



Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2018 February ; 27(2): 221–228. doi:10.1002/pds.4374.

Bias from outcome misclassification in immunization schedule safety research

Sophia R Newcomer, MPH^{1,2}, Martin Kulldorff, PhD³, Stan Xu, PhD¹, Matthew F Daley, MD^{1,4}, Bruce Fireman, MA⁵, Edwin Lewis, MPH⁵, and Jason M Glanz, PhD^{1,2}

¹Kaiser Permanente Colorado, Institute for Health Research

²University of Colorado Denver, School of Public Health, Department of Epidemiology

³Brigham and Women's Hospital and Harvard Medical School, Division of Pharmacoepidemiology and Pharmacoeconomics

⁴University of Colorado Denver, School of Medicine, Department of Pediatrics

⁵Kaiser Permanente Northern California, Division of Research, Vaccine Study Center

Abstract

Purpose—The Institute of Medicine recommended conducting observational studies of childhood immunization schedule safety. Such studies could be biased by outcome misclassification, leading to incorrect inferences. Using simulations, we evaluated 1) outcome positive predictive values (PPVs) as indicators of bias of an exposure-outcome association, and 2) quantitative bias analyses (QBA) for bias correction.

Methods—Simulations were conducted based on proposed or ongoing Vaccine Safety Datalink studies. We simulated 4 studies of 2 exposure groups (children with no vaccines or on alternative schedules) and 2 baseline outcome levels (100 and 1000/100,000 person-years), with 3 relative risk (RR) levels (RR=0.50, 1.00, and 2.00), across 1,000 replications using probabilistic modeling. We quantified bias from non-differential and differential outcome misclassification, based on levels previously measured in database research (sensitivity>95%; specificity>99%). We calculated median outcome PPVs, median observed RRs, Type 1 error, and bias-corrected RRs following QBA.

Results—We observed PPVs from 34%–98%. With non-differential misclassification and true RR=2.00, median bias was toward the null, with severe bias (median observed RR=1.33) with PPV=34% and modest bias (median observed RR=1.83) with PPV=83%. With differential misclassification, PPVs did not reflect median bias and there was Type 1 error of 100% with PPV=90%. QBA was generally effective in correcting misclassification bias.

Corresponding Author: ophia Newcomer, MPH, Kaiser Permanente Colorado, Institute for Health Research, 10065 E. Harvard Ave, Suite 300, Denver, CO, USA 80231, Phone: 303-614-1247, sophia.r.newcomer@kp.org.

Prior Presentations: Portions of this work were presented at the International Society for Pharmacoepidemiology Mid-Year Meeting, Baltimore, MD, April 10-12, 2016.

Conflict of Interest: The authors have no conflicts of interest to report.

Conclusions—In immunization schedule studies, outcome misclassification may be non-differential or differential to exposure. Overall outcome PPVs do not reflect the distribution of false positives by exposure and are poor indicators of bias in individual studies. Our results support QBA for immunization schedule safety research.

Keywords

bias (epidemiology); immunization; safety; electronic health records; database; sensitivity and specificity

INTRODUCTION

Large, linked databases, such as the Centers for Disease Control and Prevention's Vaccine Safety Datalink (VSD) and the Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring Program (PRISM), are important resources for post-market studies of vaccine safety.¹⁻³ These systems capture data on millions of individuals and billions of medical encounters from electronic health records (EHR) and medical billing claims.^{2,4} Clinical outcomes are identified with electronic data algorithms, which are typically individual or combinations of diagnosis codes.^{5,6}

In studies of acute vaccine adverse events, presumptive outcomes are identified in electronic data within short risk and control periods around vaccination. Misclassification of these presumptive outcomes has been a key challenge in this research.² Common reasons for misclassification include clinician miscoding and rule-out diagnoses.⁵ To avoid misclassification bias, researchers often chart review all presumed outcomes, and then re-analyze data with only confirmed outcomes.² The percent of presumed outcomes confirmed is typically reported as a positive predictive value (PPV) or confirmation rate.⁷ Vaccine safety studies have demonstrated considerable variability in outcome PPVs, ranging from 5% to 97%.^{8,9}

While there has been ample research on acute outcomes following vaccination, a 2013 Institute of Medicine (IOM) report called for studies of chronic outcomes, such as autoimmune and allergic diseases, following cumulative exposure to early childhood immunizations.¹⁰ The VSD has embarked on such studies of immunization schedule studies¹¹, but this research poses new challenges for addressing outcome misclassification bias. Unlike studies of acute outcomes, observation time will span years. It may not be feasible to adjudicate the hundreds or thousands of presumptive outcomes identified in electronic data.¹¹ Furthermore, outcome sensitivity is a concern, since for some non-acute conditions, parents may have varying propensity to seek care for their children, or may consult external providers not captured in electronic health records.¹²

In this study, we used simulations to evaluate one method for assessing and another method for correcting bias of an immunization schedule-outcome association due to outcome misclassification in EHR data. The first method involves using overall outcome PPVs to assess misclassification bias. PPVs are the most commonly reported measure of electronic algorithm validity in EHR-based research.¹³ For immunization schedule safety studies, researchers could validate a sample of presumptive outcomes and estimate an outcome PPV.

While it has been suggested that PPV levels >70% are sufficient for electronic-only data analysis¹⁴, the relation between overall outcome PPVs and bias of an exposure-outcome association has not been investigated for EHR-based vaccine schedule safety research. We also evaluated quantitative bias analyses (QBA), which are methods for correcting systematic error in epidemiological research.^{15,16} Traditional QBA formulas for outcome misclassification apply sensitivity and specificity estimates to a study's observed relative risk. Previous studies have reported the sensitivity and specificity for several chronic outcomes of interest in immunization schedule safety studies^{17–21}; these measures could be used as bias parameters in QBA.

For our primary objective, we evaluated whether outcome PPVs are effective indicators of bias of an immunization schedule-outcome association. To achieve this objective, we constructed simulations modeled on VSD studies that have been proposed or are ongoing, applied outcome misclassification levels previously measured in EHR data, and calculated the resulting misclassification bias and outcome PPVs. As a secondary objective, we tested the effectiveness of QBA for outcome misclassification within the same simulations.^{15,16} We examined both outcome misclassification that is independent of exposure (non-differential misclassification), and misclassification that systematically varies by exposure (differential misclassification).

METHODS

Study setting

We sought to have our simulations mimic actual VSD cohort studies of immunization schedule safety. To achieve this goal, we first identified a cohort of children born 2002–2012 from two managed care organizations (MCOs) participating in the VSD, Kaiser Permanente Colorado and Kaiser Permanente Northern California. We further limited the cohort to children continuously enrolled in their MCO from birth to their 2nd birthday, which is the period when early childhood immunizations are administered.²² We used actual birthdates and MCO enrollment time (i.e., person-time) in our analyses; all other data in this study were simulated.

Both MCOs' Institutional Review Boards approved this study; informed consent was not required.

Simulations

Within this VSD cohort, we constructed simulations of immunization schedule safety studies, where risks of chronic outcomes are compared between groups of children with different immunization patterns in early childhood (ages 0–2 years). Table 1 provides an overview of our simulations, relative risk (RR) levels, and outcome misclassification scenarios. We simulated outcomes with the formula^{23–25}:

$$p(\text{outcome}=1) = pt * \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1)}} \quad (1)$$

,where $p(\text{outcome}=1)$ is an individual's probability of having the outcome, pt is person-time contributed (in days) from each child's 2nd birthday to the first of MCO disenrollment or 8th birthday, β_0 is the log of baseline outcome rate (per day), β_1 is the log of the simulated RR, and X_1 indicates under-vaccination exposure. The second term in equation (1) represents the daily probability of experiencing an outcome. The probability of experiencing an outcome during the entire follow-up period is the product of the daily probability and pt . Simulated RR refers to the true RR representing the association between under-vaccination exposure and outcome, absent of any bias.

The 2013 IOM report requested studies comparing risk of chronic outcomes in children who receive no vaccines or are vaccinated per distinct alternative immunization schedules versus children fully-vaccinated per the U.S. Advisory Committee on Immunization Practices' (ACIP) recommended schedule.^{10,22} Therefore, for X_1 , we focused on children with no vaccines before their 2nd birthday and children whose parents choose distinct alternative schedules popularized in books or on the internet.²⁶⁻²⁸ Based on previous research, we simulated the prevalence of these two groups at 0.7% and 2.4%, respectively. The unexposed group was children fully-vaccinated per the ACIP schedule with a prevalence of 60.6%.²⁹ The remaining 36.3% of children are assumed to be missing some vaccine doses or are under-vaccinated at some point but get caught up before their 2nd birthday; these less distinct patterns of under-vaccination were not considered in this study.

We simulated two levels of baseline outcome incidence: 100 and 1000 outcomes per 100,000 person-years. We chose these rates to represent both rare (e.g., Type 1 diabetes, epilepsy) and more common (e.g., allergic conditions, asthma) conditions from a priority list of outcomes for VSD immunization schedule research.¹¹ Within each of four simulations (one for each of 2 exposure groups and 2 baseline outcome incidence levels), we separately simulated three levels of RR: 2.00, 1.00, and 0.50. For each simulation and RR level, we created 1,000 replicated datasets with a different random seed. Within each replication, exposure probabilities were applied to each child, and Bernoulli trials determined which children were assigned exposure to an under-vaccination pattern. Formula 1 was used to assign probability of outcome, and Bernoulli trials determined the outcome status for each child.

Misclassification

We reviewed published studies to identify ranges of likely outcome misclassification levels in VSD immunization schedule research. We identified validation studies of EHR-based algorithms for two outcomes of interest: asthma and Type 1 diabetes.¹¹ The best performing algorithms had sensitivity >95% and specificity >99%.^{17,18} For each simulation and RR, we made four copies of the 1,000 replicated datasets and tested two scenarios of non-differential and two scenarios of differential outcome misclassification based on these levels (Table 1, scenarios A–D). For one differential outcome misclassification scenario, we measured bias with lower outcome sensitivity among under-vaccinated children. Some parents who refuse vaccines express distrust in traditional medicine and may seek care outside the MCO¹², leading to decreased outcome sensitivity. We then measured bias from lower outcome specificity among under-vaccinated children. Clinicians may be more likely to suspect

infectious conditions in ill children who are under-vaccinated, which could lead to more “rule-out” diagnoses and higher false positive rates.

When testing the two non-differential outcome misclassification scenarios, we applied sensitivity and specificity levels from Table 1 to the simulated datasets without regard to exposure. For the two presentations of differential misclassification, sensitivity and specificity were applied separately by exposure. Bernoulli trials determined which children “flipped” to an outcome false positive or false negative status, representing EHR data misclassification.

Analysis

We calculated the observed RR, using Poisson regression with a log of person-time as the offset, within each replication. The observed RR is the immunization schedule exposure-outcome association estimated with outcome misclassification present. For each simulated immunization schedule safety study, simulated RR level, and misclassification scenario, we reported bias as the median observed RR with misclassification across replications.

For each simulation and RR level, we reported the median PPVs that resulted from each outcome misclassification scenario. When simulated RR = 1.00, we reported empirical power with and without outcome misclassification. We calculated empirical power as the percent of replications where the null hypothesis was rejected at $\alpha=0.05$ in the same direction as the simulated RR. We calculated Type 1 error for simulations with simulated RR=1.00 as the percent of replications with null hypothesis rejection at $\alpha=0.05$.

We tested the effectiveness of QBA using formulas for both QBA assuming non-differential outcome misclassification and assuming differential outcome misclassification (Appendix 1).^{15,16} To conduct the QBA, we determined the number of observed individuals that were exposed with outcome, exposed without outcome, unexposed with outcome, and unexposed without outcome in each simulated replication. We then applied QBA formulas with sensitivity and specificity measured from each replication and calculated the RR that would have been observed had misclassification not been present. We reported the median QBA-corrected RRs across replications.

All simulations and analyses were conducted using SAS 9.4®.

RESULTS

Across replications there were an average of $n=1,722$ children simulated to be completely unvaccinated (simulations #1 and #2), $n=6,117$ children simulated to be on a distinct alternative schedule (simulations #3 and #4), and $n=155,722$ children simulated to be adhering to the ACIP schedule (unexposed group in all simulations). We observed a range of bias across the simulations, levels of simulated (i.e., true) RR, and misclassification scenarios tested. Across simulations, overall median outcome PPVs ranged from 34%–98%. With non-differential misclassification, median bias across replications was towards the null and overall outcome PPVs were associated with the magnitude of median bias (Tables 2a–2b, scenarios A and B). For example, when true RR=2.00, the median observed RR was

1.33 when the median PPV=34%, and the median observed RR was 1.83 when the median PPV=85%. When the median PPV was 98%, there was virtually no bias of the median RR.

When outcome misclassification was differential to exposure, the direction of median bias varied and overall PPVs were not indicators of the direction or magnitude of median bias (Tables 2a–2b, scenarios C and D). For example, when true RR=2.00 and the baseline outcome rate was 1,000/100,000 person-years, the median observed RR was 1.65 when outcome sensitivity was lower among under-vaccinated children and 2.23 when specificity was lower among under-vaccinated children. In both scenarios, median overall outcome PPVs were 91% or 92%. In some simulations, differential misclassification caused extreme median bias. For example, in simulation #1, with lower specificity among under-vaccinated children (scenario D), the observed median RR was 2.72 when true RR=0.50, and median PPV=50%.

Power and Type 1 error

In some simulations, outcome misclassification led to reductions in power (Table 3). For example, for true RR=2.00 in simulation #1, power was reduced from 78% without any misclassification to 70% with non-differential specificity=99.9% and PPV=84% (scenario B), and further down to 40% power with non-differential specificity=99.0% and PPV=34% (scenario A). Change in power varied by true RR within the same differential misclassification scenario. For example, when specificity was lower among under-vaccinated children (scenario D), the higher outcome false positive rate in this exposed group led to consistent overestimation of the true effect. Therefore, when simulated RR=0.50, power was reduced from 100% without misclassification to 17% with misclassification, but power remained at 100% when RR=2.00.

Before misclassification was applied, Type 1 error was near 5% in all simulations where simulated RR=1.00. Non-differential outcome misclassification did not affect these rates (results not shown). However, differential misclassification led to Type 1 error up to 100%, and overall PPV levels were not associated with these rates (Table 4). For example, in simulations #3 and #4 with differential outcome sensitivity (scenario C), median PPV=51% and Type 1 error=7.1% when the baseline outcome rate was 100/100,000 person-years. In contrast, median PPV=92% and Type 1 error=67.8% when the baseline outcome rate was 1,000/100,000 person-years. While the higher false negative rate among under-vaccinated children led to bias toward observing a protective effect, the specificity of 99.5% more rapidly caused the observed effect back toward the null with the rarer outcome, leading to lower Type 1 error.

Quantitative bias analysis

In most simulations of non-differential misclassification, QBA assuming non-differential misclassification corrected bias (Tables 2a–2b, Table 4). When true RR=2.00, QBA assuming non-differential misclassification resulted in perfect or near-perfect bias correction. QBA assuming non-differential misclassification was also effective when true RR=0.50, except in Simulation #1. In that simulation with completely unvaccinated children and the rarer outcome, median bias-corrected RR was 0.67 with QBA assuming non-

differential misclassification. This median bias-corrected RR was closer to the true value of RR=0.50 than the median RR observed with misclassification of 0.84.

QBA using outcome sensitivity and specificity measurements by exposure group led to nearly perfect correction of bias in all simulations. Unsurprisingly, when misclassification was simulated to be differential by exposure, QBA assuming non-differential misclassification was ineffective, and sometimes resulted in estimates that were more biased than the estimates observed under misclassification.

DISCUSSION

While numerous studies have reported on misclassification in pharmacoepidemiological databases^{6,13,14,30,31}, there has been limited work in quantifying the bias that arises from such misclassification and on methods for correcting this bias. To our knowledge, there has been no prior work on evaluating and correcting misclassification bias in vaccine schedule research studies. Using simulations, we quantified a range of bias across plausible scenarios of non-differential and differential outcome misclassification. Our results suggest that rather than relying on overall PPVs as indicators of bias, quantitative methods should be used to account for misclassification bias.

Our results highlight several reasons why overall outcome PPVs may not be effective indicators of bias of an exposure-outcome association. First, while median PPVs were associated with the magnitude of median bias with non-differential misclassification, they were not associated with the magnitude or direction of median bias with differential misclassification. In immunization schedule research, differential outcome misclassification could occur due to parent or provider behavior.^{11,12,32} Even if non-differential outcome misclassification is presumed, exactly equal misclassification levels across exposure groups is not guaranteed.³³ Differential misclassification can lead to bias toward or away from the null, and the direction of bias is often unpredictable.^{34,35} Second, since predictive values are a function of three factors (specificity, sensitivity, and prevalence; see Appendix 2)^{5,36}, PPVs can fluctuate based on any of these factors. For example, we observed extreme differences in PPV, from 34% to 84%, from a 0.9% difference in specificity when outcome prevalence was low. Since overall outcome PPVs do not reflect the distribution of false positives by exposure (an underlying cause of bias of an exposure-outcome association), these metrics have limited utility for assessing bias in individual studies. Instead, quantitative methods to adjust for systematic error should be considered.

While QBA has been advocated as an essential tool in epidemiological research^{16,37}, these methods have had limited use in vaccine safety and EHR-based research.^{6,38,39,36} The tendency to underestimate misclassification error, along with a lack of practical examples, have been identified as barriers to implementing QBA.¹⁶ We addressed these barriers by measuring bias via simulation, and by evaluating the application of QBA in immunization schedule safety research. Our results showed that QBA was typically effective in correcting outcome misclassification bias. However, similar to previous findings by Johnson and colleagues, we found that QBA is vulnerable to assumptions made and bias parameters used.⁴⁰ For example, in simulation #1, with a rare exposure and rare outcome, QBA assuming

non-differential misclassification did not effectively correct misclassification bias when true $RR=0.50$, even though the underlying misclassification process was non-differential. This was due to differences in sensitivity and specificity by exposure that occurred due to chance.

For traditional QBA methods to be most effective in immunization schedule research, our results suggest that outcome sensitivity and specificity should be estimated by exposure. Since collecting such data may not be feasible, an alternative approach is to establish ranges of plausible sensitivity and specificity levels by exposure, and use probabilistic bias analysis to quantify a range of corrected RRs.^{15,39} Less-established quantitative bias approaches using predictive values are available.^{15,41,42} However, these methods come with challenges, including requiring positive and negative predictive values stratified by exposure¹⁵, or assuming non-differential outcome sensitivity.⁴¹ While we investigated traditional QBA methods for misclassification using sensitivity and specificity measures, evaluation of predictive value-based approaches is also warranted in EHR-based research.

Our study has several limitations. Our simulations did not incorporate all outcome misclassification levels encountered in immunization schedule safety research. Also, while our differential misclassification scenarios were based on practical concerns in immunization schedule research, it is unknown how different outcome sensitivity and specificity levels actually are across exposure groups, since these data have not been collected. As a result, our simulations may have over- or under-estimated the misclassification bias that may be encountered in this line of research. Moreover, our simulations only focused on bias from misclassification of binary outcomes. Exposure misclassification due to missing vaccine records is also a concern. Addressing measurement error of exposures, covariates, and of continuous or multi-level outcome variables, along with bias from unmeasured confounding, involves more complex bias analyses which merit further evaluation within EHR-based research.¹⁵ Finally, we only considered bias of risk ratios, and did not evaluate how outcome misclassification in EHR data would affect measures of risk difference.

Research using electronic databases are essential to U.S. vaccine safety, and improved methods for quantifying and communicating about uncertainty in this line of research are needed.⁴³ While our simulations were conducted in the context of immunization schedule safety research, our findings are broadly applicable to other EHR-based pharmacoepidemiological research. Our results serve to encourage researchers to acknowledge the potential for misclassification bias in EHR-based studies, and to use quantitative techniques for identifying, measuring, and correcting this bias.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sponsor: This study was funded by a grant from the National Institutes of Health, National Institute of Allergy and Infectious Diseases (NIH R01AI107721, "Methods for Safety Evaluation of Vaccine Schedules", Principal Investigator: Martin Kulldorff, PhD).

References

1. Baggs J, Gee J, Lewis E, et al. The Vaccine Safety Datalink: A model for monitoring immunization safety. *Pediatrics*. 2011; 127(Supplement 1):S45–S53. [PubMed: 21502240]
2. McNeil MM, Gee J, Weintraub ES, et al. The Vaccine Safety Datalink: Successes and challenges monitoring vaccine safety. *Vaccine*. 2014; 32(42):5390–5398. [PubMed: 25108215]
3. Nguyen M, Ball R, Midthun K, Lieu TA. The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring program: Strengthening the federal vaccine safety enterprise. *Pharmacoepidemiology and Drug Safety*. 2012; 21(S1):291–297. [PubMed: 22262619]
4. Platt R, Carnahan R. The US Food and Drug Administration's Mini-Sentinel Program. *Pharmacoepidemiology and Drug Safety*. 2012; 21(S1):1–303.
5. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *Journal of Clinical Epidemiology*. 2012; 65(3):343–349.e342. [PubMed: 22197520]
6. Lanes S, Brown JS, Haynes K, Pollack MF, Walker AM. Identifying health outcomes in healthcare databases. *Pharmacoepidemiology and Drug Safety*. 2015; 24(10):1009–1016. [PubMed: 26282185]
7. Xu S, Newcomer S, Nelson J, et al. Signal detection of adverse events with imperfect confirmation rates in vaccine safety studies using self-controlled case series design. *Biometrical Journal*. 2014; 56(3):513–525. [PubMed: 24402780]
8. McNeil MM, Weintraub ES, Duffy J, et al. Risk of anaphylaxis after vaccination in children and adults. *Journal of Allergy and Clinical Immunology*. 2015
9. Shui IM, Shi P, Dutta-Linn MM, et al. Predictive value of seizure ICD-9 codes for vaccine safety research. *Vaccine*. 2009; 27(39):5307–5312. [PubMed: 19616500]
10. Institute of Medicine. *The Childhood Immunization Schedule and Safety: Stakeholder Concerns, Scientific Evidence, and Future Studies*. Washington, DC: The National Academies Press; 2013. <https://doi.org/10.17226/13563>
11. Glanz JM, Newcomer SR, Jackson ML, et al. White Paper on studying the safety of the childhood immunization schedule in the Vaccine Safety Datalink. *Vaccine*. 2016; 34:A1–A29. [PubMed: 26830300]
12. Downey L, Tyree PT, Huebner CE, Lafferty WE. Pediatric vaccination and vaccine-preventable disease acquisition: associations with care by complementary and alternative medicine providers. *Maternal and Child Health Journal*. 2010; 14(6):922–930. [PubMed: 19760163]
13. van Walraven C, Bennett C, Forster AJ. Administrative database research infrequently used validated diagnostic or procedural codes. *Journal of Clinical Epidemiology*. 2011; 64(10):1054–1059. [PubMed: 21474278]
14. Carnahan RM. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. *Pharmacoepidemiology and Drug Safety*. 2012; 21(S1):90–99.
15. Lash, TL., Fox, MP., Fink, AK. *Applying quantitative bias analysis to epidemiologic data*. Springer Science & Business Media; 2011.
16. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *International Journal of Epidemiology*. 2014; dyu149.
17. Wakefield DB, Cloutier MM. Modifications to HEDIS and CSTE algorithms improve case recognition of pediatric asthma. *Pediatric Pulmonology*. 2006; 41(10):962–971. [PubMed: 16871628]
18. Lawrence JM, Black MH, Zhang JL, et al. Validation of pediatric diabetes case identification approaches for diagnosed cases by using information in the electronic health records of a large integrated managed health care organization. *American Journal of Epidemiology*. 2013; kwt230.
19. Cherepanov D, Raimundo K, Chang E, et al. Validation of an ICD-9–based claims algorithm for identifying patients with chronic idiopathic/spontaneous urticaria. *Annals of Allergy, Asthma & Immunology*. 2015; 114(5):393–398.
20. Chung CP, Rohan P, Krishnaswami S, McPheeters ML. A systematic review of validated methods for identifying patients with rheumatoid arthritis using administrative or claims data. *Vaccine*. 2013; 31:K41–K61. [PubMed: 24331074]

21. Zhong VW, Obeid JS, Craig JB, et al. An efficient approach for surveillance of childhood diabetes by type derived from electronic health record data: the SEARCH for Diabetes in Youth Study. *Journal of the American Medical Informatics Association*. 2016;ocv207.
22. Robinson CL, Advisory Committee on Immunization Practices (ACIP), ACIP Child/Adolescent Immunization Work Group. Advisory committee on immunization practices recommended immunization schedules for persons aged 0 through 18 years—United States, 2016. *MMWR*. 2016; 65(4):86–87. [PubMed: 26845283]
23. Glanz JM, McClure DL, Xu S, et al. Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations. *Journal of Clinical Epidemiology*. 2006; 59(8):808–818. [PubMed: 16828674]
24. Xu S, Zeng C, Newcomer S, Nelson J, Glanz J. Use of fixed effects models to analyze self-controlled case series data in vaccine safety studies. *Journal of Biometrics & Biostatistics*. 2012;006. [PubMed: 24707443]
25. Xu S, Zhang L, Nelson JC, et al. Identifying optimal risk windows for self-controlled case series studies of vaccine safety. *Statistics in Medicine*. 2011; 30(7):742–752. [PubMed: 21394750]
26. Dempsey AF, Schaffer S, Singer D, Butchart A, Davis M, Freed GL. Alternative vaccination schedule preferences among parents of young children. *Pediatrics*. 2011;2011–0400. peds.
27. Robison SG, Groom H, Young C. Frequency of alternative immunization schedule use in a metropolitan area. *Pediatrics*. 2012; 130(1):32–38. [PubMed: 22711719]
28. Sears, RW. *The Vaccine Book: Making the right decision for your child*. Little, Brown; 2011.
29. Daley MF, Glanz JM, Newcomer SR, et al. Assessing misclassification of vaccination status: implications for studies of the safety of the childhood immunization schedule. *Vaccine*. 2017; 35(15):1873–1878. [PubMed: 28285983]
30. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *Journal of Clinical Epidemiology*. 2011; 64(8):821–829. [PubMed: 21194889]
31. Baker MA, Nguyen M, Cole DV, Lee GM, Lieu TA. Post-licensure rapid immunization safety monitoring program (PRISM) data characterization. *Vaccine*. 2013; 31:K98–K112. [PubMed: 24331080]
32. Glanz JM, Newcomer SR, Narwaney KJ, et al. A population-based cohort study of undervaccination in 8 managed care organizations across the United States. *JAMA Pediatrics*. 2013; 167(3):274–281. [PubMed: 23338829]
33. Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *International Journal of Epidemiology*. 2005; 34(3):680–687. [PubMed: 15802377]
34. Jurek AM, Greenland S, Maldonado G. Brief Report: How far from non-differential does exposure or disease misclassification have to be to bias measures of association away from the null? *International Journal of Epidemiology*. 2008; 37(2):382–385. [PubMed: 18184671]
35. Rothman, KJ., Greenland, S., Lash, TL. *Modern Epidemiology*. Lippincott Williams & Wilkins; 2008.
36. Rosner, B. *Fundamentals of Biostatistics*. Nelson Education; 2015.
37. Greenland S. Basic methods for sensitivity analysis of biases. *International Journal of Epidemiology*. 1996; 25(6):1107–1116. [PubMed: 9027513]
38. Jonsson-Funk M, Landi S. Misclassification in administrative claims data: quantifying the impact on treatment effect estimates. *Current Epidemiology Reports*. 2014; 1(4):175–185. [PubMed: 26085977]
39. Hunnicutt JN, Ulbricht CM, Chrysanthopoulou SA, Lapane KL. Probabilistic bias analysis in pharmacoepidemiology and comparative effectiveness research: a systematic review. *Pharmacoepidemiology and Drug Safety*. 2016
40. Johnson CY, Flanders WD, Strickland MJ, Honein MA, Howards PP. Potential sensitivity of bias analysis results to incorrect assumptions of nondifferential or differential binary exposure misclassification. *Epidemiology*. 2014; 25(6):902–909. [PubMed: 25120106]

41. Brenner H, Gefeller O. Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. *American Journal of Epidemiology*. 1993; 138(11):1007–1015. [PubMed: 8256775]
42. Cai B, Hennessy S, Lo Re V, Small DS. Epidemiologic research using probabilistic outcome definitions. *Pharmacoepidemiology and Drug Safety*. 2015; 24(1):19–26. [PubMed: 25257346]
43. Lash TL, Fox MP, Cooney D, Lu Y, Forshee RA. Quantitative bias analysis in regulatory settings. *American Journal of Public Health*. 2016; 106(7):1227–1230. [PubMed: 27196652]

Key Points

- Bias from misclassification of outcomes in electronic health record data is a methodological challenge in immunization schedule safety studies.
- Using simulations, we evaluated outcome positive predictive values (PPVs) as indicators of bias of an immunization schedule-outcome association, and quantitative bias analyses for addressing this bias.
- While outcome PPVs reflected the magnitude of median bias with non-differential outcome misclassification, these metrics were not effective indicators of median bias with differential misclassification.
- With differential outcome misclassification, Type 1 error rates of 100% were observed with outcome PPVs of 90%.
- Quantitative bias analysis was effective in correcting outcome misclassification bias and should be considered in immunization schedule research.

Table 1

Description of simulations, levels of relative risk, and outcome misclassification

Simulated immunization schedule safety study*			Levels of simulated relative risk (RR)	Outcome misclassification scenarios <i>Sensitivity (SN) and specificity (SP) levels applied to each simulation and level of RR</i>
Simulation Number	Exposure Group [†] <i>Reference group: Children vaccinated per the U.S. recommended immunization schedule</i>	Baseline Outcome Rate <i>Outcome rate in reference group</i>		
1	Unvaccinated(no vaccines)	100/100,000 person-years	RR= 0.50 RR= 1.00 RR= 2.00	A: Non-differential outcome misclassification: SN=97.5%, SP=99.0% B: Non-differential outcome misclassification, with higher outcome specificity: SN=97.5%, SP=99.9% C: Differential outcome misclassification, with lower outcome sensitivity among exposed [‡] : SN ₁ =85.0%, SN ₀ =99.0%, SP ₁ =99.5%, SP ₀ =99.5% D: Differential outcome misclassification, with lower outcome specificity among exposed [‡] : SN ₁ =97.5%, SN ₀ =97.5%, SP ₁ =97.5%, SP ₀ =99.5%
2	Unvaccinated(no vaccines)	1,000/100,000 person-years		
3	Distinct alternative immunization schedules	100/100,000 person-years		
4	Distinct alternative immunization schedules	1,000/100,000 person-years		

Abbreviations: RR=relative risk; SN=overall sensitivity; SP=overall specificity; SN₁=sensitivity among exposed; SN₀=sensitivity among unexposed; SP₁=specificity among exposed; SP₀=specificity among unexposed

* 1,000 replicated datasets were generated for each simulation and relative risk level using probabilistic modeling. Four copies of these 1,000 datasets were generated to test the two non-differential and two differential outcome misclassification processes.

[†]We simulated the prevalence of completely unvaccinated children to be 0.7% and children on distinct alternative immunization schedules to be 2.4%. We simulated the unexposed group to be children fully-vaccinated per the U.S. Advisory Committee on Immunization Practices' recommended schedule, with an estimated prevalence of 60.6%. The remaining 36.3% of children are assumed to be missing some vaccine doses or are under-vaccinated at some point but get caught up on vaccines before their 2nd birthday, and these under-vaccination patterns were not considered in this study.

[‡]The lower sensitivity (scenario C) and specificity (scenario D) levels occur in a rare exposure group. Since overall outcome sensitivity and specificity levels will be a weighted average of these levels from across the exposed and unexposed groups, the overall observed sensitivity and specificity will be closer to the value in the more common unexposed group.

Tables 2a and 2b

Median outcome positive predictive value, observed relative risk (RR) with misclassification, and bias-corrected RRs with quantitative bias analysis

a. Simulated RR=0.50					
Outcome misclassification scenarios	Simulation Number		Unvaccinated children	Distinct alternative immunization schedules	
	1	2		3	4
A: <i>Nondifferential misclassification</i> SN = 97.5% SP = 99.0%	Exposure Group	100	1,000	100	1,000
	Baseline outcome rate per 100,000 person-years	821	8,206	821	8,206
	Average # of baseline outcomes across replications	34%	84%	34%	84%
	Outcome PPV	0.84	0.58	0.83	0.58
	Observed RR	0.67	0.50	0.51	0.50
	QBA-corrected RR, assuming non-differential misclassification	0.48	0.50	0.50	0.50
	QBA-corrected RR, assuming differential misclassification	84%	98%	84%	98%
	Outcome PPV	0.57	0.51	0.58	0.51
	Observed RR	0.50	0.50	0.50	0.50
	QBA-corrected RR, assuming non-differential misclassification	0.49	0.50	0.50	0.50
B: <i>Nondifferential misclassification</i> SN = 97.5% SP = 99.9%	Exposure Group	51%	92%	51%	92%
	Baseline outcome rate per 100,000 person-years	0.71	0.48	0.70	0.48
	Average # of baseline outcomes across replications	0.53	0.43	0.42	0.43
	Outcome PPV	0.48	0.50	0.50	0.50
	Observed RR	0.50	0.50	0.50	0.50
	QBA-corrected RR, assuming non-differential misclassification	50%	92%	47%	90%
	QBA-corrected RR, assuming differential misclassification	2.73	0.89	2.73	0.89
	Outcome PPV	4.58	0.88	5.04	0.88
	Observed RR	0.48	0.50	0.50	0.50
	QBA-corrected RR, assuming non-differential misclassification	0.48	0.50	0.50	0.50
C: <i>Differential sensitivity</i> SN ₁ =85.0%, SN ₀ =99.0%, SP ₁ =99.5%, SP ₀ =99.5%	Exposure Group	50%	92%	47%	90%
	Baseline outcome rate per 100,000 person-years	2.73	0.89	2.73	0.89
	Average # of baseline outcomes across replications	4.58	0.88	5.04	0.88
	Outcome PPV	0.48	0.50	0.50	0.50
	Observed RR	0.48	0.50	0.50	0.50
	QBA-corrected RR, assuming non-differential misclassification	0.48	0.50	0.50	0.50
	QBA-corrected RR, assuming differential misclassification	0.48	0.50	0.50	0.50
	Outcome PPV	0.48	0.50	0.50	0.50
	Observed RR	0.48	0.50	0.50	0.50
	QBA-corrected RR, assuming non-differential misclassification	0.48	0.50	0.50	0.50
D: <i>Differential specificity</i> SN ₁ =97.5%, SN ₀ =97.5%, SP ₁ =97.5%, SP ₀ =99.5%	Exposure Group	34%	85%	35%	85%
	Baseline outcome rate per 100,000 person-years	100	1,000	100	1,000
	Average # of baseline outcomes across replications	821	8,206	821	8,206
	Outcome PPV	0.48	0.50	0.50	0.50
	Observed RR	0.48	0.50	0.50	0.50
	QBA-corrected RR, assuming non-differential misclassification	0.48	0.50	0.50	0.50
	QBA-corrected RR, assuming differential misclassification	0.48	0.50	0.50	0.50
	Outcome PPV	0.48	0.50	0.50	0.50
	Observed RR	0.48	0.50	0.50	0.50
	QBA-corrected RR, assuming non-differential misclassification	0.48	0.50	0.50	0.50
b. When simulated RR=2.00					
Outcome misclassification scenarios	Simulation Number		Unvaccinated children	Distinct alternative immunization schedules	
	1	2		3	4
A: <i>Nondifferential misclassification</i>	Exposure Group	100	1,000	100	1,000
	Baseline outcome rate per 100,000 person-years	821	8,206	821	8,206
	Average # of baseline outcomes across replications	34%	85%	35%	85%
	Outcome PPV	0.48	0.50	0.50	0.50

b. When simulated RR=2.00		1		2		3		4	
Outcome misclassification scenarios		Unvaccinated children		Distinct alternative immunization schedules					
Simulation Number									
Exposure Group									
Baseline outcome rate per 100,000 person-years		100	1,000	100	1,000	100	1,000	100	1,000
Average # of baseline outcomes across replications		821	8,206	821	8,206	821	8,206	821	8,206
Observed RR		1.33	1.83	1.34	1.84	1.34	1.84	1.34	1.84
QBA-corrected RR, assuming non-differential misclassification		1.99	2.00	2.00	2.00	2.00	2.00	2.00	2.00
QBA-corrected RR, assuming differential misclassification		2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Outcome PPV		84%	98%	84%	98%	84%	98%	84%	98%
Observed RR		1.83	1.98	1.83	1.98	1.83	1.98	1.83	1.98
QBA-corrected RR, assuming non-differential misclassification		1.99	2.00	2.00	2.00	2.00	2.00	2.00	2.00
QBA-corrected RR, assuming differential misclassification		2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Outcome PPV		51%	92%	52%	92%	52%	92%	52%	92%
Observed RR		1.35	1.65	1.37	1.65	1.37	1.65	1.37	1.65
QBA-corrected RR, assuming non-differential misclassification		1.68	1.72	1.73	1.72	1.73	1.72	1.73	1.72
QBA-corrected RR, assuming differential misclassification		2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Outcome PPV		50%	91%	48%	91%	48%	91%	48%	91%
Observed RR		3.46	2.23	3.46	2.23	3.46	2.23	3.46	2.23
QBA-corrected RR, assuming non-differential misclassification		6.10	2.36	6.71	2.37	6.71	2.37	6.71	2.37
QBA-corrected RR, assuming differential misclassification		2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00

Abbreviations: SN=overall sensitivity; SP=overall specificity; SN₀=sensitivity among unexposed; SN₁=sensitivity among exposed; SP₀=specificity among unexposed; SP₁=specificity among exposed; PPV=positive predictive value; RR=relative risk; QBA=quantitative bias analysis

Table 3Empirical power with and without outcome misclassification[†]

Simulation Number	1	2	3	4
Exposure Group	Unvaccinated children		Distinct alternative immunization schedules	
Baseline outcome rate per 100,000 person-years	100	1,000	100	1,000
Simulated RR=0.50				
Power without misclassification	22%	100%	85%	100%
Power with misclassification:				
A: Nondifferential misclassification SN = 97.5%, SP = 99.0%	9%	100%	35%	100%
B: Nondifferential misclassification SN = 97.5%, SP = 99.9%	19%	100%	75%	100%
C: Differential sensitivity SN ₁ =85.0%, SN ₀ =99.0%, SP ₁ =99.5%, SP ₀ =99.5%,	17%	100%	65%	100%
D: Differential specificity SN ₁ =97.5%, SN ₀ =97.5%, SP ₁ =97.5%, SP ₀ =99.5%,	0%	17%	0%	52%
Simulated RR=2.00				
Power without misclassification	78%	100%	100%	100%
Power with misclassification:				
A: Nondifferential misclassification SN = 97.5%, SP = 99.0%	40%	100%	86%	100%
B: Nondifferential misclassification SN = 97.5%, SP = 99.9%	70%	100%	99%	100%
C: Differential sensitivity SN ₁ =85.0%, SN ₀ =99.0%, SP ₁ =99.5%, SP ₀ =99.5%,	34%	100%	77%	100%
D: Differential specificity SN ₁ =97.5%, SN ₀ =97.5%, SP ₁ =97.5%, SP ₀ =99.5%,	100%	100%	100%	100%

Abbreviations: SN=overall sensitivity; SP=overall specificity; SN₁=sensitivity among exposed; SN₀=sensitivity among unexposed; SP₁=specificity among exposed; SP₀=specificity among unexposed; RR=relative risk

[†] Empirical power was calculated as the percent of simulated replications where the null hypothesis was rejected at alpha=0.05 in the same direction as the simulated RR (i.e., the observed RR is >1.0 for simulated RR=2.0 and <1.0 for simulated RR=0.50).

Median outcome positive predictive value, observed relative risk (RR) with misclassification, Type 1 error rate, and bias-corrected RRs with quantitative bias analysis, when simulated RR=1.00 and misclassification is differential to exposure

Table 4

Outcome misclassification scenario	Simulation Number		2	3	4
	Exposure Group	Unvaccinated children			
C: <i>Differential sensitivity</i> SN ₁ =85.0%, SN ₀ =99.0%, SP ₁ =99.5%, SP ₀ =99.5%	Baseline outcome rate per 100,000 person-years	100	1,000	100	1,000
	Type 1 error rate without misclassification	4.6%	4.0%	5.9%	4.2%
	Outcome PPV	51%	92%	51%	92%
	Observed RR	0.91	0.87	0.92	0.87
	Type 1 error rate with misclassification	3.6%	22.2%	7.1%	67.8%
	QBA-corrected RR, assuming non-differential misclassification	0.85	0.86	0.85	0.85
	QBA-corrected RR, assuming differential misclassification	1.00	1.00	1.00	1.00
	Outcome PPV	50%	91%	47%	90%
	Observed RR	2.94	1.34	2.97	1.34
	Type 1 error rate with misclassification	100%	89.5%	100%	100%
D: <i>Differential specificity</i> SN ₁ =97.5%, SN ₀ =97.5%, SP ₁ =97.5%, SP ₀ =99.5%	QBA-corrected RR, assuming non-differential misclassification	5.04	1.37	5.58	1.38
	QBA-corrected RR, assuming differential misclassification	1.00	1.00	1.00	1.00

Abbreviations: SN=overall sensitivity; SP=overall specificity; SN₁=sensitivity among unexposed; SN₀=sensitivity among exposed; SP₁=specificity among unexposed; SP₀=specificity among exposed; PPV=positive predictive value; RR=relative risk; QBA=quantitative bias analysis