npj | vaccines

Article

Published in partnership with the Sealy Institute for Vaccine Sciences

https://doi.org/10.1038/s41541-024-00955-4

Applying causal inference and Bayesian statistics to understanding vaccine safety signals using a simulation study

Evelyn Tay[1] ✉, Michael Dymock[1], Laura Lopez[2], Catherine Glover[2], Yuanfei Anny Huang[2], K. Shuvo Bakar[3,4], Thomas Snelling[1,3,4], Julie A. Marsh[1,5] & Yue Wu[3,4]

Community perception of vaccine safety influences vaccine uptake. Our objective was to assess current vaccine safety monitoring by examining factors that may influence the availability of post-vaccination survey data, and thereby the specificity and sensitivity of existing signal detection methods. We used causal directed acyclic graphs (DAGs) and a Bayesian posterior predictive analysis (PPA) signal detection method to understand biological and behavioural factors which may influence signal detection. The DAGs informed the data simulated for scenarios in which these factors were varied. The influence of biological factors such as severity of adverse reactions and behavioural factors such as healthcare-seeking behaviour upon survey participation was found to drive signal detection. Where there was a low prevalence of moderate to severe reactions, false signals were detected when there was a strong influence of reaction severity on both survey participation and seeking medical attention. These findings provide implications for future vaccine safety monitoring.

Routinely administered vaccines are safe, but mild reactions are common and serious reactions occasionally occur, even for vaccines with otherwise excellent safety profiles. In 2010, an influenza vaccine formulation was associated with an increased risk of febrile convulsions in Australian children under 5 years old[1]; this led to the discontinued use of the particular formulation in children, but not before a temporary suspension of all influenza vaccination in young children, affecting public confidence in childhood immunisation, which reduced vaccine coverage[2]. Low vaccine coverage increases the risk of infections caused by vaccine preventable diseases[3]. There is an ongoing need to monitor, detect and address potential vaccine safety issues as soon as possible after they emerge, which is essential to public confidence in vaccination.

Monitoring can be performed for either solicited (active) or unsolicited (passive) reports of adverse events following immunisation (AEFI), which are undesirable clinical events occurring after the administration of a vaccine, irrespective of whether any causal relationship with the vaccine exists[4]. AEFI are graded according to their severity[5] ranging from mild AEFI that do not interfere with a person's activity to severe AEFI preventing normal activity and/or requiring medical attention. Australia's active vaccine safety surveillance system AusVaxSafety monitors the frequency of solicited acute AEFI for vaccines delivered through the national immunisation programme[6]. Since 2014, data have been collected through AEFI surveys sent to vaccine recipients via SMS or email 3 days after vaccination. Reports of seeking medical advice or attention are taken to be an indicator of severity, although healthcare seeking may also be influenced by vaccine concerns raised in the media and other procedural factors. The rate of survey-reported medically attended AEFI is monitored to identify potentially important vaccine safety issues that might require the suspension of a vaccine programme pending detailed investigation. In 2021, AEFI surveillance for the newly developed and rapidly deployed Coronavirus 2019 (COVID-19) vaccines required an extension of the AusVaxSafety post-vaccination surveys, including reports of any medical care or medical advice sought (attendance at a primary care clinic or emergency department or telehealth advice), the impact of reported AEFI on daily activities, and the presence of underlying health conditions[7].

A vaccine 'safety signal' is an unexpected (or unexpectedly frequent) association between a specific AEFI and a specific vaccine which requires investigation into whether a clinically important causal relationship exists[8]. To detect safety signals, AusVaxSafety employs a posterior predictive analysis (PPA) method; the PPA method is based on the posterior predicted distribution from a Bayesian logistic model adjusted for age, sex and co-morbidities[9]. The PPA method depends upon solicited reports of medical

[1]Wesfarmers Centre of Vaccines and Infectious Diseases, The Kids Research Institute Australia, Nedlands, WA, Australia. [2]National Centre for Immunisation Research and Surveillance, Westmead, NSW, Australia. [3]School of Public Health, University of Sydney, Sydney, NSW, Australia. [4]Sydney Children's Hospital Network, Westmead, NSW, Australia. [5]Centre for Child Health Research, University of Western Australia, Nedlands, WA, Australia. ✉e-mail: evelyn.tay@thekids.org.au

attentions (MAs); this, in turn, depends on survey participation which can vary across demographic groups and over time, leading to variable response rates. For example, over the 1st year of the Australian COVID-19 vaccine programme, the survey response rate declined from ~70% to 30% (unpublished data).

The challenge of vaccine safety monitoring is to identify sensitive and timely safety signals that unveil true vaccine safety issues, while minimising the frequency of false detections. To do this, survey-based vaccine safety monitoring systems must correctly interpret AEFI data from the subset of responders that exhibits non-random missingness. For example, responder rates differ between age and sex, and potentially between those who experience an AEFI and those who do not. Structured patterns of missing data may occur when vaccines are rolled out at different times to selected higher risk or specific age groups. A limitation of routine analytic approaches is the failure to properly account for these patterns of missingness in the data. Causal directed acyclic graphs (DAGs) are increasingly used to model data-generating processes in applied health research[10]. Causal DAGs depict and communicate one's understanding of complex problem domains or hypotheses, which can subsequently guide the analysis and modelling assumptions[11]. For example, behavioural researchers have applied causal and statistical modelling techniques to understand the generalisability of samples for cross-cultural comparison[12]. In order to correctly infer the true AEFI rates among the vaccinated population from the survey responder data and thereby improve the sensitivity and specificity of signal detection methods, some insight is required into the frequency of AEFI among survey non-responders.

In this paper, we approached this problem using causal DAGs created through expert elicitation to derive an assumed data-generating process, from which we simulated AusVaxSafety survey data collected under a range of scenarios. The synthetic scenarios modelled a range of biological and behavioural factors that plausibly influence the frequency of actual and reported MA following immunisation. We applied the PPA method to the simulated data under these scenarios and assessed its performance in flagging a safety signal for MA in survey responders in relation to the true rate of MA in both survey responders and non-responders in the simulated data. Using these methods, we quantified how changes in important biological and behavioural factors could influence the reporting of MAs and affect the performance of signal detection using the PPA method. Our chief objective was to assess the monitoring of vaccine safety by first understanding some of the factors affecting the sensitivity and specificity of existing signal detection methods. We obtained insights into the value of the causal DAG to account for survey non-response, to guide understanding of short-term vaccine safety, interpret the results of the PPA analysis under plausible scenarios, and review implications for future vaccine safety monitoring.

Results

Causal models

Figure 1 presents the full DAG which consists of 37 variables. Nine variables (blue nodes) depict the background risk factors for vaccine recipients. Six variables (green nodes) depict key events initiated in the health system (e.g., the distribution of vaccines). The spectrum of expected AEFI is depicted by seven variables (pink nodes), varying in expected frequency from common to rare (e.g., 16% for fever[13] and 2.7 events out of 100,000 persons for chest pain[14], following vaccination with BNT162b2). Background variables and AEFI together drive the vaccine recipient's perceived seriousness of an AEFI, which subsequently drives one's overall level of concern regarding an AEFI. The level of concern can be further influenced by a recipient's demographic background, which vaccine they received, and any contemporary factors that increase community concerns about vaccination in general, or the particular vaccine (e.g., news about vaccine safety issues). The AEFI, its perceived seriousness, and the recipient's level of concern, together drive the recipient's actions including whether they seek medical attention and/or report the AEFI if surveyed. There are 11 actions (purple nodes) and 2 temporary factors (yellow nodes) modelled in the full DAG to illustrate the problem domain. See Supplementary Table 1 for the definition of each variable and a detailed description of the DAG structure. At a high level, there are three main processes that collectively culminate in the ascertainment of an MA report: (1) the vaccine reaction (biological process), (2) the seeking of medical attention (behavioural process), and (3) the data capture process (procedural process). There are many potential interactions among these processes. The key variables extracted to form the simplified DAG are highlighted in blue text.

The simplified DAG (Fig. 2) contains the key variables that we considered to adequately capture the full causal model for the purpose of

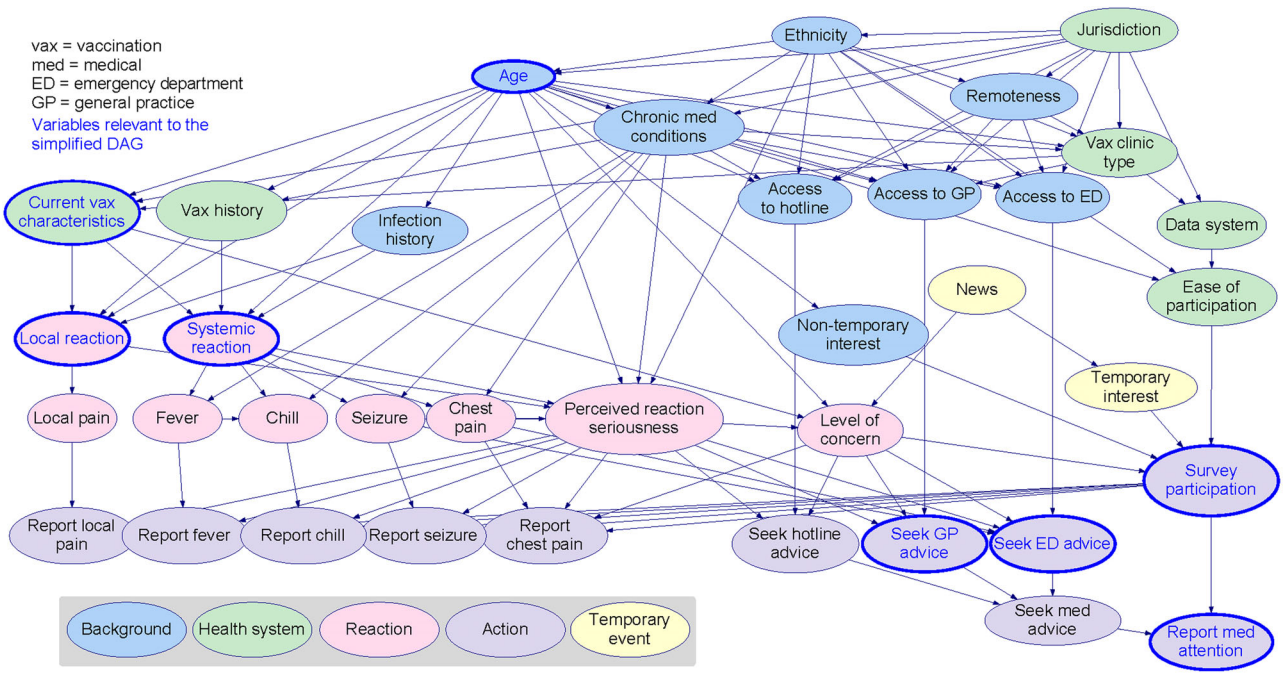

Fig. 1 | The full DAG. The figure depicts how the AusVaxSafety active surveillance system for short-term AEFI monitoring operates as a complex system. See Supplementary Table 1 for the definition of each variable and a detailed description of the DAG structure.

subsequent investigations, namely the person's age at vaccination (age), vaccine received (vaccine), whether they experience reaction within 3 days following vaccination (reaction), whether they seek medical attention (MA), whether they respond to, or participate in the survey (survey participation), and whether they report seeking medical attention (report MA). These represent a minimum set of variables required to investigate how distinct survey response behaviour in subgroups (e.g., different age groups) can interact with the underlying biological process of interest (vaccine reaction, MA) to affect the sensitivity and specificity of the PPA method when applied to observational data (reporting MA via survey participation). In the next section 'Data simulation and scenario design', we describe the design and underlying rationale for each scenario and the assumed data generation process. We assumed that the method of survey collection is constant and external influences are time invariant.

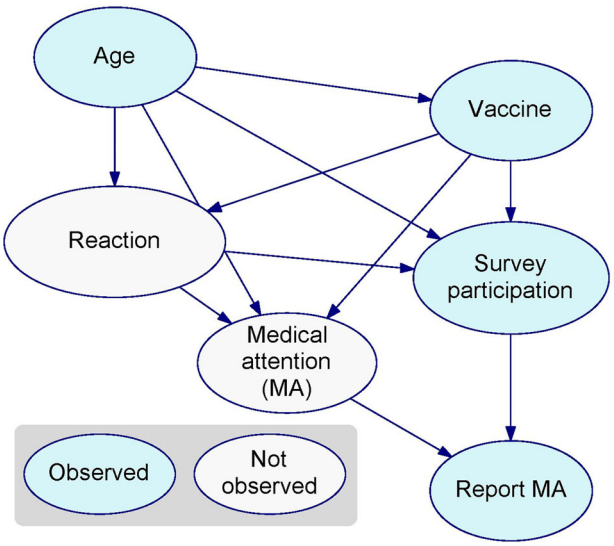

Fig. 2 | The simplified DAG. The figure depicts the relationship of a minimum set of variables required to investigate how biological, behavioural and procedural processes can interact to influence the probability of a reported MA using PPA method. See Table 1 for the definition of each variable and see Supplementary Information: Section A for how each variable is aligned with the full DAG.

Data simulation and scenario design

We simulated the five variables represented in the simplified DAG using eight parameters as defined in Table 1. The vaccine variable was not explicitly included in the statistical model as a single type of hypothetical vaccine was considered. For survey responders, only age and reported MA can be observed and the other factors are not observed. In the simulations, these parameters and their values were chosen based on several key assumptions. In brief, we assumed that the more severe an AEFI, the more likely a person is to seek medical attention and to respond to the survey. Compared to older people, younger people are assumed to be more likely to have a moderate or severe reaction[13,15], but less likely to respond to a health survey[16,17], and less likely to seek medical attention if they experience an AEFI[18]. We sampled age from a truncated Gaussian distribution with a mean of 43.5 years and a standard deviation of 18.6 years, informed by the age distribution of vaccine recipients observed by AusVaxSafety during the COVID-19 vaccine roll out in Australia. Table 1 defines variables in the simplified DAG, with parameters designed to generate data for each variable using Monte Carlo methods. Parameter values are chosen to describe a scenario for a relatively low prevalence of moderate to severe reaction, low survey participation and a weak influence of moderate to severe reaction on both survey participation and MA. We subsequently used this as a *Reference Scenario* for PPA investigation. See Supplementary Table 3 for further details about how these parameters were used to generate the event probabilities in the data simulations, including the probability of a participant reporting MA.

To facilitate the investigation of how combinations of these factors influence signal detection, we simulated data for the *Reference Scenario* consisting of 50,000 hypothetical vaccine recipients representing accumulated safety data to date, and 12 *Investigation Scenarios* consisting of 4000 hypothetical vaccine recipients representing a typical number of surveys issued to vaccine recipients in a 2-week investigation period. In contrast with the *Reference Scenario* of a relatively low probability of moderate to severe reaction (*Low R*), low survey participation (*Low P*), and a weak influence of reaction on survey participation and MA (*Weak*), we altered the value of four parameters outlined in Table 2 for each *Investigation Scenario*. These include the probability of moderate to severe reaction (θ) from a relatively low (*Scenarios 1–6, Low R*) to a relatively high (*Scenarios 7–12, High R*) arbitrarily, in other words, under the *High R Scenarios* there's a true change in the biological process compared with the *Reference Scenario*, i.e., a true increase in the risk of moderate to severe reaction as may plausibly occur due to a manufacturing issue with a particular vaccine batch. We varied the

Table 1 | Simplified DAG variables and data simulation parameters

Variable	Definition	States
Age	Age of vaccine recipient monitored for safety by an active surveillance system	<50 y, ≥ 50 y
Vaccine	Vaccine received at 3 days before the survey distribution	Yes
Reaction	The level of reaction within 3 days following the vaccination. This is often not directly observed and thus different from reported AEFI	None to mild (0), moderate to severe (1)
Survey participation	The recipient participated in the survey sent at day 3 following their vaccination at an AusVaxSafety site	No (0), yes (1)
MA	The recipient experienced a reaction within 3 days of vaccination and sought medical attention either via a general practitioner (GP) or hospital emergency department (ED)	No (0), yes (1)
Report MA	The recipient reported MA (GP and/or ED) via AEFI survey	No (0), yes (1)
Parameter	Definition	Value
θ	Probability of moderate to severe reaction in the <50 y age group	0.3
ϵ	Multiplicative change of probability of moderate to severe reaction in the ≥50 y age group	$\frac{2}{3}$
η	Probability of survey participation in the <50 y age group with a mild reaction	0.1
τ_{sp}	Multiplicative change of survey participation due to a moderate to severe reaction independent of age group	1.5
μ_{sp}	Multiplicative change of survey participation in the ≥50 y age group independent of moderate to severe reaction	1.35
ϕ	Probability of MA in the <50 y age group with a mild reaction	0.01
τ_{ma}	Multiplicative change in MA probability due to a moderate to severe reaction independent of age group	3
μ_{ma}	Multiplicative change in MA probability in the ≥50 y age group independent of severity of reaction	5

probability of survey participation (η) representing a change in the behaviour of vaccine recipients (*Medium P* for *Scenarios 2, 5, 8, 11* and *High P* for *Scenarios 3, 6, 9, 12*). Finally, we varied the influence of reaction on both the probability of survey participation (τ_{sp}) and the probability of MA (τ_{ma}) (*Strong* influence for *Scenarios 4–6, 10–12*).

PPA investigation

We paired the *Reference Scenario* with each *Investigation Scenario*, and assessed how likely the PPA method is likely to detect a signal under each *Investigation Scenario*. We inspected the diagnostics, using the *summary* function in the *cmdstanr* package in R, for the first 5 simulations for each of the 12 scenarios. All R-hat convergence diagnostic values were <1.01 and the bulk and expected sample sizes were sufficiently large (>1300). There was no evidence of nonconvergence of any chain. We present a histogram of where each simulation's number of MAs under the *Investigation Scenario* sits as a percentile of the predicted probability distribution in Figs. 3 and 4. We also presented here the percentage of simulations that generated a signal for each scenario by age group in light blue and dark blue text.

Figure 3 consists of the six *Low R Investigation Scenarios* where there is no change in the biological processes of interest compared with the *Reference Scenario*, and Fig. 4 consists of those *High R Investigation Scenarios* where there is an increase compared with the *Reference Scenario*. It is desirable for the PPA signal detection method to generate a safety signal in investigation scenarios that are set with a high prevalence (probability) of moderate to severe reaction (Fig. 4) and not to generate a signal in scenarios set with a low prevalence of moderate to severe reaction (Fig. 3). Any signals generated in simulations of high and low prevalence of reaction scenarios were therefore considered to be appropriate and inappropriate, respectively.

In *Low R Scenarios* (Fig. 3), signals were more likely to be inappropriately generated in the older age group when the influence of reactions on survey participation and MA was strong (27–78% of simulations, *Scenarios 4–6*). This is because the inflated proportion of reported MA is more likely to be detected when the number of survey participants increases due to increasing survey participation and healthcare-seeking behaviour. This same analogy applies to the relatively *High R Scenarios* (*Scenarios 10–12*). The variation in the proportion of signals detected across the scenarios is due to a combination of the precision of the posterior predictive distribution and magnitude of the bias, both related to the survey participation rate. Bias in the estimate of reported MA in the *Strong Influence Scenarios* (*Scenarios 4–6 and 10–12*) is greatest when the survey participation is low, but the precision of the posterior predictive distribution will also be lower. The absolute inflation in the proportion of respondents who report MA was higher in the older age group compared to the younger age group across all scenarios. For the younger age group, only a low proportion of simulations resulted in an inappropriate signal generation (0–4% of simulations).

In *High R Scenarios*, the probability of appropriate signal generation was lower (1–62% of simulations) in the younger age group, and in the older

Table 2 | Definition of investigation scenarios

Scenario	Scenario specification	Parameter values θ, η, (τ_{sp}, τ_{ma})
Reference	Low R, Low P, Weak	0.30, 0.1, (1.5, 3)
1	Low R, Low P, Weak	0.30, 0.1, (1.5, 3)
2	Low R, Medium P, Weak	0.30, 0.3, (1.5, 3)
3	Low R, High P, Weak	0.30, 0.4, (1.5, 3)
4	Low R, Low P, Strong	0.30, 0.1, (1.8, 5)
5	Low R, Medium P, Strong	0.30, 0.3, (1.8, 5)
6	Low R, High P, Strong	0.30, 0.4, (1.8, 5)
7	High R, Low P, Weak	0.60, 0.1, (1.5, 3)
8	High R, Medium P, Weak	0.60, 0.3, (1.5, 3)
9	High R, High P, Weak	0.60, 0.4, (1.5, 3)
10	High R, Low P, Strong	0.60, 0.1, (1.8, 5)
11	High R, Medium P, Strong	0.60, 0.3, (1.8, 5)
12	High R, High P, Strong	0.60, 0.4, (1.8, 5)

Scenarios vary with probability of moderate to severe reaction (Low or High R), probability of survey participation (Low, Medium or High P) and influence of moderate to severe reaction on the probability of survey participation and the probability of MA (Weak, Strong).

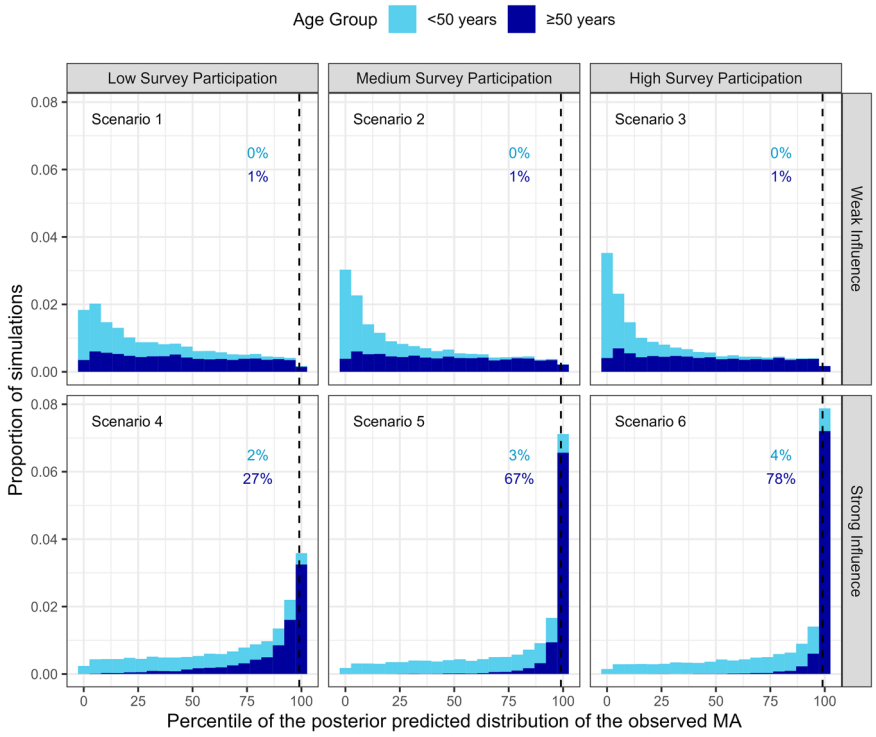

Fig. 3 | Low probability of moderate to severe reaction in investigation scenarios, no change compared to the reference scenario. Signals are flagged when the number of MAs exceeds the 99th percentile (dotted lines) upon 5000 simulations with 4000 individuals simulated for each simulation per scenario. The proportion of simulations which generated a signal within each scenario is represented as percentages within each panel in this figure. The probability of survey participation (η) was varied from *Low P* (*Scenarios 1, 4*) to *Medium P* (*Scenarios 2, 5*) and *High P* (*Scenarios 3, 6*). The influence of reaction (τ_{sp} and τ_{ma}) on both survey participation and MA respectively was varied between *Weak* for *Scenarios 1–3* and *Strong* for *Scenarios 4–6*. The exact parameters altered for each scenario are as specified in Table 2, all other parameters used are as specified in Table 1.

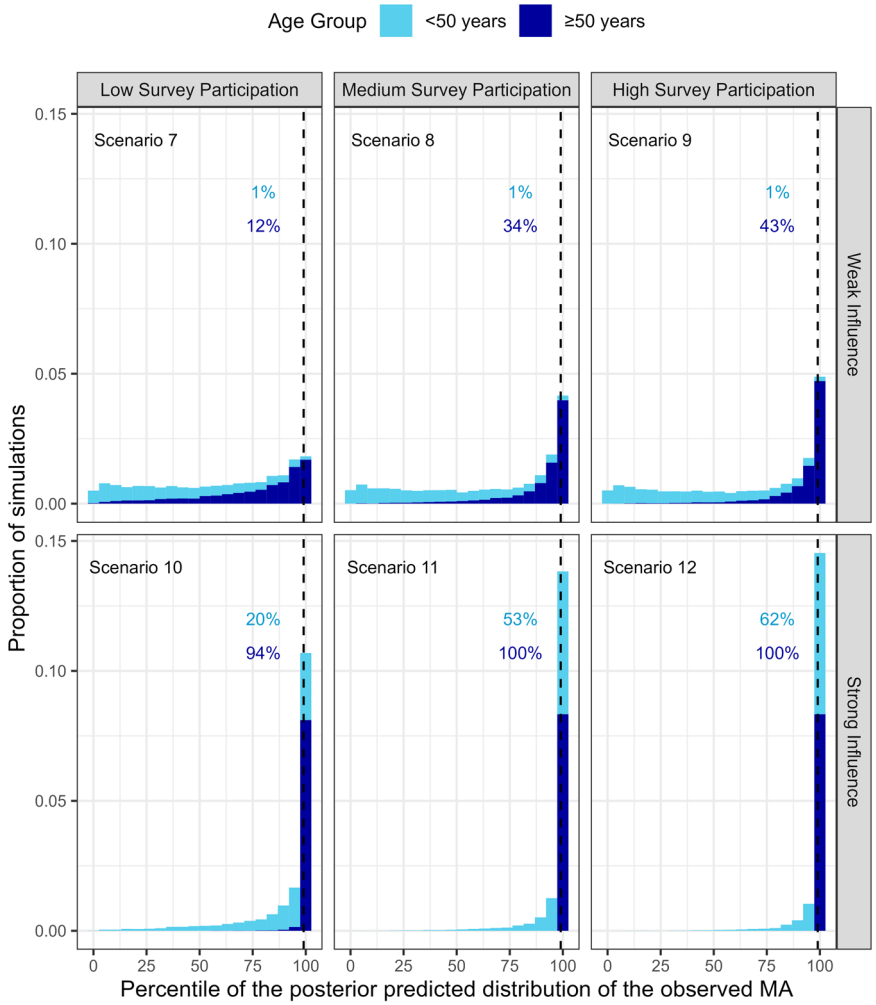

Fig. 4 | High probability of moderate to severe reaction in investigation scenarios, increase in biological processes compared to the reference scenario. Signals are flagged when the number of MAs exceeds the 99th percentile (dotted lines) upon 5000 simulations with 4000 individuals simulated for each simulation per scenario. The proportion of simulations which generated a signal within each scenario is represented as percentages within each panel in this figure. The probability of survey participation (η) was varied from *Low P* (*Scenarios 7, 10*) to *Medium P* (*Scenarios 8, 11*) and *High P* (*Scenarios 9, 12*). The influence of reaction (τ_{sp} and τ_{ma}) on both survey participation and MA, respectively, was varied between *Weak* for *Scenarios 7–9* and *Strong* for *Scenarios 10–12*. The exact parameters altered for each scenario are as specified in Table 2, all other parameters used are as specified in Table 1.

age group when the influence of reactions on survey participation and MA was weak (12–43% of simulations).

Discussion

Key challenges to vaccine safety monitoring are the low prevalence of AEFI, the relatively large number of vaccine recipients required to respond to safety surveys to ensure adequate sensitivity of the detection methods and the absence of information from vaccine recipients who are survey non-responders. There is also a need to balance sensitive signal detection methods, that detect true vaccine safety issues in a timely manner, against the frequency of false detections. To understand this problem and thereby improve current signal detection methodology, we combined causal methods and statistical modelling, and used simulation to investigate how AEFI signal detection may be affected by plausible biological and behavioural factors.

We used the novel PPA methodology, which is based on a Bayesian logistic model of the probability of reporting medical attendance following an AEFI. In brief, a PPA signal is generated by the AusVaxSafety monitoring system when the number of reported MAs for AEFI (for a specified vaccine) exceeds a threshold based on historically observed rates and the number of respondents in the current reporting period, adjusted for age. There are two key reasons why a signal might occur. First, there might be a true increase in AEFI in the vaccinated population (i.e., the biological process of interest) which is reflected in increased reports of MA in the surveyed population. Second, in the absence of a true increase in AEFI, the survey respondents might be enriched with a subset of a population whose rate of reporting MA is higher; this could be due to changes in behaviour (such as media reports) and/or procedural processes (such as an age-related roll out of a vaccine) and independent of any change in the biological processes. This possibility adds to the practical challenge of detecting a true increase in AEFI and may trigger time-consuming case-series investigations by public health researchers.

A desirable safety signal detection system should be sensitive to changes in the true prevalence of severe vaccine reactions, whilst minimising false signals due to variation in behavioural or procedural factors that might affect reporting. These include the survey participation rate and the influence of reaction severity on the propensity to participate in the survey or to seek medical attention; we found evidence that our safety method could be sensitive to both factors. For *High R Investigation Scenarios* (Fig. 4), signals failed to be generated in the younger age group if either the influence of reactions on survey participation and MA was weak or survey participation was low. For the older age group, when the reaction prevalence was high, signals were detected for low survey participation if the influence of reactions on survey participation and MA was strong. Conversely, in *Low R Investigation Scenarios* (Fig. 3), inappropriate signals can occur when there is a strong influence of reaction severity on survey participation and MA, especially when survey participation is high. We observed bias in the estimate of reported MA in the *Strong Influence Scenarios* (*4–6* and *10–12*) where there is a variation in the proportion of signals detected across the scenarios due to a combination of the precision of the posterior predictive distribution and magnitude of the bias, both related to the survey participation rate. An inflation of inappropriate or false signal detection may have a subsequent effect on investigative resources.

Inspection of the causal DAG (Fig. 2), which underpinned the data generation for this simulation study, partially explains the sensitivity of the signal detection method to changes in these behavioural factors. In essence, we wish to use data on reported MAs to make indirect inference about the (possibly changing) causal effect of vaccination on the prevalence of severe reactions. Severe reactions are only ascertained as reported medical attendances, which are conditionally dependent on both medical attendance occurring and survey participation. Independent of the influence of reaction severity, the probability of an MA is plausibly influenced by age and other participant characteristics. While the age of vaccine recipients can be measured and conditioned upon, these other factors are mostly unknown, unmeasured, and will, therefore, confound attempts to attribute any changes in MA to changes in the reactogenicity of the vaccine. Furthermore, changes in survey participation rates will also influence the rate of reported MA, independent of any true increase in MA. If these changes are driven by factors other than changes in the age distribution of vaccine recipients, there is a risk those factors will further confound the attribution of changes in reported MAs to changes in vaccine reactogenicity. Systematically monitoring survey response rates and survey participation behaviour might be used to determine the significance of an alerted signal and to improve the sensitivity of signal detection. When a signal has been alerted, it may be important to assess whether this can be explained by changes in survey participation behaviour, and this may inform the public health investigation and interpretation of the signal, *e.g.*, a signal alerted in the context of consistently high survey participation may have a greater significance than a signal alerted in the context of varying survey participation behaviour.

The PPA signal detection method is also sensitive to the quantity of data available at each analysis, especially at the start of a new surveillance period, such as the annual roll out of influenza vaccines. Low rates of survey participation reduce the quantity of data available to inform the parameters of the statistical model, and therefore the precision of the expected (predicted) number of MA in the surveillance period. As a result, even a moderately high frequency of reported MAs caused by a true increase in reactogenicity will be compatible with the statistical model and may not exceed the threshold for signal detection even when the risk of moderate to severe reactions is high. Therefore, it is desirable to increase survey participation by vaccine recipients, which requires promotion by immunisation providers at the time of vaccination.

The use of causal DAGs was informative for our study in several ways. The full DAG (Fig. 1) depicts the problem domain of vaccine safety monitoring and thus facilitates communication among people from disparate disciplines, including medical experts, public health practitioners and statisticians. It provided a common starting point for simplification of the DAG and elicitation of the parameters and scenarios necessary for the data generation process. This is a necessary simplification of the real world but captures the important components of the complex target problem domain. The full DAG also serves as a knowledge base, which can support investigations of future research questions. From a modelling perspective, how simplification should be made is driven by the purpose of modelling. In our case, the six variables included in the simplified DAG sufficiently enabled the investigation of our research question of interest. We showed, by simulation, how changes in non-biological factors might affect active surveillance, thus leading to a distorted interpretation of the biological process in the vaccinated population. Such simulations revealed how signal detection methods can be influenced by behavioural and procedural factors that affect survey participation and response, but which cannot be gleaned from the survey data alone. However, the simplified DAG presented here may not be sufficient to answer how vaccine signals should be interpreted under a specific real-world implementation scenario, where other important factors (other than age) that influence survey participation and seeking MA should also be taken into consideration. It is essential to be clear about the purpose of creating a causal DAG or any model.

There are limitations to our simulation study. For simplicity, the only variable included in the PPA method was age, although other factors are known to drive biological, behavioural and procedural processes and thus affect how a detected signal should be interpreted. The AusVaxSafety PPA model is more complex and accounts for gender, ethnicity, jurisdiction and co-morbidities, in addition to the age of the vaccine recipient. We chose to simulate age from a Gaussian distribution reflecting a plausible age distribution observed in AusVaxSafety. We appreciate the importance of considering realistic age distributions when applying the proposed method in our study to real-world scenarios, especially how such distributions can shift over time. We conducted our simulation study over two age groups (below 50 years or 50+ years), when in reality, biological and behavioural factors may have differential effects across the age range. Incorporating more appropriate PPA models (including factors such as gender and ethnicity, and greater granularity in the age groups) and a more complex causal DAG as part of a simulation study to further explore the effect of behavioural factors in vaccine safety surveillance would be of interest. Here we focused on the effect of differential survey responses among those who attend participating immunisation clinics. However, people attending participating immunisation clinics may be systematically different from those attending other immunisation services (e.g., in remote clinics or parent report from childhood immunisation services). These differences may, in turn, affect survey response and healthcare-seeking behaviours so reported AEFI rates may be over or underestimated. The current full DAG can be extended to depict this potential selection bias and reflect how the surveyed and responder population relates to the complete vaccinated population. While this may not impact the operating characteristics of our signal detection methods, future work could account for this potential selection bias in AEFI rate estimation by including the type of clinic in the statistical model. Finally, we did not differentiate between seeking phone advice for AEFI versus GP or hospital attendance, nor did we consider detection via passive (spontaneously volunteered) rather than active (solicited) AEFI reports. Greater granularity for these factors may be incorporated into future models for a better understanding of the impact of behavioural factors on vaccine safety surveillance.

Methods

The scientific causal models

A DAG consists of nodes that represent random variables that may or may not be observed (in the form of data), and arrows (arcs) that indicate a possible direct influence of predecessor (or parent) variables on their child nodes (nodes extending from other nodes). A causal DAG is one in which the arcs are intended to represent influences that are causal and not just associative. In this study, the purpose of creating the causal DAGs was to represent how the AusVaxSafety active surveillance system for short-term AEFI monitoring operates as a complex system (the *full DAG*). It addresses how this complexity may affect the reporting of AEFI, especially MA, and how this may, in turn, influence the detection of safety signals. Using a previously published causal knowledge engineering process[19], the DAGs were drafted, refined and applied for the stated purpose. Domain experts were consulted to advise on the selection of relevant variables, the causal structure and the face validity of the final DAGs. The domain experts were public health practitioners, clinical vaccinologists, programme managers familiar with AusVaxSafety's data capture processes, and statisticians responsible for the analysis of the survey data and reporting of vaccine safety signals. From the full DAG, we extracted a *simplified DAG*, which preserved all causal assumptions from the full DAG. The purpose of this simplified DAG was to facilitate the investigation of our research question of interest—how distinct survey responding behaviour in subgroups (such as different age groups) can interact with the underlying biological process of interest (vaccine reaction) to affect the sensitivity and specificity of existing signal detection methodologies.

Data simulation and scenarios

To explore factors that influence signal detection via statistical analysis, we used all variables in the simplified DAG to generate complete data (i.e., without missingness) relevant to the problem domain, including for vaccine recipients who respond to the survey (observed) and those who do not.

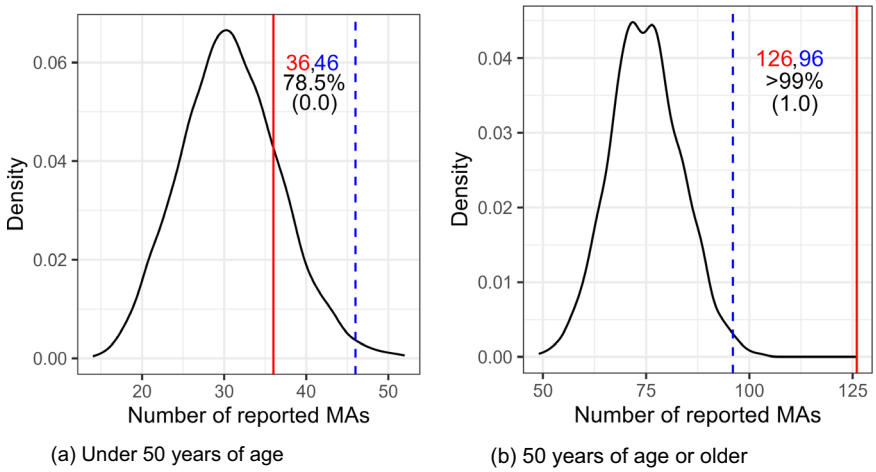

Fig. 5 | Illustrative outputs of the PPA method from a single simulation. The number of observed MAs, reported in red and the signal threshold reported in blue, are also indicated by the solid red lines and dotted blue lines, respectively. The percentile of the posterior predicted distribution is used here as an indication of the distance of the number of reported MAs from the signal threshold. As the figure illustrates the outputs of just one simulation, the proportion of simulations with an alerted signal is 0.0 in (**a**) and 1.0 in (**b**).

Binary discrete variables were sampled from Bernoulli distributions. Age was sampled from a truncated Gaussian distribution for which the statistical parameters were chosen to reflect reasonable and plausible age ranges of individuals who had received post-vaccination safety surveys. Age was then categorised as 'below 50 years old' and '50 years old and above'. Guided by the causal DAGs, we designed hypothetical investigation scenarios of scientific interest that reflect potential variations in biological, behavioural and procedural assumptions, which may lead to distinct patterns of the reporting of MA. For each scenario, we generated 5000 simulations to investigate the operating characteristics of the statistical signal detection method.

The statistical signal detection method

We were interested in how a safety signal may be generated under each hypothetical scenario of interest using the PPA method. The PPA method identifies a signal when the number of reported MAs in an investigation period exceeds a threshold. This threshold was defined as the 99th percentile of the posterior predictive distribution for the number of reported MAs under a Bayesian statistical model. The parameters of this Bayesian statistical model were informed by (historic) reference data and are regularly updated using survey data (as illustrated in Fig. 5). We summarise the simulations for each investigation scenario in four ways: (1) the mean number of reported MAs, (2) the mean threshold value (from the posterior predictive distribution for the number of MAs), (3) the mean percentile of the number of reported MAs of the posterior predicted distribution (as an indication of how closely the mean number of reported MAs approaches the mean signal threshold), and (4) the proportion of simulations where the number of reported MAs exceeds the signal threshold, i.e., the proportion of simulations in which a signal has been generated. This is illustrated in Fig. 5.

We considered a simplified version of the PPA consistent with our simplified DAG with only one explanatory variable, age, categorised into <50 years old (denoted as $g = 0$) and ≥50 years old (i.e., $g = 1$). The number of MAs was modelled as arising from a Binomial distribution:

$$y_g \sim \text{Binomial}(n_g, p_g)1$$

where:

y_g = number of reported MAs in age group g

n_g = number of survey responses for age group g

p_g = probability of an MA being reported among survey responders for age group g

The linear predictor is:

$$\text{logit}(n_g, p_g) = \alpha + 1_{[g=1]}\beta \quad (2)$$

where:

α = log-odds of an individual <50 years old reporting an MA

β = log-odds ratio of an individual ≥50 years old compared to <50 years old reporting an MA

(a) Under 50 years of age (b) 50 years of age or older

The model parameters were given the following weakly informative priors[9]:

$$\alpha \sim \text{Normal}(-4, 2^2) \quad (3)$$

$$\beta \sim \text{Normal}(0, 1)$$

The prior distribution for α induces a 95% credible interval between ~0% and 48% for the probability of an individual <50 years old reporting an MA. The prior distribution for β translates as a wide range of odds ratios avoiding extreme values of $\exp(-3) = 0.05$ and $\exp(3) = 20.09$, i.e., within three standard deviations about the mean. We used GeNIe software to build and depict the DAGs presented here, Academic Version 4.1.4016.0[20]. All simulations and analytical programming were conducted in Stan[21] (Version 0.6.1) via the *cmdstanR* package[22] in the R statistical programming language v4.2.2[23]. Posterior distributions were estimated via Markov-chain Monte Carlo (MCMC) using Stan's Hamiltonian Monte Carlo algorithm with eight MCMC chains, run in parallel, with warm-up and sampling phases each running for 1000 iterations.

Data availability

The datasets generated and analysed during the current study are available in the GitHub repository: www.github.com/ECSTay/AVSCausalModel.

Code availability

The underlying code for this study is available in GitHub and can be accessed via this link: www.github.com/ECSTay/AVSCausalModel.

Received: 12 March 2024; Accepted: 19 August 2024;
Published online: 07 September 2024

References

1. Armstrong, P. et al. Epidemiological study of severe febrile reactions in young children in Western Australia caused by a 2010 trivalent inactivated influenza vaccine. *BMJ Open* **1**, e000016 (2011).
2. Blyth, C. C. et al. The impact of pandemic A (H1N1) pdm09 influenza and vaccine-associated adverse events on parental attitudes and influenza vaccine uptake in young children. *Vaccine* **32**, 4075–4081 (2014).
3. Phadke, V. K., Bednarczyk, R. A., Salmon, D. A. & Omer, S. B. Association between vaccine refusal and vaccine-preventable

diseases in the United States: a review of measles and pertussis. *JAMA* **315**, 1149–1158 (2016).
4. World Health Organization. *Causality Assessment of an Adverse Event Following Immunization (AEFI): User Manual for the Revised WHO Classification* (WHO, 2019).
5. US Food and Drug Administration. *Guidance for Industry: Toxicity Grading Scale for Healthy Adult and Adolescent Volunteers Enrolled in Preventive Vaccine Clinical Trials* (Food and Drug Administration, Silver Spring, MD, USA, 2007).
6. AusVaxSafety. *AusVaxSafety* https://www.ausvaxsafety.org.au/ (2023) accessed 4 July 2023.
7. Deng, L. et al. The short term safety of COVID-19 vaccines in Australia: AusVaxSafety active surveillance, February–August 2021. *Med. J. Aust.* **217**, 195–202 (2022).
8. Harpaz, R. et al. Signaling COVID-19 vaccine adverse events. *Drug Saf.* **45**, 765–780 (2022).
9. Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. *Bayesian Data Analysis* (Chapman and Hall/CRC, 1995).
10. Tennant, P. W. et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int. J. Epidemiol.* **50**, 620–632 (2021).
11. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN* (Chapman and Hall/CRC, 2020).
12. Deffner, D., Rohrer, J. M. & McElreath, R. A causal framework for cross-cultural generalizability. *Adv. Methods Pract. Psychol. Sci.* **5**, 25152459221106366 (2022).
13. Polack, F. P. et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Eng. J. Med*. **383**, 2603–2615 (2020).
14. Barda, N. et al. Safety of the BNT162b2 mRNA Covid-19 vaccine in a nationwide setting. *N. Engl. J. Med.* **385**, 1078–1090 (2021).
15. Ramasamy, M. N. et al. Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. *Lancet* **396**, 1979–1993 (2020).
16. Christensen, A. I. et al. The Danish national health survey 2010. Study design and respondent characteristics. *Scand. J. Public Health* **40**, 391–397 (2012).
17. Juto, H., Gärtner Nilsson, M., Möller, M., Wennergren, D. & Morberg, P. Evaluating nonresponders of a survey in the Swedish fracture register: no indication of different functional result. *BMC Musculoskelet. Disord.* **18**, 1–9 (2017).
18. Australian Government. *Older Australians* https://www.aihw.gov.au/reports/older-people/older-australia-at-a-glance/contents/service-use/health-care-gps-and-specialists accessed 31 January 2023. (2023).
19. Mascaro, S. et al. Modeling COVID-19 disease processes by remote elicitation of causal Bayesian networks from medical experts. *BMC Med. Res. Methodol.* **23**, 76 (2023).
20. Druzdzel, M. J. SMILE: Structural Modeling, Inference, and Learning Engine and GeNIe: a development environment for graphical decision-theoretic models. In *Aaai/Iaai* 902–903 https://www.bayesfusion.com (1999).
21. Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual*, *Version 2.26.1* https://mc-stan.org/ (2022).
22. Gabry, J., Češnovar, R. & Johnson, A. *cmdstanr: R Interface to 'CmdStan'* https://mc-stan.org/cmdstanr/, https://discourse.mc-stan.org (2023).
23. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2023). https://www.R-project.org/.

Acknowledgements

This project is supported by the Australian Government Department of Health and Aged Care. Dr. Lucy Deng contributed her clinical expertise in constructing the full causal model for this study.

Author contributions

Y.W., T.S. and J.M. conceptualised this study. Causal modelling, simulation and statistical analysis were devised and conducted by E.T., Y.W., M.D. and S.B. Y.H., C.G., L.L. and T.S. provided domain expertise. All authors contributed to the interpretation of results and the writing of the manuscript. All authors approve of distribution of this manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41541-024-00955-4.

Correspondence and requests for materials should be addressed to Evelyn Tay.