Title

Bias, Precision and Timeliness of Historical (Background) Rate Comparison Methods for Vaccine Safety Monitoring: An Empirical Multi-Database Analysis

Permalink

https://escholarship.org/uc/item/1ds9g1j6

Authors

Li, Xintong
Lai, Lana YH
Ostropolets, Anna
et al.

Publication Date

2021

DOI

10.3389/fphar.2021.773875

Peer reviewed

Bias, Precision and Timeliness of Historical (Background) Rate Comparison Methods for Vaccine Safety Monitoring: An Empirical Multi-Database Analysis

Xintong Li[1], Lana YH Lai[2], Anna Ostropolets[3], Faaizah Arshad[4], Eng Hooi Tan[1], Paula Casajust[5], Thamir M. Alshammari[6], Talita Duarte-Salles[7], Evan P. Minty[8], Carlos Areia[9], Nicole Pratt[10], Patrick B. Ryan[11,12], George Hripcsak[3,13], Marc A. Suchard[4,11], Martijn J. Schuemie[4,11,12]* and Daniel Prieto-Alhambra[1,14]

[1]Centre for Statistics in Medicine, NDORMS, University of Oxford, Oxford, United Kingdom, [2]School of Medical Sciences, University of Manchester, Manchester, United Kingdom, [3]Department of Biomedical Informatics, Columbia University, New York, NY, United States, [4]Department of Biostatistics, University of California, Los Angeles, California, United States, [5]Real-World Evidence, Trial Form Support, Barcelona, Spain, [6]College of Pharmacy, Riyadh Elm University, Riyadh, Saudi Arabia, [7]Institut Universitari D'Investigació en Atenció Primària Jordi Gol (IDIAPJGol), Barcelona, Spain, [8]O'Brien Institute for Public Health, Faculty of Medicine, University of Calgary, Calgary, AB, Canada, [9]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom, [10]Quality Use of Medicines and Pharmacy Research Centre, Clinical and Health Sciences, University of South Australia, SA, Australia, [11]Observational Health Data Sciences and Informatics, New York, NY, United States, [12]Observational Health Data Analytics, Janssen R&D, Titusville, NJ, United States, [13]Medical Informatics Services, NewYork-Presbyterian Hospital, NewYork, NY, United States, [14]Health Data Sciences, Medical Informatics, Erasmus Medical Center University, Rotterdam, Netherlands

Using real-world data and past vaccination data, we conducted a large-scale experiment to quantify bias, precision and timeliness of different study designs to estimate historical background (expected) compared to post-vaccination (observed) rates of safety events for several vaccines. We used negative (not causally related) and positive control outcomes. The latter were synthetically generated true safety signals with incident rate ratios ranging from 1.5 to 4. Observed vs. expected analysis using within-database historical background rates is a sensitive but unspecific method for the identification of potential vaccine safety signals. Despite good discrimination, most analyses showed a tendency to overestimate risks, with 20%-100% type 1 error, but low (0% to 20%) type 2 error in the large databases included in our study. Efforts to improve the comparability of background and post-vaccine rates, including age-sex adjustment and anchoring background rates around a visit, reduced type 1 error and improved precision but residual systematic error persisted. Additionally, empirical calibration dramatically reduced type 1 to nominal but came at the cost of increasing type 2 error.

Keywords: incidence rate, vaccine safety, real world data, empirical - comparison, background rate

FIRST PAGE

Standfirst

Using real-world data and past vaccination data, we conducted a large-scale experiment to quantify bias, precision and timeliness of different study designs to estimate historical background (expected) compared to post-vaccination (observed) rates of safety events for several vaccines. We used negative (not causally related) and positive control outcomes. The latter were synthetically generated true safety signals with incident rate ratios ranging from 1.5 to 4.

Observed vs. expected analysis using within-database historical background rates is a sensitive but unspecific method for the identification of potential vaccine safety signals. Despite good discrimination, most analyses showed a tendency to overestimate risks, with 20–100% type 1 error, but low (0–20%) type 2 error in the large databases included in our study. Efforts to improve the comparability of background and post-vaccine rates, including age-sex adjustment and anchoring background rates around a visit, reduced type 1 error and improved precision but residual systematic error persisted. Additionally, empirical calibration dramatically reduced type 1 to nominal but came at the cost of increasing type 2 error.

Key Messages

- Within-database background rate comparison is a sensitive but unspecific method to identify vaccine safety signals. The method is positively biased, with low (≤20%) type 2 error, and 20–100% of negative control outcomes were incorrectly identified as safety signals due to type 1 error.
- Age-sex adjustment and anchoring background rate estimates around a healthcare visit are useful strategies to reduce false positives, with little impact on type 2 error.
- Sufficient sensitivity was reached for the identification of safety signals by month 1-2 for vaccines with quick uptake (e.g., seasonal influenza), but much later (up to month 9) for vaccines with slower uptake (e.g., varicella-zoster or papillomavirus).
- Empirical calibration using negative control outcomes reduces type 1 error to nominal at the cost of increasing type 2 error.

INTRODUCTION

As regulators across the world evaluate the first signals of post-marketing safety potentially associated with coronavirus disease 2019 (COVID-19) vaccines, they rely on the use of historical comparisons with so-called "background rates" for the events of interest to identify outcomes appearing more often than expected following vaccination. However, a literature gap remains on the reliability of these methods, their associated error(s), and the impact of potential strategies to mitigate them. We therefore aimed to study the bias, precision, and timeliness associated with the use of historical comparisons between post-vaccine and background rates for the identification of safety signals. We tested strategies for background rate estimation (unadjusted, age-sex adjusted, and anchored around

a healthcare visit), and studied the impact of empirical calibration on type 1 and type 2 error.

MANUSCRIPT TEXT

Background

One of the most common study designs in vaccine safety surveillance is the use of a cohort study with a historical comparison as a benchmark. This design allows the observed incidence of adverse events of the studied vaccine following immunization (AEFI) to be compared with the expected incidence of AEFI projected based on historical data (Belongia et al., 2010). Alleged strengths include greater statistical power to detect rare AEFIs, as well as improved timeliness in detecting potential safety signals by leveraging retrospective data for analysis. There are, however, also caveats with this study design (Mesfin et al., 2019). Firstly, the historical population must be similar to the vaccinated cohort to obtain comparable estimates of baseline risk. Secondly, the design is subject to various temporal confounders such as seasonality, changing trends in the detection of AEFIs, and variation in diagnostic or coding criteria over time. Thirdly, the design is highly dependent on an accurate estimation of background incidence rates of the AEFIs for comparison.

Historical rate comparison has been suggested for use in several vaccine safety guidelines, including the European Network of Centres of Pharmacoepidemiology and Pharmacovigilance (ENCePP), Council for International Organizations of Medical Sciences (CIOMS), and Good Pharmacovigilance Practices (GVP). It has also been applied extensively in various clinical domains, including the Center for Disease Control and Prevention (CDC)'s Vaccine Safety Datalink (VSD) project, which used background rates to detect safety signals for the human papillomavirus vaccine (HPV) (Gee et al., 2011), adult tetanus-diphtheria-acellular pertussis (Tdap) vaccine (Yih et al., 2009), and a broad range of paediatric vaccines (Lieu et al., 2007; Yih et al., 2011). Historical data were used in Australia to detect signals for the rotavirus vaccine (Buttery et al., 2011), and in Europe to detect signals for the influenza A H1N1 vaccine (Black et al., 2009; Wijnans et al., 2013; Barker and Snape, 2014). While this study design is widely implemented, there is high variability in the specifics of methods used to calculate historical rates, including selection of target populations, time-at-risk windows, observation time and study settings.

Uncertainties and Limitations With the use of Historical Rate Comparisons for Vaccine Safety Monitoring

Several studies have acknowledged uncertainties associated with the use of background rates relating to temporal and geographical variations. In one study that applied both historical comparisons and self-controlled methods, a signal of seizure in the 2014–2015 flu season was detected in the latter analysis but not the former. The authors explained that one possible reason was that the historical rates used might not reflect the expected baseline rate in the absence of vaccination. A second explanation was a falsely elevated

background rate because of the inclusion of events induced by a previous vaccine season. Other studies have highlighted the importance of accounting for demographic, secular and seasonal trends to appropriately interpret historical rates (Buttery et al., 2011; Yih et al., 2011). Nevertheless, the influence of such trends has not been studied systematically despite observed heterogeneity in historical incidence rates (Yih et al., 2011).

It is also essential to consider the data source since there are differences in case ascertainment. This might lead to uncertainty in background rate estimates, especially in rare events (Mahaux et al., 2016). In addition, there might also be differences in the use of dictionary or codes to define an AEFI. For example, the spontaneous reporting system generally uses the Medical Dictionary for Regulatory Activities (MedDRA), while in observational databases different codes are used (e.g., International Classification of Diseases (ICD), SNOMED-CT, READ) and the granularity of available coding can impact the sensitivity and specificity of phenotype algorithms.

There have been suggestions on how to mitigate some of the differences between the historical and observed populations, including stratifying by age, gender, geographical or calendar time (Gee et al., 2011; Yih et al., 2011). While these approaches may reduce some differences, the distribution of the observed population is rarely known unless the study uses the spontaneous case's demographic characteristics (of which the cases may be identified through the adverse event spontaneous system) as a proxy of the demographic characteristics of the observed population. This could potentially lead to a bias due to the estimation misclassification in each stratum based on the reporting rate (i.e., high vs. low reporting rates).

Large databases that link medical outcomes with vaccine exposure data provide a means of assessing signals identified, as well as estimates of a true incidence of clinical events after vaccination. However, these systems can be affected by relatively small denominators (given the rarity of the event) of vaccinated subjects, and a time lag in the availability of data. Very rare events or outcomes affecting a subset of the population might still be under-powered to assess a safety concern even when the data reflect the experience of millions of individuals (Black et al., 2009). Heterogeneity in background rates across databases and age-sex strata may also persist even after robust data harmonization using common data models (Li et al., 2021).

We therefore aimed to study the bias, precision, and timeliness associated with the use of historical comparisons between post-vaccine and background rates for the identification of safety signals. We evaluated strategies for estimating background rates and the effect of empirical calibration on type 1 and type 2 error using real-world outcomes presumed to be unrelated to vaccines (negative control outcomes) as well as imputed positive controls (outcomes simulated to be caused by the vaccines).

METHODS

Data Sources and Data Access Approval
We aimed to fill a gap in the existing literature by estimating the bias, precision and timeliness associated with the use of historical/

background compared to post-vaccination rates of safety events using "real world" (electronic health records and administrative health claims) databases from the US. Our study protocol is available in the EU PAS Register (EUPAS40259) (European Network of Centres of Pharmacoepidemiology and Pharmacovigilance, 2021), and all our analytical code is in GitHub (https://github.com/ohdsi-studies/Eumaeus). These data were previously mapped to the OMOP common data model (OHDSI, 2019). The list of included data sources, with a brief description, is available in - **Supplementary Appendix Table S1**.

The use of Optum and IBM Marketscan databases was reviewed by the New England Institution Review Board (IRB) and was determined to be exempt from broad IRB approval, as this research project did not involve human subjects research.

Exposures
We used retrospective data to study the following vaccines within the corresponding study periods: 1) H1N1 vaccination (Sept 2009 to May 2010), 2) different types of seasonal flu vaccination (Sept 2017 to May 2018), 3) varicella-zoster vaccination (Jan 2018 to Dec 2018), and 4) HPV 9-valent recombinant vaccine (Jan 2018 to Dec 2018). Vaccines were captured as drug exposure in the common data mode. Specific CVX codes and RxNorm codes, follow-up periods, and cohort construction details are available in – **Supplementary Appendix Table S2**. Post-vaccination rates were obtained for the period of 1–9 months for H1N1 and seasonal flu, and 1 to 12 for varicella-zoster and HPV vaccines. Background (historical) rates were obtained from the general population, for the same range of months 1 year preceding each of these vaccines (Unadjusted). To minimise confounding, three additional variations of background rates were estimated: 1) age-sex adjusted rates; 2) visit-anchored rates; and 3) visit and age-sex adjusted rates. In the first, background rates were stratified by age (10-years bands) and sex. In the second option, background rates were estimated using the time-at-risk following a random outpatient visit (visit-anchored). The third combined the two above to account for differences in socio-demographics and for the impact of anchoring (similar to anchoring post-vaccination in the exposed group).

Outcomes
We employed negative control outcomes as a benchmark to estimate bias (Schuemie et al., 2016; Schuemie et al., 2020). Negative controls are outcomes with no plausible causal association with any of the vaccines. As such, negative control outcomes should not be identified as a signal by a safety surveillance method, and any departure from a null effect is therefore suggestive of bias due to type 1 error. A list of negative control outcomes was pre-specified for all four vaccine groups. To identify negative control outcomes that match the severity and prevalence of suspected vaccine adverse effects, a candidate list of negative controls was generated based on similarity of prevalence and percent of diagnoses that were recorded in an inpatient setting (as a proxy for severity). Three clinical experts manually reviewed the list, which led to a final list of 93 negative control

outcomes to be included. Details of the selection process, including the candidate outcomes, reasons for exclusion, and the final negative control outcomes list are available in - **Supplementary Appendix Table S3**.

In addition, synthetic positive control outcomes were generated to measure type 2 error (OHDSI, 2019). Given the limited knowledge of such events and the lack of consistency in the true causal association amongst other problems [6], we computed synthetic positive controls with known (albeit *in silico*) causal associations with the vaccines under study [5,7]. Positive outcomes were generated by modifying negative control outcomes through injection of additional simulated occurrences of the outcome, with effect sizes equivalent to true incidence rate ratios (IRR) of 1.5, 2, and 4. With the 3 mentioned true IRR, 93 negative controls were used to construct at most $93 \times 3 = 279$ positive control outcomes, although no positive controls were synthesized if for the negative control the number of outcomes was smaller than 25. The hazard for these outcomes was simulated to be increased for the period 1 day after vaccination until 28 days after vaccination, with a constant hazard ratio during that time.

Performance Metrics

The estimated effect size for the association vaccine-outcome was based on IRR by dividing the observed (post-vaccine) over expected (historical) incidence rates. To account for systematic error, we employ empirical calibration: we firstly compute the distribution of systematic error using the estimates for the negative and positive control outcomes. We then use the distributions and their standard deviations to adjust effect-size estimates, confidence intervals, p-values, and log likelihood ratios (LLRs) to restore type 1 error to nominal. We used a leave-one-out strategy for this evaluation, calibrating the estimate for a control outcome using the systematic error distribution fitted on all control outcomes except the one being calibrated. IRR were computed both with and without empirical calibration (Schuemie et al., 2016; Schuemie et al., 2018).

Bias was measured using: 1) Type 1 error, based on the proportion of negative control outcomes identified as safety signals according to p-value < 0.05; 2) Type 2 error, based on how often positive control outcomes were missed (not identified) as safety signals ($p > 0.05$); 3) Area Under the receiver-operator Curve (AUC) for the discrimination of effect size estimate between positive and negative controls; and 4) Coverage, defined by how often the true IRR was within the 95% confidence interval of the estimated IRR.

Precision was measured using mean precision and mean squared error (MSE). Geometric mean precision was computed as 1 / (standard error)2, with higher precision equivalent to narrower confidence intervals. MSE was obtained from the log of estimated IRRs and the log of the true HR.

To understand the time it took the analysis method to identify a safety signal (aka *timeliness*), the follow-up (up to 12 months) occurring after each vaccine was divided into calendar months. For each month, the analyses were executed using the data accumulated until the end of that month, and bias and precision metrics were estimated.

Finally, we studied the proportion of controls for which IRR were not estimable due to lack of participants exposed to the vaccine of interest. We also considered as not estimable (and therefore did not report) results for negative control outcomes with a population-based incidence rate changing >50% over time during the study period.

For all the estimated metrics, we reported the results for each database – vaccine group – method group.

FINDINGS

Bias and Precision

A total of four large databases were included, most including all four vaccines of interest: IBM MarketScan Commercial Claims and Encounters (CCAE), IBM MarketScan Multi-state Medicaid (MDCD), IBM MarketScan Medicare Supplemental Beneficiaries (MDCR), and Optum© de-identified Electronic Health Record dataset (Optum EHR). The basic socio-demographics of participants registered in each of these databases are reported in **Supplementary Appenix S1**. All data sources had a majority of women, from 51.1% in CCAE to 56.23% in MDCD. As expected, data sources with older populations (e.g., IBM MDCR) had little exposure to HPV vaccination, but high numbers of participants exposed to seasonal influenza vaccination. All four data sources contributed information based on healthcare encounters in emergency rooms, outpatient as well as inpatient settings.

Historical rate comparisons were –even in their simplest form— associated with low type 2 error (0–10%), but led to type 1 errors ranging between 30% (HPV in MDCD) and 100% (H1N1 and seasonal flu in Optum EHR). Adjustment for age and sex reduced type 1 error in some but not all scenarios, and had limited impact on type 2 error (maximum 20% in all the conducted analyses). However, age and sex adjusted comparisons were still prone to type 1 error, with most (12/13) analyses still incorrectly identifying ≥40% negative controls as potential safety signals. Anchoring the estimation of background rates around a healthcare visit helped reduce type 1 error in some scenarios (e.g., H1N1 in Optum EHR went from 100 to 50%), but increased it in others (e.g., H1N1 in CCAE increased from 50% in the unadjusted to 80% in the anchored analysis). In addition, anchoring increased type 2 error in most of our analyses, although none exceeded 20% in any of the analyses. Finally, the analyses combining anchoring and age-sex adjustment led to observable reductions in type 1 error (e.g., from 70 to 30% for HPV in CCAE), with negligible increases in type 2 error in most instances (e.g., from 10 to 20% for HPV in MDCD). Detailed results for unadjusted, age-sex adjusted, and anchoring scenarios are demonstrated in **Figure 1**.

Historical rates comparison had overall good discrimination to distinguish true safety signals (i.e., positive control outcomes), with AUCs of 80% or over in all the analyses and databases. Age-sex adjustment and anchoring had little impact on this. Conversely, coverage was low, with many analyses failing to accurately measure and include the true effect of our negative and positive control outcomes (**Supplementary Appendix S2**). Coverage in unadjusted analyses ranged from 0 (H1N1 vaccines
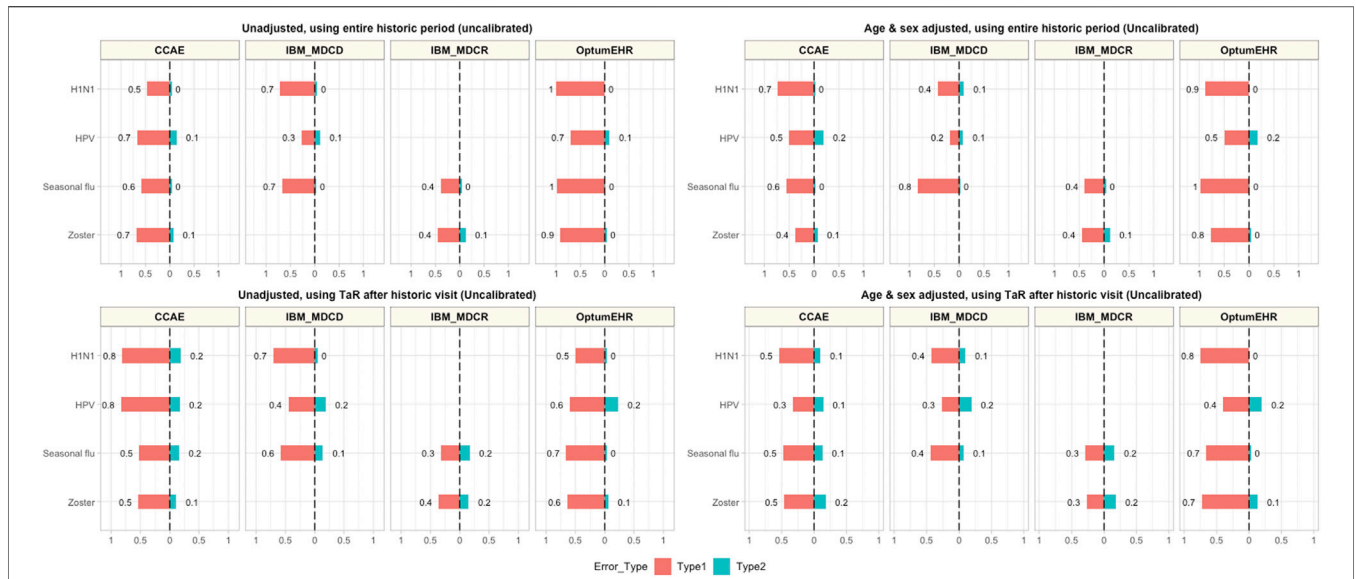
FIGURE 1 | Type 1 and Type 2 error in unadjusted, age-sex adjusted, and anchored background rate analyses CCAE: IBM MarketScan Commercial Claims and Encounters; MDCR: IBM Health MarketScan Medicare Supplemental; MDCD: IBM Health MarketScan Multi-state Medicaid; Optum EHR: Optum© de-identified Electronic Health Record Dataset.
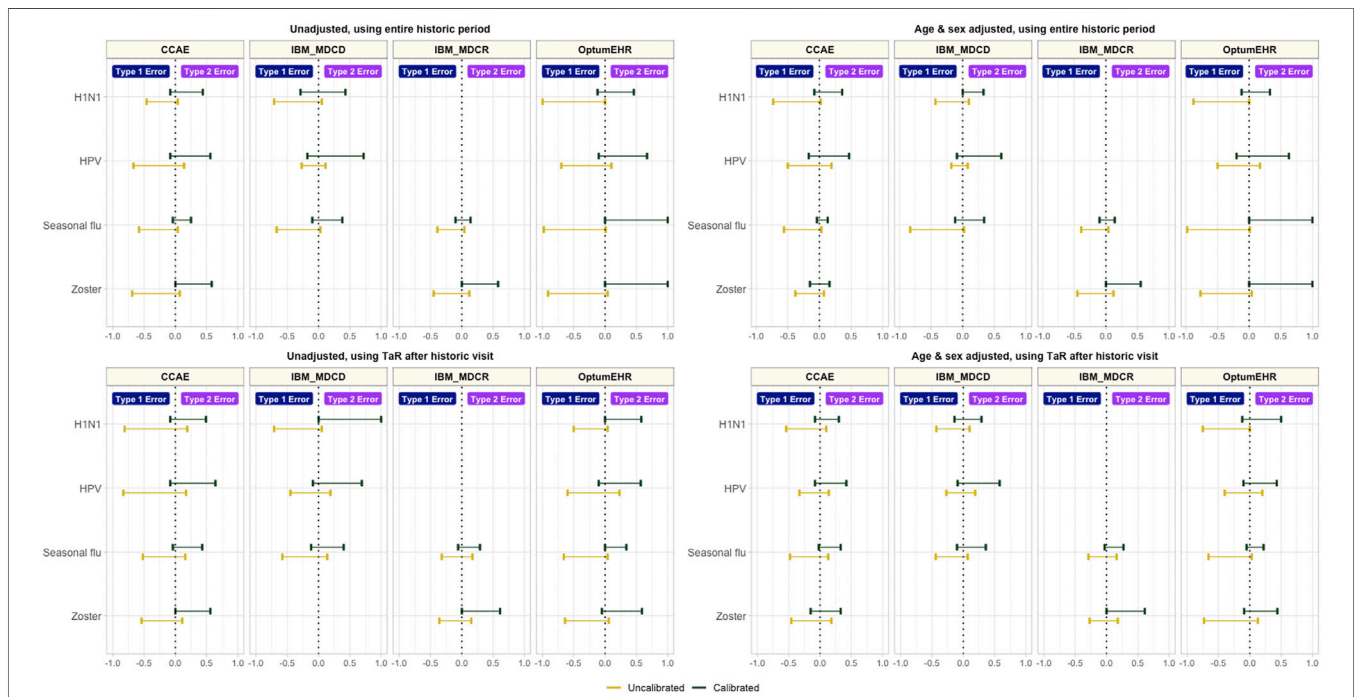

FIGURE 2 | Type 1 and type 2 error before vs after empirical calibration *CCAE: IBM MarketScan Commercial Claims and Encounters; MDCR: IBM Health MarketScan Medicare Supplemental; MDCD: IBM Health MarketScan Multi-state Medicaid; Optum EHR: Optum© de-identified Electronic Health Record Dataset.

in Optum EHR) to 0.51 (seasonal influenza vaccine in MDCR). Age-sex adjustment and anchoring had overall a positive effect on coverage, with little or no effect on discrimination (**Supplementary Appendix S2**). Precision, as measured by mean precision and MSE, varied by database and vaccine exposure as reported in **Supplementary Appendix S2**. Adjustment for age and sex and anchoring improved precision in most scenarios.
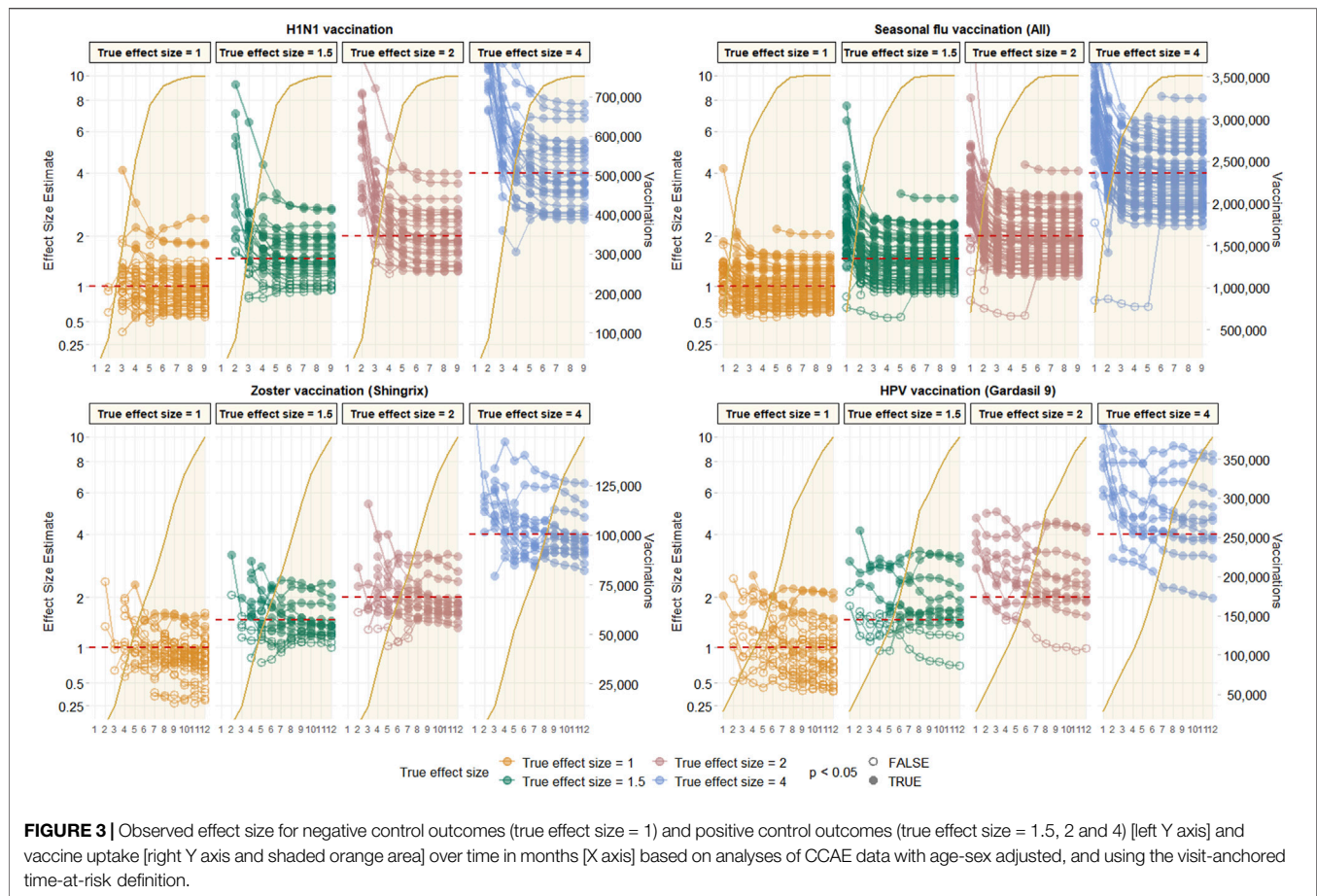
FIGURE 3 | Observed effect size for negative control outcomes (true effect size = 1) and positive control outcomes (true effect size = 1.5, 2 and 4) [left Y axis] and vaccine uptake [right Y axis and shaded orange area] over time in months [X axis] based on analyses of CCAE data with age-sex adjusted, and using the visit-anchored time-at-risk definition.

The Effect of Empirical Calibration

Empirical calibration reduced type 1 error substantially, but increased type 2 error in all the tested scenarios (see **Figure 2**). In addition to this, calibration improved coverage without impacting AUC, and decreased precision in most scenarios (**Supplementary Appendix S3**).

Timeliness

Most observed associations were unstable in the first few months of study, and stabilised around the true effect size in the first 2–3 months after campaign initiation for vaccines with rapid uptake like H1N1 or seasonal influenza. This stability was, however, not seen until much later, and sometimes not seen at all in the 12-months study period for vaccines with slower uptake like HPV or varicella-zoster. This is depicted in **Figure 3** using data from CCAE as an illustrative example, and for all other databases in **Supplementary Figures S1–S3**.

DISCUSSION

Key Results

Our study found that unadjusted background rates comparison had low type 2 error of <10% in all analyses but unacceptably high type 1 error, up to 100% in some scenarios. The method is positively biased

and uncalibrated estimates and p-values cannot be interpreted as intended; while it may be encouraging that most positive effects can be identified at a decision threshold of $p < 0.05$, this threshold will also yield a substantial proportion of false positive findings. Age-sex adjustment and anchoring background rate estimation around a healthcare visit were useful strategies to reduce type 1 error to around 50%, while maintaining sensitivity. Empirical calibration led to restoration of type 1 error to nominal but correction for positive bias necessitates increasing type 2 error. In terms of timeliness, background rate comparisons were sensitive methods for the early identification of potential safety signals. However, most associations were exaggerated and unstable in the first few months of vaccination campaign. Vaccines with higher uptake, such as H1N1 or seasonal flu, were associated with earlier identification of safety outcomes after launch in the analyses of vaccines with rapid uptake like H1N1 or seasonal influenza.

Previous studies have shown that background incidence rates of AESI vary between age and sex (Black et al., 2009). For example, the incidence of Bell's palsy in adults aged over 65 years is 4 times that in paediatric population in the United Kingdom; whereas the risk of optic neuritis is higher in females than males with the same age group in Sweden. Therefore, it is crucial that age and sex are adjusted for when using background incidence rates for comparison. Nonetheless, Li et al. (Li et al., 2021) found considerable heterogeneity in incidence rates of AESI within age-sex stratified subgroups. This suggests that

residual patient-level differences in characteristics such as comorbidities and medication use remained. Background rates comparison assumes that the background incidence in the overall population is similar to the vaccinated population. This assumption may not be valid because of confounding by indication, where the vaccinated population has more chronic conditions than the unvaccinated population. Conversely, the healthy vaccinee effect could occur, where on average healthier patients are more likely to adhere to annual influenza vaccination (Remschmidt et al., 2015).

Research in Context

Post-marketing surveillance is required to ensure the safety of vaccines, so that the public do not avoid getting life-saving vaccinations because of concerns that vaccine risks are not monitored, and that any potential risks do not outweigh the vaccine's benefits. The goal of these surveillance systems is to detect safety signals in a timely manner without raising excessive false alarms. There is an implicit trade-off between sensitivity (type 2 error) and specificity (type 1 error). Claims extending from a false positive result that is suggestive of an adverse event of a vaccine, fueled by sensationalism and unbalanced reporting in the media, could have devastating consequences on public health. A classic example of harm is the link between the MMR vaccine and autism. Although the fraudulent report by Wakefield has been retracted and many subsequent studies found no association, its lasting effects can be seen in falling MMR vaccination rates below the recommend levels from the World Health Organization (Godlee et al., 2011). Expert consensus alleged that this was a contributing factor in measles being declared endemic in the United Kingdom in 2008 (Jolley and Douglas, 2014) and sporadic outbreaks in the United States in recent years (Benecke and DeYoung, 2019). On the other hand, missing safety signals could put patients at risk as well as dampen public confidence in vaccination. Transparency is needed when communicating vaccination results to the public. However, it is a tricky balance to put both the benefits and harms of vaccination in context. The urgency to act quickly on the basis of incomplete real-world data could lead to confusion about vaccination safety. Negative perceptions about vaccination can be deeply entrenched and difficult to address. A starting point could be to include relevant background rates to provide comparison to other scenarios. As reported in our study, age and sex-adjusted rates are crucial to minimise false positive safety signals. Another form of communication could be using infographics to weigh harms versus benefits, illustrating the differential risks in various age groups as was shown by researchers from the University of Cambridge who contrasted the prevention of ICU admissions due to COVID-19 against the risk of blood clots due to the vaccine in specific age groups (Winton Centre for Risk and Evidence Communication, 2021).

Strengths and Limitations

The strength of this study lies in the implementation of a harmonised protocol across multiple databases, which allows us to compare the findings across different healthcare systems. The use of a common data model allows the experiment to be replicated in future databases while maintaining patient privacy as patient-level data will not be shared outside of each institution. Use of real negative and synthetic positive control outcomes provides an independent estimate of residual bias in the study design and data source. The fully specified study protocol was published before analysis began and dissemination of the results did not depend on estimated effects, thus avoiding publication bias. All codes used to define the cohort, exposures, and outcomes as well as analytical code are made open source to enhance transparency and reproducibility.

In our analysis, while using negative control outcomes can reflect the real confounding and measurement error, the approach of simulating positive control outcomes relied on assumptions about systematic error. It is assumed that the systematic error does not change when the true effect size is greater than 1, rather than as a function of the true effect size. Furthermore, positive control synthesis assumes the positive predictive value and sensitivity of the outcomes is the same for background outcome events and the outcome events simulated to be caused by the vaccine, which may not be true in the real world (Schuemie et al., 2020).

For the Optum EHR data, we may miss the care episodes when patients seeking care outside the respective health system, this will cause bias towards the null. All these limitations needed to be considered while interpreting our results.

Future Research and Recommendations

When using background rate comparison for post-vaccine safety surveillance, age-sex adjustment in combination with anchoring time-at-risk around an outpatient visit resulted in somewhat reduced type 1 error, without much impact on type 2 error. Residual bias, nonetheless, remained using this design, with very high levels of type 1 error observed in most analyses. Calibration is useful for reducing Type 1 error but at the expense of decreasing precision and consequently increasing type II error. Future studies using cohort and SCCS self-controlled cased series methods with empirical calibration will be evaluated.

AUTHOR'S NOTE

The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties.

DATA AVAILABILITY STATEMENT

The patient-level data used for these analyses cannot be provided due to information governance restrictions. All our analytical code is available to enable the replication of our analyses at https://github.com/ohdsi-studies/Eumaeus.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for

participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

XL is a PhD student in the Pharmacoepidemiology group at the Centre for Statistics in Medicine (CSM) (Oxford) and an MHS. She contributed to drafting the manuscript, creating tables and figures, and reviewing relevant literature. LL has experience in the analysis and interpretation of routine health data, and was responsible for leading the literature review of similar studies, and for the drafting of the first version of the manuscript. AO is a PhD student and MD at Columbia University; she contributed to the analysis and the drafting of the paper. FA is an Undergraduate Student and contributed to the drafting of the manuscript and reviewing the final version of the manuscript. MS is the guarantor of the study, and led the data analyses. DP is also study guarantor and led the preparation and final review of the manuscript. All co-authors reviewed, provided feedback, and finally gave approval for the submission of the current version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2021.773875/full#supplementary-material

REFERENCES

Barker, C. I., and Snape, M. D. (2014). Pandemic Influenza A H1N1 Vaccines and Narcolepsy: Vaccine Safety Surveillance in Action. *Lancet Infect. Dis.* 14 (3), 227–238. doi:10.1016/S1473-3099(13)70238-X

Belongia, E. A., Irving, S. A., Shui, I. M., Kulldorff, M., Lewis, E., Yin, R., et al. (2010). Real-time Surveillance to Assess Risk of Intussusception and Other Adverse Events after Pentavalent, Bovine-Derived Rotavirus Vaccine. *Pediatr. Infect. Dis. J.* 29 (1), 1–5. doi:10.1097/INF.0b013e3181af8605

Benecke, O., and DeYoung, S. E. (2019). Anti-Vaccine Decision-Making and Measles Resurgence in the United States. *Glob. Pediatr. Health* 6, 2333794X19862949–2333794X. doi:10.1177/2333794X19862949

Black, S., Eskola, J., Siegrist, C. A., Halsey, N., MacDonald, N., Law, B., et al. (2009). Importance of Background Rates of Disease in Assessment of Vaccine Safety during Mass Immunisation with Pandemic H1N1 Influenza Vaccines. *Lancet* 374 (9707), 2115–2122. doi:10.1016/S0140-6736(09)61877-8

Buttery, J. P., Danchin, M. H., Lee, K. J., Carlin, J. B., McIntyre, P. B., Elliott, E. J., et al. (2011). Intussusception Following Rotavirus Vaccine Administration: post-marketing Surveillance in the National Immunization Program in Australia. *Vaccine* 29 (16), 3061–3066. doi:10.1016/j.vaccine.2011.01.088

European Network of Centres of Pharmacoepidemiology and Pharmacovigilance (2021). EUMAEUS: Evaluating Use of Methods for Adverse Event under Surveillance (For Vaccines). 2021 [updated 15 April 2021. Available at: http://www.encepp.eu/encepp/viewResource.htm?id=40341.

Gee, J., Naleway, A., Shui, I., Baggs, J., Yin, R., Li, R., et al. (2011). Monitoring the Safety of Quadrivalent Human Papillomavirus Vaccine: Findings from the Vaccine Safety Datalink. *Vaccine* 29 (46), 8279–8284. doi:10.1016/j.vaccine.2011.08.106

Godlee, F., Smith, J., and Marcovitch, H. (2011). Wakefield's Article Linking MMR Vaccine and Autism Was Fraudulent. *BMJ* 342, c7452. doi:10.1136/bmj.c7452

Jolley, D., and Douglas, K. M. (2014). The Effects of Anti-vaccine Conspiracy Theories on Vaccination Intentions. *PLOS ONE* 9 (2), e89177. doi:10.1371/journal.pone.0089177

Li, X., Ostropolets, A., Makadia, R., Shaoibi, A., Rao, G., Sena, A. G., et al. (2021). Characterizing the Incidence of Adverse Events of Special Interest for COVID-19 Vaccines across Eight Countries: a Multinational Network Cohort Study. *medRxiv*, 2021. doi:10.1101/2021.03.25.21254315

Lieu, T. A., Kulldorff, M., Davis, R. L., Lewis, E. M., Weintraub, E., Yih, K., et al. (2007). Real-time Vaccine Safety Surveillance for the Early Detection of Adverse Events. *Med. Care* 45 (10 Suppl. 2), S89–S95. doi:10.1097/MLR.0b013e3180616c0a

Mahaux, O., Bauchau, V., and Van Holle, L. (2016). Pharmacoepidemiological Considerations in Observed-To-Expected Analyses for Vaccines. *Pharmacoepidemiol. Drug Saf.* 25 (2), 215–222. doi:10.1002/pds.3918

Mesfin, Y. M., Cheng, A., Lawrie, J., and Buttery, J. (2019). Use of Routinely Collected Electronic Healthcare Data for Postlicensure Vaccine Safety Signal Detection: a Systematic Review. *BMJ Glob. Health* 4 (4), e001065. doi:10.1136/bmjgh-2018-001065

OHDSI (2019). *The Book of OHDSI: Observational Health Data Sciences and Informatics*. OHDSI. Available at: https://ohdsi.github.io/TheBookOfOhdsi/TheBookOfOhdsi.pdf

Remschmidt, C., Wichmann, O., and Harder, T. (2015). Frequency and Impact of Confounding by Indication and Healthy Vaccinee Bias in Observational Studies Assessing Influenza Vaccine Effectiveness: a Systematic Review. *BMC Infect. Dis.* 15, 429. doi:10.1186/s12879-015-1154-y

Schuemie, M. J., Cepeda, M. S., Suchard, M. A., Yang, J., Tian, Y., Schuler, A., et al. (2020). How Confident Are We about Observational Findings in Healthcare: A Benchmark Study. *Harv. Data Sci. Rev.* 2 (1). doi:10.1162/99608f92.147cc28e

Schuemie, M. J., Hripcsak, G., Ryan, P. B., Madigan, D., and Suchard, M. A. (2018). Empirical Confidence Interval Calibration for Population-Level Effect Estimation Studies in Observational Healthcare Data. *Proc. Natl. Acad. Sci. U S A.* 115 (11), 2571–2577. doi:10.1073/pnas.1708282114

Schuemie, M. J., Hripcsak, G., Ryan, P. B., Madigan, D., and Suchard, M. A. (2016). Robust Empirical Calibration of P-Values Using Observational Data. *Stat. Med.* 35 (22), 3883–3888. doi:10.1002/sim.6977

Wijnans, L., Lecomte, C., de Vries, C., Weibel, D., Sammon, C., Hviid, A., et al. (2013). The Incidence of Narcolepsy in Europe: before, during, and after the Influenza A(H1N1)pdm09 Pandemic and Vaccination Campaigns. *Vaccine* 31 (8), 1246–1254. doi:10.1016/j.vaccine.2012.12.015

Winton Centre for Risk and Evidence Communication (2021). Communicating the Potential Benefits and Harms of the Astra-Zeneca COVID-19 Vaccine. 2021 [updated 7 April 2021. Available at: https://wintoncentre.maths.cam.ac.uk/news/communicating-potential-benefits-and-harms-astra-zeneca-covid-19-vaccine/.

Yih, W. K., Kulldorff, M., Fireman, B. H., Shui, I. M., Lewis, E. M., Klein, N. P., et al. (2011). Active Surveillance for Adverse Events: the Experience of the Vaccine Safety Datalink Project. *Pediatrics* 127 (Suppl. 1), S54–S64. doi:10.1542/peds.2010-1722I

Yih, W. K., Nordin, J. D., Kulldorff, M., Lewis, E., Lieu, T. A., Shi, P., et al. (2009). An Assessment of the Safety of Adolescent and Adult Tetanus-Diphtheria-Acellular Pertussis (Tdap) Vaccine, Using Active Surveillance for Adverse Events in the Vaccine Safety Datalink. *Vaccine* 27 (32), 4257–4262. doi:10.1016/j.vaccine.2009.05.036