European Economic Review 170 (2024) 104882

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Using double-debiased machine learning to estimate the impact of Covid-19 vaccination on mortality and staff absences in elderly care homes.

Sourafel Girma [a], David Paton [b,*]

[a] Nottingham University School of Economics, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
[b] Nottingham University Business School, Jubilee Campus, Nottingham NG8 1BB, UK

ARTICLE INFO

JEL classification:
C1, I18

Keywords:
Machine learning
Vaccines
Care homes
Covid-19

ABSTRACT

Machine learning approaches provide an alternative to traditional fixed effects estimators in causal inference. In particular, double-debiased machine learning (DDML) can control for confounders without making subjective judgements about appropriate functional forms. In this paper, we use DDML to examine the impact of differential Covid-19 vaccination rates on care home mortality and other outcomes. Our approach accommodates fixed effects to account for unobserved heterogeneity. In contrast to standard fixed effects estimates, the DDML results provide some evidence that higher vaccination take-up amongst residents, but not staff, reduced Covid mortality in elderly care homes. However, this effect was relatively small, is not robust to alternative measures of mortality and was restricted to the initial vaccination roll-out period.

1. Introduction

Many countries experienced significant numbers of Covid-19 related deaths in elderly care homes throughout the pandemic (see for example Chen et al., 2021; Bjoerkheim and Tabarrok, 2022). In response, many governments, including in the UK, focused vaccination efforts on such care homes including priority allocation of resources, vaccination promotion campaigns and, in a number of jurisdictions, compulsory vaccination for care workers.

The primary aim of the vaccination programme was to reduce mortality amongst elderly residents, a group particularly vulnerable to Covid-19. This aim might be achieved in a number of ways: vaccination of workers leading indirectly to lower rates of transmission amongst residents via less transmission from workers; vaccination of residents leading directly to fewer infections; vaccination of residents leading to lower mortality rates amongst those who still get infected. A secondary aim of the programme was to reduce staff absences. This aim might be achieved by high rates of vaccination of workers leading to lower infection rates and hence fewer Covid-related absences.

At the same time, there have been concerns that some of these benefits might be limited by side effects caused by vaccination and which may have contributed to non-Covid related staff absences and even potentially to non-Covid mortality amongst residents (Lv et al., 2021).

The key research questions of this paper are whether vaccination efforts in elderly care homes led to reductions in resident mortality and staff absences and, if so, what was the magnitude of such effects.

* Corresponding author.
E-mail address: David.Paton@nottingham.ac.uk (D. Paton).

https://doi.org/10.1016/j.euroecorev.2024.104882
Received 23 July 2024; Received in revised form 1 October 2024; Accepted 1 October 2024
Available online 6 October 2024

There is evidence from randomised controlled trials (see UKHSA, 2022; Huang and Kuan, 2022 and Rahmani et al., 2022 for reviews of this literature) that Covid vaccination can provide significant protection against serious illness and mortality relative to unvaccinated people who have not previously been infected with Covid. This provides an obvious route whereby higher vaccination rates for care home might reduce mortality from Covid.

Early on in the programme, there were hopes that vaccination would also prove highly effective against infection and transmission. Over time, it became clear that vaccination provided only limited and short-term effectiveness against infection (Andrews et al., 2022; Goldberg et al., 2021) and this raises doubt about the potential for the vaccination programme to prevent transmission from staff to residents.

Further, there are reasons why even the direct impact of vaccination on mortality may be lower in real-world situations than that observed in RCTs. These include cohort and behavioural effects rooted in the decision-making process regarding vaccination.

Cohort effects occur if groups facing lower risk from Covid, such as those with immunity from a prior infection, are less likely than others to choose to get vaccinated. It is well established that prior immunity provides protection that is at least as strong and durable as vaccination (Goldbert et al., 2021; Covid-19 Forecasting Team, 2023).[1] If the vaccinated population includes a relatively low proportion of previously infected people, then even if vaccination provides protection on an individual basis, at a population level we may observe infection or mortality rates that are more similar between vaccinated and unvaccinated populations.

Behavioural effects may include people who are vaccinated responding to their perceived lower risk by taking fewer precautions against catching Covid than those who are unvaccinated. For example, Antonelli et al. (2022) provide evidence that vaccinated people who get infected were more likely to experience very mild symptoms. This may be particularly important in the context of elderly care homes if vaccinated staff remain unaware of their infection and continue to work even when they are at risk of infecting residents.

As a result of these effects, even if vaccination reduces the mortality risk on an individual basis, the overall impact of higher vaccination rates on mortality may be more ambiguous. Such factors can help to explain why observational data often reveal infection rates no different and sometimes higher amongst vaccinated groups compared to those unvaccinated, even controlling for age (UKHSA, 2022).

The empirical evidence on whether, in fact, vaccination rates do have causal effects in reducing mortality is mixed. Looking at aggregate data, observed mortality rates during infection waves were significantly lower in many countries after the vaccination roll out relative to before, suggesting vaccination may be an important factor in reducing mortality. However, this association was not consistent. Places such as Hong Kong, New Zealand and Taiwan that had seen relatively low mortality during 2020, all saw increased Covid-related mortality rates during 2022, well after most people in their populations had been vaccinated. For example, the 7-day average mortality rate in Hong Kong reached 40 deaths per million in March 2022. This was a rate far above the highest peak seen in the US during the whole pandemic despite cumulative vaccination rates at that time being similar.[2]

There are two important trends that may act as confounders in such temporal associations. One is the growth in importance of Covid-19 variants such as Omicron that are highly infectious but less lethal than variants dominant prior to the vaccination rollout. The second is the cumulative growth over time of those who have experienced but survived Covid infection and, as noted above, have a high level both of immunity from subsequent re-infection as well as protection from serious illness in the event of a re-infection.

The co-existence of these trends with the vaccine rollout makes it hard to identify causal effects of vaccination in aggregate data and this is reflected in the limited and inconsistent evidence documenting causal impacts. Subramanian and Kumar (2021) use both country-level and US county level data and find "no discernable relationship between percentage of population fully vaccinated and new COVID-19 cases" (p.1237). Gibson (2023) similarly finds "no apparent relationship between vaccination rates and changes in Covid-19 cases or changes in mortality" (p. 2999). In contrast, Agrawal et al. (2023) use an event study framework on a sample of 43 countries and conclude that Covid-19 vaccination rates had a significant impact in reducing excess deaths. Extrapolating their results, the authors estimate that the vaccination campaign averted over two million deaths. These findings are consistent with some other population-level studies such as Chen (2023) and Hoxha et al. (2023). However, crucially none of these studies control for the effect of immunity from prior infections, nor do they address other issues of causality. A notable exception is Gibson (2022) who exploits exogenous interruptions in vaccine supply in New Zealand to estimate instrumental variable models concluding that the booster roll-out led to a significant <u>increase</u> in overall excess mortality.

One possible explanation for the difficulty in identifying a strong vaccination effect is that much of the existing research has examined impacts at the whole population level. A feature of Covid-19 is its discrimination by age and underlying health. For those aged under 60 and without underlying health problems, Covid-19 induced mortality was very low meaning there is little potential for vaccination to have material impacts on mortality for a large proportion of the population. As a result, studies testing for vaccination effects at a population level, face a high bar in achieving statistical significance.

Given this, focusing on the most vulnerable groups such as elderly care home residents is potentially a more fruitful source of improving our understanding of the impact of vaccination on mortality. To date there has been only limited research on the impact of Covid vaccination on mortality amongst the elderly care home resident population or on staff absences.

Maltezou et al. (2021; 2022) find some evidence that Covid-19 vaccination can reduce absenteeism amongst healthcare workers, though Chrissian et al. (2022) reports a high level of staff absences due to side effects from Covid-19 vaccination amongst health

[1] Vaccination may still have provided marginal additional protection to those previously infected.

[2] Our World in Data https://ourworldindata.org/coronavirus/country/hong-kong

workers.[3] Most recently, Girma and Paton (2023) find that a vaccination mandate for care homes in England had no observable impact on resident mortality, though it did lead both to a notable increase in take-up amongst staff but also to a net reduction in the number of staff working in elderly care homes.[4] Although the vaccine mandate itself appears to have been unsuccessful in reducing mortality, it remains possible that higher vaccination rates amongst staff led to at least some short-term impact on infection and transmission and hence some reduction in resident mortality.

The aim of this paper is to provide evidence on the impact of vaccination take-up of both staff and elderly residents on Covid and non-Covid mortality for residents as well as on other outcomes such as staff absences.

The measurement of vaccination take-up in statistical work is not a trivial issue for at least two reasons. First, the effectiveness of Covid-19 vaccination is known to decay significantly over time (Covid-19 Forecasting Team, 2023). Second, most vaccination campaigns involved people receiving a second, and sometimes third or fourth dose with the aim of restoring vaccine effectiveness and it is not obvious how to combine these additional doses into a single 'treatment' measure. Here we build on existing work to propose a new 'vaccine stock' measure that takes account of both these issues.

We use data from English care homes at the local authority level. Our basic statistical approach is to estimate the association between vaccination levels and subsequent outcomes such as mortality. Establishing causality in such a framework is notoriously challenging. The linear fixed effects model has been a workhorse in the applied economics literature, mainly because of its ability to control for unobserved time-invariant heterogeneity which might otherwise compromise causal inference. In our context, by controlling for time- and area-specific effects, we can examine whether areas that experienced relatively large increases in vaccination take-up also experienced relatively small increases (or large decreases) in subsequent mortality. Time-varying differences in areas can potentially be accounted for by including as controls confounding variables such as community infection rates.

A shortcoming of this approach is that establishing causality depends on making strong assumptions about the role the confounding variables play in the association between the key variables of interest. There is also the difficulty in standard regression models of dealing with a high number of potential control variables. Here, we apply the double-debiased machine learning (DDML) approach, a method that obviates the need to assume specific-functional forms regarding the relationship between the outcome variable and covariates on the one hand, and treatment variables and covariates on the other.

Machine learning is still relatively novel in the context of healthcare in general and vaccination in particular. Exceptions include Olenski and Sacher (2024) who use variational inference to help identify the impact of nursing home quality on Covid-19 outcomes and Xiangxiang et al. (2023) who use DDML to estimate the causal impact of the take-up of anti-dementia drugs on patient outcomes whilst Cuadros et al. (2023) and Chernozhukov et al. (2023) both use DDML to examine the causal determinants of vaccination uptake. Magazzino, Mele and Coccia (2022) use machine learning to examine the impact of vaccination on infection fatality rates across countries, but at a whole population level and only considers the initial vaccine rollout period. To our knowledge, our paper is the first to use DDML to estimate the causal impact of vaccination itself on healthcare outcomes in elderly care homes.

The rest of the paper is set out as follows. In the next section, we provide an overview of Covid vaccination in English care homes. In Section 3, we set out our statistical method and describe our data. We present our results in Section 4 and in the final sections discuss the implications of our findings and conclude.

2. Overview of Covid vaccination in English care homes

The English national covid-19 vaccination programme was the first in the world and was launched on 8th December 2021.[5] The initial aim of the programme was to reduce mortality and pressure on the National Health Service (NHS). To this end, the programme focused on a series of priority groups set out by the Joint Committee on Vaccination and Immunisation (JCVI). The first JCVI priority group was "Residents in a care home for older adults and staff working in care homes for older adults" (UKHSA, 2023, p.19).

The care home vaccination effort was carried out largely by GPs and through Community Trust services (Mountier-Jack et al., 2023). Figs. 1a and 1b illustrate vaccination trends for elderly care home staff and residents respectively. We show the trend for each of the first dose (from December 2020), second dose (from March 2021), the first booster (from autumn 2021) and finally the second booster (from March 2022). Note that the latter was offered only to residents and not staff.

Take up amongst residents (Fig. 1a) was rapid and high for the first two doses, eventually reaching around 96 % coverage. There was, however, considerable variation across local authorities in both speed and level of take-up. For example, two weeks after the start of second dose rollout, coverage varied from as low as 34 % in Thurrock to a high of 94 % in Hammersmith and Fulham. Two months after the roll-out, coverage had reached over 90 % but ranged from 78 % (in Blackburn) to 96 % (Plymouth).

Take-up of the first booster was much slower. Two months from the booster roll out, average take-up was 67 % but ranging from under 30 % (Rutland and Reading) to over 85 % in Richmond-upon-Thames. Average take-up eventually reached just under 90 %. The second booster (only offered to residents) was slower yet, reaching around 40 % after two months.

Staff take-up (Fig. 1b) was consistently slower than for residents. For this reason, in June 2021, the UK Government announced its intention to introduce a vaccine mandate for care home workers in England. The mandate made it a condition of employment at residential care homes, that workers should have two doses of a recognised Covid vaccine no later than 11th November 2021. As a

[3] Earlier research such as ten Doesschate et al, (2022) and Verelst (2021) examines the impact of vaccination for other diseases on workplace absenteeism amongst healthcare workers.

[4] Abouk et al (2024) similarly find that vaccine mandates for healthcare workers in the US exacerbated workforce shortages.

[5] Mounier-Jack et al. (2023) provide an overview of the Covid-19 vaccination roll-out in England.

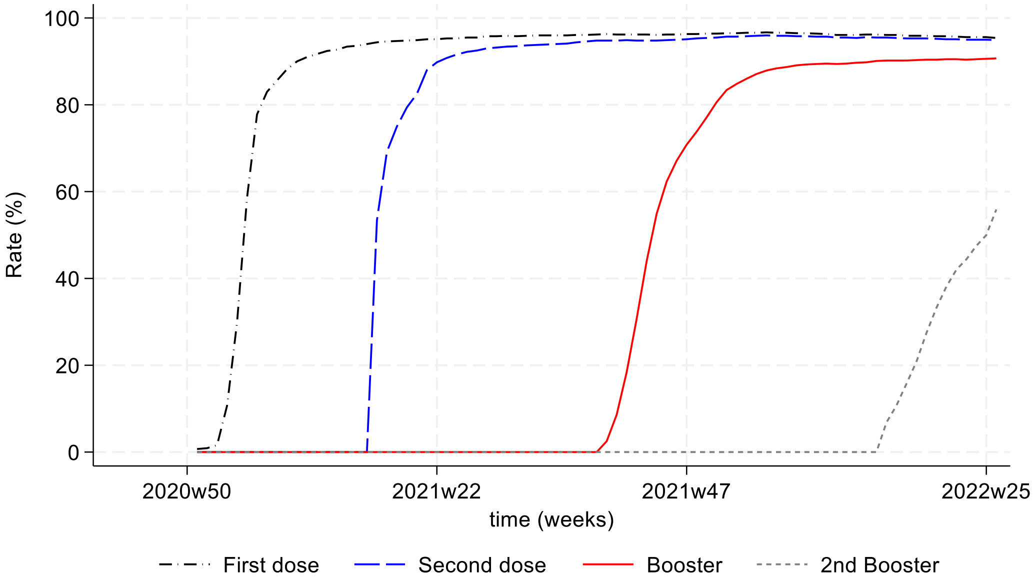

Fig. 1a. Vaccination rates for elderly care home residents in England.
Notes: (i) Sample period is from 2020 week 50 until 2022 week 26.
(ii) Data are from the Department for Health and Social Care, link provided in Appendix Table A1.

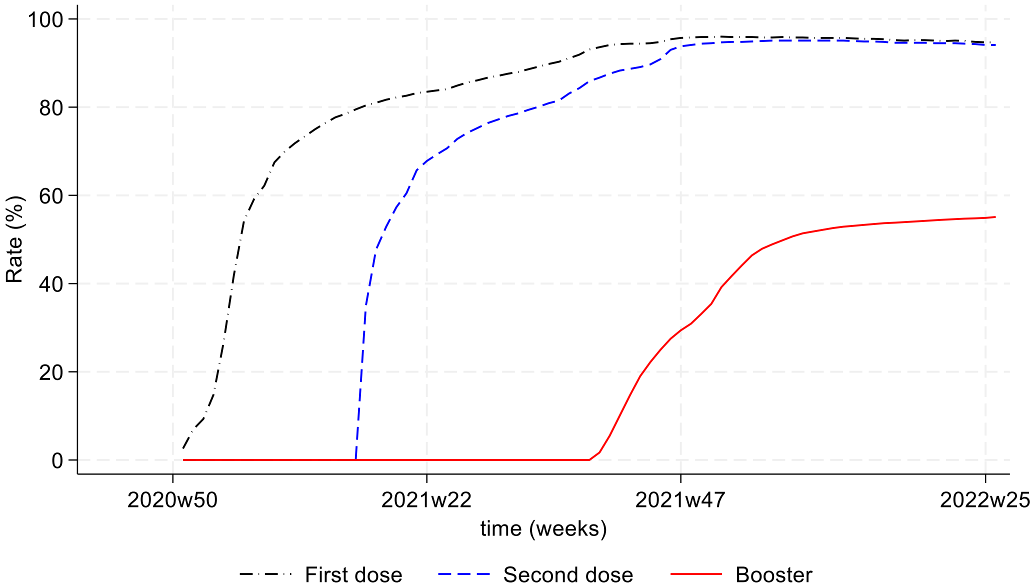

Fig. 1b. Vaccination rates for elderly care home staff in England.
Notes: (i) Sample period is from 2020 week 50 until 2022 week 26.
(ii) Data are from the Department for Health and Social Care, link provided in Appendix Table A1.

result, takeup amongst staff of the first and second doses rose to about 95 % with a range of 89 % to 100 %. However, although the mandate was effective in increasing take-up, it also led to many workers leaving the care sector. Girma and Paton (2023) estimate that, as a result, net staffing in elderly care homes in England reduced by between 3 and 4 per cent.[6]

The mandate did not require staff to take part in the autumn 2021 booster programme and, as a result, staff take-up of the booster stalled at around 50 % and as low as 26 % in some local authorities. As noted above, the second booster campaign in spring 2022 was

[6] Butler, Butler, Farnell and Simmons (2024) provide an interesting analysis of worker performance in the context of Covid-19 infections.

rolled out only to care home residents and not staff.

3. Methodology and data

3.1. Statistical methodology: doubly debiased machine learning (DDLM)

Classical regression strategies have been widely used in applied econometrics due to their ability to control for unobserved heterogeneity through fixed effects and their capacity for making causal inferences about specific parameters. However, traditional regression approaches face two major limitations that can be problematic in practice.

First, they break down in the presence of high-dimensional data, as they lack a systematic mechanism for selecting the most relevant confounders which capture potential nonlinearities and interactions. Second, they rely on assumptions about the functional relationship between the outcome, treatment variables, and confounders. These assumptions, such as linearity and additivity, are often made without substantive economic justification.

One might be tempted to include as many control variables as possible and use a highly flexible functional form, but this risks including irrelevant variables and compromising causal inference. Moreover, the confidence intervals for the causal parameters of interest may end up with incorrect coverage.

The double-debiased machine learning (DDML) approach offers an effective solution, allowing researchers to control for confounders even in high-dimensional data without relying on subjective judgments about functional form. DDML combines the predictive power of machine learning with traditional econometric methods, yielding estimates that can be interpreted as causal. Specifically, covariate adjustment functions are estimated using machine learning models, making them robust to functional form misspecifications and overfitting, while average treatment effects are obtained through orthogonalized regressions, as explained below.

We are seeking to evaluate the impact of Covid-19 vaccination (treatment or policy variables denoted by the vector d) on mortality and staff absences in elderly care homes (the outcome variables denoted y) conditional on a set of *potential* control variables or covariates x. All outcome variables are measured relative to their value six weeks earlier because such differencing has the merit of removing time-invariant unobserved confounders, if any.[7]

Adopting the notation of Chernozhukov et al. (2018), we consider a partially linear model as follows:

$$y_{it} = \theta_0 d_{it} + g_0(x_{it}) + \zeta_{it}, \tag{1}$$

$$d_{it} = m_0(x_{it}) + v_{it}, \tag{2}$$

In the above, i is used to index (148) local authorities in England and t denotes time measured at weekly frequency. The effective estimating sample runs from the week 23 of 2020 to week 26 of 2022.

The equation of interest is [1] and θ_0 is the vector of regression coefficients that we would like to estimate. The covariates are hypothesised to be related to the outcome variable through the unknown function $g_0(x_{ti})$.

The second equation specifies the nature of the confounding, where the unknown function $m_0(x_{it})$ captures the possible dependence of d on x. ζ_{it} and v_{it} represent equation errors.

We consider two treatment variables; residents and staff vaccines stocks. For each of these, we consider contemporaneous and lagged effects up to six weeks[8] (giving a total of 7 sub-treatment variables in each case). This means we have a 14 $\times 1$ vector of treatment variables.

When looking for causal effects of vaccination, we need to be aware of the likely timing of impacts. For example, mortality from Covid occurs, on average, around three weeks after infection. When considering the impact of staff vaccination on resident mortality, we can expect a further lag to allow time for reduced infection of residents from staff. The lagged effect is likely to be shorter for other outcome variables such as staff absences but even in that case, an absence due to a positive Covid test will reflect infection some time earlier. For this reason, estimated effects of current vaccination and (for Covid-related mortality at least) the first few lags can be thought of as an *ad hoc* placebo test with any observed impact unlikely to be causal.

The vector of *potential* covariates x to be used in the nonparametric functions $g_0(x_{ti})$ and $m_0(x_{ti})$ are listed as follows: life expectancy of the population aged 90 or over, local unemployment rate, earnings, population density, public health expenditure per capita, the number of care home residents aged 65 or over, the number of residents aged 85 or over, the percentage of the total population that is over 90 years old, Black, Asians and non-British White, local authority level infection rates amongst the general population lagged by three weeks, cumulative Covid death rates in care homes lagged by three weeks, weekly time trend and local authority level fixed effects.

The motivation for these controls is rooted in the epidemiological evidence on the determinants of mortality from Covid. For example, there is evidence (e.g. Mackey et al., 2020; Pathk et al., 2022) that mortality rates were higher in areas of economic deprivation and amongst particular ethnic and demographic groups. Acharya et al. (2021) argue that differences in public health expenditure may help to explain infection trends. Community infection rates accounts for differential timing across regions of Covid

[7] Note that using a shorter lag would risk the effects being influenced by the treatment, whilst a longer lag would mean dropping more observations. In our supplementary analysis, reported in the Appendix, we report results in which the outcomes are measured in levels.

[8] See the appendix for a detailed explanation of the construction of the vaccination stocks.

waves. Finally, the inclusion of cumulative death rates is intended to control for the fact that high previous infection rates in an area is likely to result in lower mortality rates from subsequent infections.[9]

A number of the controls are measured at the local authority level and as such may not reflect demographics specific to the care home population. Including local authority fixed effects as potential controls provides a more generalised approach that can reflect unobservable factors that may be specific to residential care homes in particular local authorities.

Employing as many control variables as possible and choosing a highly flexible functional form risks overfitting and compromising causal inference. For instance, the confidence intervals for θ_0 might end up having the wrong coverage (Chernozhukov et al., 2018). On the other hand, adopting a parsimonious approach risks underfitting the conditioning functions and biasing the treatment effects estimator. The role of machine learning algorithms, therefore, is to select the optimal set of conditioning variables and associated functional form via the functions $g_0(x_{it})$ and $m_0(x_{it})$.

However, as shown by Chernozhukov et al. (2018), a direct application of off-the-shelf ML methods to estimate the causal parameter, θ_0 is not generally valid. This is because doing so introduces so-called regularisation bias on account of the fact that ML methods trade off variance for bias by design.

In order to overcome (*debias)* this regularisation bias, Chernozhukov et al. (2018) propose to estimate two (*double*) equations using suitable ML learners. In this paper we use the random forest algorithm with repeated 5-fold cross validation clustered at local authority level. Throughout, the root mean squared error is used as the loss function. In brief, our DDML method can be described as follows:

1. Apply a random forest learner of y on x to estimate $\hat{g}_0(\mathrm{x})$ and partial out the effect of x from y to obtain $\boldsymbol{z_{it}} = y_{it} - \hat{g}_0(x_{it})$.
2. Use a separate random forest learner of each element of the treatment vector d, say $\mathrm{d}_{(k)}; k = 1,\ldots,14$ on x to estimate $\hat{m}_{0(k)}(x)$ and partial out the effect of x and obtain the orthogonalised regressor $\boldsymbol{v_{(k)it}} = d_{(k)it} - \hat{m}_{(k)0}(x_{it})$. Once this is done, form the vector of orthogonalised regressors as $(\boldsymbol{v}_{(1)\boldsymbol{it}} \quad \cdots \quad \boldsymbol{v}_{(14)\boldsymbol{it}})'$
3. Regress $\boldsymbol{z_{it}}$ on $\boldsymbol{v_{it}}$ to obtain the vector of causal parameter estimates.

If the nuisance models $\hat{g}_0(\mathrm{x})$ and $\hat{m}_0(\mathrm{x})$, and the final parameter $\hat{\theta}_0$ are estimated *on the same* observations this would lead to poor out-of-sample performance. A creative solution to this problem is sample splitting. This involves estimating the nuisance models $\hat{g}_0(\mathrm{x})$ and $\hat{m}_0(\mathrm{x})$ on one part of the data; and then obtain $\hat{\theta}_0$ using the other part of the data. In this paper, we follow best practice by employing cross-fitting which switches the role of the model selection and estimation samples to ensure that the entire sample is used for the final estimation.

An important contribution of Chernozhukov et al. (2018) is in showing that this OLS estimator converges to the true value as the sample size increases (i.e. it is consistent) thereby solving the regularisation bias associated with the naive ML approach.

As already stated, our baseline models utilize random forest machine learners, a popular type of supervised machine learning model.[10] Random forests combine multiple decision trees into powerful ensemble models. These tree-based models use nonparametric algorithms to recursively partition the data into non-overlapping regions. A key advantage of random forests is their random selection of a fraction of features (conditioning variables) when constructing each tree, which helps to de-correlate the trees and enhance predictive performance.

We also explore the robustness of this approach in our supplementary analysis by estimating the effects using the parametric elastic net machine learner instead of the non-parametric random forest algorithm. We implement the procedure using the R package, ddml (Ahrens et al., 2024).

We compare the results using the DDML approach with estimates from more standard panel data regressions with local authority and week-specific fixed effects. Again, we estimate different models for various potential lags of our treatment variables.[11]

3.2. Data

We use weekly data from the Department of Health and Social Care (DHSC) on vaccination take-up of each dose for each English upper-tier local authority for elderly care home residents and staff. Data are based on returns provided weekly by individual care homes.

The data are published for upper tier local authorities. These are local government regions that vary in size but, have a mean population of just over 400,000 and cover a mean area of around 300 miles^2. In terms of population size, they are, on average, somewhat bigger than US counties and similar to French Departments. Each local authority contains an average of about 75 elderly care homes. Due to boundary changes mid-sample, we dropped Northamptonshire from the sample leaving a total of 148 English authorities. Weekly data on staffing and vaccinations were collected from the start of the vaccination roll out at the end of 2020,

[9] Cumulative prior infections in care homes would be a more direct measure but these data are not available consistently for care homes.

[10] James et al (2021) provides a helpful overview of various machine learning algorithms.

[11] The presence of local-authority fixed effects means time-invariant covariates drop out of the model. The remaining co-variates are infection rates amongst the general population and cumulative Covid death rates in care homes each lagged by three weeks.

though response rates from care homes were relatively low for the first few weeks of the rollout.[12] The vaccination data include rates for each of the first dose, second dose, first booster and (for residents only) second booster.

Our baseline estimates use data from 2020, week 23.[13] Although some data are available before this time, there are concerns over its reliability due to rapidly changing testing and diagnostic practices in the early part of the pandemic.[14]

We collect data from the Office of National Statistics (ONS) on weekly deaths each week in each local authority. The ONS report includes the total number of deaths as well as those in which Covid-19 is mentioned as a cause on the death certificate and each series is broken down by those occurring in elderly care homes and those occurring elsewhere.[15] We use the ONS series by week of death occurrence (rather than week of registration).

From these data, we construct three weekly measures of mortality:

(i) Deaths from Covid in care homes per 1000 resident population.
(ii) Total deaths in care homes per 1000 resident population.
(iii) Deaths from Covid in care homes as a percentage of all Covid deaths.

We see these measures as complementary ways of assessing the vaccination impact. Looking at total deaths in care homes gets around any potential problems with the classification of deaths as from Covid. The motivation for measure (iii) is that care home staff and residents were top of the UK Governments priority groups for vaccination. As a result, we would expect to see a particular reduction in Covid deaths in care homes as a proportion of the total as the vaccine was rolled out and in those areas where take up in care homes was relatively high. As a caveat to this point, the elderly and other at-risk adults were also on the vaccination priority list, albeit below care home staff and residents.

Data on care home staffing are taken from the DHSC workforce statistics. We use two different outcome measures, each using weekly data for upper tier local authorities. The first measure is the percentage of directly employed staff who are absent due to Covid-19. Unfortunately, these data are only available from week 6 of 2021 so we are unable to track absences before and during the start of the initial vaccination roll out. Note that throughout the period we study, both staff and residents were subject to systematic Covid-19 testing so absences should not be affected by variations in testing levels.

Weekly staff absences for non-Covid reasons are not published.[16] However, DHSC do publish data on the percentage of care homes each week that are either in breach or at risk of being in breach of agreed staffing levels that are deemed to be safe. This measure is likely to reflect absences due both to Covid but to other illnesses (including those due to vaccination side effects) as well as any staff recruitment or retention issues, for example those related to the vaccine mandate. The staffing breach data are available from 2020 week 50, i.e. the time when the initial vaccination programme was launched.

For each local authority, we also collected data on the other potential confounding variables: life expectancy of the population aged 90 of over, unemployment rate, earnings and public health expenditure, demographic information on population density, population by age, and ethnicity, the lagged weekly community infection rate in each local authority, calculated as the number of positive Covid tests reported to the UKHSA per 100,000 total population and, finally, the lagged cumulative rate of care home deaths from Covid.

The best way to measure our treatment (vaccination) is not obvious given the relatively fast waning of vaccine effectiveness noted above. Agrawal et al. (2023) deal with this issue by estimating a 'vaccine stock' at each time period that allows for waning. We follow their approach by calculating the vaccination stock as being the percentage of each group (residents or staff) who have been vaccinated, subject to a 3 % weekly decay rate.[17]

This approach, however, does not take account of the fact that most of those vaccinated were provided with second and booster doses after a period that may have countered the waning effect of the original dose. Here, we adjust our stock variable to account for second and booster doses by assuming that each dose results in maximum effectiveness but then decays again by 3 % per week. Our precise calculation is provided in the Appendix. We show the national trend of resident and staff vaccination stocks in Fig. 2. The boost in stock levels following each rollout, followed by the gradual decay is clearly visible.

For comparison, in our supplementary analysis, we report estimates using a slower (2 %) and faster (4 %) weekly decay rate as well as an alternative, simpler, measure - the percentage of the relevant population (staff or residents) who have received at least one vaccination dose in each week.

In Table 1, we present some descriptive statistics for our treatment and outcome variables and covariates. For the two treatment variables (staff and resident vaccination stocks), we report statistics lagged by up to 6 weeks to reflect the fact that we can expect some

[12] The potential issue here is if response rates are systematically related to vaccination take-up rates, for example if areas with initially low response rates from care homes are also those where take-up was slower. In our supplementary analysis, we report results with the observation period starting 10 weeks after the initial roll out when response rates were much higher.

[13] The ONS weekly data are reported using a system of 52 weeks for most years (for leap years, they use 53 weeks). In 2021, Week 1 is the week ending 8th January 2021 and Week 52 is the week ending 31st December.

[14] In particular, community infection rates, which are a potentially important co-variate, are likely to be unreliable earlier than this. Using this sample period means we have about 6 months of data before the start of the vaccination rollout.

[15] This does not necessarily mean that Covid-19 was the underlying cause of deaths.

[16] If available, this would provide an additional potential control variable.

[17] A 3% rate is reasonably consistent with reported estimates of vaccine effectiveness, e.g. those published in the UKHSA (2022) Vaccine Surveillance reports.

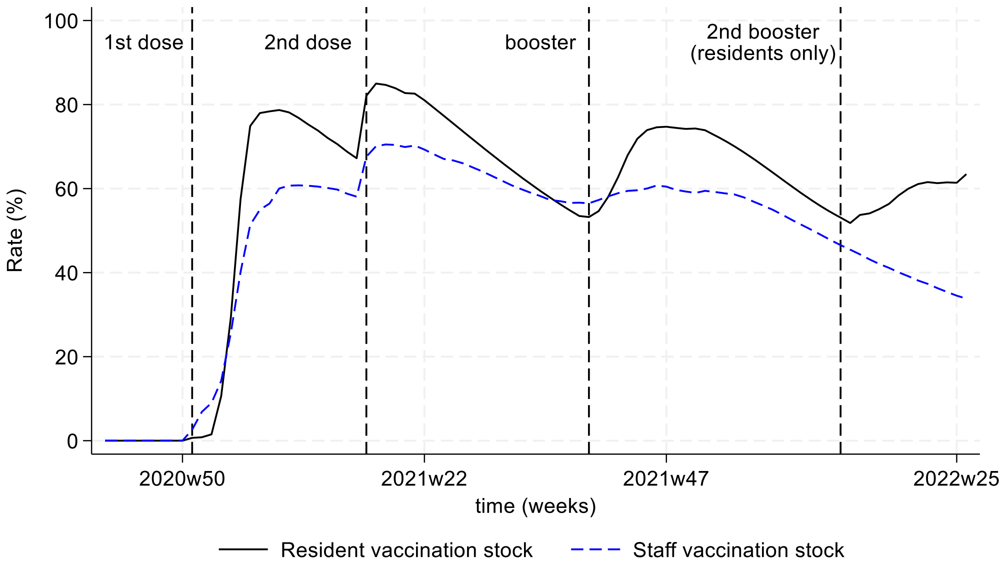

Fig. 2. Vaccination stock for elderly care home residents and staff in England.
Notes: (i) Sample period is from 2020 week 50 until 2022 week 26.
(ii) Data are from the adult social care statistics for England, link provided in Appendix Table A1.
(iii) Stock values are calculated as described in the text and data appendix.

time to elapse between the treatment (vaccination) and outcomes such as infection, staff absence or mortality. Details on the definition of each variable and data sources are provided in Appendix Table A1.

4. Results

4.1. Univariate trends

We present simple time trends of our key outcome variables in Figs. 3 to 5. In Fig. 3a, we show trends in care home deaths in total and also broken down by Covid and non-Covid. Covid deaths peak and then decrease a few weeks following the roll out of the first dose, a period that coincides with the expected seasonal peak in deaths. From spring of 2021, Covid deaths remain low for most of the rest of the period. After a smaller peak in spring 2022, Covid deaths decline again a few weeks after the second booster roll out.

Remembering that deaths will be associated with infections some weeks prior, lower Covid mortality in the post-vaccination period is consistent with vaccination reducing Covid mortality. However as noted earlier, other trends make it difficult to infer a causal effect. First is the gradual growth in the cumulative proportion of residents who will have immunity from having survived a previous Covid infection. Second is the emergence variants such as Omicron that appear to have been highly infectious but less dangerous. A final point is that, to the extent that a high number of particularly vulnerable people had died during the first (pre-vaccine) Covid wave, we would expect the remaining pool of vulnerable residents to be lower during the vaccine rollout.[18] Each of these factors may have contributed to lower post-vaccination mortality even in times of high infection rates.

We explore further whether mortality trends follow a pattern that would have been observed even without vaccination by showing (in Fig. 3b) care home Covid deaths as a percentage of all Covid deaths. Elderly care homes were the first priority group for vaccination. Elderly not in care homes were the next priority and the broader population were vaccinated only from around April 2021. Given this, if vaccination was driving the reduction in mortality, we would expect to see Covid-related deaths in care homes decrease relative to those in the rest of the population after the first dose roll out. If anything, we see the opposite effect about 3–4 weeks from the start of the rollout. The trend is mixed with the percentage declining after the second dose (a period when the broader population were receiving their first dose), then increasing significantly in late 2021 before decreasing again following the rollout of the 2nd booster dose.

Overall, the descriptive data provide a mixed picture on the link between the vaccine roll out and care home resident mortality.

The final two graphs display trends in staff absences from Covid. In Fig. 4 we show the percentage of staff reported as being absent from Covid each week. The data are only available from spring 2020. Absence rates decline in the run up to the second dose and shortly after. However, this trend soon levels off before increasing again. There then follows a further decrease during summer and autumn of 2021, but there is no sign of the trend being affected by the rollout of the first booster dose. Absence rates do decline sharply after the 2nd booster, but recall that this was not targeted at staff.

[18] We thank an anonymous referee for this point.

Table 1
Summary statistics of treatment and outcome variables and covariates.

	mean	median	sd	min	max
Treatment variables					
Resident vaccine stock	47.35	60.13	31.65	0.00	94.62
Staff vaccine stock	39.37	51.68	26.42	0.00	82.42
Resident percentage at least 1 vaccination	66.05	93.70	43.03	0.00	100.00
Staff percentage at least 1 vaccination	61.63	84.90	40.99	0.00	100.00
Outcome variables (relative to values six weeks earlier)					
Covid deaths in care per 1000 residents	−0.33	0.00	3.09	−80.00	60.00
Total deaths in care per 1000 residents	−0.50	0.00	6.49	−140.00	100.00
% Covid deaths occurring in care homes	−0.29	0.00	29.50	−100.00	100.00
% staff absence from Covid	−0.12	−0.10	1.04	−5.70	6.20
% care homes in staffing breach/risk	0.04	0.00	3.26	−33.30	33.30
Outcome variables (levels)					
Covid deaths in care per 1000 residents	0.74	0.00	1.95	0.00	60.00
Total deaths in care per 1000 residents	7.86	7.27	5.00	0.00	100.00
% Covid deaths occurring in care homes	15.79	0.00	24.30	0.00	100.00
% staff absence from Covid	1.05	0.80	0.86	0.00	7.20
% care homes in staffing breach/risk	2.96	0.00	4.20	0.00	33.30
Covariates					
Life expectancy	9.43	9.12	1.45	6.94	15.48
Earnings	622.04	598.10	83.83	470.30	976.20
Unemployment rate	4.76	4.60	1.61	1.90	9.30
% of population aged 90 or over	0.84	0.82	0.27	0.24	1.68
Care home residents aged 65 or over	1766.68	1155.50	1669.68	50.00	8221.00
Care home residents aged 85 or over	1008.78	624.50	1025.16	29.00	5122.00
Population density	2810.33	1566.32	3220.06	64.06	15,794.50
% of population White	81.99	88.79	17.25	27.12	98.32
% of population White non-British	5.21	3.36	4.98	0.60	28.46
% of population Black	4.37	1.45	6.29	0.10	28.34
% of population Asian	9.30	5.39	10.25	0.58	44.91
Community infection rate	301.50	180.50	394.52	0.00	3882.23
Cumulative care home Covid death rate	109.65	104.32	70.79	4.39	880.00
Public health expenditure per capita	0.08	0.07	0.03	0.03	0.19
$N = 16{,}132$					

Notes:.
(i) The summary statistics are from the start of the vaccination rollout in 2020 week 23 until 2022 week 26. Exceptions are staffing breach data (starting point = 2020 week 50, $N = 11{,}988$) and absences (starting point = 2021 week 6,$N = 11{,}396$).
(ii) For covariates that are not measured weekly (life expectancy, earnings, unemployment, public health expenditure and population measures), statistics are for 2021.
(iii) Data sources are as listed in Appendix Table A1.

In Fig. 5 we present data on the percentage of care homes at risk or in breach of overall staffing standards. This should reflect absences from all causes, not just Covid. These data will reflect absences caused by vaccine side effects as well as reductions in staffing caused by the vaccine mandate. The percentage of care homes affected increases from the rollout of the first dose before decreasing sharply. It then trends up again from mid-2021 until early-2022, the period in which the vaccine mandate was being implemented.

In Table 2, we report partial correlations between each of the outcome variables and the vaccination stocks. Some correlations are significant and consistent with a vaccination effect. For example, resident vaccination is associated with a significantly lower Covid death rate when using the 4- and 5-week lags. However, the correlation is not significant for the 6-week lag. For other outcomes, and similar to the univariate time trends, there is little consistent pattern in the correlations either in terms of the statistical significance or even the direction of correlations.

Overall, the univariate time trends and partial correlations present only limited and inconsistent evidence of vaccination effects on mortality or staff absences. That does not necessarily mean vaccination had no effect – the trends are likely to have been influenced by a range of factors (including the prevalence of community transmission) and which may serve to mask a vaccination effect. To this end, we now turn to the fixed effects and DDML estimates of the average treatment effect parameters which allow us to exploit differences in vaccination take-up across local authorities to help identify causal effects.

4.2. Results from machine learning techniques

4.2.1. Variable importance (VIP) analysis

By way of exploratory data analysis, we conducted some variable importance (VIP) analysis using standard machine learning techniques. In particular, we build models that predict the outcome variables based on the random forest learner which uses the treatment and control variables as predictors. In general, VIP analysis consists of computing the increase in mean squared errors (MSE) of predictions as a result of excluding each predictor variable in turn. The larger the percentage increase in MSE associated with excluding a variable, the greater is the model's predictive accuracy and the more important the variable is deemed to be. Thus, by

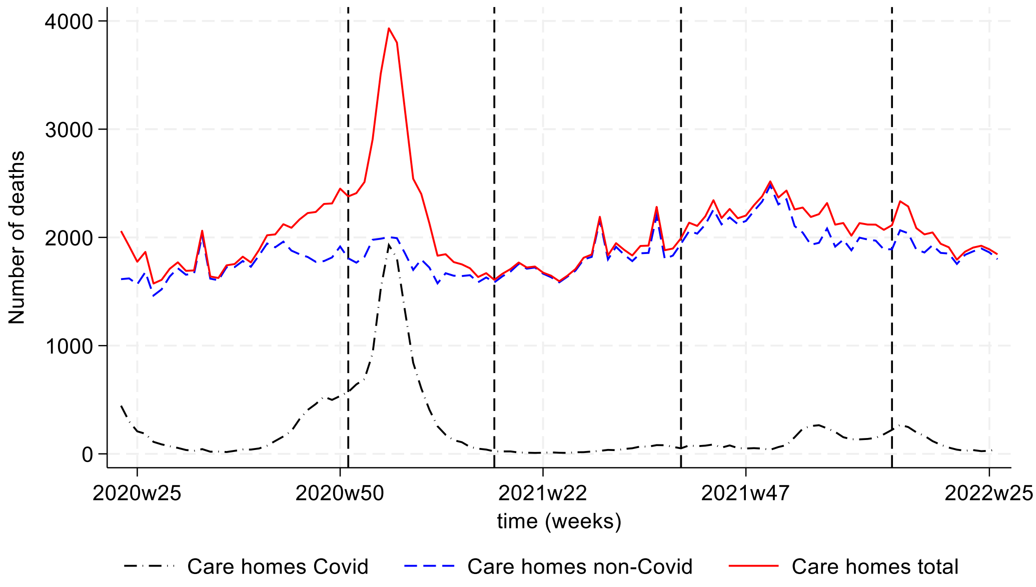

Fig. 3a. Care home deaths in England: Covid, non-Covid and total.
Notes: (i) Sample period is from 2020 week 23 until 2022 week 26.
(ii) Data are from the ONS, link provided in the Appendix Table A1.

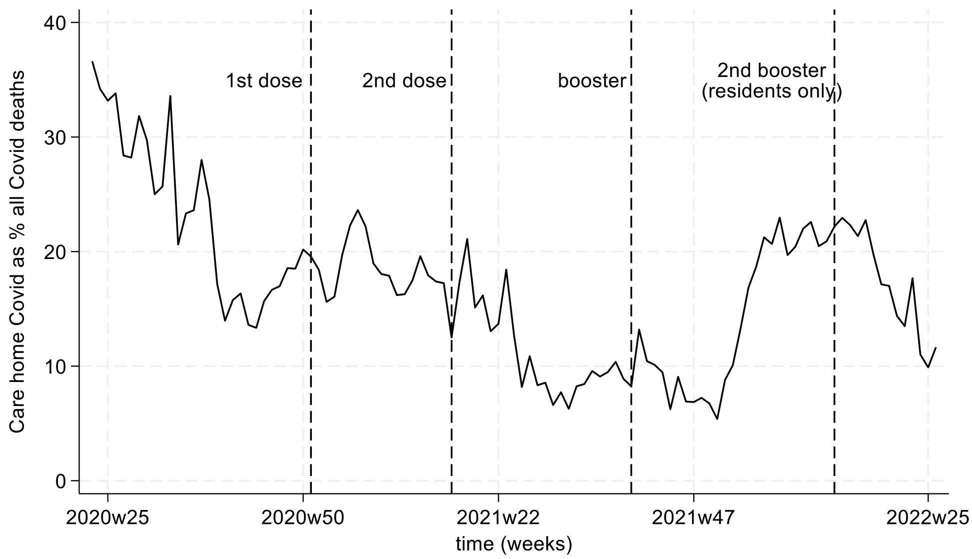

Fig. 3b. Care home Covid deaths in England as % of all Covid deaths.
Notes: (i) Sample period is from 2020 week 23 until 2022 week 26.
(ii) Data are from the ONS, link provided in the Appendix Table A1.

ranking variables' importance according to the value of percentage increase, one can get a sense of the order of importance of the predictors.

We report results from the VIP analysis in Table 3. For most outcome variables, the community infection rate and demographic variables emerge as important predictors. The various lags of resident and staff vaccination stocks generally rank low in the list. Although helpful in indicating the relative importance statistically of variables, the VIP analysis does not directly answer our main research question regarding the direction and magnitude of the causal relationships (if any) between vaccination rates and our outcome variables.

4.2.2. Fixed effects and double debiased machine learning (DDML) results

We start by reporting, in Table 4, standard fixed effects panel data estimates. As noted above, we would not expect to observe any

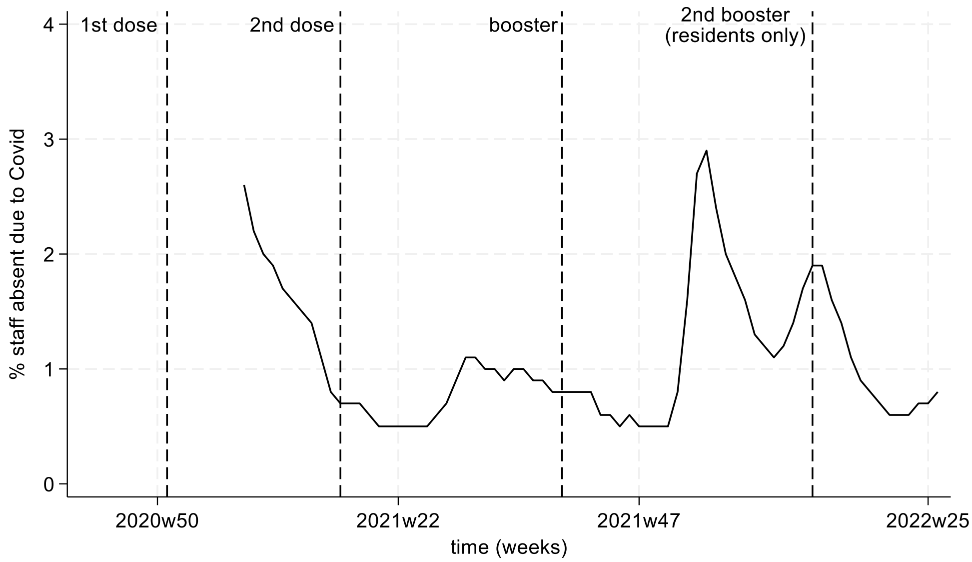

Fig. 4. Elderly care home staff absences for Covid in England.
Notes: (i) Sample period is from 2021 week 6 until 2022 week 26.
(ii) Data are from the Department for Health and Social Care, link provided in the Appendix Table A1.

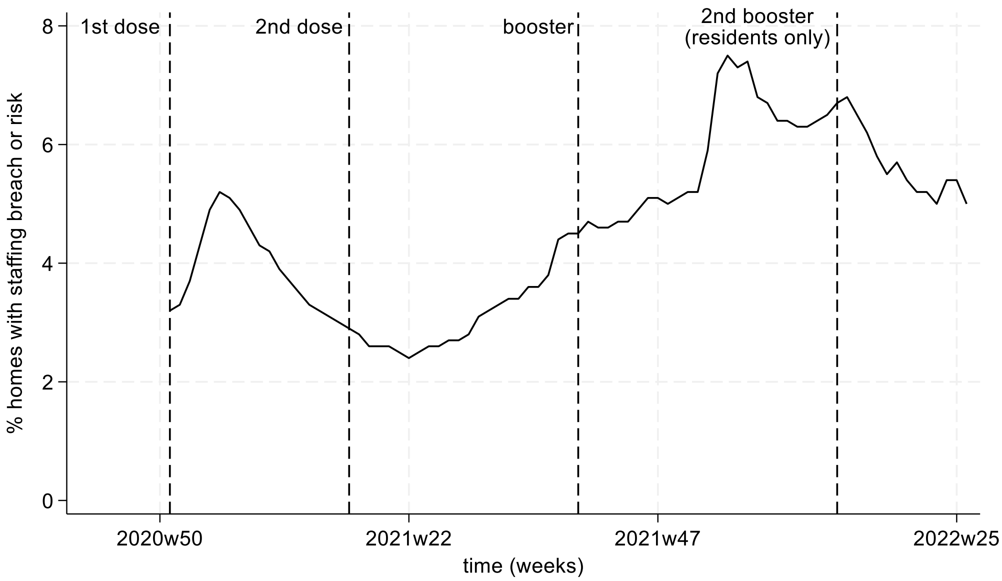

Fig. 5. Elderly care homes at risk or in breach of staffing standards in England.
Notes: (i) Sample period is from 2020 week 50 until 2022 week 26.
(ii) Data are from the Department for Health and Social Care, link provided in the Appendix Table A1.

causal impact of vaccination in the first three weeks after vaccination. We would also expect for any impact on mortality to be greater for resident than staff vaccination. For resident vaccination, there is little evidence of any significant (at the 5 % level) effects for any outcome. The exception is a significantly negative impact on the percentage of Covid deaths occurring in care homes when using the current vaccination stock value. This effect is unlikely to be causal given the time for vaccination to take effect on observed mortality,

Table 2
Partial correlation of each of outcome variables with treatment variables (one dose only).

	Outcome variable				
	Covid death rate in care homes	Total death rate in care homes	% of all Covid deaths occurring in care homes	% staff absence from Covid	% care homes in staffing breach/risk
Resident vaccine stock					
Current	0.048***	0.084***	−0.132*	−0.014***	−0.052***
Lagged 1 week	0.001	−0.021	0.001	−0.001	0.036**
Lagged 2 weeks	−0.019*	−0.052**	−0.041	−0.001	−0.007
Lagged 3 weeks	−0.004	−0.027	0.183	0.004	0.034**
Lagged 4 weeks	−0.030***	0.009	−0.188	0.010*	−0.013
Lagged 5 weeks	−0.025**	−0.034	0.003	0.022***	−0.009
Lagged 6 weeks	−0.010	−0.017	0.129	−0.018***	−0.009
Staff vaccine stock					
Current	0.048***	0.024	0.317**	−0.027***	0.043**
Lagged 1 week	−0.025	−0.003	−0.016	0.024**	−0.004
Lagged 2 weeks	−0.011	0.008	0.029	0.008	0.017
Lagged 3 weeks	−0.021	0.057	−0.285	−0.004	−0.040*
Lagged 4 weeks	0.012	−0.095**	0.226	−0.013	−0.006
Lagged 5 weeks	−0.083***	−0.070*	−0.107	−0.032***	−0.021
Lagged 6 weeks	0.133***	0.165***	−0.066	0.059***	0.043***
N	15,984	15,984	10,956	9916	10,952

Notes:.
(ii) Significance level: * 0.10, ** 0.05, *** 0.01.
(ii) Sample period for mortality is 2020 week 23 until 2022 week 26. For absences the period starts in 2021 week 6 and for staffing breach in 2020 week 51.
(iii) Data sources as described in Appendix Table 1.

though it is possible that it reflects patients who are both unable to be vaccinated due to frailty or illness but whose condition also makes them particularly vulnerable to death following infection.[19] It is worth noting that the point estimates for most of the mortality measures using the later lags (when a vaccination effect is more likely to show) are generally positive.

For staff vaccination, we observe some negative estimates for mortality measure (iii) but statistical significance is not robust. The estimates indicate vaccination is positively associated with staff absence for some lags but negatively associated with staffing breach risks.

In our supplementary analysis (Tables A2a and A2b), we split the sample into two time periods, the first from the start of our sample until the end of the primary course roll out (2020 week 23 until 2021 week 39) and the second for the rollout of the booster doses (2021 week 39 until 2022 week 26). In neither period do we find evidence that vaccination take-up significantly reduced mortality. There is some evidence that staff vaccination is associated with fewer care homes in breach of staffing levels for early, but not the later period.

Turning to the doubly debiased machine learning (DDML) approach, we report our baseline estimates in Table 5. Then in Tables 6a and 6b we present separate estimates for each of the primary and booster rollouts.

Looking first at the estimated causal impact on mortality, in contrast to the standard panel fixed effects regression estimates, the DDML results indicate that resident vaccination led to a reduction in resident Covid and total mortality rates. No such reduction is found for our other mortality measure (percentage of total Covid deaths that occurred in care homes) or for staff vaccination on any mortality outcome. This is unsurprising given the evidence noted above that vaccination has only limited impact against infection and, hence, transmission. Indeed, we find significantly positive impacts on total mortality from staff vaccination. However, the fact these are also observed for current and very early lags (a period when it is implausible that staff vaccination could impact resident mortality) means we are cautious about inferring causality here, even when using DDML.

When we split the sample into the two time periods, the estimates suggest any reduction in mortality from resident vaccination was seen only in the roll out of the primary course, i.e. until September 2021. After that point, if anything resident vaccination is associated with higher Covid-related mortality. For our third mortality measure, although the point estimates for the later lags are negative, we find no significant impacts even in the early period.

Staff vaccination is associated with higher mortality in the early period, though again the fact that we observe that effect even when using early lags (i.e. before a vaccination effect could be observed), suggests this may not be a true causal effect.

In order to appreciate the clinical significance of the effect of resident vaccination, recall that the treatments are measured in percentages (times 100). The average point estimate for the 4th to 6th lag for the impact of resident vaccination on the total death rate

[19] More general concerns about potential reverse due to unobserved heterogeneity are mitigated to some extent by the inclusion of LA-specific fixed effects and trends in the conditioning. In step 2 of the DDML algorithm, we use an ML learner on the treatment vector (d) to obtain the orthogonalized regressor, purging the treatment variables of potential endogeneity. Furthermore, most of our treatment variables are lagged, making it less likely that future infection rates could influence past vaccine stock levels.

Table 3
VIP analysis, % Inc MSE increases.

	Covid death rate in care homes	Total death rate in care homes	% of all Covid deaths occurring in care homes	% staff absence from Covid	% care homes in staffing breach/risk
Community infection rate	35.08	33.30	41.01	48.53	43.48
% of population aged 90 or over	31.04	27.22	30.82	46.68	33.65
Population density	30.08	28.77	32.37	42.14	30.20
% of population Black	23.22	24.71	28.22	32.57	29.94
Cumulative care home Covid death rate	23.03	16.18	44.25	43.88	41.62
Unemployment rate	20.43	15.75	33.24	36.01	32.46
% of population Asian	20.19	21.24	28.09	34.27	28.75
Staff vaccine stock current	18.57	10.30	22.13	14.37	21.42
Time trend	18.00	16.03	26.06	72.94	24.86
Staff vaccine stock lag 1	16.90	8.59	21.39	12.97	18.03
% of population White	16.41	18.60	28.79	36.71	26.11
Earnings	16.00	10.00	23.10	44.62	38.20
% of population White non-British	14.03	14.21	24.54	51.05	30.28
Public health expenditure per capita	13.25	13.19	27.11	39.42	32.59
Resident vaccine stock current	11.40	14.27	25.88	33.16	30.00
Staff vaccine stock lag 6	11.32	8.33	19.22	21.14	17.99
Staff vaccine stock lag 2	10.44	11.24	24.97	17.48	13.22
Care home residents aged 85 or over	9.77	9.10	29.15	45.09	34.23
Resident vaccine stock lag 5	9.51	9.55	21.13	25.56	31.15
Staff vaccine stock lag 3	9.28	7.65	25.42	19.63	14.43
Life expectancy	9.21	9.31	26.39	37.37	33.41
Resident vaccine stock lag 2	8.76	14.56	25.26	18.36	30.60
Resident vaccine stock lag 1	8.13	13.62	28.45	24.64	29.91
Care home residents aged 65 or over	7.34	8.76	30.93	39.01	35.05
Resident vaccine stock lag 3	7.12	7.77	24.82	16.38	22.87
Staff vaccine stock lag 4	6.79	7.40	24.53	22.77	13.70
Resident vaccine stock lag 4	6.76	7.50	24.32	18.49	29.93
Resident vaccine stock lag 5	5.88	10.08	23.16	22.44	30.86
Staff vaccine stock lag 5	4.80	7.20	18.37	21.39	18.39

Notes:.
(i) Sample period for mortality is 2020 week 23 until 2022 week 26. For absences the period starts in 2021 week 6 and for staffing breach in 2020 week 51.
(ii) Data sources as described in Appendix Table 1.
(iii) Figures are the percentage increase in the mean square error (MS) associated with excluding a variable. Higher values indicate greater importance of the variable as a predictor.

during the primary dose period is 0.022. That implies a 10 % increase in the resident vaccination stock from the mean value for that period (38.3 %) is associated with a reduction in the mortality rate per 1000 residents of 0.084 per week. With a total population of about 260 thousand residents, that is equivalent to about 22 deaths per week nationwide, close to 1 % of total care home deaths.

The DDML estimates of the impact of vaccination on our two staffing measures are inconsistent. Some results suggest staff vaccination levels were associated with more, rather than fewer, staff absences. For staffing breaches, very few of the estimates are statistically significant at the 5 % level.[20]

In our supplementary analysis we explore how robust these results are to various different specifications. First (see Table A3) we find estimates using the elastic net learner approach. These are qualitative similar to the baseline estimates.

Next (Table A4), we report estimates using the actual percentage of residents and staff who had received at least one vaccination. The main difference with our baseline results is that we no longer find a significant reduction in mortality rates from resident vaccination.

We then repeat (Table A5) our baseline estimates but starting the sample only from 2021 week 10. As noted above, the rationale for this experiment is that response rates from care homes regarding vaccination levels were much higher meaning the vaccination stock measures are more precisely measured from that point on. These results are qualitatively similar to our baseline specification, but the point estimates for the impact of resident vaccination on mortality rates are somewhat lower.

In Table A6, we report estimates with our outcome variables measured in levels rather than relative to the values six weeks

[20] Note that staff absences due to vaccination side effects might be observed even from current staff (but not current resident) vaccination.

Table 4
Panel data estimates of vaccination on outcomes in elderly care homes.

	Covid death rate in care homes	Total death rate in care homes	% of all Covid deaths occurring in care homes	% staff absence from Covid	% care homes in staffing breach/risk
Residents					
Current	−0.0137	0.0130	−0.1525**	−0.0016	−0.0103
	(0.013)	(0.019)	(0.072)	(0.003)	(0.016)
One week	0.0102	0.0056	−0.0062	0.0004	0.0397**
	(0.015)	(0.027)	(0.074)	(0.003)	(0.017)
Two weeks	−0.0143	−0.0496*	−0.0560	−0.0010	−0.0243*
	(0.013)	(0.025)	(0.102)	(0.003)	(0.014)
Three weeks	0.0134	−0.0102	0.2112**	−0.0004	0.0277**
	(0.015)	(0.021)	(0.090)	(0.003)	(0.013)
Four weeks	−0.0140	0.0248	−0.2039*	−0.0015	−0.0208**
	(0.018)	(0.023)	(0.106)	(0.003)	(0.010)
Five weeks	0.0225	0.0223	0.0530	0.0040	−0.0060
	(0.014)	(0.025)	(0.084)	(0.003)	(0.010)
Six weeks	0.0007	0.0090	0.0380	−0.0019	0.0012
	(0.015)	(0.021)	(0.068)	(0.004)	(0.009)
Staff					
Current	−0.0274	−0.0674**	0.0166	−0.0032	0.0048
	(0.025)	(0.031)	(0.111)	(0.006)	(0.019)
One week	0.0033	0.0349	0.0754	0.0055	−0.0107
	(0.015)	(0.038)	(0.148)	(0.005)	(0.020)
Two weeks	0.0129	0.0556	0.0274	−0.0031	0.0109
	(0.014)	(0.037)	(0.168)	(0.005)	(0.018)
Three weeks	−0.0031	0.0812**	−0.2459	−0.0002	−0.0355*
	(0.027)	(0.040)	(0.178)	(0.005)	(0.019)
Four weeks	0.0424	−0.0684	0.2480	0.0013	0.0004
	(0.039)	(0.042)	(0.178)	(0.005)	(0.015)
Five weeks	−0.0362	−0.0244	−0.1886	−0.0064	−0.0029
	(0.034)	(0.032)	(0.148)	(0.006)	(0.015)
Six weeks	0.0182	0.0074	0.0291	0.0123**	0.0072
	(0.035)	(0.030)	(0.115)	(0.006)	(0.017)

Notes:.
(i) * $p < 0.1$; ** $p < 0.05$, *** $p < 0.01$; standard errors in brackets.
(ii) Sample period for mortality is 2020 week 23 until 2022 week 26. For absences the period starts in 2021 week 6 and for staffing breach in 2020 week 51.
(iii) Data sources as described in the data appendix.
(iv) Treatment variable is "vaccination stock" as described in the text.
(v) Models are estimated separately for each lag.
(vi) All models include the following co-variates: local authority level infection rates amongst the general population lagged by three weeks, cumulative Covid death rates in care homes lagged by three weeks, local authority level fixed effects and time (weekly) fixed effects.

previously. The main difference with these results are the significantly negative estimates found for staff vaccination. However, the significant estimates for the early lags (i.e. before we should be likely to observe a true effect) suggest that at least some of those estimates are unlikely to be causal.

Finally, we report estimates (in Tables A7a and A7b) using slower (2 %) and faster (4 %) weekly decay rates to estimate our vaccination stock variables. For our mortality measures at least, these results are substantively similar to those using our baseline specification.

Taken together, our DDML estimates reveal some evidence that higher resident vaccination rates may have reduced Covid mortality in elderly care homes at least in the period of the initial rollout (i.e. up to about September 2021). There is little evidence of any reduction in mortality in the booster period or from staff vaccination. Further, we do not find evidence of a reduction in Covid deaths in elderly care homes relative to deaths in the broader community, despite the much higher vaccination rates in care homes.

5. Discussion

Some commentary about the pandemic takes as read that high rates of vaccination played a significant role in reducing Covid-associated mortality. Although there is a negative temporal association between the vaccination roll out and mortality rates, it is difficult to identify a causal impact of vaccination separately from other explanations for lower mortality rates such as immunity from previous infection and/or later variants of the virus being less fatal.

We used a doubly debiased machine learning (DDML) approach to identify causal relationships between vaccination rates and elderly care home mortality and other outcomes. The elderly care sector contributed to a large proportion of Covid deaths in many countries. We had access to detailed data on the evolution of vaccination coverage amongst staff and residents in care homes in

Table 5
DDLM estimates of vaccination on outcomes in elderly care homes.

	Covid death rate in care homes	Total death rate in care homes	% of all Covid deaths occurring in care homes	% staff absence from Covid	% care homes in staffing breach/risk
Residents					
Current	0.010**	0.013***	−0.010	−0.016***	−0.009
	(0.004)	(0.005)	(0.026)	(0.002)	(0.006)
One week	−0.003	−0.006	−0.007	−0.009***	−0.002
	(0.004)	(0.006)	(0.024)	(0.002)	(0.007)
Two weeks	−0.019***	−0.026***	−0.009	−0.002	−0.002
	(0.004)	(0.006)	(0.025)	(0.002)	(0.007)
Three weeks	−0.028***	−0.035***	−0.002	0.004**	−0.004
	(0.004)	(0.006)	(0.023)	(0.002)	(0.006)
Four weeks	−0.030***	−0.033***	−0.009	0.009***	−0.011*
	(0.004)	(0.005)	(0.023)	(0.002)	(0.006)
Five weeks	−0.024***	−0.023***	0.013	0.012***	−0.009
	(0.004)	(0.005)	(0.021)	(0.002)	(0.005)
Six weeks	−0.010***	−0.006	0.031	0.012***	−0.003
	(0.004)	(0.005)	(0.020)	(0.002)	(0.005)
Staff					
Current	0.011***	0.024***	0.068**	−0.005***	0.002
	(0.003)	(0.004)	(0.029)	(0.002)	(0.006)
One week	0.007**	0.021***	0.058*	0.000	0.004
	(0.003)	(0.004)	(0.030)	(0.002)	(0.006)
Two weeks	0.002	0.015***	0.044	0.004**	0.002
	(0.003)	(0.005)	(0.031)	(0.002)	(0.005)
Three weeks	−0.001	0.013***	0.026	0.007***	−0.003
	(0.003)	(0.005)	(0.030)	(0.002)	(0.005)
Four weeks	0.001	0.013***	0.025	0.010***	−0.003
	(0.003)	(0.004)	(0.029)	(0.002)	(0.006)
Five weeks	0.005	0.019***	0.018	0.013***	−0.000
	(0.003)	(0.005)	(0.026)	(0.002)	(0.006)
Six weeks	0.014***	0.030***	0.023	0.019***	0.005
	(0.003)	(0.004)	(0.024)	(0.002)	(0.005)

Notes:.
(i) * $p < 0.1$; ** $p < 0.05$, *** $p < 0.01$; standard errors in brackets.
(ii) Observation period for mortality is 2020 week 23 until 2022 week 26. For absences, period starts in 2021 week 6 and for staffing breach in 2020 week 50.
(iii) Data sources as described in the data appendix.
(iv) Treatment variable is "vaccination stock" as described in the text.
(v) The *potential* covariates are as described in the text.
(vi) Estimates use the random forest algorithm.

England. The advantage of the DDML approach is that it can leverage the predictive power of machine learning methods while exploiting traditional econometric approach. The resulting estimates should be more robust than traditional approaches to functional form misspecification and overfitting.

Standard panel data regression estimates do not indicate that higher vaccination take-up reduced mortality in elderly care home. In contrast, using the DDML approach, we are able to identify some evidence that vaccination may have reduced Covid-related mortality to some extent. This finding is, however, somewhat equivocal: it applies only to two of our three mortality measures and even for those two measures, the effect is only found for the period of the first course of Covid-19 vaccination (i.e. up to September 2021). Even using DDML, we are unable to identify strong evidence that vaccination rates amongst care home staff reduced mortality or that resident vaccination reduced mortality during booster roll out period (from September 2021). Indeed, in the later period, we find some evidence that higher vaccination rates are associated with higher Covid mortality.

The lack of evidence that vaccination of care home staff has any causal effect in reducing mortality amongst residents is unsurprising in the light of research noted above suggesting vaccination has modest impacts on contagiousness and, hence transmission. The limited effect of resident vaccination on mortality requires more explanation given evidence from some randomised controlled trials (RCTs) that, other thing equal, Covid vaccination lowers the risk of serious illness and death. However, there are good reasons why such trial outcomes may not be replicated in real world data.

First, is the well-known difficult of controlling for healthy vaccinee effects even within RCTs and which may lead to effectiveness being overstated. Second, RCTs typically compare mortality rates amongst those vaccinated with those unvaccinated and who have not previously had Covid. Even at the start of the vaccine rollout in the UK, many residents and staff in care homes had previously recovered from a Covid infection. Previous infection is known to reduce very significantly the chance both of subsequent infection and of mortality given subsequent infection. As a result, the potential for any beneficial mortality impact of vaccination amongst this group will be correspondingly smaller. Further, if the previously infected group are less likely than others to be vaccinated, this would

Table 6a
DDLM estimates of vaccination on outcomes in elderly care homes: primary course.

	Covid death rate in care homes	Total death rate in care homes	% of all Covid deaths occurring in care homes	% staff absence from Covid	% care homes in staffing breach/risk
Residents					
Current	0.021***	0.019***	0.021	−0.026***	−0.024***
	(0.006)	(0.006)	(0.031)	(0.002)	(0.009)
One week	0.005	−0.005	0.016	−0.019***	−0.018**
	(0.005)	(0.007)	(0.031)	(0.002)	(0.009)
Two weeks	−0.017***	−0.029***	−0.006	−0.011***	−0.015*
	(0.005)	(0.007)	(0.030)	(0.003)	(0.009)
Three weeks	−0.029***	−0.038***	−0.023	−0.006*	−0.012
	(0.006)	(0.007)	(0.030)	(0.003)	(0.008)
Four weeks	−0.034***	−0.037***	−0.045	0.002	−0.017**
	(0.006)	(0.007)	(0.032)	(0.004)	(0.007)
Five weeks	−0.028***	−0.025***	−0.028	0.007**	−0.014**
	(0.006)	(0.006)	(0.029)	(0.003)	(0.007)
Six weeks	−0.011**	−0.004	−0.015	0.009***	−0.006
	(0.005)	(0.006)	(0.027)	(0.003)	(0.005)
Staff					
Current	0.031***	0.044***	0.027	−0.009**	−0.019**
	(0.005)	(0.007)	(0.043)	(0.004)	(0.009)
One week	0.024***	0.038***	−0.001	0.009**	−0.012
	(0.004)	(0.007)	(0.045)	(0.004)	(0.009)
Two weeks	0.019***	0.028***	−0.017	0.019***	−0.005
	(0.004)	(0.006)	(0.045)	(0.004)	(0.008)
Three weeks	0.014***	0.027***	−0.042	0.027***	−0.002
	(0.004)	(0.007)	(0.045)	(0.004)	(0.008)
Four weeks	0.013***	0.025***	−0.032	0.035***	−0.002
	(0.004)	(0.007)	(0.040)	(0.004)	(0.007)
Five weeks	0.012***	0.027***	−0.021	0.040***	−0.001
	(0.004)	(0.007)	(0.038)	(0.004)	(0.007)
Six weeks	0.019***	0.035***	−0.009	0.040***	0.005
	(0.004)	(0.007)	(0.034)	(0.004)	(0.006)

Notes:.
(i) * $p < 0.1$; ** $p < 0.05$, *** $p < 0.01$; standard errors in brackets.
(ii) Observation period for mortality is 2020 week 23 until 2021 week 39. For absences, period starts in 2021 week 6 and for staffing breach in 2020 week 51.
(iii) Data sources as described in the data appendix.
(iv) Treatment variable is "vaccination stock" as described in the text.
(v) The potential covariates are as described in the text.
(vi) Estimates use the random forest algorithm.

provide another mechanism that pushes up population-level mortality rates amongst the vaccinated relative to unvaccinated cohorts.

Apart from the novel application of DDML techniques, a key strength of our analysis is the construction of a measure of 'vaccination stock' that takes account both of waning vaccination effectiveness over time as well as the impact of second and booster vaccination doses.

The limited effectiveness of vaccination on Covid mortality is consistent with research (for example, Fabiani et al. (2020), Simonsen et al., 2007 and Verhees et al., 2019) on the impact of influenza vaccination and which has similarly found it difficult to establish a clear causal impact on mortality rates.

There are of course limitations and caveats to our analysis. Our statistical approach relies for identification on differences in vaccination rates across local authority areas. Although there were quite significant differences in levels and trends in take-up rates across England, nearly all local authorities saw a relatively high level of take-up, particularly amongst residents. That means our results can be interpreted as suggesting a marginal increase in take-up has at best limited impact on mortality rates. We do not directly estimate the effect on mortality of average vaccination rates relative to no vaccination at all.

That said, our DDML results are consistent with descriptive trends. Had vaccination had a very large impact on mortality rates (as suggested by some of the modelling research), we would expect to observe at least some signal in the trend data. The difficulty in identifying any consistent signal during the various rollout periods is indicative at least that any mortality effect is modest.

6. Conclusion

A good deal of attention has recently been directed to the potential of using machine learning in quantitative research. In this paper, we explore this potential in trying to identify a causal impact of Covid-19 vaccination in reducing mortality in elderly care homes.

Using standard panel data regression analysis and doubly debiased machine learning (DDML) techniques suggests both that

Table 6b
DDLM estimates of vaccination on outcomes in elderly care homes: booster period.

	Covid death rate in care homes	Total death rate in care homes	% of all Covid deaths occurring in care homes	% staff absence from Covid	% care homes in staffing breach/risk
Residents					
Current	−0.008**	0.022**	−0.244***	−0.007***	0.001
	(0.004)	(0.010)	(0.073)	(0.003)	(0.010)
One week	−0.009**	0.013	−0.207***	−0.003	0.008
	(0.004)	(0.012)	(0.077)	(0.003)	(0.012)
Two weeks	−0.005	−0.012	−0.119	0.006*	0.010
	(0.004)	(0.015)	(0.079)	(0.003)	(0.013)
Three weeks	0.001	−0.018	0.023	0.015***	0.019
	(0.002)	(0.016)	(0.075)	(0.003)	(0.014)
Four weeks	0.005**	−0.014	0.075	0.024***	0.022
	(0.002)	(0.015)	(0.070)	(0.002)	(0.013)
Five weeks	0.009***	−0.010	0.177***	0.031***	0.025*
	(0.003)	(0.014)	(0.068)	(0.002)	(0.014)
Six weeks	0.013***	−0.008	0.271***	0.034***	0.024*
	(0.002)	(0.012)	(0.064)	(0.002)	(0.013)
Staff					
Current	0.013***	0.055***	0.076	0.013***	0.032***
	(0.004)	(0.012)	(0.072)	(0.003)	(0.012)
One week	0.015***	0.054***	0.118	0.013***	0.032***
	(0.005)	(0.012)	(0.083)	(0.003)	(0.012)
Two weeks	0.012***	0.052***	0.122	0.010***	0.031**
	(0.003)	(0.012)	(0.083)	(0.003)	(0.012)
Three weeks	0.009***	0.038**	0.140	0.006*	0.022*
	(0.003)	(0.015)	(0.090)	(0.003)	(0.012)
Four weeks	0.006*	0.013	0.155	0.002	0.018
	(0.003)	(0.029)	(0.095)	(0.003)	(0.013)
Five weeks	0.001	0.013	0.125	0.000	0.017
	(0.006)	(0.025)	(0.095)	(0.003)	(0.014)
Six weeks	0.004	0.001	0.185*	0.003	0.020
	(0.003)	(0.032)	(0.104)	(0.004)	(0.015)

Notes:.
(i) * $p < 0.1$; ** $p < 0.05$, *** $p < 0.01$; standard errors in brackets.
(ii) Observation period is 2021 week 39 until 2022 week 26.
(iii) Data sources as described in the data appendix.
(iv) Treatment variable is "vaccination stock" as described in the text.
(v) The potential covariates are as described in the text.
(vi) Estimates use the random forest algorithm.

vaccination had only a limited impact on care home mortality and that any impact was restricted to the initial rollout period of the vaccine.

Our analysis casts doubt on the hypothesis that high rates of vaccination were a particularly important factor in reducing Covid mortality after the initial waves. In turn, this has implications for public policy relating to Covid vaccination. In particular, it may be appropriate to look again at the case for continuing to expend resources on offering regular booster vaccination doses to vulnerable populations such as care home residents.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank two anonymous referees, the Editor and participants at a Nottingham University Business School staff seminar for a number of useful comments and suggestions.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.euroecorev.2024.104882.

References

Abouk, R., MacLean, J.C., Earle, J.S., Park, S., 2024. Promoting public health with blunt instruments: evidence from vaccine mandates. NBER Work. Pap. (March), 32286. www.nber.org/papers/w32286.

Acharya, A., Wolfson, C., Matta, S., Cardona, C., Lamba, S., Bishai, D., 2021. The role of public health expenditures in COVID-19 control: evidence from local governments in England. SSM – Public Health 15, 100861.

Agrawal, V., Sood, N., Whaley, C.M., 2023. The impact of the global Covid-19 vaccination campaign on all-cause mortality. NBER Work. Pap. (October), 31812. www.nber.org/papers/w31812.

Ahrens, A., Hansen, C., Schaffer, M., Wiemann, T., 2024. ddml: double/Debiased Machine Learning. R Package Version 0.2.2.9000. https://thomaswiemann.com/ddml/.

Andrews, N., Stowe, J., Kirsebom, F., et al., 2022. Covid-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. New Engl. J. Med. 386 (April), 1532–1546.

Antonelli, M., Penfold, R.S., Merino, J., et al., 2022. Risk factors and disease profile of post-vaccination SARS-CoV-2 infection in UK users of the Covid Symptom Study app: a prospective, community-based, nested, case-control study. The Lancet 22 (1, Jan), 43–55.

Bjoerkheim, M.B., Tabarrok, A., 2022. Covid in the nursing homes: the US experience. Oxford Rev. Econ. Policy 38 (4), 887–911.

Butler, D., Butler, R., Farnell, A., Simmons, R., 2024. COVID-19 infections and short-run worker performance: evidence from European football. Eur J Oper Res 315 (2, June), 750–763.

Chen, M.K., Chevalier, J.A., Long, E.F., 2021. Nursing home staff networks and COVID-19. PNAS 118 (1), e2015455118.

Chen, Y.T., 2023. Effect of vaccination patterns and vaccination rates on the spread and mortality of the COVID-19 pandemic. Health Policy Technol. 12 (1), 100699.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J., 2018. Double/debiased machine learning for treatment and structural parameters. Econom J 21 (1, Feb), C1–C68.

Chernozhukov, V., Demirer, M., Duflo, E., Fernández-Va, I., 2023. Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. NBER Work. Pap. 24678. https://www.nber.org/papers/w24678.

Chrissian, A.A., Oyoyo, U.E., Patel, P., et al., 2022. Impact of COVID-19 vaccine-associated side effects on health care worker absenteeism and future booster vaccination. Vaccine 40 [online].

Covid-19 Forecasting Team, 2023. 'Past SARS-CoV-2 infection protection against re-infection: a systematic review and meta-analysis. The Lancet 401, 833–842. www.thelancet.com/journals/lancet/article/PIIS0140-6736(22)02465-5/fulltext.

Cuadros, D.F., Gutierrez, J.D., Moreno, C.M., et al., 2023. Impact of healthcare capacity disparities on the COVID-19 vaccination coverage in the United States: a cross-sectional study. The Lancet Region. Health 18, 100409.

Fabiani, M., Volpe, E., Faraone, M., Bella, A., Pezzotti, P., Chini, F., 2020. Effectiveness of influenza vaccine in reducing influenza-associated hospitalizations and deaths among the elderly population; Lazio region, Italy, season 2016-2017. Expert Rev. Vacc. 19 (5), 479–489.

Gibson, J., 2022. The rollout of COVID-19 booster vaccines is associated with rising excess mortality in New Zealand. Univ. Waikato Work. Pap. Econ. 11/22.

Gibson, J., 2023. Jabbing the economy back to life. Appl Econ Lett 30 (21), 2999–3005.

Girma, S., Paton, D., 2023. Covid-19 vaccines as a condition of employment: impact on uptake, staffing and mortality in elderly care homes. Manag. Sci forthcoming.

Goldberg, Y., Mandel, M., Bar-On, Y.M., Bodenheimer, O., Freedman, L., Haas, E., Milo, R., Alroy-Preis, S., Ash, N., Huppert, A., 2021. Waning immunity after the BNT162b2 vaccine in Israel. New Engl. J. Med. 385 (e.85, Dec).

Hoxha, I., Agahi, R., Bimbashi, A., Aliu, M., Raka, L., et al., 2023. Higher COVID-19 vaccination rates are associated with lower COVID-19 mortality: a global analysis. Vacc. (Basel) 11 (1), 74.

Huang, Y.Z., Kuan, C.C., 2022. Vaccination to reduce severe COVID-19 and mortality in COVID-19 patients: a systematic review and meta-analysis. Eur Rev Med Pharmacol. Sci. 26 (5), 1770–1776.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An Introduction to Statistical learning: With Applications in R, 2nd Edition. Springer.

Lv, G., Yuan, J., Xiong, X., Li, M., 2021. Mortality rate and characteristics of deaths following COVID-19 vaccination. Front. Med (Lausanne) 8, 670370.

Mackey, K., Ayers, C.K., Kondo, K.K., Saha, S., Advani, S.M., Young, S., Spencer, H., Rusek, M., Anderson, J., Veazie, S., Smith, M., Kansagara, D., 2020. Racial and ethnic disparities in COVID-19–related infections, hospitalizations, and deaths: a systematic review. Ann. Intern. Med. 174 (3), 362–373.

Magazzino, C., Mele, M., Coccia, M., 2022. A machine learning algorithm to analyse the effects of vaccination on COVID-19 mortality. Epidemiol. Infect. 150 (e168), 1–12.

Maltezou, H., Gamaletsou, M.N., Koukou, D-M, et al., 2022. Association between Covid-19 vaccination status, time elapsed since the last vaccine dose, morbidity, and absenteeism among healthcare personnel: a prospective, multicenter study. Vaccine 40 (52), 7660–7666. https://doi.org/10.1016/j.vaccine.2022.10.049.

Maltezou, H.C., Panagopoulos, P., Sourri, F., et al., 2021. COVID-19 vaccination significantly reduces morbidity and absenteeism among healthcare personnel: a prospective multicenter study. Vaccine 39 (48), 7021–7027. https://doi.org/10.1016/j.vaccine.2021.10.054.

Mounier-Jack, S., Paterson, P., Bell, S., et al., 2023. Covid-19 vaccine roll-out in England: a qualitative evaluation. PLoS One 18 (6), e0286529.

Olenski, A., Sacher, S., 2024. Estimating nursing home quality with selection. Rev. Econ. Stat. (Feb). https://doi.org/10.1162/rest_a_01449.

Pathak, E.B., Menard, J.M., Garcia, R.B., Salemi, J.L., 2022. Joint effects of socioeconomic position, race/ethnicity, and gender on COVID-19 mortality among working-age adults in the United States. Int. J. Environ. Res. Public Health 19 (9), 5479.

Rahmani, K., Shavaleh, R., Forouhi, M., Disfani, H.F., Kamandi, M., et al., 2022. The effectiveness of COVID-19 vaccines in reducing the incidence, hospitalization, and mortality from COVID-19: a systematic review and meta-analysis. Front. Public Health 10, 873596.

Simonsen, L., Taylor, R.J., Viboud, C., Miller, M.A., Jackson, L.A., 2007. Mortality benefits of influenza vaccination in elderly people: an ongoing controversy. Lancet Infect. Dis. 7, 658–666.

Subramanian, S.V., Kumar, A., 2021. Increases in COVID-19 are unrelated to levels of vaccination across 68 countries and 2947 counties in the United States. Eur. J. Epidemiol. 36, 1237–1240.

ten Doesschate, T., van der Vaart, T.W., Debisarun, P.A., et al., 2022. BCG vaccine to reduce healthcare worker absenteeism in COVID-19 pandemic, a randomized controlled trial. Clin. Microbiol. Infect. 28 (9).

UKHSA (2022), *Vaccine Surveillance reports* (various) (www.gov.uk/government/publications/covid-19-vaccine-weekly-surveillance-reports).

UKHSA (2023), *COVID-19: the Green Book*, chapter 14a, Sept https://www.gov.uk/government/publications/covid-19-the-green-book-chapter-14a.

Verelst, F., Beutels, P., Hens, N., Willem, L., 2021. Workplace influenza vaccination to reduce employee absenteeism: an economic analysis from the employers' perspective. Vaccine 39 (14).

Verhees, R.A.F., Thijs C, C., Ambergen, T., Dinant, G.J., Knottnerus, J.A., 2019. Influenza vaccination in the elderly: 25 years follow-up of a randomized controlled trial. No impact on long-term mortality. PLoS ONE 14 (5), e0216983.

Xiangxiang, J., Lv, G., Li, M., Lu, K., 2023. Causal effect of anti-dementia drugs on patient' economic burden using double/debiased machine learning approach. Innov. Aging 7 (S1), 1000.