/j.ijid.2024.02.019>

Hussain Z. (2022). England and Wales have seen rise in excess deaths in 2022. *BMJ (Clinical research ed.)*, *378*, o2283. <https://doi.org/10.1136/b>

Norman, G. & Streiner, D. (2015). Biostatistics: The Bare Essentials, : PMPH-USA.

Office for National Statistics (2023). Dataset Deaths by vaccination status, England. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsbyvaccinationstatusengland>

Office for National Statistics (2023). Deaths registered weekly in England and Wales, provisional: week ending 26 May 2023. <https://www.ons.gov.uk/releases/deathsregisteredweeklyinenglandandwalesprovisionalweekending26may2023>

Office for Statistics (2022). Population and household estimates, England and Wales: Census 2021, unrounded data . <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationandhouseholdestimatesenglandandwales/census2021unroundeddata>

Our World in Data (2024). Coronavirus (COVID-19) Vaccinations. <https://ourworldindata.org/covid-vaccinations>

Pinsky, P. F., Miller, A., Kramer, B. S., Church, T., Reding, D., Prorok, P., Gelmann, E., Schoen, R. E., Buys, S., Hayes, R. B., & Berg, C. D. (2007). Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. American journal of epidemiology, 165(8), 874–881. <https://doi.org/10.1093/aje/kwk075>

Riou, J., Hauser, A., Fesser, A., Althaus, C. L., Egger, M., & Konstantinoudis, G. (2023). Direct and indirect effects of the COVID-19 pandemic on mortality in Switzerland. *Nature communications*, *14*(1), 90. https://doi.org/10.1038/s41467-022-35770-9

Rosso, A., Flacco, M. E., Soldato, G., Di Martino, G., Acuti Martellucci, C., Carota, R., De Benedictis, M., Di Marco, G., Di Luzio, R., Fiore, M., Caponetti, A., & Manzoli, L. (2023). COVID-19 Vaccination Effectiveness in the General Population of an Italian Province: Two Years of Follow-Up. Vaccines, 11(8), 1325. <https://doi.org/10.3390/vaccines11081325>

Schöley, J., Aburto, J. M., Kashnitsky, I., Kniffka, M. S., Zhang, L., Jaadla, H., Dowd, J. B., & Kashyap, R. (2022). Life expectancy changes since COVID-19. Nature human behaviour, 6(12), 1649–1659. <https://doi.org/10.1038/s41562-022-01450-3>

Skowronski, D. M., & De Serres, G. (2021). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. The New England journal of medicine, 384(16), 1576–1577. <https://doi.org/10.1056/NEJMc2036242>

Soliani L. (2008). Statistica Applicata. : Uni.Nova.

Vail, E. A., Gershengorn, H. B., Wunsch, H., & Walkey, A. J. (2021). Attention to Immortal Time Bias in Critical Care Research. *American journal of respiratory and critical care medicine*, *203*(10), 1222–1229. <https://doi.org/10.1164/rccm.202008-3238CP>

Walkowiak, M. P., Domaradzki, J., & Walkowiak, D. (2023). Unmasking the COVID-19 pandemic prevention gains: excess mortality reversal in 2022. Public health, 223, 193–201. <https://doi.org/10.1016/j.puhe.2023.08.004>

World Health Organization (2022). WHO COVID-19 dashboard. <https://data.who.int/dashboards/covid19/cases?n=c>

Wright, L., Steptoe, A., & Fancourt, D. (2021). Predictors of self-reported adherence to COVID-19 guidelines. A longitudinal observational study of 51,600 UK adults. The Lancet regional health. Europe, 4, 100061. https://doi.org/10.1016/j.lanepe.2021.100061

Yao, Y., Hu, B., & Liu, X. (2023). Life expectancy trends in China in the post-COVID-19 era. *The Lancet. Public health*, *8*(12), e906–e907. <https://doi.org/10.1016/S2468-2667(23)00074-9>

**Supplementary Materials:**

|  |  |
| --- | --- |
| A graph with a bar graph  Description automatically generated  Figure S1: Box Plot of RRs by Dose Category (18-39 Age Group) | A graph with white squares and black background  Description automatically generated  Figure S2: Box Plot of RRs by Dose Category (40-49 Age Group) |
| A graph with a black background  Description automatically generated  Figure S3: Box Plot of RRs by Dose Category (50-59 Age Group) | A graph with squares and lines  Description automatically generated  Figure S4: Box Plot of RRs by Dose Category (60-69Age Group) |
| A graph with white squares and lines  Description automatically generated  Figure S5: Box Plot of RRs by Dose Category (70-79 Age Group) | A graph with squares and lines  Description automatically generated  Figure S6: Box Plot of RRs by Dose Category (80-89 Age Group) |
| A graph with a black background  Description automatically generated  Figure S7: Box Plot of RRs by Dose Category (90+ Age Group) |  |

**Data Availability Statement**

The data presented in this study are openly available on the UK ONS web page entitled ” Deaths by vaccination status, England”, available online: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsbyvaccinationstatusengland>

Supplementary Data

Geogle Drive: Data Analytics Project Supplementary Materials,

<https://drive.google.com/drive/folders/1-Sdlq8HwKATzCygxnOzkF69a5pSmY44c?usp=sharing>

This folder contains the following supplementary dataset:

* Box plot for different age groups 1.R
* Box Plot for different age groups2.R
* Negative-Binomial Distribution Forest Plot.R
* Negative-Binomial Distribution Stargazer.R
* Poisson Distribution for Age group (Unvacin and Vacin).R
* Supplementary Tables S1-S7 (RR Values).xlsx
* Table 2.xlsx