METHOD ARTICLE

Looking in the medicine cabinet: methods for using real-world data to assess the impact of measles, mumps and rubella (MMR) and recombinant adjuvanted varicella-zoster vaccines on coronavirus disease 2019 (COVID-19) prevention and case fatality [version 1; peer review: 1 not approved]

Douglas McNair, Hao Hu, Casey Selwyn

Bill & Melinda Gates Foundation, Seattle, USA

v1 First published: 30 Jul 2021, 5:115
<https://doi.org/10.12688/gatesopenres.13329.1>
Latest published: 30 Jul 2021, 5:115
<https://doi.org/10.12688/gatesopenres.13329.1>

Abstract

Background: Analysis of real-world data can be used to identify promising leads and dead ends among products being repurposed for clinical practice for coronavirus disease 2019 (COVID-19). This paper uses real-world data from Cerner Labs collected from 90 source institutions in the United States to assess the potential impact of two viral vaccines on COVID-19 case fatality rates.

Methods: We identified 373,032 polymerase chase reaction (PCR)-positive COVID-19 cases in the Cerner Labs database between 01-MAR-2020 and 31-DEC-2020 and identified patients that had received measles, mumps and rubella (MMR) or a recombinant adjuvanted varicella-zoster vaccine within the previous 5 years. We calculated heterogeneity scores to support interpretation of results across institutions, and used stepwise forward variable selection to construct covariable-based propensity scores. These scores were used to match cases and control for biasing and confounding issues inherent in observational data.

Results: Neither the recombinant adjuvanted varicella-zoster vaccine nor MMR showed significant efficacy in prevention of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. We could not derive clinically significant results on the impact of MMR for case fatality rates due to persistently high rates of heterogeneity between institutions. However, we were able to achieve acceptable levels of heterogeneity for the analysis of the recombinant adjuvanted varicella-zoster vaccine, and found a clinically meaningful benefit of reduced case fatality rate, with an odds ratio of 0.43 (95% confidence interval [CI]: 0.38 – 0.48).

Conclusions: Using propensity score matching and heterogeneity

Open Peer Review

Approval Status

1

version 1

30 Jul 2021

[view](#)

1. Sara Khalid , University of Oxford, Oxford, UK

Any reports and responses or comments on the article can be found at the end of the article.

statistics can help guide our interpretation of real-world data, and rigorous statistical methods are needed to reduce bias or disparities in data interpretation. Applying these methods to the impact of viral vaccines on COVID-19 case fatalities yields actionable findings for further analysis.

Keywords

COVID-19 vaccines, repurposing, non-specific vaccine effects, real world data, real world evidence

This article is included in the [Coronavirus \(COVID-19\)](#) collection.

Corresponding author: Casey Selwyn (casey.selwyn@gatesfoundation.org)

Author roles: **McNair D:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Hu H:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Selwyn C:** Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: All authors are employees of the Bill and Melinda Gates Foundation.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2021 McNair D *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: McNair D, Hu H and Selwyn C. **Looking in the medicine cabinet: methods for using real-world data to assess the impact of measles, mumps and rubella (MMR) and recombinant adjuvanted varicella-zoster vaccines on coronavirus disease 2019 (COVID-19) prevention and case fatality [version 1; peer review: 1 not approved]** Gates Open Research 2021, 5:115 <https://doi.org/10.12688/gatesopenres.13329.1>

First published: 30 Jul 2021, 5:115 <https://doi.org/10.12688/gatesopenres.13329.1>

Introduction

Novel therapeutic and vaccine development for coronavirus disease 2019 (COVID-19) has advanced at extraordinary speed and scale. Against all historical precedent, multiple products have progressed from discovery through emergency regulatory approval in less than one year. In addition to these novel discoveries, scientific efforts have also focused on assessing the arsenal of existing products to determine whether off-label use of products could impact COVID-19 acquisition, disease progression and mortality. Within weeks of the identification of COVID-19, myriad claims were made about the potential efficacy of approved products, which consequently provoked supply shortages of multiple essential drugs and the initiation of many uncoordinated clinical trials.

Background on non-specific vaccine effects

As early as March 2020, observational studies and opinion pieces pointed to the potential impact of existing, broadly implemented viral vaccines on overall COVID-19 epidemiological patterns (Aaby *et al.*, 2020; Fidel & Noverr, 2020; Gold *et al.*, 2020), and multiple clinical trials were initiated across the globe to test these hypotheses¹. The claims built on existing but contested literature that has examined immunological theory, observational data and clinical trial evidence to determine whether certain vaccines have beneficial or deleterious effects on all-cause mortality. One side of the debate hypothesizes that viral vaccines such as Bacillus Calmette-Guerin vaccine (BCG), measles-containing vaccine (MCV), oral polio vaccine (OPV) and Rabies reduce mortality; that non-live vaccines such as inactivated polio vaccine (IPV) and Diphtheria, Tetanus, Pertussis (DTP) increase mortality; and that stronger findings in both directions are observed among females (Higgins *et al.*, 2016; Jensen *et al.*, 2016; Kandasamy *et al.*, 2016; Messina *et al.*, 2019; Pollard *et al.*, 2017). The logic is that certain vaccines might induce cross-protective ‘trained’ innate immune responses for some period after vaccine administration, while the adaptive system builds the more specific response to the target immunogen (Netea *et al.*, 2016). Heterologous protection against other viruses thus occurs via induction of interferons and activated natural killer cells. Vaccine components such as adjuvants are also hypothesized to induce other immunological effects among T cells, including potential epigenetic changes which impact cell responses to future pathogens (Goodridge *et al.*, 2016; Pollard *et al.*, 2017).

A 2014 SAGE review found that the evidence could not ‘exclude nor confirm the possibility of beneficial or deleterious non-specific immunological effects of the vaccines under study on all-cause mortality’, and that findings were not sufficient to warrant a policy change in recommendation (Higgins *et al.*, 2016; SAGE, 2014). A follow-on systematic review examined the relevant clinical trials, cohort and case-control studies on the mortality effects of BCG, DTP and MCV, and again suggested

that while there was observed evidence of impact on all-cause mortality, more study was needed to identify whether there was any real mechanistic link (Higgins *et al.*, 2016). In particular, reviews were concerned about the quality of the studies and potential for bias (Higgins *et al.*, 2016), and others have previously noted overall methodological issues in epidemiological studies of non-specific effects (Farrington *et al.*, 2009; Fine & Chen, 1992). Multiple studies also point to the multiple health-seeking behavioral attributes associated with vaccination that are likely to confound overall mortality outcomes (Fireman *et al.*, 2009; Pollard *et al.*, 2017).

Assessing existing vaccines for protection against COVID-19

Amidst this debate, the possibility of cross-protective innate immunity offered by other vaccines led us to undertake analyses of real-world data for signals of non-specific heterologous vaccine efficacy (VE) against COVID-19. If such VE were found and determined to be substantial, it would be relevant as an ancillary public health intervention for controlling the COVID-19 pandemic through vaccine repurposing and off-label human use in a short-time frame. In the present work, we directed attention to electronic health record-derived real-world data on MMR and recombinant adjuvanted varicella-zoster vaccines received in U.S. adult populations within 5 years prior to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreak in 2020. Additionally, as outlined below, we undertook this analysis to demonstrate a research methodology for using real world evidence to test hypotheses about the potential off-label activity of vaccines. Thus, the observations and conclusions we draw in this article are not intended to be used for clinical or other treatment purposes, and consequently the authors and their sponsoring organizations disclaim any and all liability for such uses.

Measles, Mumps & Rubella Vaccine (MMR). One hypothesized mechanism of heterologous protection involves overlapping of vaccine epitopes between the Spike (S) glycoprotein of the SARS-CoV-2 virus, the Fusion (F1) glycoprotein of measles virus and the envelope (E1) glycoprotein of the rubella virus (Sidiq *et al.*, 2020). MMR is a live attenuated vaccine indicated for prevention of measles, mumps, and rubella infection, marketed in the U.S. since 1971. MMR vaccination in adults is often required in connection with employment in hospitals, health services and clinics, schools, and other settings and roles having high contact with the public, especially children. In many instances, employers requiring MMR vaccination of employees is conditioned on test results showing low or undetectable titer of antibodies against measles, mumps, and rubella viruses.

Recombinant adjuvanted varicella-zoster vaccine. Another hypothesized protection mechanism involves induction of augmented adaptive immune responses through dendritic cell activation and innate immune cell recruitment with certain contemporary adjuvants such as Adjuvant System AS01, which is a liposome-based vaccine adjuvant containing 3-O-desacyl-4'-monophosphoryl lipid A (MPLA) and the saponin QS-21 (Didierlaurent *et al.*, 2014).

¹ Identified through press releases and a review of announced trials on ClinicalTrials.gov, via a search with keywords ‘Covid-19’ and ‘MMR’; ‘MR’; ‘OPV’; ‘zoster’; ‘BCG’

Bacille Calmette-Guérin (BCG). Some groups (see e.g. [Curtis et al., 2020](#); [Escobar et al., 2020](#); [Gupta, 2020](#); [O'Neill & Netea, 2020](#)) have hypothesized a COVID-protective effect for BCG vaccination. However, we were not able to assess the impact of BCG using our dataset, as BCG is not widely used in the U.S., and when indicated it is primarily administered as an adjuvant treatment in bladder cancer. We found that this usage is associated with a variety of confounding factors that limit the relevance of data from a cohort of bladder cancer patients receiving BCG treatment towards decision-making for a general population.

Using real word data

Randomized control trials rightfully remain the gold standard in determining safety and efficacy of products for licensure and widespread use, and as noted above, multiple scholars have emphasized the need for further studies assessing the non-specific effect of vaccines, on all-cause mortality and now against COVID-19. However, clinical development is lengthy and costly, and during a global pandemic, there are limitations in financial and human resources and infrastructure, and trials ultimately compete for patients needed to reach an answer on a product's value. In this context, real-world data can generate a set of evidence to help prioritize constrained clinical trial resources and test products with the highest chance of success, provided that adequate sample size and endpoint event counts are available and appropriate adjustments are undertaken to mitigate biases and confounding.

Real world data (RWD) has recently been advanced to improve the responsiveness of research and regulatory agencies to the needs of the public and policymakers. The [21st Century Cures Act \(2015\)](#) attempted to address the limitations of interpreting tightly controlled RCTs by incorporating real-world evidence (RWE) into the drug-labeling process. Similar programs have been initiated in various countries around the world, and in multi-national consortia, including the International Conference on Harmonization. When RWD is analyzed rigorously and used to generate package of RWE, this evidence can be used for hypothesis testing that separates the signal from the noise in claims of existing product effectiveness against COVID-19. This can, in turn, help to inform clinical trial targets and designs, and monitor real-world effectiveness of products as they are introduced.

In this context, analyzing real-world data can help optimize trial resources by assessing outcomes among large datasets and identifying leads and dead-ends among the array of on-formulary products. To demonstrate this, we applied a novel statistic method to a real word database of health records in the United States to assess the potential impact of twenty-five vaccines to see which, if any, merited detailed analysis. Based on this initial triage, we reviewed two viral vaccines (measles, mumps and rubella [MMR] and a recombinant adjuvanted varicella-zoster vaccine) against COVID-19 in the United States. We assessed these vaccines for potential impact on COVID-19 disease prevention and case fatality.

Methods

We utilized a commercially available electronic health record (EHR) derived de-identified privacy-protected HIPAA-compliant protected health information (PHI)-free secondary-use assented real-world data (RWD) repository (Cerner Healthe DataLabs RWD™, Cerner Corporation, Kansas City, MO, USA). All data was fully anonymized and there was no direct human subject research, and we thus did not seek approval from an Institutional Review Board. The repository comprises records from 2010 to present, with over 500 million care-related records and 160 million distinct patients. Material is encrypted end-to-end and pre-cleaned upon arrival into a data warehouse. Each data element is timestamped, and multiple episodes for a single person are longitudinally tracked with an encrypted key. The instance of the Cerner RWD used for the present study contains longitudinally-linked health records for 100% of patients in the care of 90 U.S. institutions and their affiliated ambulatory care clinics and physician offices, accessed via a private cloud environment with Python and Pandas software and ANSI-standard SQL. The data cleaning, transformation and further analyses were performed with R 4.0.4 (The R Foundation) and RStudio 1.4.1106 (RStudio, PBC).

At the time of our analysis, there were 373,032 PCR-proven, distinct COVID-19 patients in the Cerner Healthe DataLabs RWD™ repository between 01-MAR-2020 and 31-DEC-2020 in 90 tenants (extended data, Table S1; [McNair et al., 2021](#)). We retained 347,570 records in the 38 tenants whose contributed cases comprise more than 0.5% of the aggregate sample size (extended data, Figure S1; [McNair et al., 2021](#)). Demographic characteristics of all patients are available in Table S1 ([McNair et al., 2021](#)).

After an initial review of 25 vaccines in common use in the US, we narrowed the analysis to two vaccines with signals warranting further attention: an MMR vaccine, and a recombinant adjuvanted varicella-zoster vaccine. For MMR-related analyses, we confined the records to persons aged 25 to 64 years old inclusive at the time of their 01-MAR-2020 and 31-DEC-2020 index episode (Table 1). In general, such subjects received MMR in connection with their employment as health care workers, first responders, teachers, and other professions in which employers require periodic titer measurement and revaccination with MMR, and the vast majority of vaccinated MMR subjects are female. Not surprisingly, there are relatively few prevalent persons aged 65 and older at the time of their 01-MAR-2020 and 31-DEC-2020 index COVID-19 episodes who received MMR vaccination within the preceding 5 years. In similar fashion, we confined our analyses related to the recombinant adjuvanted varicella-zoster vaccine just to persons 50 years of age or older and excluded subjects receiving that same vaccination at younger ages (Table 1).

For each PCR-proven COVID-19 subject in the real-world data repository selected by SQL query inclusion-exclusion criteria, we utilized vaccination for influenza within the lookback time horizon as a proxy for vaccine non-hesitancy. This was

Table 1. Descriptive statistics for the COVID-19 PCR positive individuals used in MMR and recombinant adjuvanted varicella-zoster vaccine-related analyses in the 38 tenants whose contributed COVID-19 cases comprise more than 0.5% of the aggregate sample size. SARS-CoV-2=severe acute respiratory syndrome coronavirus 2; COVID-19=coronavirus disease 2019; PCR=polymerase chain reaction.

n (Percent)	SARS-CoV-2+, 25-64years (n=207,331)	SARS-CoV-2+, 50years+ (n=169,006)
Gender		
Female	113,399 (54.7)	87,428 (51.7)
Male	93,415 (45.1)	81,155 (48.0)
Deceased		
Yes	4,225 (2.04)	15,305 (9.1)
Vaccination	<i>MMR</i>	<i>Recombinant adjuvanted varicella-zoster vaccine</i>
Total	3,648 (1.8)	7,267 (4.3)
Female	3,441 (1.7)	4,247 (2.5)
Male	198 (0.1)	3,011 (1.8)
Seasonal influenza vx history		
Yes	39,530 (19.1)	45,636 (27.0)
Encounter: emergency/inpatient		
Yes	151,416 (73.0)	138,112 (81.7)

done to account for potential confounding due to health-seeking behavioral attributes that might be associated with vaccination (Fireman *et al.*, 2009; Pollard *et al.*, 2017). We required ICD-10-CM U07.1 code denoting COVID-19 disease and a positive PCR result for SARS-CoV-2. We recognize that PCR sensitivity is less than 100% and this composite selection criterion may therefore be overly conservative. We designed SQL to retrieve records from the real-world data repository with selection criteria to provide comparability of cases and controls in terms of (a) a diverse range of comorbid ambulatory care-sensitive conditions (ACSCs) and (b) time span of longitudinal follow-up subsequent to the index episode, as well as likelihood and timing of contact with the health system enabling endpoint ascertainment. The subjects included in our analysis are patients whose EHR records showed that they received at least one medication prescription in connection with their care episodes. We recognize that this selection criterion excludes some people without known ACSCs who presented to an ambulatory clinic/office or who may have been dismissed with no prescribed medications or only with recommended OTC medications, and no follow-up plan.

Persons who presented to the institutions' emergency departments or who were admitted to hospital as inpatients had, in general, clinical findings and physician orders indicative of more severe COVID-19 disease than persons whose care was

delivered in ambulatory settings. Emergency Department patients and inpatients tended to have multiple symptoms, larger numbers of physician orders, larger numbers of episodes of care preceding the index episode, larger numbers of comorbid conditions, and laboratory result values in the records of the index episode consistent with severity of illness such as would require care in an acute-care venue. By contrast, records for many of the ambulatory outpatient episodes lacked any SNOMED-CT observation or finding denoting a COVID-19 symptom, despite having PCR-positive test results associated with the U07.1 ICD-10-CM code.

Heterogeneity

The magnitude of a treatment benefit, if one exists, may depend on the patient's phenotypic and other characteristics. It is therefore important to investigate the heterogeneity and consistency of the treatment effect across subgroups to improve the reliability of study findings and their projectability to other populations. An observed heterogeneity in treatment effect across subgroups can arise because of chance alone, whereas true heterogeneity may be difficult to detect by standard statistical tests because of low power. For this reason, we measured heterogeneity across source institutions using the I^2 statistic, with higher values reflecting increasing heterogeneity (Higgins & Thompson, 2002; Higgins *et al.*, 2003; Lin *et al.*, 2017). Sources of heterogeneity were assessed by subgroup analysis and by

meta-regression (Baker *et al.*, 2009; Higgins & Thompson, 2002; Van Houwelingen *et al.*, 2002). Subgroup analyses included mean age and gender ratio by site.

Propensity score matching

Some recent papers (e.g. Gold *et al.*, 2020) about possible protective effects of vaccines, including MMR, BCG, and varicella-zoster vaccines have done little to control for bias and confounding in observational data. By contrast, we recommend that propensity score adjustment, heterogeneity estimation, and meta-analytic average treatment effect (ATE) estimation are essential when performing hypothesis testing on real-world data. Using stepwise forward variable selection and the purposeful variable selection method of Bursac and coworkers (Bursac *et al.*, 2008), we constructed covariable-based propensity scores [CBPS] (Fong *et al.*, 2018; Imai & Ratkovic, 2014; Ning *et al.*, 2020; Schneeweiss *et al.*, 2009) and performed 1:4 matching vaccine-exposed cases to controls based on these scores.

Results

On preliminary analysis of the prevalent PCR-proven COVID-19 cases in the real-world data (McNair *et al.*, 2021), prior vaccinations with a recombinant adjuvanted varicella-zoster vaccine or MMR do not appear to have significant protective effects against SARS-CoV-2 infections (extended data, Table S3; McNair *et al.*, 2021), but they do appear to show substantial effects reducing COVID-19 case fatality rate, with the MMR result rendered inconclusive when using additional statistical controls described below (extended data, Table S2; McNair *et al.*, 2021). Because of the null result for a protective effect, for the subsequent analyses we focus on the effects of prior vaccinations in reducing COVID-19 case fatality rate. However, the case fatality ratio (CFR) when expressed as an effect size exhibits substantial heterogeneity among the institutions contributing the cases to be analyzed. I^2 values indicate that 99% of the total variance in MMR- and recombinant adjuvanted varicella-zoster vaccine-associated mortality-prevention effect-size across studies is due to systematic (site-wise by source institution) heterogeneity rather than due to random sampling error. Bias of I^2 estimates is expected to be negligible for a collection of 90 source institutions as in the Cerner repository of COVID-19 cases and contemporaneous populations of patients receiving care during the pandemic (von Hippel, 2015).

We determined the absence of disparities in vaccine provisioning by race and gender by the Cochran-Mantel-Haenszel test (Landis *et al.*, 1978; Mantel, 1963; Mantel & Haenszel, 1959) noting that the exact version of the generalized Cochran-Mantel-Haenszel test does not necessarily result in the same p-value as Fisher's exact test (Davis, 1986). Evidence-combining was performed by standard methods of meta-analysis (Lipsey & Wilson, 2001; Singh *et al.*, 2005; Van Houwelingen *et al.*, 2002).

Our several attempts to use covariables available in the EHR-derived RWD repository to devise a suitable propensity score predicting MMR exposure were unsuccessful in achieving acceptable balance. Hosmer-Lemeshow calibration was poor across propensity score quartiles for MMR exposure. Furthermore, the

best of our provisional propensity scores for MMR exposure demonstrated residual heterogeneity $I^2 = 43\%$, a value that we believe to be too large to permit reliable interpretation of analyses of efficacy endpoints in real-world data (ideally below 10%; see (Austin & Stuart, 2015; Berger *et al.*, 2017)). The average treatment effect size calculated for the available data is not significant, and the sample size is not presently sufficient to power detection of so small an effect size (Munoz & Rosner, 1984; Nam, 1992; Wallenstein & Wittes, 1993; Woolson *et al.*, 1986). For this reason, we plan to defer further analysis of MMR non-specific protective efficacy until a larger sample size has accrued or pooling of data and meta-analysis with other repositories (such as the Oxford University – OHDSI Scylla repository) becomes possible. In general, findings were consistent that the presence of MMR vaccination did not yield lower case fatality rates than controls, but given the limitations described above, we could not derive clinically significant results from this dataset.

By contrast to MMR, the available data on the recombinant adjuvanted varicella-zoster vaccine yielded a propensity score having acceptable performance in terms of covariable balance and reduction in heterogeneity after matching (Figure 1–Figure 4). Here, the overall odds ratio for CFR was 0.43 (95%CI: 0.38 - 0.48). Thus, the case fatality-related non-specific vaccine efficacy (VE) for a recombinant adjuvanted varicella-zoster vaccine vaccination prior to SARS-CoV-2 infection is approximately $100 \times (1.00 - 0.43) = 57\%$.

Discussion

In the present real-world data analysis we did not find evidence that recent prior vaccination of adults with either a recombinant adjuvanted varicella-zoster vaccine or MMR vaccine conferred substantial protection of individuals against acquiring SARS-CoV-2 infection (extended data, Table S4; McNair *et al.*, 2021). For case fatality rate, inadequate MMR sample size and covariable distributions prevented construction of a propensity score having adequate performance for weighting or matching on covariables to reduce bias and heterogeneity sufficiently to enable reliable analysis. Sample size and event counts were, however, sufficient for the cohort who received a recombinant adjuvanted varicella-zoster vaccine to enable construction of a suitable propensity score, and analysis of propensity score matched cases and controls revealed a clinically meaningful benefit in reduced case fatality rate.

Regarding limitations of the present analysis, we recognize that there may likely be a temporal relation of non-specific protection by vaccines against COVID-19 or other infections. We conducted exploratory analyses [using methods reported by (Zhang *et al.*, 2020)] of the recency of recombinant adjuvanted varicella-zoster and MMR vaccinations to the date of presentation for the index COVID-19 episode. However, in the real-world data presently available for analysis there were insufficient patients exposed to a recombinant adjuvanted varicella-zoster vaccine between 2015 and 2017 to enable detailed analysis of the effect of the time interval between vaccination and SARS-CoV-2 infection. Therefore, our analysis

Figure 1. Standardized effect-sizes, with and without covariate propensity score (PS) adjustment, recombinant adjuvanted varicella-zoster vaccine.

Figure 2. Propensity score probability distributions of pre-coronavirus disease 2019 (COVID-19) period exposure to a recombinant adjuvanted varicella-zoster vaccine.

relied on the recombinant adjuvanted varicella-zoster vaccine exposures between 2017 and 2019, inclusive. Further analysis of temporal relationships and durability of non-specific heterologous protection awaits the accrual of a larger cohort or pooling of data with other investigators' real-world data repositories.

Beyond the concerns of the present work, we believe that this study illustrates the importance of certain best practices for reliable analysis of real-world data [see (Berger *et al.*, 2017;

Fang *et al.*, 2020; Stürmer *et al.*, 2020). In particular, we recommend the following six rules and practices:

- Only retain those sites whose contributed cases comprise at least 0.5% of the aggregate sample size.
- Perform propensity score adjustment to reduce bias and confounding.
- Assess possible residual bias or disparities with the Cochran-Mantel-Haenszel test or other suitable approach.

Figure 3. Covariables standardized difference after propensity score matching, recombinant adjuvanted varicella-zoster vaccine.

Figure 4. Forest plot, Recombinant adjuvanted varicella-zoster vaccine. CI=confidence interval.

- Proceed with multivariable inferential or causal modeling only if the residual heterogeneity is less than 10% and the residual post-propensity score adjustment maximum standardized mean difference on the covariables is less than 0.10.
- Perform meta-analysis to determine the average treatment effect (ATE) across the retained sites and examine forest plots and related outputs of meta-analysis.
- Verify the statistical power afforded by the available sample size for measuring the modeled ATE.

The first of these practices reflects considerations of the difficulty of collecting real-world data involving a large number of mapped, curated variables for propensity score adjustment and inferential modeling. Typically, real-world data tend to consist of dozens or hundreds of variables but a smaller number of data-contributing sites, which means $p > n$ (p is the number of predictor variables and n is the number of observed sites). In addition, exposure counts and outcome events often have

over-disperse distributions and have low prevalence. Thus, the data are sparse and unbalanced. This makes it challenging to apply classical statistical analysis methods (Hastie *et al.*, 2009). Under such conditions, a threshold is applied to focus on patterns and statistical associations within sites having substantial prevalent cases and controls, such that data sparsity will be less influential on analyses and analyses will not be dominated by one or two contributing sites having large numbers of prevalent cases. Minimum prevalence thresholding in RWD analysis is similar to the practice of applying sample selection filters for microbiome analyses of diversity based on minimum operational taxonomic unit (OTU) abundance and minimum total number of 16S rRNA reads (Bokulich *et al.*, 2013).

Conclusion

In summary, speculation about possible non-specific heterologous vaccine efficacy likely can only be conclusively resolved by appropriate randomized control trials. However, valuable evidence may be obtained from real-world data, provided that adequate samples size and endpoint event counts are

available and appropriate adjustments are undertaken to mitigate biases and confounding. Such evidence includes estimates of effect-size and average treatment effectiveness, which may help to inform statistical analysis plans for prospective trials (Levenson, 2020). While there are several theoretical hypotheses regarding the concept of heterologous trained immunity possibly conferred by ~~ok-to-leave~~ live here attenuated vaccines and vaccines with novel adjuvants, well-designed controlled studies are needed, before considering heterologous vaccination as a modality for preventing COVID-19 infection or mitigating its severity or outcomes.

Data availability

Underlying data

Zenodo: Looking in the medicine cabinet: Methods for using real-world data to assess the impact of MMR and Recombinant adjuvanted varicella-zoster vaccine on COVID-19 prevention and case fatality. <http://doi.org/10.5281/zenodo.5093688> (McNair *et al.*, 2021).

This project contains the following underlying data:

- 03_COVID_all_cleaned.csv (CSV file with all COVID+ subjects in Cerner institutions)
- 04_COVID_mmr_data_cleaned.csv (CSV file with COVID+ patients between 25–64 years old, including institution id, age category, gender, whether patient is in emergency department or inpatient, flu vaccine history, MMR vaccine history and mortality outcomes)
- 05_COVID_mmr_data_matched.csv (CSV file matching MMR vaccine-exposed cases to controls based on propensity scores)
- 06_COVID_zoster_data_cleaned.csv (CSV file with COVID+ patients above 50 years old, including institution id, age category, gender, whether patient is in emergency department or inpatient, flu vaccine history, zoster vaccine history and mortality outcomes)
- 07_COVID_zoster_data_matched.csv (CSV file matching zoster vaccine-exposed cases to controls based on propensity scores)

- 08_General_25_64_data_cleaned.csv (CSV file, all patients in Cerner institutions between 25 – 64 years old, including institution id, age category, gender, whether patient is in emergency department or inpatient, flu vaccine history, MMR vaccine history, SARS-CoV-2 infection and COVID-19 mortality outcomes).
- 09_General_over50_data_cleaned.csv (CSV file, all patients in Cerner institutions above 50 years old, including institution id, age category, gender, whether patient is in emergency department or inpatient, flu vaccine history, zoster vaccine history, SARS-CoV-2 infection and COVID-19 mortality outcomes).
- 10_Included_tenants.csv (CSV file, institution IDs whose contributed cases comprise at least 0.5% of the aggregate sample size).

Extended data

Zenodo: Looking in the medicine cabinet: Methods for using real-world data to assess the impact of MMR and Recombinant adjuvanted varicella-zoster vaccine on COVID-19 prevention and case fatality. <http://doi.org/10.5281/zenodo.5093688> (McNair *et al.*, 2021).

This project contains the following extended data:

- 01_Basic_Analysis.R (descriptive analysis script in R, for use with cleaned data files 3, 4 and 6)
- 02_Cleaning.R (Script demonstrating how Cerner data was cleaned upon download)
- 11_MMR_ps.R (R script to run for MMR-related files analysis)
- 12_Zoster_ps.R (R script to run for zoster-related files analysis)
- S1_20210712_vx_off_target_pubdraft_supplemental.docx (Tables, Graphs and scripts associated with publication)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

References

- Aaby P, Benn CS, Flanagan KL, *et al.*: **The non-specific and sex-differential effects of vaccines.** *Nat Rev Immunol.* 2020; **20**(8): 464–470.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Austin PC, Stuart EA: **Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies.** *Stat Med.* 2015; **34**(28): 3661–3679.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Baker W, Michael White C, Cappelleri JC, *et al.*: **Understanding heterogeneity**

- in meta-analysis: the role of meta-regression.** *Int J Clin Pract.* 2009; **63**(10): 1426–1434.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Berger ML, Sox H, Willke RJ, *et al.*: **Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making.** *Value Health.* 2017; **20**(8): 1003–1008.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bokulich NA, Subramanian S, Faith JJ, *et al.*: **Quality-filtering vastly improves**

- diversity estimates from Illumina amplicon sequencing. *Nat Methods*. 2013; 10(1): 57–59.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bursac Z, Gauss CH, Williams DK, et al.: **Purposeful Selection of Variables in Logistic Regression**. *Source Code Biol Med*. 2008; 3: 17.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Curtis N, Sparrow A, Ghebreyesus TA, et al.: **Considering BCG vaccination to reduce the impact of COVID-19**. *Lancet*. 2020; 395(10236): 1545–1546.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Davis LJ: **Exact Tests for 2 × 2 Contingency Tables**. *Am Stat*. 1986; 40(2): 139–141.
[Publisher Full Text](#)
- Didierlaurent AM, Collignon C, Bourguignon P, et al.: **Enhancement of adaptive immunity by the human vaccine adjuvant AS01 depends on activated dendritic cells**. *J Immunol*. 2014; 193(4): 1920–1930.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Escobar LE, Molina-Cruz A, Barillas-Mury C: **BCG vaccine protection from severe coronavirus disease 2019 (COVID-19)**. *Proc Natl Acad Sci U S A*. 2020; 117(30): 17720–17726.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fang Y, Wang H, He W: **A statistical roadmap for journey from real-world data to real-world evidence**. *Ther Innov Regul Sci*. 2020; 54(4): 749–757.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Farrington CP, Firth MJ, Moulton LH, et al.: **Epidemiological studies of the non-specific effects of vaccines: II—Methodological issues in the design and analysis of cohort studies**. *Trop Med Int Health*. 2009; 14(9): 977–985.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Fidel PL Jr, Noverr MC: **Could an unrelated live attenuated vaccine serve as a preventive measure to dampen septic inflammation associated with COVID-19 infection?** *mBio*. 2020; 11(3): e00907–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fine PE, Chen RT: **Confounding in studies of adverse reactions to vaccines**. *Am J Epidemiol*. 1992; 136(2): 121–135.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Fireman B, Lee J, Lewis N, et al.: **Influenza vaccination and mortality: differentiating vaccine effects from bias**. *Am J Epidemiol*. 2009; 170(5): 650–656.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fong C, Hazlett C, Imai K: **Covariate balancing propensity score for a continuous treatment: application to the efficacy of political advertisements**. *Ann Appl Stat*. 2018; 12(1): 156–177.
[Publisher Full Text](#)
- Gold JE, Tilley LP, Baumgartl WH: **MMR vaccine appears to confer strong protection from COVID-19: few deaths from SARS-CoV-2 in highly vaccinated populations**. ResearchGate, 2020.
[Publisher Full Text](#)
- Goodridge HS, Ahmed SS, Curtis N, et al.: **Harnessing the beneficial heterologous effects of vaccination**. *Nat Rev Immunol*. 2016; 16(6): 392–400.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gupta PK: **New disease old vaccine: Is recombinant BCG vaccine an answer for COVID-19?** *Cell Immunol*. 2020; 356: 104187.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hastie T, Tibshirani R, Friedman J: **The elements of statistical learning: data mining, inference, and prediction**. (Springer Science & Business Media). 2009.
[Publisher Full Text](#)
- Higgins JPT, Soares-Weiser K, López-López JA, et al.: **Association of BCG, DTP, and measles containing vaccines with childhood mortality: systematic review**. *BMJ*. 2016; 355: i5170.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Higgins JPT, Thompson SG: **Quantifying heterogeneity in a meta-analysis**. *Stat Med*. 2002; 21(11): 1539–1558.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Higgins JPT, Thompson SG, Deeks JJ, et al.: **Measuring inconsistency in meta-analyses**. *BMJ*. 2003; 327(7414): 557–560.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- H.R.6 - 21st Century Cures Act. 2015.
<https://clinicaltrials.gov/>
- Imai K, Ratkovic M: **Covariate balancing propensity score**. *J R Statist Soc B*. 2014; 76: 243–263.
[Reference Source](#)
- Jensen, KJ, Benn CS, van Crevel R: **Unravelling the nature of non-specific effects of vaccines—A challenge for innate immunologists**. *Semin Immunol*. 2016; 28(4): 377–383.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kandasamy R, Voysey M, McQuaid F, et al.: **Non-specific immunological effects of selected routine childhood immunisations: systematic review**. *BMJ*. 2016; 355: i5225.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Landis JR, Heyman ER, Koch GG: **Average partial association in three-way contingency tables: a review and discussion of alternative tests**. *Int Stat Rev*. 1978; 46(3): 237–254.
[Publisher Full Text](#)
- Levenson MS: **Regulatory-grade clinical trial design using real-world data**. *Clin Trials*. 2020; 17(4): 377–382.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lin L, Chu H, Hodges JS: **Alternative measures of between-study heterogeneity in meta-analysis: reducing the impact of outlying studies**. *Biometrics*. 2017; 73(1): 156–166.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lipsey MW, Wilson DB: **Practical meta-analysis**. (SAGE publications, Inc). 2001.
[Reference Source](#)
- Mantel N: **Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure**. *J Am Stat Assoc*. 1963; 58(303): 690–700.
[Publisher Full Text](#)
- Mantel N, Haenszel W: **Statistical aspects of the analysis of data from retrospective studies of disease**. *J Natl Cancer Inst*. 1959; 22(4): 719–748.
[PubMed Abstract](#) | [Publisher Full Text](#)
- McNair D, Hu H, Selwyn C: **Looking in the medicine cabinet: methods for using real-world data to assess the impact of measles, mumps and rubella (MMR) and recombinant adjuvanted varicella-zoster vaccines on coronavirus disease 2019 (COVID-19) prevention and case fatality**. [Data set]. Zenodo. 2021.
<http://www.doi.org/10.5281/zenodo.5093688>
- Messina N, Zimmermann P, Curtis N: **The impact of vaccines on heterologous adaptive immunity**. *Clin Microbiol Infect*. 2019; 25(12): 1484–1493.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Munoz A, Rosner B: **Power and sample size for a collection of 2 x 2 tables**. *Biometrics*. 1984; 40(4): 995–1004.
[Publisher Full Text](#)
- Nam J: **Sample size determination for case-control studies and the comparison of stratified and unstratified analyses**. *Biometrics*. 1992; 48(2): 389–395.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Netea MG, Joosten LA, Latz E, et al.: **Trained immunity: a program of innate immune memory in health and disease**. *Science*. 2016; 352(6284): aaf1098.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ning Y, Sida P, Imai K: **Robust estimation of causal effects via a high-dimensional covariate balancing propensity score**. *Biometrika*. 2020; 107(3): 533–554.
[Publisher Full Text](#)
- O'Neill LAJ, Netea MG: **BCG-induced trained immunity: can it offer protection against COVID-19?** *Nat Rev Immunol*. 2020; 20(6): 335–337.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pollard AJ, Finn A, Curtis N: **Non-specific effects of vaccines: plausible and potentially important, but implications uncertain**. *Arch Dis Child*. 2017; 102(11): 1077–1081.
[PubMed Abstract](#) | [Publisher Full Text](#)
- SAGE W: **Evidence based recommendations on nonspecific effects of BCG, DTP-containing and measles-containing vaccines on mortality in children under 5 years of age**. Background Paper for SAGE Discussions. 2014.
[Reference Source](#)
- Schneeweiss S, Rassen JA, Glynn RJ, et al.: **High-dimensional propensity score adjustment in studies of treatment effects using health care claims data**. *Epidemiology*. 2009; 20(4): 512–22.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sidiq KR, Sabir DK, Ali SM, et al.: **Does early childhood vaccination protect against COVID-19?** *Front Mol Biosci*. 2020; 7: 120.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Singh K, Xie M, Strawderman WE: **Combining information from independent sources through confidence distributions**. *Ann Stat*. 2005; 33(1): 159–183.
[Publisher Full Text](#)
- Stürmer T, Wang T, Golightly YM, et al.: **Methodological considerations when analysing and interpreting real-world data**. *Rheumatology (Oxford)*. 2020; 59(1): 14–25.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Van Houwelingen HC, Arends LR, Stijnen T: **Advanced methods in meta-analysis: multivariate approach and meta-regression**. *Stat Med*. 2002; 21(4): 589–624.
[PubMed Abstract](#) | [Publisher Full Text](#)
- von Hippel PT: **The heterogeneity statistic I(2) can be biased in small meta-analyses**. *BMC Med Res Methodol*. 2015; 15: 35.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wallenstein S, Wittes J: **The power of the Mantel-Haenszel test for grouped failure time data**. *Biometrics*. 1993; 49(4): 1077–1087.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Woolson RF, Bean JA, Rojas PB: **Sample size for case-control studies using Cochran's statistic**. *Biometrics*. 1986; 42(4): 927–932.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zhang W, Li J, Liu L: **A unified survey on treatment effect heterogeneity modeling and uplift modeling**. ArXiv Preprint ArXiv: 2007.12769. 2020.

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 18 February 2022

<https://doi.org/10.21956/gatesopenres.14572.r31660>

© 2022 Khalid S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sara Khalid

Centre for Statistics in Medicine, NIHR Biomedical Research Centre Oxford, NDORMS, University of Oxford, Oxford, UK

The authors have presented an important topic and interesting methodology.

The main observations are:

- a) Insufficient clarity in description of analysis and corresponding results.
- b) Lack of reporting detail. The authors are highly encouraged to consider and report guidelines for observational research e.g. STROBE or similar.

Examples:

- Sample size consideration is mentioned in the methods but it is not obvious/present in the results?
- Justification for choice of 1:4 matching is lacking.
- Use for variable selection within PS matching is interesting - the justification of using this as opposed to an established method such as comparison of standardised mean differences is lacking. Furthermore the results for this step were not clear.
- One ICD10 code is mentioned. There are known to be several other icd10 codes for covid and positive PCR. Were those considered? If not, the analysis remains selective and should be explained in the discussion.
- Unclear what aggregate sample size here refers to and justification for 0.5% threshold: "38 tenants whose contributed cases comprise more than 0.5% of the aggregate sample size".
- "we recommend the following six rules and practices": recommendations should as a rule be supported by evidence. These recommendations were not the objective of this study nor was specific analysis presented to support these claims. The authors are strongly

recommended to reconsider presenting them within this paper.

Is the rationale for developing the new method (or application) clearly explained?

Partly

Is the description of the method technically sound?

No

Are sufficient details provided to allow replication of the method development and its use by others?

No

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Observational research, machine learning, epidemiology, health informatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.