ORIGINAL REPORT

WILEY

Bayesian hierarchical methods in the detection of potentially teratogenic first-trimester medications

Alana Cavadino[1,2] | David Prieto-Merino[3,4] | Joan K. Morris[1,5]

[1]Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK

[2]Section of Epidemiology and Biostatistics, School of Population Health, The University of Auckland, Auckland, New Zealand

[3]Faculty of Epidemiology & Population Health, London School of Hygiene and Tropical Medicine, London, UK

[4]Applied Statistics in Medical Research Group, Catholic University of Murcia (UCAM), Murcia, Spain

[5]Population Health Research Institute, St George's, University of London, London, UK

Correspondence
Joan K. Morris, Population Health Research Institute. St George's, University of London, Cranmer Terrace SW17 0RE, London, UK.
Email: jmorris@sgul.ac.uk

Funding information
Medical Research Council, Grant/Award Number: 1504916

Abstract

Purpose: Bayesian hierarchical models (BHMs) have been used to identify adverse drug reactions, allowing information sharing amongst adverse reactions and drugs expected to have similar properties. This study evaluated the use of BHMs in the routine signal detection analyses of potential first-trimester teratogens, where these models have not previously been applied.

Methods: Data on 15 058 malformed foetuses exposed to first trimester medications (1995-2011) from 13 European congenital anomaly (CA) registries were analysed. The proportion of each CA in women taking a specific medication was compared with the proportion of that CA in all other women in the dataset (55 CAs × 523 medications). BHMs were grouped by either medications or CAs or by both simultaneously, and the results compared with analysing each medication-CA combination separately and adjusting for multiplicity using a double false discovery rate (FDR) procedure. The proportions of "high-risk" medications (medications which have been shown to carry a moderate to high risk of foetal malformations) identified as potential signals were compared, as well as the total number of potential signals requiring follow up (the effective workload).

Results: BHMs identified more high-risk medications than the double FDR method, but the effective workload was larger. A BHM grouping both medications and CAs, for example, identified 23% of high-risk medications compared with 14% by the double FDR; however, there was an increase from 16 to 71 potential signals requiring follow up.

Conclusion: For comparable effective workloads, BHMs did not outperform the double FDR, which is comparatively straightforward to implement and is therefore recommended for continued use in teratogenic signal detection analyses.

KEYWORDS
Bayesian hierarchical models, birth defects, congenital anomalies, EUROmediCAT, false discovery rate, multiple testing, pharmacoepidemiology, pharmacovigilance, signal detection

1 | INTRODUCTION

Pregnant women are excluded from the majority of new medication safety studies, so information about potential risks to a foetus is lacking. Routine screening of congenital anomaly (CA) data for potentially teratogenic medications taken during the first trimester of pregnancy is performed by EUROmediCAT.[1] EUROmediCAT compares the odds of exposure to a specific medication for a specific CA with the odds of exposure to the same medication for all other medication-exposed CAs[2] using a double false discovery rate (FDR) procedure to incorporate groupings of similar medications when adjusting for multiple testing.[3]

In the field of pharmacovigilance, large spontaneous reporting databases detect signals of adverse events (AEs) using disproportionality analyses to identify drug-AE combinations with more observed reports than expected under the assumption of independence.[4] Bayesian hierarchical models (BHMs) have been applied to these spontaneous reporting datasets,[5] with the aim of improving estimated associations for any drug-AE combination by incorporating information from other similar drugs (or AEs) using specified groupings.[6] BHMs can explicitly allow (without imposing) for the possibility that different medications in the same group might be related and that AE rates are more likely to be similar within than across these groups. A BHM with information sharing in two dimensions has also been proposed to incorporate groupings of both medications and AEs simultaneously.[7] Simulation studies and application to a sample of the World Health Organisation pharmacovigilance database have demonstrated that a two-dimensional model of information sharing can produce a more powerful BHM to detect true adverse drug reactions, compared with sharing information only in one dimension.[7] We aimed to ascertain whether BHMs that share information between medications and/or CAs using existing coding hierarchies have the potential to improve the effectiveness of signal detection methods using EUROmediCAT data.

2 | METHODS

2.1 | Study population

EUROCAT is a network of European population based CA registries whose data is obtained through both active case finding and voluntary reporting,[8] with multiple sources of information used to ascertain CA cases including live birth, foetal death, and termination of pregnancy for foetal anomaly. Data quality indicators are used to assess consistency of inclusion criteria, data collection, and recording across registries.[9] The International Classification of Diseases coding version 10 with British Paediatric Association extension (IDC10-BPA) is used to code CAs according to EUROCAT guidelines. EUROmediCAT comprises EUROCAT registries that collect information on first-trimester (defined as the first day of a woman's last menstrual period up to her 12th gestational week) medication use. Maternal medication exposure data is primarily obtained through prospectively recorded maternity records, with additional sources including maternal interviews and general practitioner records.[10,11] Data on 31 197 malformed foetuses

Key Points

- Bayesian hierarchical models have the potential to improve teratogenic signal detection by incorporating information sharing between similar medications and/or congenital anomalies.
- In our analysis, Bayesian hierarchical models demonstrated a potential to detect signals with fewer exposed cases than the current frequentist signal detection procedure.
- In our analysis of prospectively collected registry data on congenital anomalies, Bayesian hierarchical models did not outperform the currently used double false discovery rate (FDR) method of adjusting for multiple testing in signal detection; continued use of double FDR methods are therefore recommended for teratogenic signal detection.

with first-trimester medication exposures from 1995 to 2011 were available for this study, covering a population of around 7 million births across 13 registries in 11 countries. For some registries, there was considerable data loss where it was not possible to verify the timing of reported medication exposures, which has been discussed previously (6). However, the distribution of types of CA were similar for those pregnancies included and excluded due to unknown timing, suggesting that cases remaining in the dataset for these registries should not be prone to selection biases in this respect. Ethical and data access approvals were obtained for each database from the relevant governance infrastructures. This EUROmediCAT dataset was used previously to compare different FDR methods for multiple testing adjustment,[3] and a large proportion of this data was also previously analysed for signal detection purposes.[2]

2.2 | Congenital anomalies and medication exposures

Cases with chromosomal anomalies, skeletal dysplasia, genetic syndromes, and microdeletions were excluded as they are unlikely to be caused by teratogenic medications. The aetiology of congenital dislocation of the hip is mechanical, so foetuses with this being their only recorded major CA were also excluded (n = 905). EUROCAT's hierarchical coding system categorises the 54 nonchromosomal CA subgroups into 11 main organ system groups and five "other anomalies/syndromes" subgroups https://eu-rd-platform.jrc.ec.europa.eu/sites/default/files/Full_Guide_1_4_version_28_DEC2018.pdf.[12] Five hundred forty-two foetuses with a congenital heart defect (CHD) but no information regarding the specific type of CHD were combined into an additional subgroup ("unspecified CHD"; Table S1.

The Anatomical Therapeutic Chemical (ATC) system[13] is used to code unlimited exposures in EUROmediCAT data. The ATC coding

system has a hierarchical structure, grouping medications at five levels; the highest (ATC1) groups medications into 14 main anatomical groups, the next three (ATC2-4) use three to five digit codes, respectively, to represent further therapeutic, pharmacological, and/or chemical classifications, and the most detailed level (ATC5) uses seven digits to identify the chemical substance. Older ATC codes subject to alterations over time[14] were updated to the newer codes. Foetuses exposed only to vitamins, minerals, and/or folic acid were excluded. Foetuses were excluded if exposed only to topical medications, codes with less than five digits (detail below ATC4 level; n = 1219), second/third trimester medications (n = 1490), or with unknown timing (n = 12 073). Further description of these exclusions by registry have been described in more detail in our previous study.[3] After these exclusions, a total of 15 058 malformed foetuses were included in the analysis dataset. ATC5 codes were analysed, but where information was only available to ATC4 level, this was also included. Five hundred twenty-three medications with at least three exposed foetuses were investigated. As in the previous EUROmediCAT analysis, foetuses exposed only to medications with fewer than three exposures in total were included in the dataset as controls (since medications with fewer than three exposures are not analysed for signal detection).[2,3]

2.3 | Statistical analysis

Results from BHMs were directly compared with those obtained previously for the double FDR procedure on the same dataset.[3] Briefly, the double FDR comprises two steps[15,16]: first, a representative minimum P value is calculated for each group and only those groups with a representative P value below the specified FDR threshold are included in the next step. In the second step, a Simes FDR procedure is applied across all combinations belonging to those groups passing the first step. All data management and calculations for Fisher's exact test and the double FDR procedure were performed using Stata 12.[17]

For the BHMs, a Gamma Poisson Shrinker (GPS) and a BHM with a Poisson distribution were combined to model the CA and medication counts.[7] The expected count E_{ij} for each observed count c_{ij} was calculated using the marginal totals for medication i and CA j assuming no association between i and j. The proportional reporting ratio (PRR)[18] for medication i and CA j was the ratio of the observed to expected counts, $PRR_{ij} = \frac{c_{ij}}{E_{ij}}$. The data structure for two-dimensional information sharing by medications and CAs is displayed in Table 1. Here, d represents ATC3 medication codes with D groups and $i = 1, ..., n_d$ medications within each group d and a denotes groupings of CAs according to the EUROCAT organ system classes with A groups of CAs and $j = 1, ..., n_a$ CAs within each group. The lighter grey shading in Table 1 represents the set of the $d = 2$ group of medications crossed with the $a = 2$ group of CAs, and the dark grey cell in Table 1 c_{2122} denotes the observed count for the combination of medication $i = 1$ in the $d = 2$ medication group with CA $j = 2$ in the $a = 2$ CA group. Each set in Table 1 has a group distribution, such that each medication-CA combination within that two-way group shares a common prior distribution. There is also a prior distribution for the set of all top-level sets, that is, an average across all CAs and/or medications.

Table S2 presents notation for the BHMs, for which a Poisson distribution was used to model the observed counts c_{ij} for each combination of a medication i and a CA j according to four models of information sharing. Model 1 is a separate BHM for each medication-CA combination, with no grouping of medications or CAs. Models 2 and 3 are one-dimensional models of information sharing, with

TABLE 1 Information sharing by grouping of both medications and CAs

CAs														
			1				2				...	A			
ATC medications	ATC3 (d)	Organ class (a) CA (j) ATC5 (i)	1	2	...	n_1	1	2	...	n_2		1	2	...	n_A
	1	1													
		2													
		⋮													
		n_1													
	2	1													
		2													
		⋮													
		n_2													
	⋮	⋮													
	D	1													
		2													
		⋮													
		n_D													

Abbreviations: ATC, Anatomical Therapeutic Chemical; CA, congenital anomaly.

grouping for medications only (using ATC3 codes, model 2) or CAs only (using EUROCAT organ system classes, model 3). In model 2, the effects for each group of medications are calculated separately for each CA, allowing a different distribution for each CA. Conversely, a different distribution is allowed for each medication when grouping by CAs in model 3. As a sensitivity analysis, an alternative formulation of model 2 was also considered where CAs were treated as coming from one overall group, imposing a common distribution of effects across the group of all CAs, separately for each group of ATC3 medications (and vice-versa for model 3). For model 2, this allowed a more direct comparison with the double FDR method, which groups medications using ATC3 codes but adjusts for multiple testing across all CAs. Model 4 is two-dimensional in that groupings of both medications and CAs are incorporated, as displayed in Table 1. The minimally informative priors used throughout are described in Section S1; normal distributions were used for estimation of means (eg, average PRRs for each medication/CA or group of medications/CAs), and uniform distributions were used for variance parameters.[19] Minimally informative priors were used, as our main aim was to assess the effect of the groupings themselves on the model results, with the main source of "informative" prior information therefore coming from the groupings that were used. Different choices of values for the parameters of prior distributions were assessed for their effect on model fit and results. BHMs were implemented using JAGS via R package rjags.[20,21] The code used to specify these models in JAGS is presented in Table S2. The coda package[22] in R was used to assess model convergence and to summarise the posterior distribution for each parameter, including convergence statistics and visual inspection of trace, density and auto-correlation plots for the parameters in each model. These measures were also used to determine the required number of total iterations and thinning.

Any medication-CA combination with a posterior 2.5th percentile (ie, the lower limit of a one-sided 97.5% posterior confidence interval [PCI]) greater than 1 for the PRR was considered a potential signal. As this choice of threshold is somewhat arbitrary, the effect of choosing a stricter 0.5th percentile as a cut-off was also assessed. As the purpose of signal detection is to screen for potential teratogens, combinations are only flagged as potential signals if they represent an increased in reporting (ie, PRR > 1). The number of associations with a PRR < 1 (medications relatively "less harmful" for a specific CA than the average medication in the dataset) and corresponding 97.5th (or 99.5th) percentile less than 1 were monitored to determine how often these occurred.

2.4 | Evaluation and comparison of signal detection methods

Although risk classification systems have been implemented and used in a number of countries including Australia, the United States, and Sweden,[23-25] a key challenge in the assessment of signal detection methods for CA data is that there is no "gold standard" for classifying risks according to specific CAs.[26,27] The

Australian government Department of Health provides an online database of recorded pregnancy related risks associated with medicines,[25] and this categorisation system was used to independently identify medications for which there has been evidence of high teratogenic risk. All medications are divided into five lettered categories, with category A medications being considered safe for use during pregnancy. Medications in category B are those which have not shown evidence of harmful effects to human foetuses; category C medications may carry harmful effects to human foetuses, but with no evidence of causing CAs. Categories D and X medications are believed to increase the frequency of human foetal malformations, carrying a moderate to high-risk. The Australian classification system does not distinguish risks for specific CAs. The total number of "high-risk" medications identified by each model was compared, as well as the proportion of medications identified as potential signals out of the total number of high-risk medications in the data. This is called the identification rate, and is defined as follows:

Identification rate
$$= \frac{\text{Number of "high-risk" medications identified as potential signals}}{\text{Total number of "high-risk" medications in the data}}$$

The proportion of category A medications being identified as potential signals was also considered, as a measure of the likely number of "false positive" associations. The total number of unique medications in the set of potential signals identified by each method is referred to as the "effective workload." Note that we refer only to *potential* signals as the aim of this study is to assess signal generation methods for CA data, and we do not further evaluate the potential signals here (since this has already been done for this dataset[28]).

3 | RESULTS

3.1 | Description of signal detection dataset

Data on 15 058 malformed foetuses were available for analysis with 55 CAs in 16 organ system groups (see Table S1): an average of 4.5 CAs per group. Half of the groups had only one CA, and the largest group (congenital heart disease) included 17 CAs (specific heart anomalies). There were 1.6 recorded medication exposures per pregnancy on average, ranging from one (in 65% of cases) up to 16 (in one case). The number of ATC medications with at least three exposures in the data was 523, of which 39 (7.5%) were coded only to ATC4. The total number of recorded exposures to these medications was 22 624. There were 28 765 potential medication-CA combinations (523 medications × 55 CAs) and 116 ATC3 groups, with an average of nine (range 1-20) medications and 487 (range 53-1086) medication-CA combinations per group.

Of the 523 medications in the signal detection analyses, 44 (8.4%) were high-risk, 297 (57%) were "low risk" (of which 77 were category A medications), and the other 182 medications (35%) were

not present in the Australian categorisation system database (coded as "unclassified risk"). Three medications mapped to a code in both the low-risk and high-risk group depending on their dosage; there is no information on dosage in EUROmediCAT data, so these medications were assigned to the unclassified risk category. Of 116 ATC3 groups, 94 (81%) contained no high-risk medications, 13 (11%) contained one high-risk medication, and nine (8%) groups contained at least two high-risk medications.

3.2 | Signal detection analysis

Table 2 and Figure 1 present key results from the four BHMs and the double FDR procedure, according to the two thresholds used to define potential signals for BHMs (95% and 99% PCIs) and for FDR cut-offs varying from 5% to 50%. A cut-off of 5% (FDR 5%), for example, means that up to 5% of the potential signals in the double FDR analysis might be expected to be false positive associations. The number of potential high-risk medication signals identified by each method is displayed in Figure 1, which plots the identification rate against the effective workload. Table 2 shows that the number of ATC3 groups with at least one potential signal decreased as the cut-off level for each method became stricter and fewer potential signals were identified. The effective workload is also shown for each method (in bold), including a breakdown by risk categories. Table 2 and Figure 1 show that a one-dimensional BHM with grouping by ATC3 and a 95% PCI cut-off resulted in the most potential signals overall and identified the most high-risk DX medications as potential signals (identification rate = 48%). However, this model also resulted in a very high effective workload (n = 160), with 30% of all medications identified as potential signals and at least one potential signal in over half the ATC3 groups. Individual BHMs and those grouping only by CA gave similar results for a 95% PCI cut-off, though with fewer high-risk signals identified and higher effective workloads. Using a stricter 99% PCI cut-off to define potential signals always resulted in a lower identification rate, but a higher proportion of the potential signals being for high-risk medications. The proportion of category A medications identified as potential signals was lower than the high-risk proportion for double FDR, but higher than the high-risk proportion for the majority of BHMs. The "strictest" model was double FDR, especially at lower thresholds, for example, double FDR 5% identified only three potential signals (two high-risk). A BHM grouping both medications and CAs with a 95% PCI threshold identified four more high-risk medications (identification rate 23%) than double FDR 50% (identification rate 14%); however, this was at the expense of more than a 4-fold increase in effective workload, from 16 medications to 71, for a gain in identification rate of only 9%. When using the 99% PCI cut-off to define potential signals, the two-dimensional BHM did not perform as well as the double FDR 50%.

Table S3 presents the overlap between the 16 potential medication signals from double FDR 50% and potential medication signals in BHMs. BHMs with a 95% PCI and grouping by CAs or by both CAs

and ATC3 each did not include one of the double FDR medication signals, although the double FDR signal not present was for a different medication in the two BHMs. The 99% PCI cut-off excluded more double FDR signals, except for a BHM with no grouping and a 99% PCI cut-off, which included all 16 potential signals from double FDR. Results from the alternative one-dimensional models (averaging over the ungrouped dimension) are presented in Table S4. These models resulted in low effective workloads and identification rates, more comparable with those using a double FDR. Double FDR 50% and its "equivalent" one-dimensional BHM with grouping by ATC3 and averaging over CAs resulted in 16 and 15 medications being identified as potential signals, respectively (Table S4). However, two more high-risk medications (six vs four) were identified by double FDR.

The number of "less harmful" associations according to each method is presented in Table S5, along with the total number of combinations identified as potential signals (as shown in Table 2) for comparison purposes. Double FDR resulted in only three "less harmful" associations; however, all of the BHMs resulted in a considerable number (up to 23 or 69 for models using a 99% or a 95% PCI threshold, respectively).

3.3 | Different potential signals according to different methods: The effect of shrinkage in BHMs

As well as the overall number of potential signals, differences in which medication-CA combinations were identified as potential signals in the different methods were apparent. One situation where this occurred was due to shrinkage to the null, where a potential signal attenuates in some BHMs due to the influence of other combinations in that group, as demonstrated in Figure S1. Shrinkage to the mean can also occur if a strong association influences other combinations in its group to create additional potential signals that are not present without this shrinkage, for example, see Figure S2.

4 | DISCUSSION

Many BHMs considered in this study identified more of the high-risk medications (higher identification rate) than the double FDR; however, these improvements came at the expense of a substantial increase in the effective workload, and therefore lower proportions of the potential signals being for high-risk medications. Considering the simplicity of the double FDR method, we recommend that the double FDR method continue to be used in practice for the detection of potential signals of teratogenic medications using EUROmediCAT data.

4.1 | Different potential signals according to FDR and BHM approaches

BHMs incorporating information sharing could identify a greater number of potential high-risk medication signals than double FDR and

TABLE 2 Summary of results from Fisher's exact test with various adjustments for multiple testing and from BHMs with a Poisson distribution, according to different groupings of ATC-coded medications, and/or CAs

Type of Model and Grouping	Cut-off Level/ Threshold Used to Define Potential Signals	Medication-CA Combinations Identified as Potential Signals (Out of n = 28 765)[a]	Number of Unique Medication Associations (Out of n = 523 Medications)				Effective workload[a]	"High-risk" Proportion[b] (8% of all medications)	Identification Rate[c]	"False Positive" Proportion[d] (15% of all medications)	CAs with at Least One Signal (out of n = 55 CAs)	ATC3 Groups with at Least One Signal (out of n = 116 ATC3 groups)
			"Low Risk"		"High-Risk" D or X (n = 44)	"Unclassified Risk" (n = 182)						
			A (n = 77)	B or C (n = 220)								
Individual BHMs: no grouping	95% PCI	223	25	73	17	44	159	11%	39%	16%	48	69
	99% PCI	53	9	9	7	18	43	16%	16%	21%	24	28
One-dimensional BHM; grouping by ATC3	95% PCI	210	24	52	21	46	143	15%	48%	17%	52	81
	99% PCI	55	9	8	6	15	38	16%	14%	24%	26	27
One-dimensional BHM; grouping by CA	95% PCI	216	18	66	18	58	160	11%	41%	11%	46	64
	99% PCI	50	6	8	9	15	38	24%	20%	16%	25	31
Two-dimensional BHM; grouping by ATC3 and CA	95% PCI	112	14	19	10	28	71	14%	23%	20%	36	46
	99% PCI	24	4	2	3	7	16	19%	7%	25%	14	11
Fisher's test and double FDR with ATC3 grouping	FDR 50%[e]	25	3	1	6	5	16	38%	14%	19%	14	7
	FDR 30%[e]	19	3	0	6	5	15	40%	14%	20%	11	7
	FDR 20%[e]	13	3	0	5	3	11	45%	11%	27%	8	4
	FDR 10%	7	0	0	4	1	5	80%	9%	0%	5	2
	FDR 5%	5	0	0	2	1	3	67%	5%	0%	4	2

Abbreviations: ATC, Anatomical Therapeutic Chemical; BHM, Bayesian hierarchical model; CA, congenital anomaly; FDR, false discovery rate; PCI, posterior confidence interval.

[a]Note that the effective workload is lower than the total number of combinations identified as potential signals, since a medication can be a potential signal in combination with more than one CA.

[b]"High-risk" proportion: the proportion of all potential medication signals that are high-risk.

[c]Identification rate: the proportion of high-risk medications that are identified as potential signals, out of the 44 high-risk medications in the dataset.

[d]"False positive" proportion: the proportion of potential signals that were marked as category A medications (considered to be safe for use during pregnancy).

[e]FDR cut-off levels were assessed in 5% increments from 5% to 50%, but some cut-off levels resulted in the same number of signal were identified. within the ranges 15%-25%, 30%-40%, and 45%-50% the same values for this table;
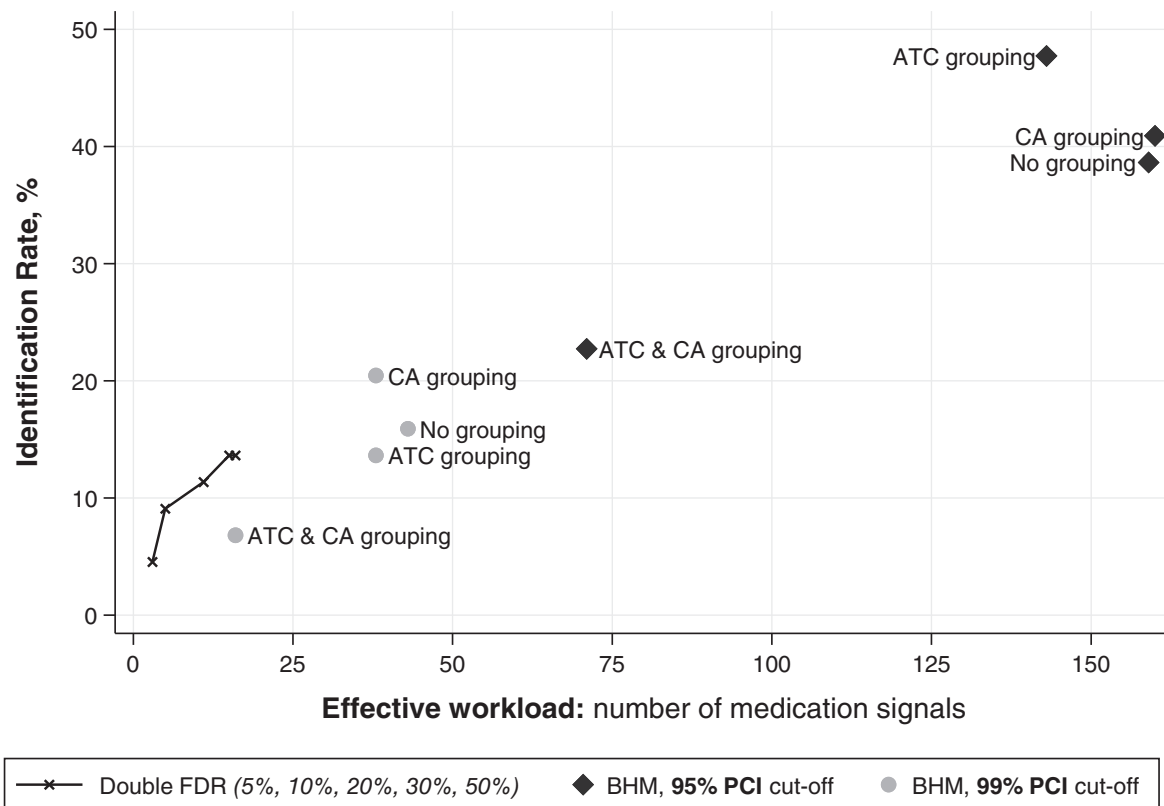
FIGURE 1   Identification rate vs effective workload: comparing the use of the double FDR procedure with Poisson Bayesian hierarchical models (BHMs) using a 95% PCI and a 99% PCI as a cut-offs for definition of signals, for four types of grouping by congenital anomaly (CA) and/or ATC medications

strengthened the analysis of combinations with low cell counts by using information in the surrounding cells, allowing them to be included in the set of resulting potential signals (Table S5). In contrast, the frequentist EUROmediCAT approach requires at least three exposures to identify any association as a potential signal. This detection of a potential medication signal for a newer drug with little data is one of the potential advantages of BHMs in signal detection. Further investigation is warranted to determine how likely it is that the additional potential signals detected by BHMs are true associations and whether alternative model specifications could improve the power of BHMs in this context. BHMs will be more powerful if similar drugs do have similar teratogenic effects or if CAs in the same organ system are affected by the same drugs.

4.2 | Evaluation and comparison of methods

The lack of existing knowledge regarding the teratogenic effect of medications used during pregnancy makes it difficult to evaluate how many teratogens are missed by each method and the possible reasons for any lack of detection. A key limitation of using the Australian classification system is that high-risk medications are not identified in association with a specific CA. In addition, categorisation of medications as B or C in the Australian database may indicate a lack of

evidence rather than meaning these medications are really low risk for CAs. Furthermore, almost a third of the medications in the EUROmediCAT data were not present in the Australian risk classification database. In practice, teratogenic risk is nearly always specific to certain CAs.[29] This may have affected our "identification rate," which does not reflect the number of different CAs that a medication is associated with. In this analysis, any associations arising due to confounding by indication cannot be identified. Our use of the risk categorisation system here was not to judge the absolute strengths of a signal detection procedure, but rather to directly compare methods in terms of the volume of potential signals and assessment of resulting workload for follow up of identified associations; the limitations identified above should therefore be present across all models considered.

4.3 | Methodological considerations

An important assumption of the Poisson distribution is that events in the data occur independently. However, in EUROmediCAT, a malformed foetus often has multiple CAs and/or medication exposures and certain CAs may be more likely to co-occur within pregnancies. Similarly, exposure to a specific medication may increase the likelihood of exposure to another medication, for example, it is common to take several different asthma medications together. It may also be the

case that if one particular medication is taken, any other medications in that group will not be taken. In addition, twins with the same anomalies will violate the independence assumption. Approximately 3% of foetuses in EUROmediCAT data are twin births, 5% of whom have the same anomalies.[30] Therefore, the occurrence of twins will not be a serious violation of the independence assumption here. The occurrence of twins is further considered when following up identified associations in greater detail.

Another potential consideration for our two-dimensional BHMs is the way in which information sharing is specified around the medication-CA combination of interest. It may be argued that our BHM allows the count for the medication-CA combination of interest (ie, the dark grey cell in Table 1) to contribute the model twice: first, to generate a hypothesis about the data via the prior distribution, and second, to test it. One solution might be to consider a model where the count for the combination of interest itself is removed from the prior distribution for that combination. Table 1 presented our two-dimensional model, where the prior for any combination includes information from all mediations and CAs within that set; an alternative formulation could be to include only those combinations with either the medication or CA in common with the combination of interest. These possibilities require further investigation.

In any signal detection analysis using disproportionality measures, reporting biases for a common medication may lead to inflation in the overall rates for that medication, meaning that other associations in the database are masked.[27,31,32] This is not thought likely to occur in the EUROmediCAT as medication exposure is generally collected before the CA diagnosis. Masking may also be an issue in EUROmediCAT data through the use of malformed controls if a proportion of the control group is related to the medication of interest. Studies have demonstrated that the removal of a masking effect may help lead to new signals of public health relevance being discovered.[31,33] It is also thought, however, that significant masking is not common in large spontaneous reporting databases, and where present it mostly affects rarely reported AEs.[34-36] Confounding by co-reported medications can also occur if two medications are frequently prescribed together but only one causes the CA of interest.[32] We may expect a teratogen to act in a similar way regardless of where it is taken; however, certain medications may have varying usages and/or availability in different EUROmediCAT registries and countries. As many medication-CA combinations have very small numbers, the best approach to an ongoing signal detection process is considered to be investigation of any potential registry effects at a later stage in the analysis; as such, after potential signals are generated, the next step of the EUROmediCAT signal detection process includes the adjustment of estimates for confounding by registry.[28]

This analysis excludes all chromosomal anomalies; these anomalies could theoretically be analysed as a negative control outcome as no medications are expected to be associated with any chromosomal anomalies. However, the risk of a chromosomal anomaly is strongly associated with maternal age, and methods to adjust for this confounder in signal detection analyses would need to be developed.

4.4 | Strengths and limitations of EUROmediCAT data

The existing EUROCAT network, upon which EUROmediCAT is based[11], ensures that CAs are coded in a detailed and standardised manner across all registries. Good agreement between medication exposures recorded in the EUROmediCAT database and those actually used has also been demonstrated.[37] As maternal medication exposure data in EUROmediCAT registries is primarily obtained through prospectively recorded maternity records, confounding by the time of pregnancy registration of adverse outcomes is unlikely to have occurred. On the other hand, there is known under ascertainment for certain medication exposures in EUROmediCAT data, which may reduce the sensitivity of any signal detection analysis.[10,38] There is also a lack of information regarding the dosage and precise timing of medication exposures. Although the critical period of development for most major CAs occurs during the second and third gestational months, the exact timing can differ according to the type of CA and some CAs may also develop after the first trimester of pregnancy.[39-41] In this study, we cannot determine whether medications were taken during the particular critical period for development of each specific CA. As only malformed foetuses exposed to at least one medication are included, it is not possible to estimate the relative risk of a CA for medication exposures compared with a healthy (ie, non-exposed and non-malformed) control population.[42]

5 | SUMMARY AND CONCLUSIONS

Despite the difficulties in assessing the performance of the signal detection methods, we recommend the double FDR method for continued use in signal detection analyses of EUROmediCAT data.

CONFLICT OF INTERESTS
The author(s) have no conflicts of interest to declare.

ORCID
Alana Cavadino https://orcid.org/0000-0002-5709-367X

REFERENCES

1. Morgan M, De Jong-van den Berg LT, Jordan S. Drug safety in pregnancy—monitoring congenital anomalies. *J Nurs Manag*. 2011;19 (3):305-310.

2. Luteijn JM, Morris JK, Garne E, Given J, de Jong-van den Berg L, Addor MC, et al. EUROmediCAT signal detection: a systematic method for identifying potential teratogenic medication. *Br J Clin Pharmacol*, 2016, 82, 1110, 1122.

3. Cavadino A, Prieto-Merino D, Morris JK. Identifying signals of potentially harmful medications in pregnancy: use of the double false discovery rate method to adjust for multiple testing. *Br J Clin Pharmacol*. 2019;85(2):356-365.

4. Suling M, Pigeot I. Signal detection and monitoring based on longitudinal healthcare data. *Pharmaceutics*. 2012;4(4):607-640.

5. Crooks CJ, Prieto-Merino D, Evans SJ. Identifying adverse events of vaccines using a Bayesian method of medically guided information sharing. *Drug Saf*. 2012;35(1):61-78.

6. Deshpande G, Gogolak V, Smith SW. Data mining in drug safety: review of published threshold criteria for defining signals of disproportionate reporting. *Pharmaceutical Medicine*. 2010;24(1):37-43.

7. Brook M. *Bayesian Hierarchical Methods for Detection of Adverse Reactions to Drugs in Large Databases using Hierarchies of Drugs and Adverse Events*. London: London School of Hygiene and Tropical Medicine; 2011.

8. Greenlees R, Neville A, Addor MC, et al. Paper 6: EUROCAT member registries: organization and activities. *Birth Defects Res A Clin Mol Teratol*. 2011;91(Suppl 1):S51-s100.

9. Loane M, Dolk H, Garne E, Greenlees R. Paper 3: EUROCAT data quality indicators for population-based registries of congenital anomalies. *Birth Defects Res A Clin Mol Teratol*. 2011;91(Suppl 1):S23-S30.

10. Bakker M and Jonge Ld. EUROCAT Special Report: Sources of Information on Medication Use in Pregnancy. 2014; Available from: http://www.eurocat-network.eu/content/Special-Report-Medication-Use-In-Pregnancy.pdf.

11. Boyd PA, Haeusler M, Barisic I, Loane M, Garne E, Dolk H. Paper 1: the EUROCAT network—organization and processes. *Birth Defects Res A Clin Mol Teratol*. 2011;91(Suppl 1):S2-S15.

12. EUROCAT Central Registry. EUROCAT Guide 1.4: instructions for the registration and surveillance of congenital anomalies. 2013, EUROCAT Central Registry, University of Ulster

13. WHO Collaborating Centre for Drug Statistics Methodology. ATC Structure and principles; 2011; Available from: http://www.whocc.no/atc/structure_and_principles/ (accessed April 21, 2017).

14. WHO Collaborating Centre for Drug Statistics Methodology. ATC alterations from 1982-2016. December 18, 2015 2015 [cited August 11, 2016]; Available from: http://www.whocc.no/atc_ddd_alterations__cumulative/atc_alterations/.

15. Mehrotra DV, Heyse JF. Use of the false discovery rate for evaluating clinical safety data. *Stat Methods Med Res*. 2004;13(3):227-238.

16. Mehrotra DV, Adewale AJ. Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals. *Stat Med*. 2012;31(18):1918-1930.

17. StataCorp. Stata Statistical Software. *Release 12*. StataCorp LP: College Station, TX; 2011.

18. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*. 2001;10(6):483-486.

19. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on an article by Browne and Draper). *Bayesian Anal*. 2006;1(3):515-533.

20. Plummer M. JAGS. A program for analysis of Bayesian graphical models using Gibbs sampling. *The 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*; Vienna, Austria: Published online in the DSC 2003 proceedings; 2003.

21. Development Core R, Team R. *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2008.

22. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. *R News*; Austria: The R Foundation; 2003:7-11.

23. Tronnes JN, Lupattelli A, Nordeng H. Safety profile of medication used during pregnancy: results of a multinational European study. *Pharmacoepidemiol Drug Saf*. 2017;26(7):802-811.

24. Briggs GG, Freeman RK, and Yaffe SJ. Drugs in Pregnancy and Lactation: a Reference Guide to Fetal and Neonatal Risk. 2012: Lippincott Williams & Wilkins.

25. Australian Government Department of Health. Prescribing medicines in pregnancy database. 2016 (cited 2016 19th August); Available from: https://www.tga.gov.au/prescribing-medicines-pregnancy-database.

26. Stephenson WP, Hauben M. Data mining for signals in spontaneous reporting databases: proceed with caution. *Pharmacoepidemiol Drug Saf*. 2007;16(4):359-365.

27. Almenoff J, Tonning JM, Gould AL, et al. Perspectives on the use of data mining in pharmacovigilance. *Drug Saf*. 2005;28(11):981-1007.

28. Given JE, Loane M, Luteijn JM, et al. EUROmediCAT signal detection: an evaluation of selected congenital anomaly-medication associations. *Br J Clin Pharmacol*. 2016;82(4):1094-1109.

29. Mitchell AA. Research challenges for drug-induced birth defects. *Clin Pharmacol Ther*. 2016;100(1):26-28.

30. Boyle B, McConkey R, Garne E, et al. Trends in the prevalence, risk and pregnancy outcome of multiple births with congenital anomaly: a registry-based study in 14 European countries 1984–2007. *BJOG*. 2013;120(6):707-716.

31. Gould AL. Practical pharmacovigilance analysis strategies. *Pharmacoepidemiol Drug Saf*. 2003;12(7):559-574.

32. Hauben M, Madigan D, Gerrits CM, Walsh L, and Van Puijenbroek EP. The role of data mining in pharmacovigilance. *Expert Opin Drug Saf*, 2005. 4(5): p. 929–48.

33. Pariente A, Avillach P, Salvo F, et al. Effect of competition bias in safety signal generation: analysis of a research database of spontaneous reports in France. *Drug Saf*. 2012;35(10):855-864.

34. Maignen F, Hauben M, Hung E, Van Holle L, and Dogne JM. Assessing the extent and impact of the masking effect of disproportionality analyses on two spontaneous reporting systems databases. *Pharmacoepidemiol Drug Saf*, 2014. 23(2): p. 195–207.

35. Wang HW, Hochberg AM, Pearson RK, Hauben M. An experimental investigation of masking in the US FDA adverse event reporting system database. *Drug Saf*. 2010;33(12):1117-1133.

36. Zeinoun Z, Seifert H, Verstraeten T. Quantitative signal detection for vaccines: effects of stratification, background and masking on GlaxoSmithKline's spontaneous reports database. *Hum Vaccin*. 2009; 5(9):599-607.

37. de Jonge L, de Walle HE, de Jong-van den Berg LT, van Langen IM, Bakker MK. Actual use of medications prescribed during pregnancy: a cross-sectional study using data from a population-based congenital anomaly registry. *Drug Saf*. 2015;38(8):737-747.

38. de Jonge L, Garne E, Gini R, et al. Improving information on maternal medication use by linking prescription data to congenital anomaly registers: a EUROmediCAT study. *Drug Saf*. 2015;38(11):1083-1093.

39. Czeizel AE. Drug exposure in pregnant women. *Lupus*. 2004;13(9): 740-745.

40. Czeizel AE. Specified critical period of different congenital abnormalities: a new approach for human teratological studies. *Congenit Anom (Kyoto)*. 2008;48(3):103-109.

41. Scheuerle AE, Aylsworth AS. Birth defects and neonatal morbidity caused by teratogen exposure after the embryonic period. *Birth Defects Res A Clin Mol Teratol*. 2016;106(11):935-939.

42. Prieto L, Martinez-Frias ML. Case-control studies using only malformed infants who were prenatally exposed to drugs. What do the results mean? *Teratology*. 2000;62(1):5-9.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.