CrossMark

ORIGINAL RESEARCH ARTICLE

Drug Safety Monitoring in Children: Performance of Signal Detection Algorithms and Impact of Age Stratification

Osemeke U. Osokogu[1] · Caitlin Dodd[1] · Alexandra Pacurariu[1,3] · Florentia Kaguelidou[1,2] ·
Daniel Weibel[1] · Miriam C. J. M. Sturkenboom[1]

Abstract

Introduction Spontaneous reports of suspected adverse drug reactions (ADRs) can be analyzed to yield additional drug safety evidence for the pediatric population. Signal detection algorithms (SDAs) are required for these analyses; however, the performance of SDAs in the pediatric population specifically is unknown. We tested the performance of two SDAs on pediatric data from the US FDA Adverse Event Reporting System (FAERS) and investigated the impact of age stratification and age adjustment on the performance of SDAs.

Methods We tested the performance of two established SDAs: the proportional reporting ratio (PRR) and the empirical Bayes geometric mean (EBGM) on a pediatric dataset from FAERS (2004–2012). We compared the performance of the SDAs with a published pediatric-specific reference set by calculating diagnostic test-related statistics, including the area under the curve (AUC) of receiver operating characteristics. Impact of age stratification and age-adjustment on the performance of the SDAs was assessed. Age adjustment was performed by pooling (Mantel-Hanszel) stratum-specific estimates.

Results A total of 115,674 pediatric reports (patients aged 0–18 years) comprising 893,587 drug–event combinations (DECs) were analysed. Crude values of the AUC were similar for both SDAs: 0.731 (PRR) and 0.745 (EBGM). Stratification unmasked four DECs, e.g., 'ibuprofen and thrombocytopenia'. Age adjustment did not improve performance.

Conclusion The performance of the two tested SDAs was similar in the pediatric population. Age adjustment does not improve performance and is therefore not recommended to be performed routinely. Stratification can reveal new associations, and therefore is recommended when either drug use is age-specific or when an age-specific risk is suspected.

Osemeke U. Osokogu and Caitlin Dodd contributed equally to this article.

✉ Caitlin Dodd
c.dodd@erasmusmc.nl

[1] Department of Medical Informatics, Erasmus University Medical Center, Rotterdam 3015 CN, The Netherlands

[2] Department of Pediatric Pharmacology and Pharmacogenetics, Hôpital Robert Debré, APHP, Univ Paris 7-Diderot, Sorbonne Paris Cité, EA08, INSERM CIC1426, Paris 75019, France

[3] Dutch Medicines Evaluation Board, Utrecht, The Netherlands

Key Points

Detection of drug safety signals in children, who represent a heterogeneous population, where age may be a confounder or effect modifier, is an area in which only limited research has been carried out.

The signal detection algorithms (SDAs) showed good performance on pediatric data and can be utilized for pediatric signal detection.

Age adjustment did not improve the performance of the SDAs.

Age stratification showed that some signals may be detected only in specific pediatric age groups. For routine surveillance, checking for effect modification across age strata may generate useful information.

△ Adis

1 Introduction

Spontaneous reports of suspected adverse drug reactions (ADRs) can yield important information regarding the safety of drugs [1]. Usually, such reports are screened for emerging safety issues by applying statistical methods called signal detection algorithms (SDAs). Current SDAs compare the reporting rate of a drug–event combination (DEC) of interest with the expected count calculated from the overall reporting rate of that reaction in the entire database [1, 2]. Although SDAs are routinely applied to reports pertaining to the general population, the performance of SDAs in the pediatric population specifically has not been investigated to date. Compared with adults, the pattern of drug use and occurrence of ADRs in pediatric patients may differ [3–5] since the latter population comprises a heterogeneous group of subjects at various stages of development with age-dependent organ maturation and hormonal changes [6]. Several studies investigating ADR reporting in children have identified different reporting patterns in this population than in adults [3, 5, 7, 8]. Since ADRs may be age specific, adjustment for age seems to be a logical step when investigating pediatric ADRs and has been advocated by some researchers [4]. The major aim of stratification is verification of confounding and effect modification which otherwise may mask true signals [9]. Confounding by age can be dealt with by stratifying for age categories and pooling stratum-specific estimates. However, if age-specific estimates differ (in case of effect modification) pooling/adjustment should not be done; instead, a verification of each individual stratum should be performed. While stratification has been investigated by some researchers [10], adjustment is routinely implemented in some Bayesian but not in frequentist SDAs [11–13]. Few studies have systematically addressed the impact of age stratification or adjustment and the results are contradictory [9, 14, 15].

Within the context of the Global Research in Pediatrics (GRiP) Network of Excellence [16], we aimed to evaluate the performance of two well-established SDAs in the pediatric population and determine if age stratification or adjustment impacts signal detection in this population.

2 Methods

2.1 Data Source

Data were retrieved from the publicly available version of the US FDA Adverse Event Reporting System (FAERS), which comprises spontaneous reports of suspected ADRs submitted by manufacturers, healthcare professionals, and patients. FAERS is one of the largest repositories of spontaneous reports in the world [17, 18]. In this study, we analyzed reports received from the first quarter of 2004 through to the third quarter of 2012.

For performance analysis, only reports of ADRs occurring in children and adolescents (<18 years of age) were retained. The ADRs in FAERS are coded according to the Medical Dictionary for Regulatory Activities (MedDRA®) [19].

To improve the quality of the dataset, we excluded reports with missing age, the main variable in our study. Also, reports with reported age equal to zero and with a MedDRA® preferred term indicating prenatal exposure were removed, as these imply in utero drug exposure and were therefore not relevant for our study. We minimized the number of duplicates (i.e., the same report submitted by different reporters) by applying an algorithm based on case identifier, report identifier, and drug and event names. For multiple reports (i.e., the same report is reported at a later time, with additional and updated information) [20], the most recent (and most updated) report was retained for analysis.

As drug names included in FAERS are not standardized, a harmonization procedure was implemented. Briefly, this consisted of removing superfluous characters and applying a generalized edit distance matching algorithm [21] to map free-text drug names to synonyms and finally to the corresponding active substance and World Health Organization–Anatomic Therapeutic Chemical (WHO-ATC) code.

In this study, only those drugs reported as the primary or secondary suspect in the FAERS database were retained for analysis. Analysis was performed at DEC level, meaning that within each report, every suspect drug was combined with all reported ADRs. Thus, one report may comprise more than one DEC.

2.2 Signal Detection Algorithms (SDAs)

We tested two well-established SDAs that are routinely used by various national and international regulatory and/or research institutions for signal detection: the proportional reporting ratio (PRR) [2] and the empirical Bayes geometric mean (EBGM) [13] (see Table 1). We also tested count of reports as a positive control. In order to define a signal of disproportionate reporting [22, 23], we selected thresholds that are currently applied in routine practice. We applied the SDAs at the end of the study period, when the maximum number of reports had accrued.

2.3 Performance Assessment Measures

The performance of the SDAs was assessed by calculating diagnostic test-related statistics, namely specificity and

Table 1 Signal detection algorithms and corresponding thresholds applied

Signal detection algorithm	Applied threshold[a]	Institution where the method and the respective threshold is currently used
Number of reports	≥5	NA
PRR	PRR lower bound 95 % CI ≥1 and $n \geq 5$ reports	European Medicines Agency (EMA)
EBGM	EB05 CI ≥1.8, $n \geq 3$ reports, and EBGM ≥2.5	Medicines and Healthcare products Regulatory Agency (MHRA)

CI confidence interval, *EB05* lower bound of the 95 % confidence interval, *EBGM* empirical Bayes geometric mean, *NA* not available, *PRR* proportional reporting ratio

[a] Thresholds were obtained from Candore et al. [23]

sensitivity, positive predictive value (PPV), and negative predictive value (NPV) [24, 25]. Sensitivity is the ability of the method to identify true signals correctly, while specificity is the ability to exclude false signals correctly. PPV and NPV are posterior probabilities, describing how many of the signals classified as positive or negative are indeed correctly classified [24, 25].

Since diagnostic test-related statistics are dependent on the threshold choice, their individual comparison has only a limited, albeit practical, value. Therefore, we also estimated the area under the curve (AUC) of receiver operating characteristics (ROC) in order to compare the performance of the SDAs [26]; the AUC incorporates both sensitivity and specificity across all the possible values for a certain SDA. Calculation of AUCs was conducted by varying only the point estimate of each SDA and did not take into account the other components of the SDA.

For the purpose of performance evaluation, a previously constructed pediatric-specific GRiP reference set of positive and negative DECs was used. It consists of 37 positive and 90 negative DECs and includes drugs that are administered to children and events that are regarded as important for this population. The positive DECs are those that were confirmed to occur based on evidence from Summary of Product Characteristics (SmPC) and the published literature, while the negative DECs are those that could not be confirmed at the time of literature review by either the SmPC or the published literature. For a full description of the reference set, see Osokogu et al. [27].

2.4 Stratification and Adjustment for Age

The impact of age stratification and adjustment on the performance of the SDAs was investigated. First, we checked for possible effect modification across age strata, by stratifying the data according to age categories defined by the International Conference on Harmonization (ICH) [28] and calculating stratum-specific measures for each SDA. Secondly, we calculated age-adjusted estimates for PRR and EBGM by combining the stratum-specific estimates in an overall measure [29]. The performance of each SDA was reassessed after adjustment.

2.5 Statistical Analysis

Differences in the performance (AUC) of each SDA, crude versus age-adjusted and crude versus count of reports (positive control) were tested using paired chi-squared tests. Stratum-specific contingency tables were tested for homogeneity using the Breslow Day Tarone test [30]. The Mantel-Haenszel approach was used for pooling and calculating age-adjusted estimates [29]. The lower bound of the EBGM 95 % confidence interval (EBGM05) was calculated using the lower bound of the 95 % confidence interval (EB05) for each stratum and then computing a Mantel-Haenszel average based on Zeinoun et al. [31]. Statistical significance was defined by $p < 0.05$.

Analysis was performed using SAS® software version 9.2 (SAS Institute, Cary, NC, USA). Graphs were made in SAS® software version 9.2 and R version 3.1.3.

3 Results

3.1 Descriptive Analysis

For the study period (first quarter of 2004 through to the third quarter of 2012), a total of 4,285,088 reports were retrieved from FAERS. After eliminating duplicates ($n = 43,125$) and removal of adult reports ($n = 2,686,530$) and reports with missing age ($n = 1,419,524$) or age equal to zero with a MedDRA® preferred term indicating pre-natal exposure ($n = 20,235$), 115,674 reports corresponding to 893,587 individual DECs were retained for analysis of pediatric spontaneous reports (see Table 2).

The total number of pediatric reports that included the investigated drugs and ADRs from the reference set can be observed in Fig. 1, which also shows data regarding adults (for comparison purposes). The number of children

Table 2 Description of pediatric reports by age categories

Age group	Number of reports [n (%)]
Neonates: 0–27 days	5091 (4.40)
Infants: 28 days–23 months	12,566 (10.86)
Children: 2–11 years	49,982 (43.21)
Adolescents: 12–17 years	48,035 (41.53)
Total	115,674 (100)

exposed to the drugs of interest, for whom any of the investigated ADRs was reported, varied from 26 patients (for praziquantel) to 7535 patients (for ibuprofen), with a median of 781 patients exposed across all drugs. The number of events of interest in FAERS ranged from 164 reports (ventricular arrhythmia) to 14,777 (anaphylaxis), with a median of 1004 reports across all events. For a more detailed description of reports counts please refer to Electronic Supplementary material Table 1.

3.2 Overall Performance of SDAs

Both SDAs showed high specificity and low sensitivity. They both had similar specificity values (PRR: 83.8 % and EBGM: 91.9 %), while sensitivity was lower for EBGM than for PRR (17.2 vs. 37.9 %). The NPV and PPV were similar for both SDAs. When we applied the threshold-independent (AUC-based) approach, the tested SDAs showed similar performance in the pediatric population, although the AUC value for EBGM (0.745) was slightly higher than for PRR (0.731). None of the SDAs performed better than the simple report count (AUC = 0.634; $p = 0.27$ for PRR and $p = 0.14$ for EBGM)

3.3 Stratification and Adjustment for Age and its Impact on Performance

Upon calculating SDA values per age stratum and testing for heterogeneity across strata, we observed effect modification for some associations. Some false negatives (positive DECs that failed to be highlighted as signals when analyzing data pertaining to the entire pediatric population) were unmasked in some strata. Four DECs were unmasked in total: ibuprofen–thrombocytopenia and isoniazid–seizure (by PRR) and clarithromycin–erythema multiforme and ibuprofen–erythema multiforme (by EBGM). Conversely, 'ibuprofen–acute liver injury', also a positive DEC, was highlighted when we analyzed data pertaining to the entire pediatric population, but it became clear after stratifying that this DEC was highlighted only in older children (adolescents) and not in younger children (see Fig. 2). For an overview of SDA values across age strata and results of heterogeneity tests please refer to Electronic Supplementary Material Figures 1A and 1B.

We evaluated the performance of the methods within individual age strata (see Table 3). On average, performance of the SDAs was lower within age strata than in the entire pediatric population and performance improved with increasing stratum size. For infants and neonates, the performance was very low, not better than chance ($p > 0.5$ for both SDAs). The adolescent group exhibited the best performance, which was similar to the overall performance.

After adjusting for age by pooling the stratum-specific estimates, the performance of the SDAs decreased, although not significantly (see Fig. 3; crude vs. adjusted AUC for PRR: 0.731 vs. 0.688, $p = 0.267$; crude vs. adjusted AUC for EBGM: 0.745 vs. 0.683, $p = 0.216$).

4 Discussion

In this study, we have demonstrated that age stratification for detection of drug safety signals in children may unmask some signals that do not appear in either crude or adjusted analysis. Adjustment for age does not improve performance of the PRR and EBGM.

For the investigated events, similar reporting patterns were observed for children and adults, while the investigated drugs appeared to have different reporting patterns (see Fig. 1). Different drug-related reporting patterns in children versus adults have been reported previously [5]. Consequently, reported DECs for children may differ from adults [3, 5], underlining the need for pediatric-specific approaches to signal detection, especially when we consider that reported drugs may vary by age group even within the pediatric population [3, 32].

Overall, the PRR and EBGM showed good performance, although results were slightly lower than results reported on other (not pediatric-specific) reference sets [32, 33]. The similarity in performance between PRR and EBGM is in accordance with recent results from the PROTECT (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium) project [23]. The fact that the performance (based on AUC) of PRR and EBGM was not statistically significantly better than simple report count may be due to the lack of power. Within age strata, performance seemed to correlate with stratum size: the poorest results were observed for infants and neonates (the smaller groups), slightly improving for children, while the best performance was observed for adolescents, the age stratum with the highest number of tested DECs. Decrease in power due to fewer reports and therefore DECs may account for this observation. The fact that we used lower bounds of confidence intervals for signaling instead of point estimates might have exacerbated the influence of sample size on the results, since smaller strata will have

Fig. 1 Count of reports in the pediatric and adult population for the investigated adverse drug reactions (**a**) and drugs (**b**), cumulatively for the period quarter 1 2004 to quarter 3 2012. The number of reports in children is represented by *bars* and plotted on the *left axis*, while the number of reports in adults is represented by the *red line* and plotted on the *right axis*; reports with missing age or age = 0 were excluded. Only reports mentioning any of the drugs or events in the reference set were considered
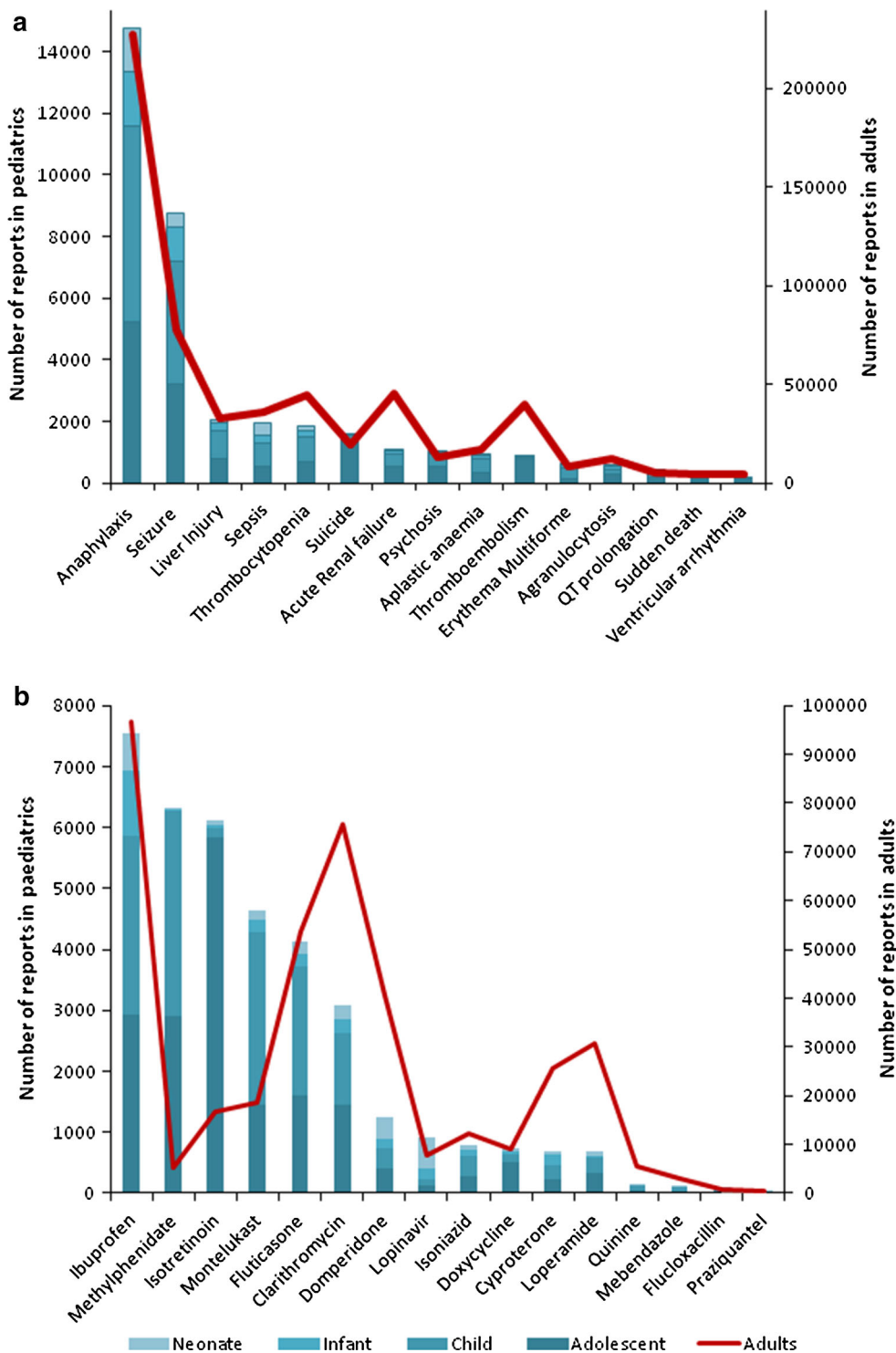
higher variability. In neonates and infants for whom expected counts were difficult to calculate because of few reports, we observed that simple report counts performed similar or even better than the SDAs and might be an alternative to commonly used SDAs. The fact that simple report count performed better than SDAs may have been because the reference set comprised known

DECs (which in turn may have influenced reporting) rather than emerging safety issues, a hypothesis proposed by Norén et al. [34].

Inspection of SDA values across child-specific strata (age stratification) revealed some heterogeneity in estimates, pointing to some effect modification. For example, 'ibuprofen–thrombocytopenia' was found as a signal in the
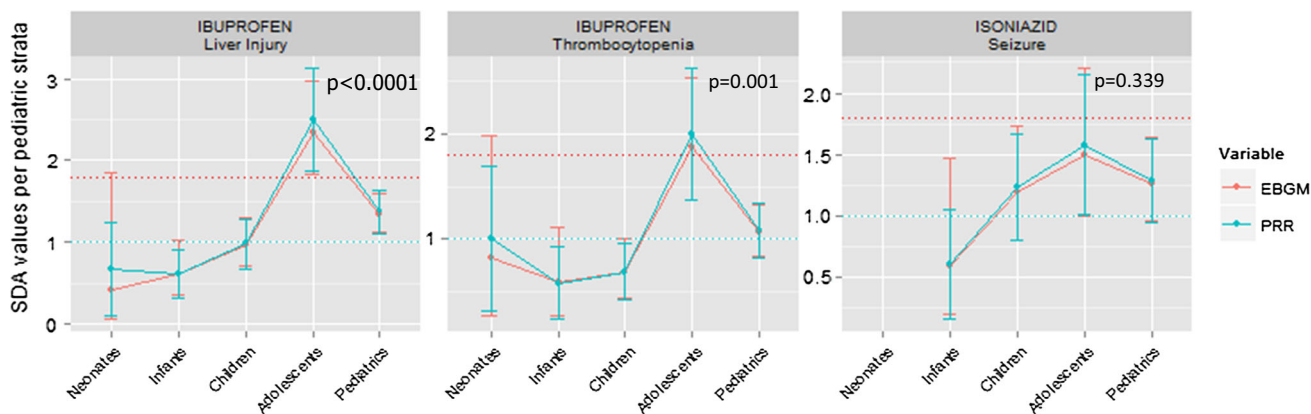
p-values were calculated with Breslow Day Tarone test for homogeneity

Fig. 2 Variation of proportional reporting ratio and empirical Bayes geometric mean estimates across pediatric specific strata—selected examples. *EBGM* empirical Bayes geometric mean, *PRR* proportional reporting ratio, *SDA* signal detection algorithm

Table 3 Performance of signal detection algorithms across age strata

Age groups and signal detection algorithms	Size of the age stratum (number of reports)	AUC
Neonates	5091	
Number of reports		0.625
EBGM		0.600
PRR		0.65
Infants	12,566	
Number of reports		0.667
EBGM		0.548
PRR		0.554
Children	49,982	
Number of reports		0.654
EBGM		0.698
PRR		0.649
Adolescents	48,035	
Number of reports		0.698
EBGM		0.771
PRR		0.718
Entire pediatric population		
Number of reports	115,674	0.634
EBGM		0.746
PRR		0.733

AUC area under the curve, *EBGM* empirical Bayes geometric mean, *PRR* proportional reporting ratio
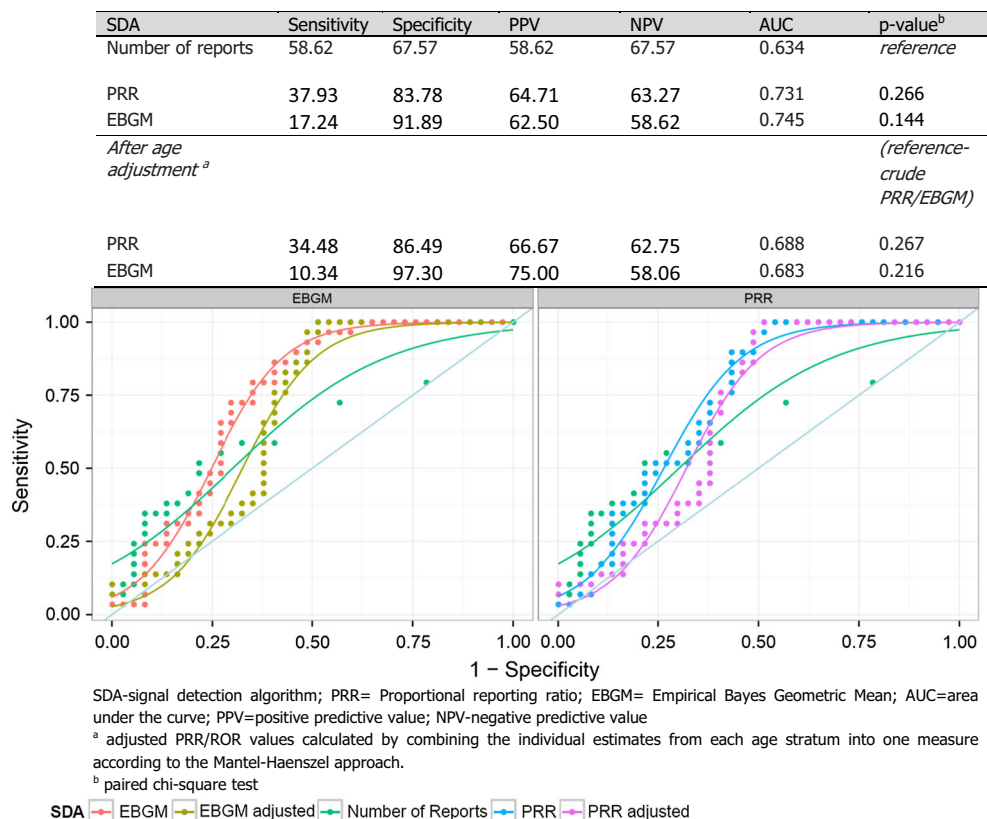
adolescents' group but not detected in the entire pediatric population or the younger age categories. This suggests that age-specific SDA calculations are sometimes needed, rather than age-adjusted SDA estimates. The age-adjusted estimates did not improve performance; in fact, even PPV unexpectedly decreased. Simulation studies have shown that when adjusted for strata, Bayesian methods such as EBGM tend to be underestimated when there are sparse strata [15]; this was also the case in our study. Previous studies in adults show contradictory results, with some showing a beneficial effect [9] while others did not [15]. The reason for our finding is not entirely clear; a possible explanation is that age is not a strong confounder for the investigated DECs. Also, the method of weighting (Mantel-Haenszel approach) may have played a role since more weight was assigned to age groups with more reports (adolescents and children). This may have masked signals occurring in age groups with fewer reports.

The limitations of data mining in FAERS include those inherent to spontaneous reporting databases: under-reporting, lack of denominator data and control group, biases in reporting, as well as missing and poor-quality data [35]. Missing information regarding age substantially reduced the study sample size since we could not determine whether these reports described patients aged less than 18 years old. While these biases are well acknowledged and have a definite impact, they cannot be completely avoided. Compared with adults, there are fewer reports and different reporting patterns for children [3, 36, 37], which may complicate signal detection in the pediatric population.

Evaluating performance of SDAs is a constant challenge due to lack of standard methodologies, imperfect reference standards, and uncertainty regarding the best thresholds (see the Electronic Supplementary Material for measures of performance using alternative thresholds). Some of the drugs and events in the reference set are specific to one age group within pediatrics and this is obvious in Fig. 1, even though the reference set was designed to be relevant for the entire pediatric population. We acknowledge that the reference set used, although specifically constructed for this

Fig. 3 Performance of signal detection algorithms within the entire pediatric population

SDA	Sensitivity	Specificity	PPV	NPV	AUC	p-value[b]
Number of reports	58.62	67.57	58.62	67.57	0.634	*reference*
PRR	37.93	83.78	64.71	63.27	0.731	0.266
EBGM	17.24	91.89	62.50	58.62	0.745	0.144
After age adjustment [a]						*(reference-crude PRR/EBGM)*
PRR	34.48	86.49	66.67	62.75	0.688	0.267
EBGM	10.34	97.30	75.00	58.06	0.683	0.216

SDA-signal detection algorithm; PRR= Proportional reporting ratio; EBGM= Empirical Bayes Geometric Mean; AUC=area under the curve; PPV=positive predictive value; NPV-negative predictive value
[a] adjusted PRR/ROR values calculated by combining the individual estimates from each age stratum into one measure according to the Mantel-Haenszel approach.
[b] paired chi-square test

SDA — EBGM — EBGM adjusted — Number of Reports — PRR — PRR adjusted

purpose, does not include all the ADRs that are highly specific for pediatrics. This highlights the need for pediatric-specific approaches to signal detection, accounting for not just the entire pediatric population but also the different age strata within pediatrics. Still, the reference set captures various drug use and ADRs patterns [38] and is currently the only available pediatric-specific reference set. The thresholds applied to define a signal were obtained from previous publications and other cut-off points may generate better results; further research on pediatric-specific thresholds should be encouraged.

5 Conclusion

Our study revealed that age adjustment did not improve the performance of the SDAs. However, stratification revealed some variation in the values of SDAs across strata (effect modification) and inspection of stratum-specific estimates might sometimes yield useful information during routine surveillance.

References

1. Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiol Drug Saf. 2009;18:427–36.
2. Evans SJW, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf. 2001;10:483–6.
3. Blake KV, Zaccaria C, Domergue F, La Mache E, Saint-Raymond A, Hidalgo-Simon A. Comparison between paediatric and adult suspected adverse drug reactions reported to the European Medicines Agency: implications for pharmacovigilance. Paediatr Drugs. 2014;16:309–19.
4. Star K, Edwards IR. Pharmacovigilance for children's sake. Drug Saf. 2014;37:91–8.
5. Star K, Norén GN, Nordin K, Edwards IR. Suspected adverse drug reactions reported for children worldwide: an exploratory study using VigiBase. Drug Saf. 2011;34:415–28.
6. Kearns GL, Abdel-Rahman SM, Alander SW, Blowey DL, Leeder JS, Kauffman RE. Developmental pharmacology—drug disposition, action, and therapy in infants and children. N Engl J Med. 2003;349:1157–67.
7. Aagaard L, Weber CB, Hansen EH. Adverse drug reactions in the paediatric population in Denmark: a retrospective analysis of reports made to the Danish Medicines Agency from 1998 to 2007. Drug Saf. 2010;33:327–39.
8. Kimland E, Rane A, Ufer M, Panagiotidis G. Paediatric adverse drug reactions reported in Sweden from 1987 to 2001. Pharmacoepidemiol Drug Saf. 2005;14:493–9.
9. Woo EJ, Ball R, Burwen DR, Braun MM. Effects of stratification on data mining in the US Vaccine Adverse Event Reporting System (VAERS). Drug Saf. 2008;31:667–74.
10. Orre R, Lansner A, Bate A, Lindquist M. Bayesian neural networks with confidence estimations applied to data mining. Comput Stat Data Anal. 2000;34(4):473–93.
11. Almenoff JS, LaCroix KK, Yuen NA, Fram D, DuMouchel W. Comparative performance of two quantitative safety signalling methods. Drug Saf. 2006;29(10):875–87.
12. Norén GN, Bate A, Orre R, Edwards IR. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. Stat Med. 2006;25:3740–57.
13. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. Am Stat. 1999;53(3):177–90.
14. Almenoff JS, Pattishall EN, Gibbs TG, DuMouchel W, Evans SJW, Yuen N. Novel statistical tools for monitoring the safety of marketed drugs. Clin Pharmacol Ther. 2007;82(2):157–66.
15. Hopstadius J, Norén GN, Bate A, Edwards IR. Impact of stratification on adverse drug reaction surveillance. Drug Saf. 2008;31(11):1035–48.
16. Global Research in Paediatrics. Network of Excellence. http://www.grip-network.org/index.php/cms/en/home. Accessed 28 Feb 2016.
17. Moore TJ, Furberg CD. The safety risks of innovation: the FDA's Expedited Drug Development Pathway. JAMA. 2012;308(9):869–70.
18. Weiss-Smith S, Deshpande G, Chung S, Gogolak V. The FDA drug safety surveillance program: adverse event reporting trends. Arch Intern Med. 2011;171:591–3.
19. MedDRA MSSO. Introductory guide: MedDRA version 15.1. 2012. http://www.meddra.org/sites/default/files/guidance/file/intguide_15_1_English_0.pdf. Accessed 28 Feb 2016.
20. Poluzzi E, Piccinni C, Raschi E, De Ponti F. Data mining techniques in pharmacovigilance: analysis of the publicly accessible FDA Adverse Event Reporting System (AERS). Rijeka: INTECH Open Access Publisher; 2012.
21. SAS. COMPGED Function. Returns the generalized edit distance between two strings. Computing the Generalized Edit Distance. In: SAS® 9.2 language reference: dictionary, fourth edition. http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a002206133.htm. Accessed 3 Mar 2015.
22. Hauben M, Aronson JK. Defining 'signal' and its subtypes in pharmacovigilance based on a systematic review of previous definitions. Drug Saf. 2009;32(2):99–110.
23. Candore G, Juhlin K, Manlik K, Thakrar B, Quarcoo N, Seabroke S, et al. Comparison of statistical signal detection methods within and across spontaneous reporting databases. Drug Saf. 2015;38(6):577–87.
24. Zhou X-H, McClish DK, Obuchowski NA. Statistical methods in diagnostic medicine. Hoboken: Wiley; 2009.
25. Strom BL, Kimmel SE, Hennessy S. Pharmacoepidemiology. 5th ed. Chichester: Wiley-Blackwell; 2012.
26. van Puijenbroek EP, Bate A, Leufkens HGM, Lindquist M, Orre R, Egberts ACG. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiol Drug Saf. 2002;11:3–10.
27. Osokogu OU, Fregonese F, Ferrajolo C, Verhamme K, de Bie S, Catapano M, et al. Pediatric drug safety signal detection: a new drug–event reference set for performance testing of data-mining methods and systems. Drug Saf. 2015;38(2):207–17.
28. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Clinical Investigation of Medicinal Products in the Pediatric Population E11. Current Step 4 version. 20 Jul 2000. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E11/Step4/E11_Guideline.pdf. Accessed 28 Feb 2016.
29. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst. 1959;22:719–48.
30. Breslow NE, Day NE, editors. Statistical methods in cancer research. Volume I—the analysis of case-control studies. IARC Scientific Publications No. 32. Lyon: International Agency for Research on Cancer; 1980.
31. 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.
32. de Bie S, Ferrajolo C, Straus SMJM, Verhamme KMC, Bonhoeffer J, Wong ICK, et al. Pediatric drug safety surveillance in FDA-AERS: a description of adverse events from GRiP Project. PLoS One. 2015;10(6):e0130399.
33. Harpaz R, DuMouchel W, LePendu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA Adverse Event Reporting System. Clin Pharmacol Ther. 2013;93:539–46.
34. Norén GN, Caster O, Juhlin K, Lindquist M. Zoo or savannah? Choice of training ground for evidence-based pharmacovigilance. Drug Saf. 2014;37:655–9.

35. Almenoff J, Tonning JM, Gould AL, Szarfman A, Hauben M, Ouellet-Hellstrom R, et al. Perspectives on the use of data mining in pharmacovigilance. Drug Saf. 2005;28(11):981–1007.

36. Aagaard L, Strandell J, Melskens L, Petersen PSG. Holme Hansen E. Global patterns of adverse drug reactions over a decade: analyses of spontaneous reports to VigiBase™. Drug Saf. 2012;35:1171–82.

37. Morales-Olivas FJ, Martínez-Mir I, Ferrer JM, Rubio E, Palop V. Adverse drug reactions in children reported by means of the yellow card in Spain. J Clin Epidemiol. 2000;53:1076–80.

38. Impicciatore P, Choonara I, Clarkson A, Provasi D, Pandolfini C, Bonati M. Incidence of adverse drug reactions in paediatric in/out-patients: a systematic review and meta-analysis of prospective studies. Br J Clin Pharmacol. 2001;52:77–83.