Title: Socioeconomic and temporal heterogeneity in SARS-CoV-2 exposure and disease in England from May 2020 to February 2023

Authors: Christian Morgenstern[1, #], Thomas Rawson[1], Wes Hinsley[1], Pablo N. Perez Guzman[1], Samir Bhatt[1,2, *], Neil M. Ferguson[1, *]

Affiliations:

[1]MRC Centre for Global Infectious Disease Analysis, Jameel Institute, School of Public Health, Imperial College London, UK

[2]University of Copenhagen, Denmark

[*]Contributed equally

[#]Corresponding author: c.morgenstern@imperial.ac.uk

Abstract

Background: The COVID-19 pandemic had resulted in over 20.5 million confirmed cases and 175,000 deaths in England by December 2023. The pandemic's impact varied significantly across different population groups, influenced by deprivation, ethnicity, and policy measures.

Methods: We analysed individual-level data on SARS-CoV-2 testing, hospitalisations, deaths, and vaccination records in England from May 2020 to February 2022. We used Poisson regression models to estimate incidence rate ratios (IRRs) for first pillar 2 PCR positive cases, associated hospitalisations, and deaths, adjusting for sex, ethnicity, deprivation, geographic region, age, and epidemiological week. Model selection was based on cross-validation and performance metrics (AIC, R^2).

Findings: The data analysed included 12,310,485 first SARS-CoV-2 pillar 2 PCR-confirmed infections, 79,315 hospitalisations, and 107,823 deaths associated with the first SARS-CoV-2 infection. Significant differences were observed across IMD quintiles, with the most deprived areas showing higher incidence rate ratios (IRRs) for death (1.64, 95% CI: 1.60-1.67) and hospitalisation (1.80, 95% CI: 1.75-1.85) compared to the least deprived areas as the reference group for the entire study period. Ethnic disparities were also notable, with higher IRRs for death and hospitalisation for all non-White ethnicities relative to White ethnicities as the reference group. We note that the magnitude of IRRs, for both deprivation and ethnicities, declined from the wild-type to the omicron periods for severe outcomes. For cases, we observed IRRs above one for non-White ethnicities during the wild-type and alpha periods only. Vaccine effectiveness (VE) was also assessed, with models indicating a significant reduction in risk post-vaccination across all outcomes of interest.

Interpretation: Deprivation and ethnicity significantly influenced COVID-19 outcomes in England. For severe outcomes, pre-existing health inequalities lead to large and persistent disparities. For infections, both protective and support measures need to be structured with ethnicity and deprivation in mind in the early parts of a pandemic.

Funding: UK Medical Research Council, Schmidt Foundation, National Institute of Health Research, Community Jameel.

Keywords: SARS-CoV-2, risk factors, deprivation, ethnicity, vaccine effectiveness.

Research in Context

Evidence before this study

The COVID-19 pandemic has significantly impacted global health, and previous research has highlighted the role of socioeconomic factors, such as deprivation and ethnicity, in influencing outcomes. McGowan et al. found in a scoping review that 91% of studies showed significantly higher COVID-19 mortality in areas of social disadvantage relative to more affluent areas. A systematic review by Pan et al. [2] showed the impact of ethnicity on clinical outcomes in COVID-19. However, the impact of deprivation and ethnicity over different time periods of the COVID-19 pandemic and its interplay with public health measures is poorly understood.

Added value of this study

This study leverages extensive data from multiple sources, including PCR-confirmed COVID-19 cases, hospitalisation records, death registries, and vaccination databases, to comprehensively analyse COVID-19 outcomes in relation to deprivation and ethnicity across England. Using Poisson regression, we provide estimates of incidence rate ratios (IRR) associated with different levels of deprivation and ethnic backgrounds. This study also incorporates data on periods defined by dominant variants and public health measures, allowing for a more detailed examination of how these factors interact with local socioeconomic contexts to influence health outcomes.

Implications of all the available evidence

The findings underscore the importance of considering disparities in COVID-19 outcomes by deprivation and ethnicity and highlight the need for targeted strategies to address these inequalities. Policymakers should prioritise resource allocation and tailored interventions in high-risk areas to mitigate the adverse impacts of the pandemic. Furthermore, the effectiveness of vaccination programs should be continuously evaluated in the context of these socioeconomic determinants to enhance their efficacy and reach. This research contributes to the growing body of evidence supporting the need for an equitable public health response that addresses the specific needs of diverse populations across different localities before the pandemic by reducing health inequalities and in the pandemic response.

Introduction

By December 2023, the COVID-19 pandemic had caused over 20.5 million confirmed cases and over 175,000 deaths in England [3,4]. The pandemic did not impact individuals equally; individual-based studies have explored the links between deprivation, ethnicity, and other factors on health outcomes to varying levels [1,2,5–8].

Past studies have explored heterogeneity in risk in England [9] and Scotland [10] using the Index of Multiple Deprivation (IMD) [11] at the Lower-Tier Local Authority (LTLA) level. LTLAs are areas where local government provides services, and the UK Office of National Statistics (ONS) provides an extensive set of data, including population age distribution, ethnicity, and measures of deprivation. These measures vary substantially across LTLAs, as did the number of infections, hospitalisations, and deaths reported across England over the course of the pandemic (Figure 1). We aim to understand heterogeneity in COVID-19 outcomes and how this evolved to help inform policy planning [3,12,13].

SARS-CoV-2 testing in England [14] had four "pillars": hospital testing for health and care workers and individuals with clinical needs (pillar 1), freely available testing for the general population (pillar 2), serology (pillar 3) and targeted surveillance (pillar 4). Pillar 2 testing increased significantly from 10th May 2020, and most testing was halted on 27th February 2022 with the introduction of the 'Living with COVID' strategy [15]. Several public health measures were introduced over this period, including stay-at-home orders (see SI B.5) [3,13] at local and national levels.

We additionally investigate vaccine effectiveness (VE) in the post-vaccination period and the impact of vaccination on risk heterogeneity. The SARS-CoV-2 immunisation programme in England was one of the most rapid globally, with 89.9% of the adult population aged 20 or over receiving at least one dose and 86.6% receiving at least two doses by 27th February 2022 [16]. Test-negative case-control studies (TNCCs) were the standard epidemiological tool for evaluating SARS-CoV-2 VE during the pandemic [17–19], though TNCCs can potentially suffer from biases [20]. The main alternative to TNCCs used for VE estimation during the COVID-19 pandemic is large population observational cohort studies that use individual-level healthcare and surveillance data linked to vaccination status [21–24].

[Figure 1]

Methods

Data sources

The UK Health Security Agency (UKHSA) maintains a single unified database of all SARS-CoV-2 positive test results in the country, categorised by pillar. The resulting database contains a hashed version of a unique identifier (NHS number), the age, sex, ethnic group, home LTLA, symptom status, testing pillar, and date of specimen for each positive case.

Second, linked to the case database by (hashed) National Health Service (NHS) number, UKHSA maintains a database of all deaths within 28 days of a positive PCR test for SARS-CoV-2. This database records the date of death, the date of hospital admission, and how the death was identified (e.g., via hospital reporting, death registration, or both).

Third, two NHS data sources on hospitalisation were also linked by NHS number to the case database: the Secondary Uses Service (SUS) data on hospital episodes and the Emergency Care Data Set (ECDS) on attendance at Accident and Emergency departments.

Lastly, England maintains a national vaccination register, the National Immunisation Management System (NIMS) [16]. Every SARS-CoV-2 vaccine dose given in England is recorded at the time of vaccine administration. The anonymised version of the dataset used here contained a unique identifier (hashed NHS number), age, sex, ethnic group, LTLA and administration date.

We extracted data from the versions of these databases with data up to 25th July 2023. Cases with a valid NHS number were then linked to the NIMS vaccination dataset. Where a match was found, the case was classified as having received a vaccine (either before or after testing positive). Cases with no matching immunisation record were classified as having not been vaccinated.

Population denominator data was used to calculate the number of people who have neither received a vaccine nor been diagnosed with SARS-CoV-2 infection. We used the 2021 Census population estimates generated by the ONS [25], stratified by age, sex and LTLA [26]. We made the simplifying assumption that in the absence of COVID-19, the size of the English population and its age distribution would have been at a steady state for our analysis period consistent with the Census 2021 estimates. Individuals with missing information on sex, age, or residence location were not included in the analysis.

We use the Index of Multiple Deprivation (IMD) dataset released in 2019 [11] by the UK Ministry of Housing, Local Communities, and Local Government, aggregated at the LTLA level and its subcomponents (SI B.4).

170 **Statistical Analysis**

171 Survival models were used to estimate Incidence Rate Ratios (IRR), an estimate of the increased risk
172 of a particular outcome for a specific population group relative to a reference group. We used Poisson
173 regression using an offset to adjust for person-days at risk to generate IRRs for the different strata of
174 interest, proving a good approximation to continuous time survival models (see SI A)[27]. Poisson
175 regressions are more general than Cox proportional hazard survival models, allowing for non-
176 proportional hazards. We investigated using negative binomial regressions to account for
177 overdispersion. We parametrically captured variation by systematically examining evidence for
178 interactions between covariates. We conducted model selection to identify the model which performed
179 best across all three outcomes of interest (infection, hospitalisation, and deaths) by considering the
180 average Akaike's information criterion (AIC) and coefficient of determination (R^2) [28,29] across the three
181 respective models.

182 We included terms for sex, ethnic group, vaccination status and IMD in all models. We used a
183 categorical IMD variable, ranking LTLAs by average IMD score and then binning them into quintiles
184 (1 being least deprived, five being most). The three other covariates included in all models were
185 geographic region (either upper tier local authority or region), age (in 10-year bands with the last band
186 being 85+ or 10-year bands with all individuals below 40 grouped and a max-age of 80+) and
187 epidemiological week. We examined all pairwise interactions between these three variables. The level
188 of public health and social measures was computed from ONS LTLA restriction level data and the
189 Oxford COVID-19 Government Response Tracker [3] and used as a parametric covariate (see SI Section
190 A and C.4) as restriction levels changed several times, creating non-continuous time intervals.

191 All survival analyses used the test specimen date as the outcome event date. Three SARS-CoV-2-related
192 events were defined: death within 28 days of any positive PCR test, hospitalisation where the individual
193 hospitalised tested PCR-positive between 14 days before admission and the day before discharge, and
194 any PCR-confirmed infection within pillar 2. We only consider the first SARS-CoV-2 infection for all
195 three events and associated hospitalisations and deaths.

196 We fitted the model for all three outcomes of interest. Individuals were treated as censored after a
197 positive SARS-CoV-2 pillar 2 PCR test to exclude post-primary infections and avoid bias since the UK
198 vaccination programme excluded people from being vaccinated within 28 days of a positive test.

199 Since the UKHSA/NHS and ONS data used different ethnicity categories, we aggregated ethnicities
200 into six groups that could be identified consistently across all datasets: White, South Asian, Asian
201 (other), Black, and Mixed/Other to be consistent across the two datasets (SI B.1). Details are provided
202 in Supplementary Information (SI B.1). The computation of age for each individual is outlined in the
203 SI B.2.

204 We estimated VE accounting for the type of vaccine (mRNA or adenovirus-based), time since

vaccination, and outcome of interest (death, hospitalisation, Pillar 2 PCR positive case). Details are provided in SI section B.3. By comparing currently vaccinated individuals against people who would be vaccinated in the future, we controlled for potential differences in behaviour and exposure risk between those who are never vaccinated and those who eventually are. Censoring after the first recorded positive test partially mitigated estimates from being affected by the accumulation of naturally acquired immunity in the population.

To investigate how estimates varied over time, for example, in periods dominated by particular variants, we restricted model fitting to data from the specific period of interest (censoring individuals with positive tests before that period).

We used negative binomial regressions as a sensitivity analysis to explore over-dispersion in the count data. In all figures, we display only results with a p-value of 0.05 or less.

All analyses were undertaken in R version 4.3.1 using the H2O.ai machine learning package version 3.42.0.2 [30], which offers high-performance parallelised algorithms for fitting general linear models to large datasets. A 32-core Intel-based server with 128GB RAM was used to conduct the analyses.

Data access

While all data used in this analysis were anonymised, the individual-level nature of the data used risks individuals being identified or being able to self-identify if it is released publicly. Requests for access to the underlying source data should be directed to UKHSA.

Ethical approval

Surveillance of COVID-19 testing and vaccination is undertaken under Regulation 3 of The Health Service (Control of Patient Information) Regulations 2002 to collect confidential patient information (www.legislation.gov.uk/uksi/2002/1438/regulation/3/made) under Sections 3(i) (a) to (c), 3(i)(d) (i) and (ii) and 3(3). Data were shared with the investigators as part of the UK's emergency response to the COVID-19 pandemic via the SPI-M subcommittee of the UK Scientific Advisory Group for Emergencies (SAGE). Ethics permission was sought for the study via Imperial College London's standard ethical review processes, and the study was approved by the College's Research Governance and Integrity Team (ICREC reference: 21IC6945).

Role of the funding source

The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Results

We present the breakdown of the England population (56,427,863 people in England [Census 2021], 56,417,353 are in our synthetic population, and 56,344,410 are included in the analysis) in Table 1, including the number of first episodes of SARS-CoV-2 infections (12,310,485) and associated hospitalisations (79,315) or deaths (107,823) linked to that first infection from the week beginning 10th May 2020 (isoweek 18) to 27th February 2022 (isoweek 8) confirmed by a pillar 2 PCR positive test. These 95 weeks cover the full period for which population-wide testing was available.

[Table 1]

We examined a total of 360 model variants, which varied by the covariates and interactions included (see SI A.2 for covariate definitions considered): age, definition of deprivation, regional disaggregation, ethnicity disaggregation, and which interaction terms to include in the model. SI Table A.3–A.5 lists the top 88 selected models for each outcome of interest. We found that less granular age categories had better 10-fold cross-validation performance (under 40s, 10-year age bands up to 70, single group for 80+). Similarly, coarse spatial disaggregation (by NHS region of England) was favoured, and deprivation was quantified by IMD quintiles. We chose ethnicity disaggregation into six groups as the results were very close without a clear favoured disaggregation.

We compared the estimates from the preferred model against the data, and found a good qualitative fit examining marginal IRRs (SI A.2).

We investigated the impact of overdispersion for each outcome of interest using negative binomial regression models. For Pillar 2 PCR confirmed cases, we found a moderate level of overdispersion (k=0.33) during the wild-type period. We found that either very high values of k (indicating marginal overdispersion) or the Poisson model were preferred for all other periods. For hospitalisation and death outcomes, the Poisson model was always preferred (SI section A and Table A.1).

Figure 2 presents the modelled IRRs for deprivation and ethnicity categories from the preferred model for each endpoint and by period defined by variant dominance. IRRs for sex, age and region are presented in SI Figure (D.12).

[Figure 2]

When we consider the entire period (May 2020-February 2022), the highest IMD quintile, corresponding to the most deprived LTLAs, has an IRR (relative to the least deprived IMD quintile) of 1.64 (95% CI: 1.60-1.67) for death as the outcome, 1.80 (95% CI: 1.75-1.85) for hospitalisation, 1.02 (95% CI: 1.02-1.02) for a positive pillar 2 PCR test. For the Pillar 2 positive test outcome, we observed significant differences between periods defined by variant dominance. The biggest differences were observed for the Alpha variant with an IRR (for the highest quintile relative to the lowest) for that endpoint of 1.45 (95% CI: 1.45-1.46).

For deaths, over the entire study period, South Asian ethnicity had an IRR 1.63 (95% CI: 1.59-1.68), Asian (other) ethnicity an IRR 1.24 (95% CI: 1.17-1.32), and Black ethnicity an IRR 1.49 (95% CI: 1.43-1.54), with White ethnicity as the reference. For hospitalisation, over the entire study period, South Asian ethnicity had an IRR 1.90 (95% CI: 1.84-1.96), Asian (other) ethnicity an IRR 1.38 (95% CI: 1.30-1.47), and Black ethnicity an IRR 1.61 (95% CI: 1.55-1.68). Using pillar 2 PCR positive tests as the endpoint, over the entire study period, South Asian ethnicity had an IRR 0.89 (95% CI: 0.88-0.89), Asian (other) ethnicity an IRR 0.78 (95% CI: 0.78-0.79), and Black ethnicity an IRR 0.89 (95% CI: 0.89-0.90). The IRRs for Mixed/Other ethnicity are reported in SI Table D.3.

Ethnicity differences declined for the Pillar 2 PCR positive test endpoint throughout the pandemic. IRR for pillar 2 PCR positive cases declines for all non-White ethnicities (relative to White) over time, with $IRR_{WT} > IRR_{Alpha} > IRR_{Delta}$. During the Omicron period, non-White IRRs were broadly similar to the Delta period but typically slightly above IRR_{Delta}. For death and hospitalisation, we observed a declining IRR over time for South Asian ethnicity, with IRR_{WT}, IRR_{Alpha}, and IRR_{Delta} exhibiting overlapping confidence intervals but $IRR_{Omicron}$ statistically significantly lower. Other Asian ethnicities had high IRRs (overlapping confidence intervals) for the WT and Alpha periods and significantly lower IRRs for Delta and Omicron. Black ethnicities exhibited high IRRs throughout the entire period in the range of 1.01-1.45 for death and 1.10-1.76 for hospitalisation.

Results for deprivation were consistent for severe outcomes (hospitalisation and death) throughout the pandemic, with IRRs monotonically increasing as deprivation increased, but with IRRs during the Omicron period lower than other periods. This result also applied to a more granular representation of deprivation, using deciles (Figure D.12(A) in SI) and for the different subcomponents of the IMD measure (SI section E.1-E.7). For Pillar 2 PCR positive cases, we observed an increasing IRR with increasing deprivation for the WT and Alpha periods but not for the Delta and Omicron periods. This resulted in the IRR for the whole period being close to 1. This result was consistent with IMD deciles and the subcomponents of the IMD index.

[Figure 3]

In Figure 4 we present the VE results (SI section B.3 and Table D.2). We obtain a VE (defined as 1-IRR) estimate for mRNA vaccines ranging from 86.8% (over 18 weeks after second dose, 95% CI: 86.2-87.3) to 97.1% (less than two weeks after booster dose, 95% CI: 96.6-97.6) for protection against death for individuals with at least two doses, 90.1% (over 18 weeks after second dose, 95% CI: 89.5-90.5) to 96.7% (2-10 weeks after booster dose, 95% CI: 96.5-97) against hospitalisation and 30.4% (over 18 weeks after booster dose, 95% CI: 29.3-31.5) to 84.5% (2-10 weeks after second dose, 95% CI: 84.4-84.6) against pillar 2 PCR positive confirmed infections for the full study period.

For adenovirus-based vaccines, we estimated VE for individuals with at least two doses, ranging from 77.2% (over 18 weeks after second dose, 95% CI: 76.2-78) to 95.9% (less than two weeks after booster

308 dose, 95% CI: 95.3-96.5) for protection against death, 84.3% (over 18 weeks after second dose, 95%
309 CI: 83.6-85) to 96.2% (2-10 weeks after booster dose, 95% CI: 95.9-96.5) against hospitalisation, and
310 29.6% (over 18 weeks after booster dose, 95% CI: 27.6-31.5) to 68.5% (2-10 weeks after second dose,
311 95% CI: 68.3-68.6) against pillar 2 PCR-positive confirmed infections for the full study period.

312 We found that VE was consistently lower for the Omicron period. There was little to no difference in
313 VE estimates for severe outcomes for the Alpha and Delta time periods.

314 We present the estimated VE results for all variants, vaccine status, and vaccine types in SI Table D.2
315 with their associated 95% CIs. The confidence intervals are narrow due to the large number of events
316 observed (Table 1B); confidence intervals are wider for the variant period-specific estimates.

317 [Figure 4]

318 Figure 4 shows the estimated IRRs for the restriction level covariate. We estimated the IRR for the full
319 time period of 1.14 (95% CI: 1.12-1.15) for deaths, 1.13 (95% CI: 1.12-1.15) for hospitalisation, and
320 1.17 (95% CI: 1.17-1.17) for Pillar 2 PCR positive cases. The estimated IRR for restriction level for the
321 WT and Alpha periods was lower for deaths and hospitalisation. However, for Pillar 2 PCR positive
322 cases, we found that the IRR for restriction level was higher during the WT period and lower for the
323 Alpha period (Figure 4). The IRRs for the Delta period are one, as no substantial restrictions were in
324 place at that time.

325

326 **Discussion**

327 We considered the impact of sex, age, region, deprivation, level of restrictions, and ethnicity on the
328 relative risk of a SARS-CoV-2 pillar 2 PCR-positive first infection, hospitalisation, or death. Our
329 proposed model was selected to provide a good fit across all three outcomes of interest. Our results
330 extend previous work on the impact of deprivation [1] and evidence risk heterogeneity by ethnicity in
331 England [6–8] and globally [2].

332 We find that modelling overdispersion for pillar 2 PCR positive cases was necessary for the wild-type
333 period, consistent with Morrissey et al. [12], and the value for the overdispersion we find aligns with other
334 estimates outside China [31,32]. We did not find overdispersion to be present for severe outcomes, and the
335 Poisson model provides the best fit to the data. Even though pillar 2 PCR tests were available widely,
336 this indicates that some groups, defined by higher levels of deprivation or non-White ethnicity, may
337 have accessed testing less, leading to the observed overdispersion. This was less likely to be the case
338 for more severe outcomes as testing would be more likely conducted due to the more severe nature of
339 the infection.

340 The differences in IRRs we observe by outcome (infection, hospitalisation, or death) indicate a
341 difference between severe outcomes and infections. For severe outcomes, we observed a significant

risk heterogeneity by age, with older age groups at higher risks than younger age groups at an exponential rate (Figure SI D.10(B)) consistent with the variation of the infection fatality ratio by age in the pre-vaccination era [33]. Increased risk is also positively associated with higher levels of deprivation (Figure 3A), and the difference between the most deprived quintile of LTLAs and the least deprived quintile is significant for all periods considered. Similarly, we find that non-White ethnicities are at increased risk. For pillar 2 PCR-positive cases, the picture is more nuanced. The risk by age is driven by contact patterns, as defined by the average number of contacts per person (SI Figure C.6) for the different age groups. We observed that the difference in risk between age groups was on a similar scale compared to the large differences we observed for severe outcomes. We saw similar risk heterogeneity patterns for both deprivation and ethnicity but less strongly than for severe outcomes. The results for pillar 2 PCR-positive cases were consistent with other studies [34] and that individuals exhibit self-protective behaviours [35].

For severe outcomes, we observed less variation throughout the pandemic, pointing to existing health inequalities driving the strong association between deprivation, ethnicity, and risk of severe outcomes. This held consistently for all sub-components of the index of multiple deprivation, except for the Living Environment component, and particularly strongly for the Income, Employment, Health, and Education sub-components (Figure SI sections E.1-E.7). This observation is consistent with known socio-economic and ethnic disparities in the UK for type-2 diabetes [36], which is a known risk factor for SARS-CoV-2 infections. Interestingly, we observed a high IRR for South Asian ethnicity, for both severe outcomes and infections, which reduced over time (Figure 3.B), pointing to the metabolic component impacting outcomes being reduced for omicron and delta as individuals of South Asian ethnicity are at higher risk of metabolic disease [37].

We assessed vaccine effectiveness, and our results broadly align with previously published VE estimates for England by UKHSA[17] for similar periods and VE definitions. We find vaccine effectiveness for the mRNA vaccines is consistent with UKHSA estimates of 88.0% (95% CI: 85.3-90.1) for two doses of BNT162b2 for infections for persons with the Delta variant. Similarly, we find vaccine effectiveness for the adenovirus vaccines is in line with UKHSA estimates of 67.0% (95% CI: 61.3-71.8) for persons with two doses of the ChAdOx1 nCoV-19 vaccine. For the Omicron variant, estimates vary more widely[38] with VE against hospitalisation, ranging from 89.6% to 95.3 compared to 67.8% to 92.4% from the literature. Similarly, VE against pillar 2 PCR confirmed infection ranged from 17% to 69.5% compared to 11.7% to 75.3% in the literature. We note that we only consider first infections for the Omicron period, which may bias our VE estimates to be higher.

The policy implications of our study are that both protective and support measures need to be structured with ethnicity and deprivation in mind, particularly in the early stages of a pandemic. Deprivation had a particularly significant impact during the Alpha variant era for infections, and for ethnicity, we observed a reduction of IRRs as the pandemic progressed (Figure 3). This suggests that more deprived

areas were less able to isolate effectively, potentially due to work, during heightened restrictions. Ethnic minorities, in particular Bangladeshi and Pakistani men, worked in occupations directly impacted by lockdown measures, and more broadly, ethnic minorities had less access to savings [39], potentially exposing them to a higher risk of community transmission.

Considering different levels of restrictions, we found IRRs for infections and severe outcomes in the wild-type and Alpha periods to be above one. This highlights the impact of the different timing of restrictions between regions, as we already account for week and region effects in the model. We note that other results were not sensitive to whether restrictions were included as a predictor in the model.

A limitation is that we only considered first infections in this study and could not make any statements about reinfections. However, we note that up to November 2021, with the arrival of the Omicron variant of SARS-CoV-2, we have observed predominantly first infections. We only quantify the incidence among those seeking tests, not the underlying infection incidence. This limits our ability to model the impact of comorbidities, such as obesity[40], as we cannot differentiate if the outcome is worse due to pre-existing comorbidities or those groups being more exposed.

Future research priorities include a multi-outcome survival model to incorporate infections and severe outcomes in the same model. Integrating serology and infection prevalence data is essential in addressing questions such as the impact of pre-existing comorbidities on the risk heterogeneity by ethnicity or socio-economic group.

Declarations

Funding All authors acknowledge funding from the Medical Research Council (MRC) Centre for Global Infectious Disease Analysis (MR/X020258/1) funded by the UK MRC and carried out in the frame of the Global Health EDCTP3 Joint Undertaking supported by the EU; the NIHR for support for the Health Research Protection Unit in Modelling and Health Economics, a partnership between the UK Health Security Agency (UKHSA), Imperial College London, and London School of Hygiene & Tropical Medicine (grant code NIHR200908); a philanthropic donation from Community Jameel supporting the work of the Jameel Institute S.B. acknowledges support from the Novo Nordisk Foundation via The Novo Nordisk Young Investigator Award (NNF20OC0059309). SB acknowledges the Danish National Research Foundation (DNRF160) through the chair grant. S.B. acknowledges support from The Eric and Wendy Schmidt Fund for Strategic Innovation via the Schmidt Polymath Award (G-22-63345), which also supports C.M. The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. For the purpose of open access, the author has applied a 'Creative Commons Attribution' (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Availability of data and materials While all data used in this analysis were anonymised, the individual-level nature of the data used risks individuals being identified or being able to self-identify if it is released publicly. Requests for access to the underlying source data should be directed to UKHSA.

Code availability https://github.com/cm401/covid19_risk_heterogeneity_england

Competing interests. The authors declare no competing interests.

Authors' Contributions NMF, CM, and SB conceptualised the study. CM, WH, PPG, and NMF curated and validated the data. CM formally analysed and visualised the data. SB and NMF acquired funding and supervised the study. CM wrote the original manuscript draft. All authors were responsible for the methodology and review and editing of the manuscript. All authors discussed, edited, and approved the final version of the manuscript. All authors had final responsibility for the decision to submit the manuscript for publication.

References

1 McGowan VJ, Bambra C. COVID-19 mortality and deprivation: pandemic, syndemic, and endemic health inequalities. *Lancet Public Health* 2022; **7**: e966–75.

2 Pan D, Sze S, Minhas JS, *et al.* The impact of ethnicity on clinical outcomes in COVID-19: A systematic review. *EClinicalMedicine* 2020; **23**: 100404.

3 Hale T, Angrist N, Goldszmidt R, *et al.* A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav* 2021; **5**: 529–38.

4 Phillips T. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). https://github.com/OxCGRT/covid-policy-dataset.

5 Kontopantelis E, Mamas MA, Webb RT, *et al.* Excess deaths from COVID-19 and other causes by region, neighbourhood deprivation level and place of death during the first 30 weeks of the pandemic in England and Wales: A retrospective registry study. *Lancet Reg Health - Eur* 2021; **7**: 100144.

6 Ward H, Atchison C, Whitaker M, *et al.* SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. *Nat Commun* 2021; **12**: 905.

7 Lo C-H, Nguyen LH, Drew DA, *et al.* Race, ethnicity, community-level socioeconomic factors, and risk of COVID-19 in the United States and the United Kingdom. *eClinicalMedicine* 2021; **38**: 101029.

8 Mathur R, Rentsch CT, Morton CE, *et al.* Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. *The Lancet* 2021; **397**: 1711–24.

9 Padellini T, Jersakova R, Diggle PJ, *et al.* Time varying association between deprivation, ethnicity and SARS-CoV-2 infections in England: A population-based ecological study. *Lancet Reg Health - Eur* 2022; **15**: 100322.

10 Wood AJ, Sanchez AR, Bessell PR, Wightman R, Kao RR. Assessing the importance of demographic risk factors across two waves of SARS-CoV-2 using fine-scale case data. *PLOS Comput Biol* 2023; **19**: e1011611.

11 UK Ministry of Housing, Communities and Local Government. The English Indices of Deprivation 2019. 2019; published online Sept 26. https://assets.publishing.service.gov.uk/media/5d8e26f6ed915d5570c6cc55/IoD2019_Statistical_Release.pdf.

12 Morrissey K, Spooner F, Salter J, Shaddick G. Area level deprivation and monthly COVID-19 cases: The impact of government policy in England. *Soc Sci Med* 2021; **289**: 114413.

13 Phillips T. Oxford Covid-19 Government Response Tracker (OxCGRT), Blavatnik School of Government, University of Oxford. https://github.com/OxCGRT/covid-policy-tracker.

14 Department of Health and Social Care. Coronavirus (COVID-19): scaling up testing programmes. 2020 https://assets.publishing.service.gov.uk/media/5e888f05e90e0707799498b3/coronavirus-covid-19-testing-strategy.pdf.

15 UK Health Security Agency. Testing Times: striving for quality in the COVID-19 national testing service, April 2020 to March 2022. https://www.gov.uk/government/publications/coronavirus-covid-19-scaling-up-testing-programmes.

16 NHS England. National vaccination programmes. https://www.england.nhs.uk/contact-us/privacy-notice/national-flu-vaccination-programme/.

17 Lopez Bernal J, Andrews N, Gower C, *et al.* Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *N Engl J Med* 2021; **385**: 585–94.

18 Lopez Bernal J, Andrews N, Gower C, *et al.* Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. *BMJ* 2021; : n1088.

19 Chua H, Feng S, Lewnard JA, *et al.* The Use of Test-negative Controls to Monitor Vaccine Effectiveness: A Systematic Review of Methodology. *Epidemiology* 2020; **31**: 43–64.

20 WHO. Evaluation of COVID-19 vaccine effectiveness: interim guidance, 17 March 2021. 2021 https://iris.who.int/handle/10665/340301.

21 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**: 1393–400.

22 Chemaitelly H, Yassine HM, Benslimane FM, *et al.* mRNA-1273 COVID-19 vaccine effectiveness against the B.1.1.7 and B.1.351 variants and severe COVID-19 disease in Qatar. *Nat Med* 2021; **27**: 1614–21.

23 Vasileiou E, Simpson CR, Shi T, *et al.* Interim findings from first-dose mass COVID-19 vaccination roll-out and COVID-19 hospital admissions in Scotland: a national prospective cohort study. *The Lancet* 2021; **397**: 1646–57.

24 Nafilyan V, Bermingham CR, Ward IL, *et al.* Risk of death following COVID-19 vaccination or positive SARS-CoV-2 test in young people in England. *Nat Commun* 2023; **14**: 1541.

25 ONS. Census. https://www.ons.gov.uk/census.

26 Office of National Statistics. Office of National Statistics Census 2021 Create a custom dataset. https://www.ons.gov.uk/datasets/create.

27 Laird N, Olivier D. Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques. *J Am Stat Assoc* 1981; **76**: 231–40.

28 Stoica P, Selen Y. Model-order selection. *IEEE Signal Process Mag* 2004; **21**: 36–47.

29 Kvalseth TO. Cautionary Note about R 2. *Am Stat* 1985; **39**: 279.

30 H2O. H2O.ai. h2o: R Interface for H2O. R package version 3.42.0.2 2023. https://github.com/h2oai/h2o-3.

31 Endo A, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res* 2020; **5**: 67.

32 Hodcroft EB, Wohlfender MS, Neher RA, Riou J, Althaus CL. Estimating R_e and overdispersion in secondary cases from the size of identical sequence clusters of SARS-CoV-2. 2024; published online May 27. DOI:10.1101/2024.05.26.24307940.

33 COVID-19 Forecasting Team. Variation in the COVID-19 infection–fatality ratio by age, time, and geography during the pre-vaccine era: a systematic analysis. *The Lancet* 2022; **399**: 1469–88.

34 Rawson T, Hinsley W, Sonabend R, Semenova E, Cori A, Ferguson NM. The impact of health inequity on spatial variation of COVID-19 transmission in England. *PLOS Comput Biol* 2024; **20**: e1012141.

35 Papageorge NW, Zahn MV, Belot M, *et al.* Socio-demographic factors associated with self-protecting behavior during the Covid-19 pandemic. *J Popul Econ* 2021; **34**: 691–738.

36 Nagar SD, Nápoles AM, Jordan IK, Mariño-Ramírez L. Socioeconomic deprivation and genetic ancestry interact to modify type 2 diabetes ethnic disparities in the United Kingdom. *eClinicalMedicine* 2021; **37**: 100960.

37 Ramachandran A. Trends in prevalence of diabetes in Asian countries. *World J Diabetes* 2012; **3**: 110.

38 Stowe J, Andrews N, Kirsebom F, Ramsay M, Bernal JL. Effectiveness of COVID-19 vaccines against Omicron and Delta hospitalisation, a test negative case-control study. *Nat Commun* 2022; **13**: 5736.

39 Platt L, Warwick R. COVID-19 and Ethnic Inequalities in England and Wales [*]. *Fisc Stud* 2020; **41**: 259–89.

40 Yates T, Summerfield A, Razieh C, *et al.* A population-based cohort study of obesity, ethnicity and COVID-19 mortality in 12.6 million adults in England. *Nat Commun* 2022; **13**: 624.

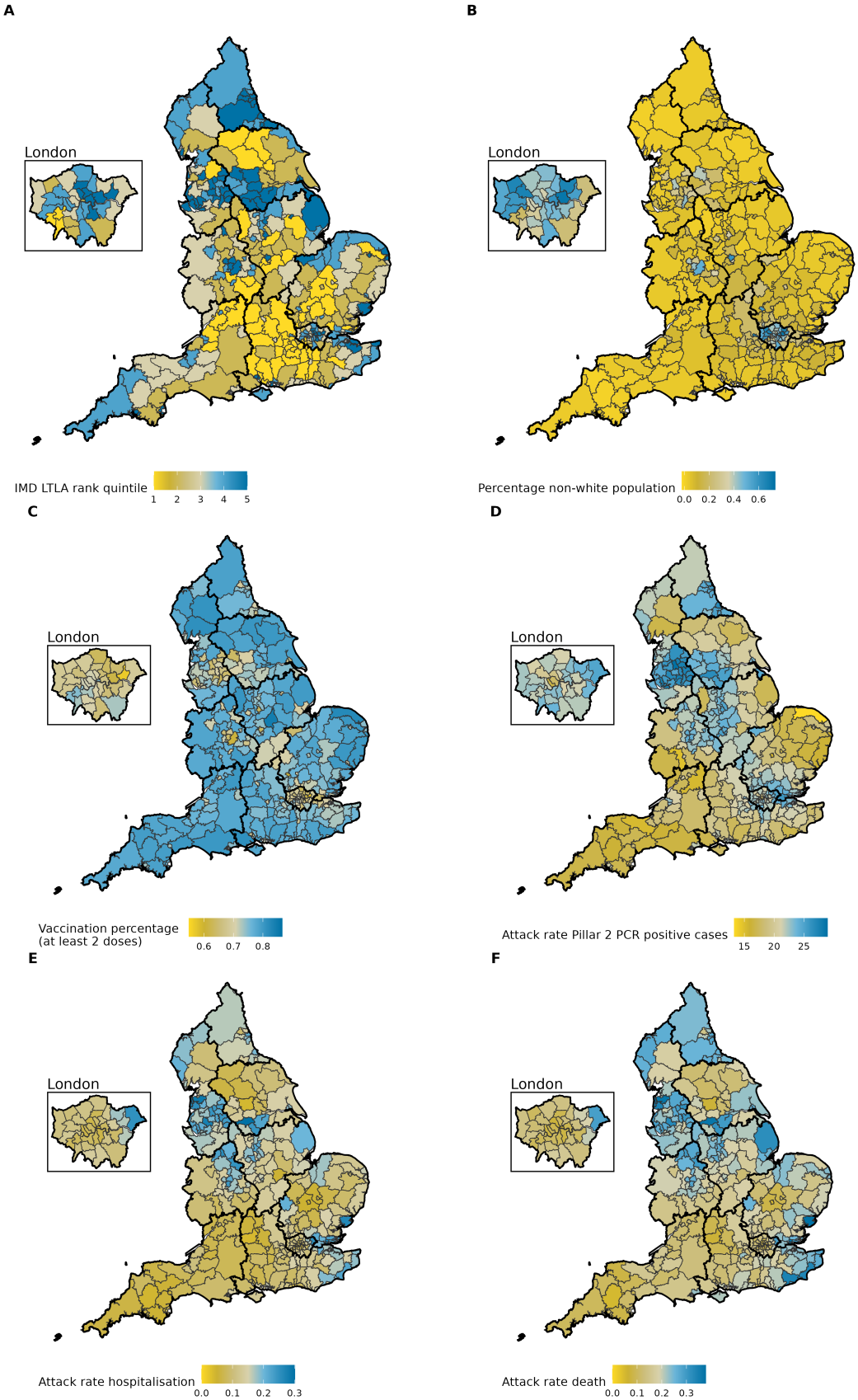

Figure 1: Maps of the population of England at the LTLA level between 10 May 2020 – 27 February 2022. (A) Index of Multiple Deprivation quintile by LTLA, (B) percentage of non-White population by LTLA, (C) percentage of individuals vaccinated with at least two doses for all ages, (D) percentage of positive Pillar 2 PCR confirmed first cases by LTLA, (E) percentage of hospitalisation by LTLA, (F) percentage deaths by LTLA

532 **A**

Variable	Category	n	Person days at risk	Cases (Pillar 2 PCR)	Hospitalisations	Deaths
Sex	Female	28,685,081	17,345,400,702	6,517,293	32,884	49,389
	Male	27,659,329	16,807,116,305	5,793,192	46,431	58,434
Age	under_40	28,863,746	17,378,043,704	7,986,079	951	1,246
	40_49	7,126,246	4,298,749,448	1,832,839	1,909	2,211
	50_59	7,574,556	4,665,987,550	1,400,202	5,251	5,989
	60_69	5,734,391	3,547,069,498	635,915	11,240	12,753
	70_79	4,559,170	2,762,316,669	284,970	21,571	26,508
	above_80	2,486,301	1,500,350,138	170,480	38,393	59,116
Region	East Midlands	4,842,841	2,923,964,911	1,086,738	7,759	10,572
	East of England	6,278,538	3,821,826,715	1,296,183	8,816	12,411
	London	8,837,013	5,309,194,394	1,937,334	10,494	12,285
	North East	2,651,524	1,599,576,314	638,786	4,244	6,407
	North West	7,420,784	4,455,050,061	1,849,257	14,878	19,004
	South East	9,264,652	5,678,938,650	1,890,892	10,965	16,179
	South West	5,680,938	3,491,446,658	1,050,950	4,586	7,439
	West Midlands	5,894,953	3,564,918,434	1,296,108	9,891	12,840
	Yorkshire and The Humber	5,473,167	3,307,600,870	1,264,237	7,682	10,686
IMD Quintile	IMD1	7,674,923	4,699,769,599	1,569,158	7,821	11,696
	IMD2	9,445,584	5,762,949,346	1,921,063	11,564	16,629
	IMD3	10,876,412	6,615,134,330	2,302,369	15,088	20,659
	IMD4	11,801,363	7,137,336,955	2,586,990	16,830	22,278
	IMD5	16,546,128	9,937,326,777	3,930,905	28,012	36,561
Ethnicity	Asian (other)	1,548,640	913,540,586	322,769	1,101	1,209
	Black	2,307,809	1,356,279,568	601,101	2,688	3,152
	Mixed/Other	2,698,916	1,643,470,031	395,571	1,012	1,143
	South Asian	4,058,419	2,381,064,042	984,164	4,955	5,381
	White	45,730,626	27,858,162,780	10,006,880	69,559	96,938

533

534

535 **B**

Variable	Category	n	Person days at risk	Cases (Pillar 2 PCR)	Hospitalisations	Deaths
Vaccination status	Not Vaccinated	12,524,706	21,652,863,421	7,030,900	59,942	80,601
	Over 3w after first dose Adenovirus	343,341	1,168,792,975	97,771	922	2,419
	Over 18w after second dose Adenovirus	2,781,388	1,561,139,076	1,068,220	4,299	5,879
	2-10w after booster dose >2w Adenovirus	7,434,225	772,914,618	428,088	806	966
	10-18w after booster dose Adenovirus	9,188,109	332,906,007	137,555	931	1,522
	Over 18w after booster dose Adenovirus	508,392	12,907,790	5,238	51	148
	Less than 3w after first dose mRNA	253,760	438,532,528	210,475	2,010	2,364
	Over 3w after first dose mRNA	2,462,228	1,187,931,423	542,990	1,081	2,388
	Less than 2w after second dose mRNA	331,407	234,175,466	47,748	66	66
	2-10w after second dose mRNA	1,206,396	829,053,400	113,626	78	132
	10-18w after second dose mRNA	967,000	919,526,650	373,924	492	398
	Over 18w after second dose mRNA	4,066,262	1,129,558,164	751,049	2,853	3,765
	Less than 2w after booster dose mRNA	301,068	178,424,050	168,635	227	151
	2-10w after booster dose mRNA	6,054,894	586,137,083	294,073	764	770
	10-18w after booster dose mRNA	5,512,367	326,242,322	207,494	1,279	1,809
	Over 18w booster dose mRNA	1,954,193	43,244,152	16,974	236	385

536

537

538 **Table 1: Population for England with numbers of first Pillar 2 PCR confirmed SARS-CoV-2, hospitalisations and**
539 **deaths between 10th May 2020 and 27th February 2022. (A) Population breakdown by sex, age, region, deprivation,**
540 **and ethnicity. (B) Breakdown by vaccination status for categories with more than 200,000 doses at the end of the**
541 **study period (n is the count of individuals in this category at the end of the study period, person days at risk and**
542 **counts of Pillar 2 PCR positive cases, hospitalisations and deaths are over the entire study period). Vaccination status**
543 **categories are by vaccination type (Adenovirus or mRNA), vaccine dose, and time since vaccination. The whole table**
544 **for panel B is available in SI Figure D.9.**

545

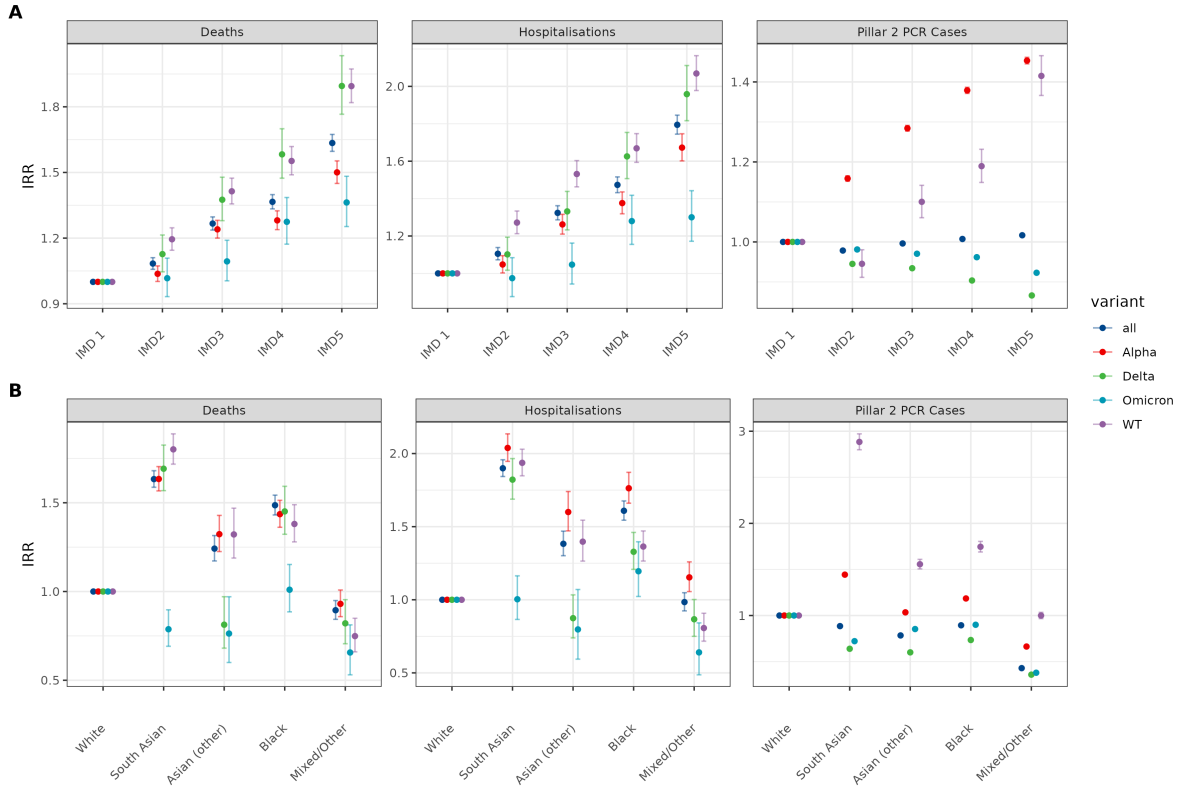

Figure 2: Estimated IRRs for (A) deprivation and (B) ethnicity covariates from preferred models. Deprivation was categorised by quintiles of the Index of Multiple Deprivation score defined at the LTLA level (results for sub-components are available in the SI), with the first (least deprived) quintile (IMD1) being the reference group (IRR=1). Results are shown for the whole pandemic ('all' - test dates between May 2020 and February 2022) and for the time intervals within that period where specific viral variants dominated (see SI for details). WT=Wild-Type (pre-December 2020).

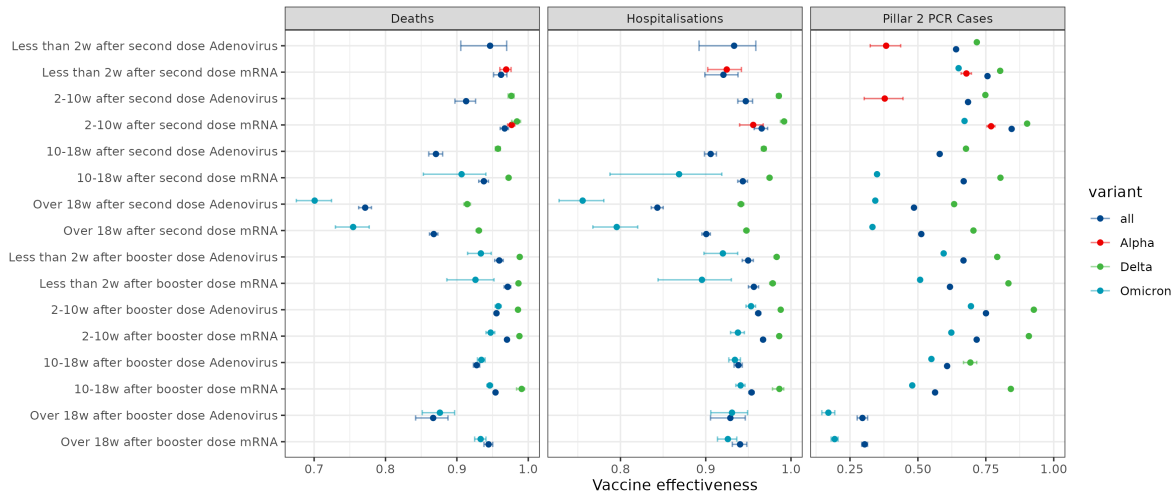

Figure 3: Vaccine effectiveness (VE=1-RR) for different vaccination statuses with not vaccinated as the reference group (IRR=1 / VE=0). Results are shown for the whole pandemic ('all' - test dates between May 2020 and February 2022) and for the time intervals within that period where specific viral variants dominated (see SI for details). Results for the Second Dose and Booster Dose categories are displayed for mRNA and Adenovirus vaccines (all other results are in SI Table D.3 and Figure D.11). WT=Wild-Type (pre-December 2020).

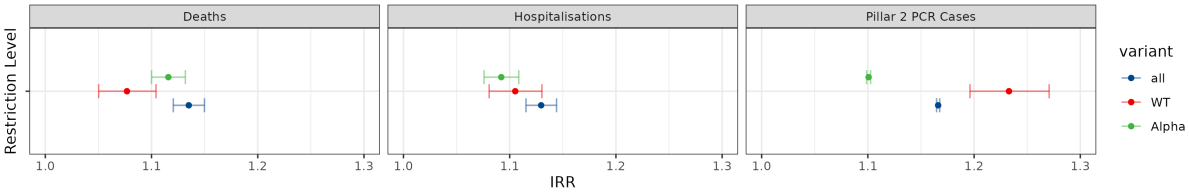

Figure 4: Estimated IRRs for restriction levels. Results are shown for the whole pandemic ('all' - test dates between May 2020 and February 2022) and for the time intervals within that period where specific viral variants dominated (see SI for details). Restriction level is a numeric covariate in the model.