University Medical Center
Utrecht

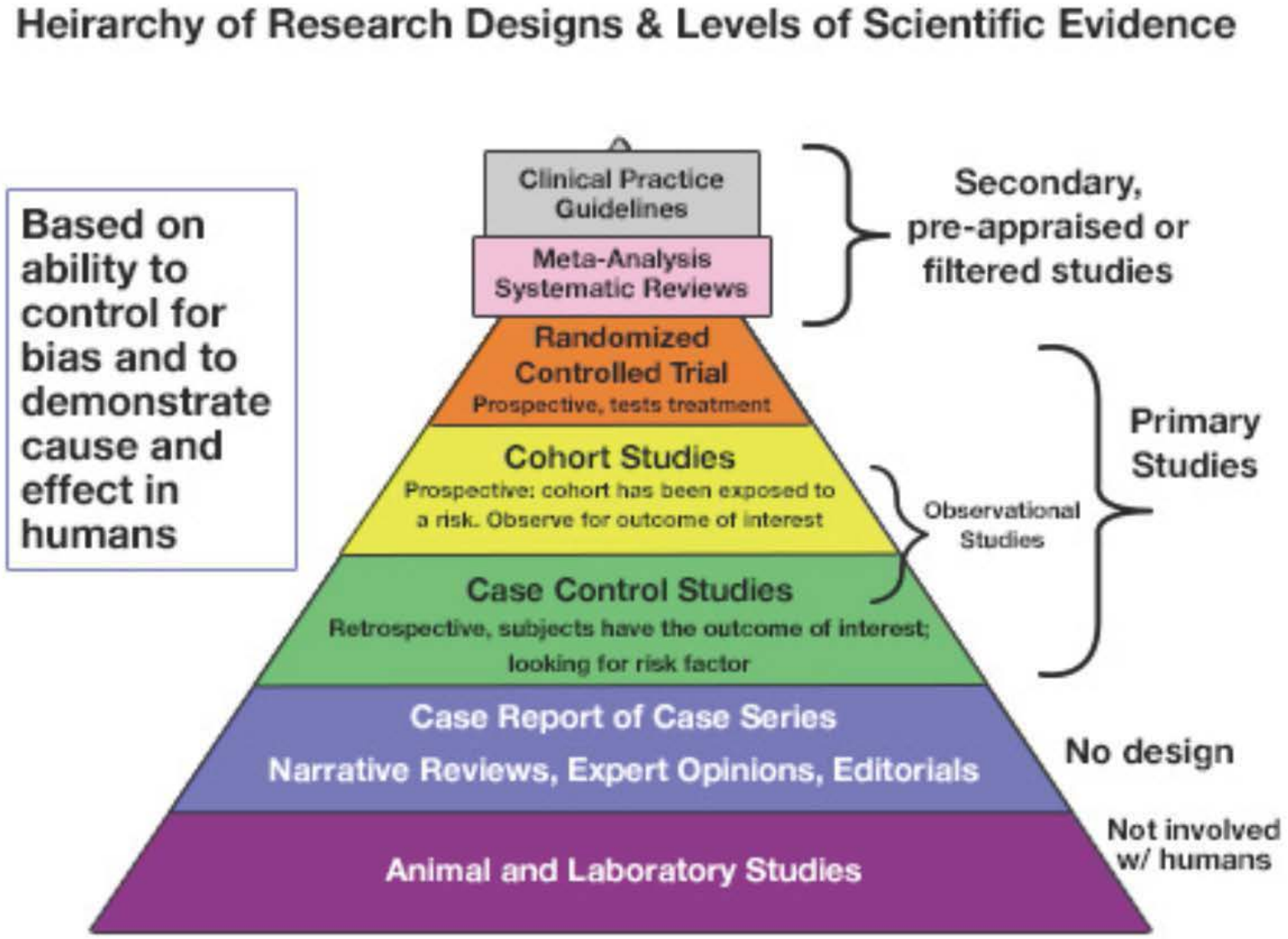
The need for big data & big data methods in pediatrics:

lessons from the Global Research in Pediatrics Network of Excellence

Prof. dr. Miriam Sturkenboom

Utrecht University Medical Center,
The Netherlands

The traditional way to look at evidence



There's a gap between what we know and what we do…

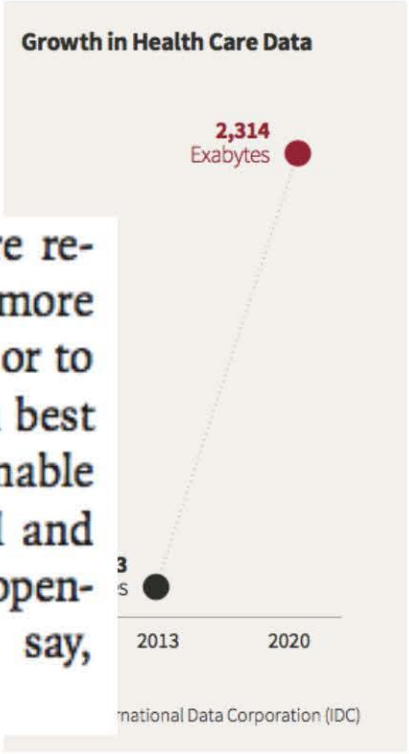
45.1% of medicine is not evidence based;[1] it takes 17 years to translate science to practice[2]

It's humanly impossible to keep up with the knowledge and the data…

Doctors would have to read approximately 29 hours each workday to keep up with new professional insights;[2] 80% of data is unstructured and each of us will produce 300M books of health-related data in our lifetime

1.) http://www.nejm.org/doi/full/10.1056/NEJMsa022615#t=abstract
2.) http://www.nejm.org/doi/full/10.1056/NEJMsa022615#t=abstract

Shifting landscape?



There will always be an argument for more research and for better data, but waiting for more data is often an implicit decision not to act or to act on the basis of past practice rather than best available evidence. The goal must be actionable data — data that are sufficient for clinical and public health action that have been derived openly and objectively and that enable us to say, "Here's what we recommend and why."

Data is fundamentall changing the researc enterprise and creati new extraordinary opportunities to learn things that were either un-learnable or would have taken generations.

- Stanford Faculty

REVIEW ARTICLE

THE CHANGING FACE OF CLINICAL TRIALS

Jeffrey M. Drazen, M.D., David P. Harrington, Ph.D., John J.V. McMurray, M.D., James H. Ware, Ph.D., and Janet Woodcock, M.D., *Editors*

Evidence for Health Decision Making — Beyond Randomized, Controlled Trials

Thomas R. Frieden, M.D., M.P.H.

Are there areas for 'big' data use in regulatory and clinical decision making?

The EMA said:

"Technological advances in both science and information technology are generating ever-increasing amounts of data on health and medicines. The objective of this workshop was to increase understanding of how big data will impact on our understanding of disease and facilitate medicines development, so that the regulatory community can identify opportunities and address challenges in its use for medicines decision-making. In his opening remarks, Professor Guido Rasi (Executive Director, EMA) emphasised the clear potential of big data to benefit patients.

"However, it is challenging to incorporate these data in a meaningful way into routine regulatory decision-making and importantly to understand how to determine whether the conclusions and associations arising from multiple analyses across varied data sets are causal and not simply spurious coincidence. Workshop participants included patient representatives, healthcare professionals, and representatives from government, industry, and academia, as well as regulators from across the globe."

http://www.ema.europa.eu/docs/en_GB/document_library/Report/2017/02/WC500221938.pdf

According to US Food and Drug Administration What does big data offer?

- **Breadth** – large numbers of individuals get us closer to the underlying source population –
- **Depth** – increasing amount of data on each individual increases the chance that we will have measures of likely confounders
- **Diversity** – different types of data offer the potential to "cross check" findings for any particular data source
- **FDA-Sentinel system:** more than 100 million patient health care data

From: D Martin EMA big data workshop

Need for bigger data and big data approaches in pediatrics to support decision making: some lessons from the Global Research in Pediatrics project (FP-7 EC)

www. grip-network.org

This project has received funding from the European Union's Seventh Framework Programme for technological development and demonstration under grant agreement n° 261060

search...

Global Research in Paediatrics
Network of Excellence

home | Who we are | Objectives and results | Events and initiatives | Patient organizations | privacy

OVERVIEW

OUR WORK

- Training and education
- Epidemiological and post-marketing studies
- Tools for interoperability
- Paediatric clinical studies
- Paediatric formulations
- Drug development in neonates

GRiP was created to address the **lack of appropriate testing** and **information** on **paediatric drugs**. GRiP partners are working to reduce the current fragmentation of the efforts to study and develop the use of medicine in children.

NEWS AND EVENTS

2017-06-04

EVENTS

Final GRiP meeting - Padua, 6-7 June 2017

read more

2017-03-21

EVENTS

NEWSLETTER

The GRiP Newsletter offers insight on GRiP's activities and results, a chance to get to know GRiP members better, and all the updates on the world of paediatric clinical pharmacology.
To subscribe, enter your details below

Name:

FOCUS

GRiP Webinars

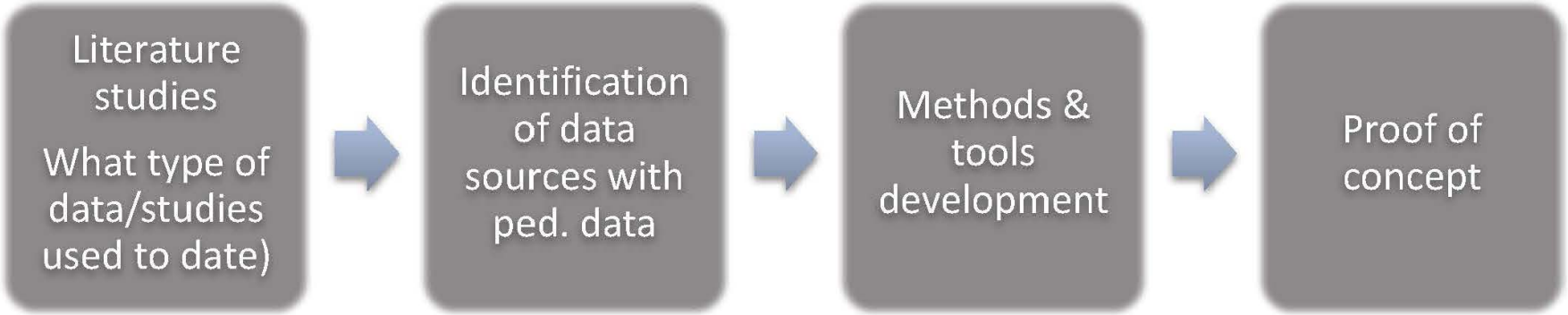
***Meet the expert in Paediatric Formulation* series**

In spite of all new trials following pediatric regulations

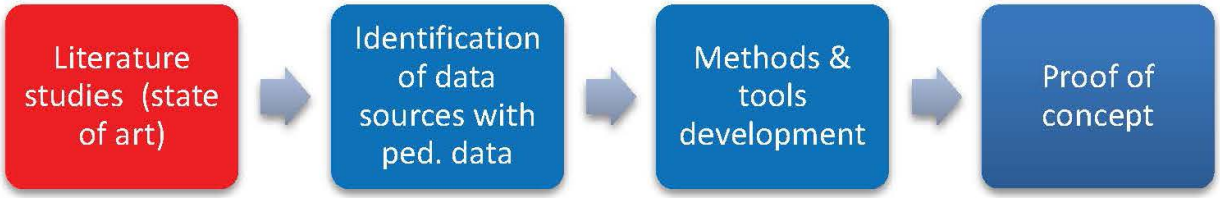
- Big Health Data in children are generated every day in routine healthcare
  - Spontaneous reports of adverse events
  - Medical records (GPs, paediatricians)
  - Registries (vaccinations)
  - Claims records (pharmacy dispensings, hospitalizations …)
  - …..

- These should be used to study the effects of drugs in children and learn about use, benefits and safety

Attempt to establish global pediatric pharmacoepidemiological platform



GRIP e-learning module in pediatric pharmacoepidemiology & pharmacovigilance



Literature review on safety and effectiveness studies to study current state of art

Inventory: current state of the art in pediatric pharmacoepidemiology ?

- Which designs are applied?
- Which data sources are used
- Which methods are used?

Drug safety studies

- *Safety studies (268)*

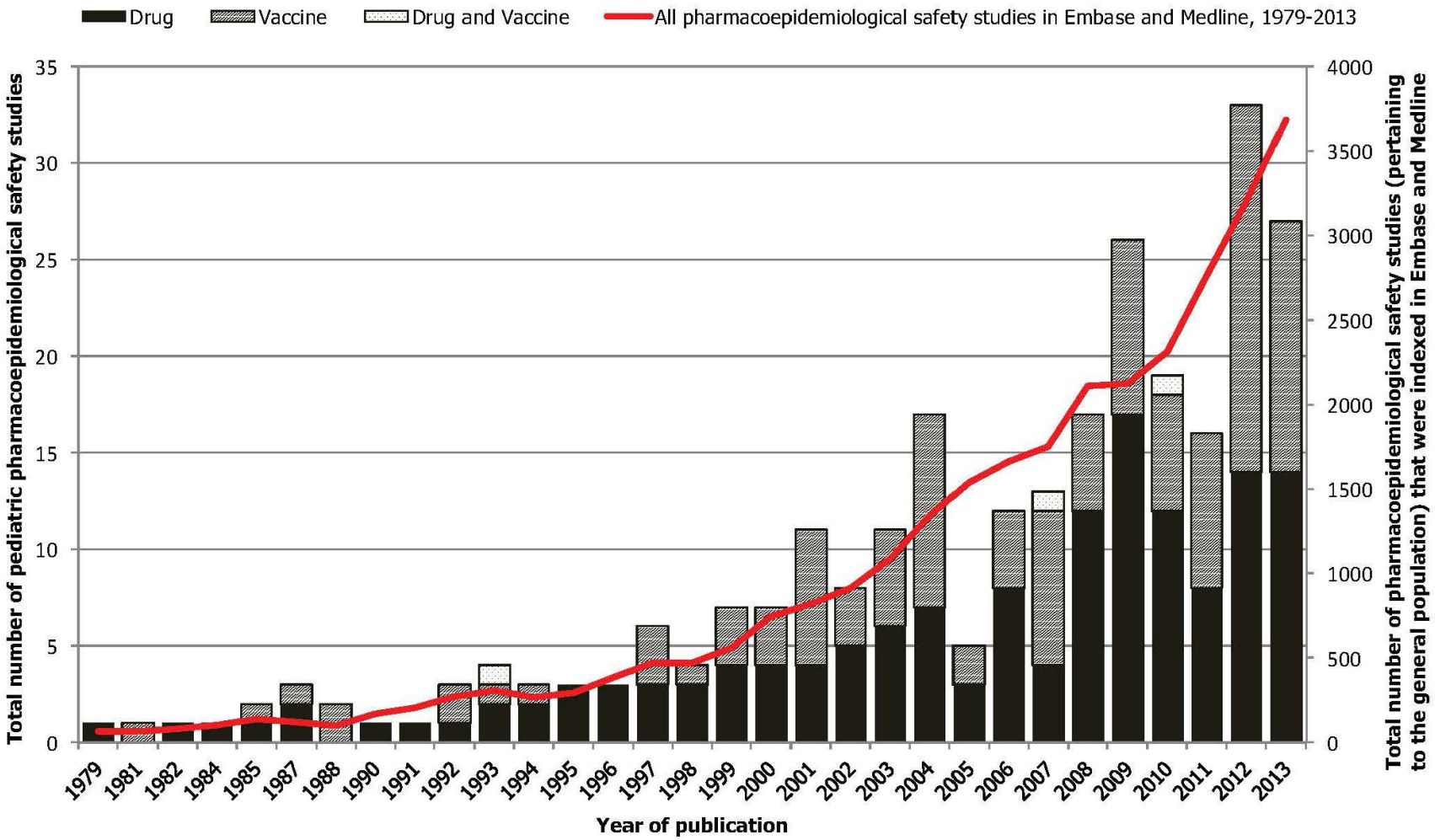
Pharmacoepidemiological safety studies in children: a systematic review

Osemeke U. Osokogu[1]*, Julijana Dukanovic[1], Carmen Ferrajolo[1], Caitlin Dodd[1], Alexandra C. Pacurariu[1], Wichor M. Bramer[2], Geert 't Jong[3], Daniel Weibel[1], Miriam C. J. M. Sturkenboom[1] and Florentia Kaguelidou[1,4]

[1] *Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands*
[2] *Medical Library, Erasmus University Medical Center, Rotterdam, The Netherlands*
[3] *Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Canada*
[4] *Department of Pediatric Pharmacology and Pharmacogenetics, Hôpital Robert Debré, APHP, Univ Paris 7-Diderot, Sorbonne Paris Cité, EA08, INSERM CIC1426, Paris, France*

Osokogu et al. Pharmacoepidemiol Drug Saf. 2016 Aug; 25(8): 861–870.

Results literature review paediatric safety studies



Osokogu et al. Pharmacoepidemiol Drug Saf. 2016 Aug; 25(8): 861–870.

Results literature review paediatric safety studies

- Location: North America (154 [57.5%]) or Europe : (92 [34.3%])
- Only 75 child only studies

- Type of compounds: 147 [54.9%] small molecules, rest vaccines

- Data source:
  - Studies utilizing secondary data: have larger sample sizes
  - Paper medical charts: Main source for
    - exposure (85 [31.7%]) and outcome (122 [45.5%]) data

- Design:
  - Cohort studies: most common (174 [64.9%])
  - SCCS - 30 (11.2%)
  - Case crossover - 4 (1.5%)

Osokogu et al. Pharmacoepidemiol Drug Saf. 2016 Aug; 25(8): 861–870.

Conclusions safety review

Key points

- The number of pharmacoepidemiological safety studies is steadily increasing in pediatrics
- We identified various challenges including funding, design, type and source of data, mode of data collection, age and geographic spread of the investigated population, studied drugs and outcomes, sample size, control of confounding and reporting of results.
- Pharmacoepidemiological safety studies in children can be improved in several ways including global collaboration.

Osokogu et al. Pharmacoepidemiol Drug Saf. 2016 Aug; 25(8): 861–870.

Results literature review paediatric effectiveness studies

- *Effectiveness studies (164)*

Dukanovic J et al. Manuscript submitted for publication

Results literature review paediatric effectiveness studies

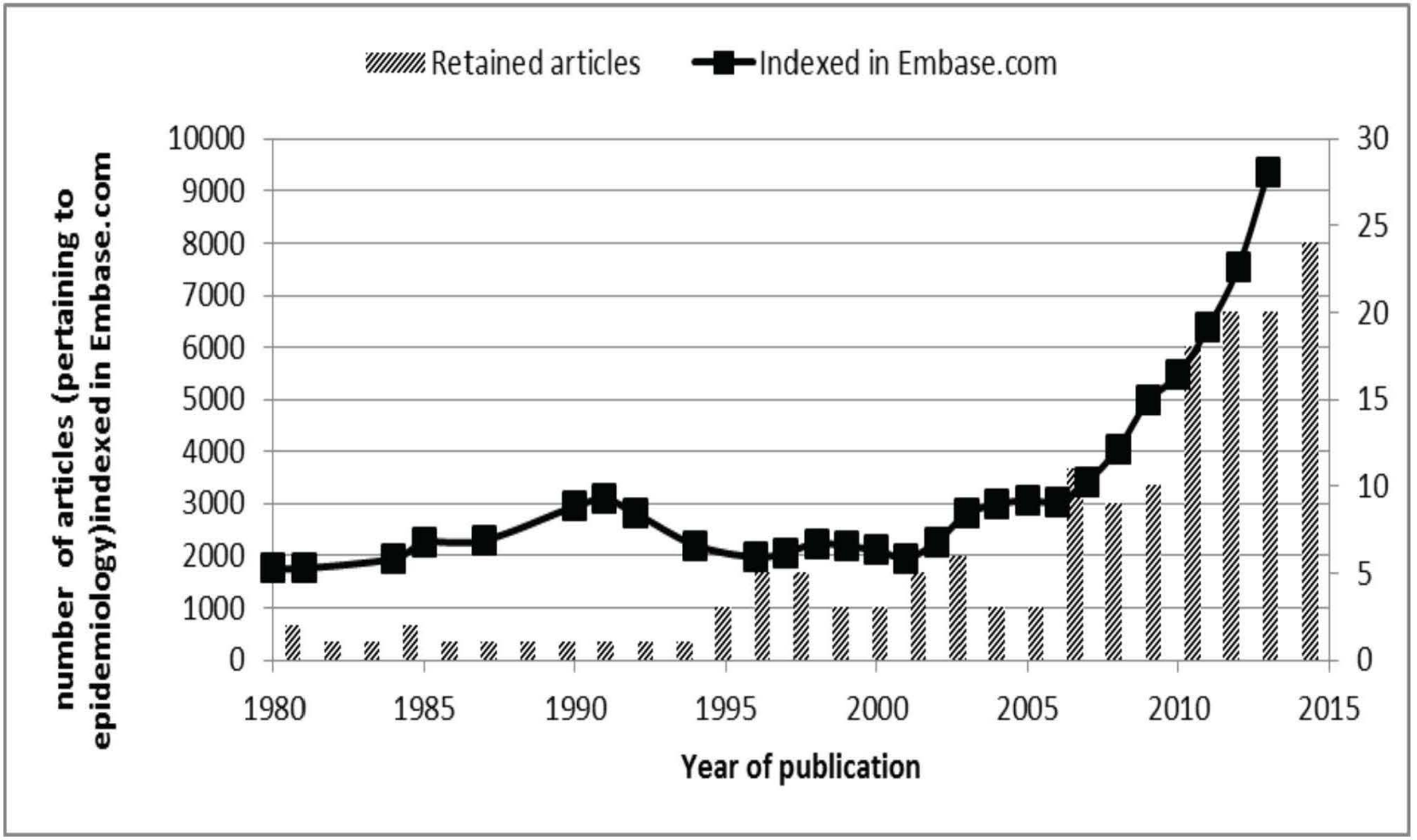
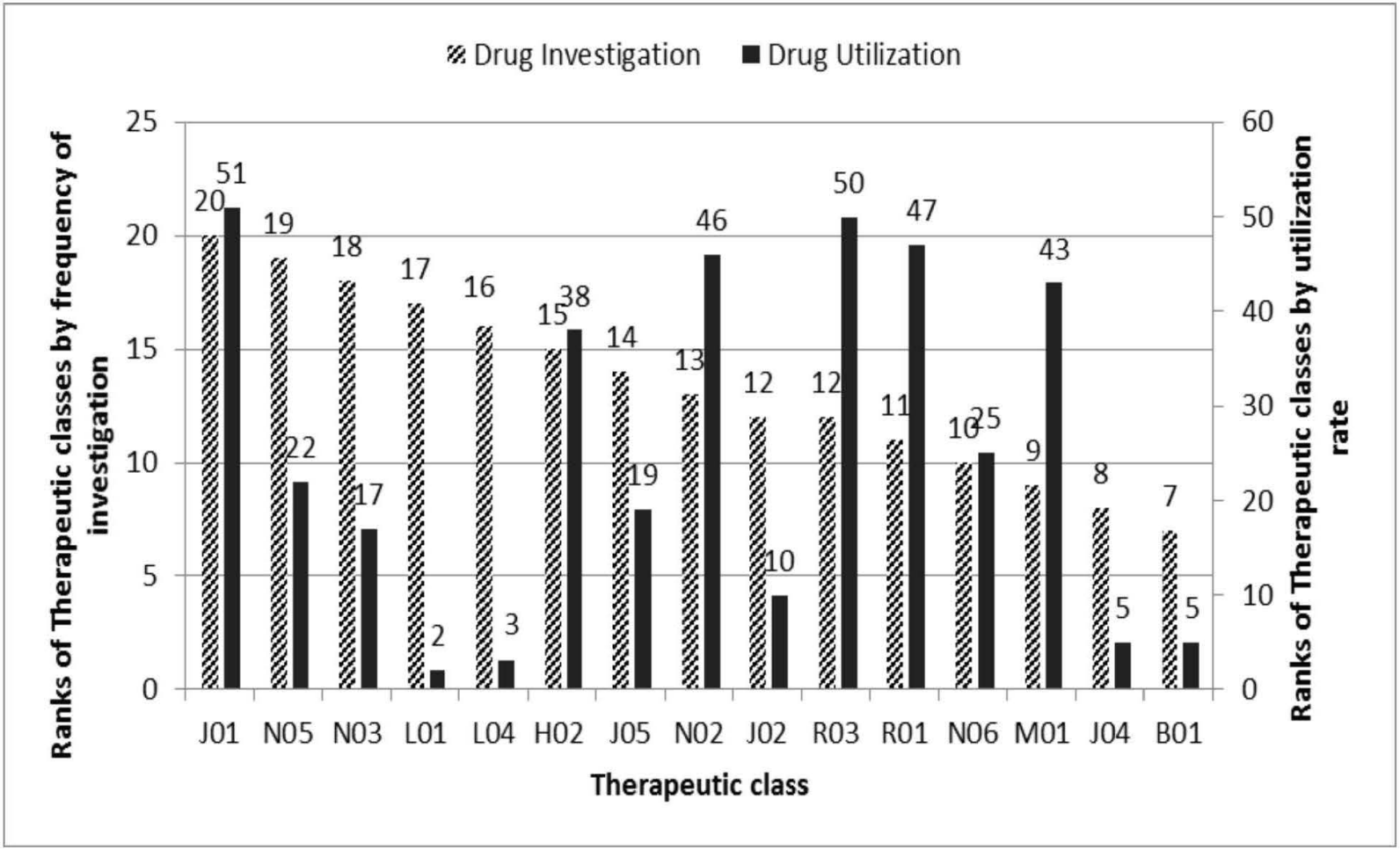


Fig. 2: Number of pharmacoepidemiological effectiveness studies

Dukanovic J et al. Manuscript submitted for publication

Discrepancy between use of drugs in pediatrics and studies



Comparison between evaluated and routinely utilized drugs

Dukanovic J et al. Manuscript submitted for publication

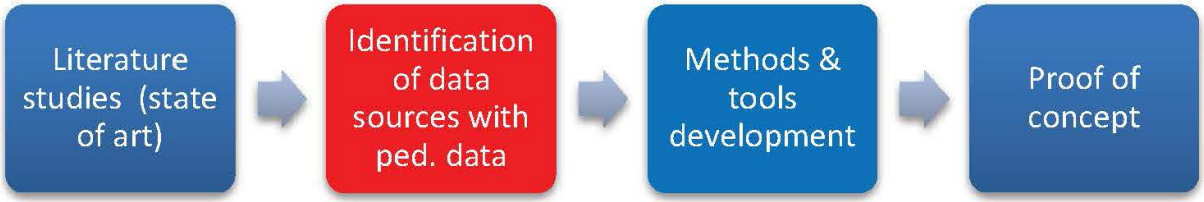
Conclusions paediatric safety & effectiveness studies

- Studies are conducted mainly in developed countries
- Increased number of studies following the introduction of the BPCA (US) and pediatric legislation (EU)

- Use of more modern methods (propensity scores) especially for effectiveness studies
- Many intermediate outcomes instead of clinical outcomes
- Most studies rely on traditional data collection, opportunity for use of electronic health record data
- Need to use more modern methods (propensity scores) for confounding

- Data pooling needed to achieve desired sample size and ability to look at hard outcomes
- Increased capacity needed for conduct of these studies

Dukanovic J et al. Manuscript submitted for publication



Are there available big health data sources that can be used to generate evidence on the effects of drugs in children?

Spontaneous reporting databases

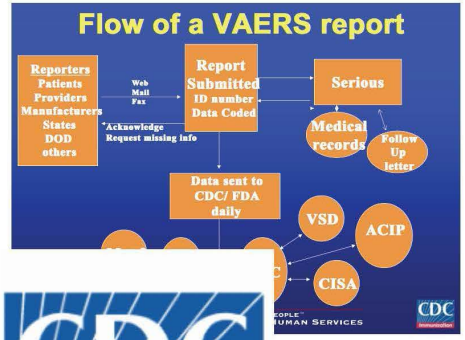
Spontaneous reporting*

'Spontaneous' (or voluntary) reporting of adverse effects is when health professionals or patients decide that they will report suspected harm from a medicine to their local or national pharmacovigilance centre.

Uppsala Monitoring Centre

World Health Organization

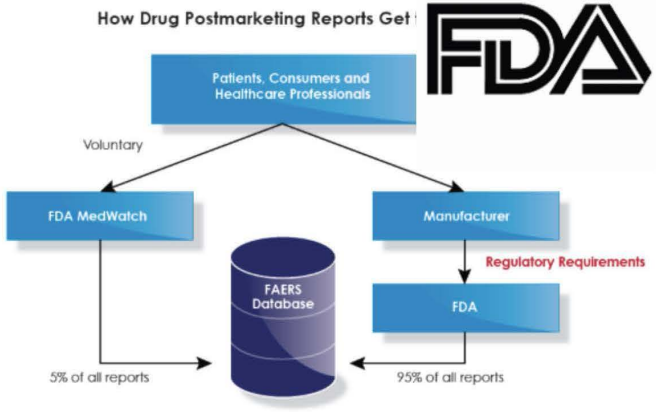
VigiBase, the WHO database of individual case safety reports



CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

EUDRAVIGILANCE
Pharmacovigilance in the European Economic Area

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH



Spontaneous reports: FAERS (public version)

Distribution of pediatric ICSRs (N = 106,122) within FAERS according to age-category.

	Total	≤ 27 days	28 days-23 months	2–11 years	12–17years
	N = 106,122 (%)	N = 4,717 (4.4%)	N = 16,096 (15.2%)	N = 47,248 (44.5%)	N = 38,061 (35.9%)
Males	54,768 (54.5%)	2,114 (54.1%)	7,921 (55.3%)	27,075 (59.9%)	17,658 (47.7%)
Mean age (95%CI)	9.1 (9.0–9.1)				
Reported drugs	236,491	12,180 (5.2%)	34,575 (14.6%)	103,988 (44.0%)	85,748 (36.3%)
Drugs/ICSR [median (IQR)]	1 (1–3)	1 (1–3)	1 (1–3)	1 (1–3)	1 (1–3)
Reported events	397,220	21,265 (5.4%)	59,306 (14.9%)	173,395 (43.7%)	143,254 (36.1%)
Events/ICSR [median (IQR)]	1 (1–1)	1 (1–2)	1 (1–2)	1 (1–1)	1 (1–1)

PLoS One. 2015; 10(6): e0130399.Published online 2015 Jun 19.

Spontaneous reports EUDRAVIGILANCE (Academic version)

Table 1 Description of pediatric ADR reports by age categories in EUDRAVIGILANCE

Age group	Number of DECs, n (%) full set	Number of DECs, n (%) vaccines
Infants: 0 days-23 months	402,817	208,658
Children: 2-11 years	406,136	72,271
Adolescents: 12-17 years	368,422	60,064
Total	**1,177,375**	**340,993**

Dodd CN et al., manuscript in preparation

Spontaneous reports VAERS: public version

Age group	Number of vaccine-event combinations, n (%) Vaccines
Infants: 0 days-23 months	848,365 (54%)
Children: 2-11 years	437,082 (28%)
Adolescents: 12-17 years	271,216 (17%)
Total	**1,556,663 (100%)**

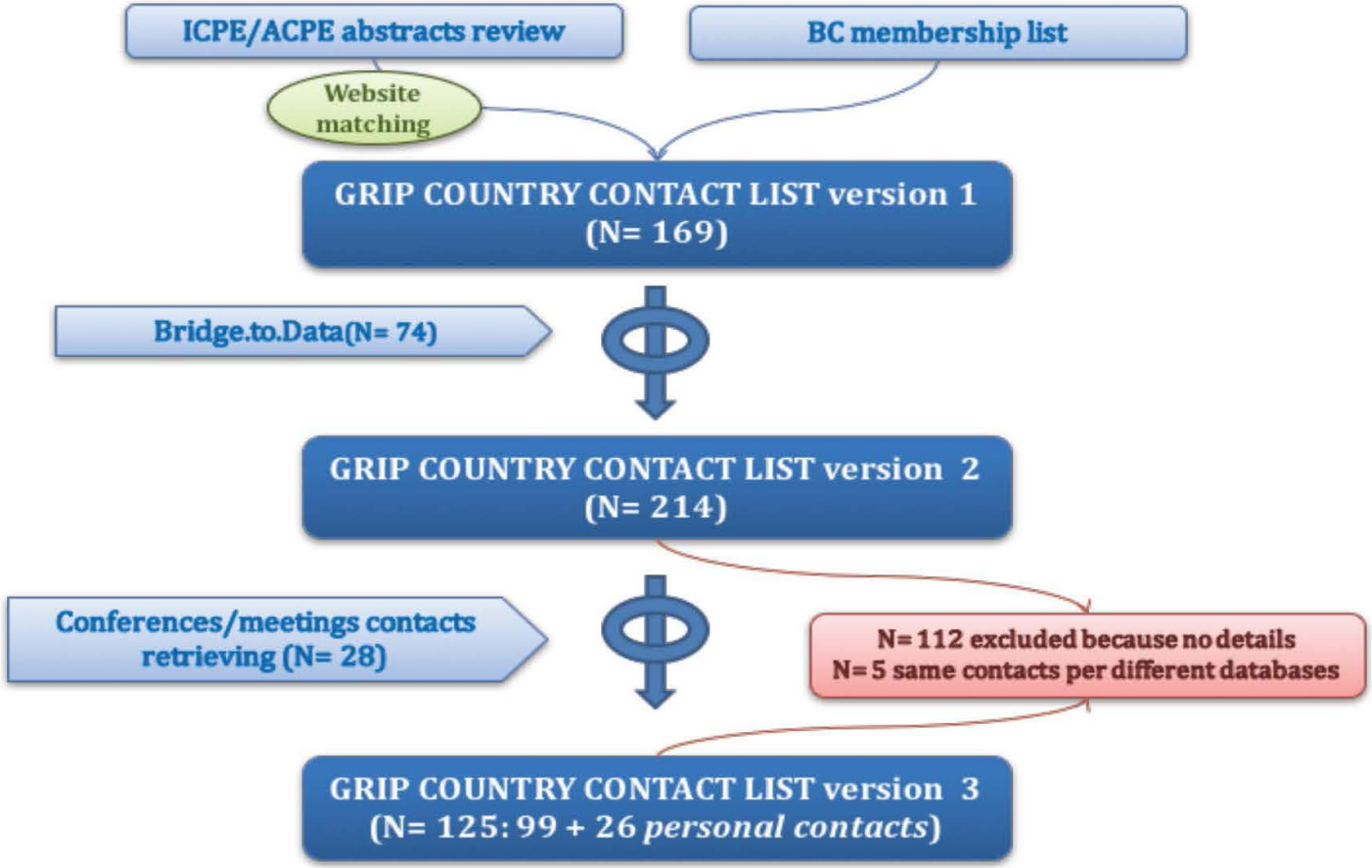
Conclusion SRS

- Millions of spontaneous reports are available for pediatrics in publicly accessible datasources
- Each source has different structure
- Methods for cleaning, deduplication and pooling of data might improve ability to do data mining in pediatrics specifically

Global Research in Paediatrics

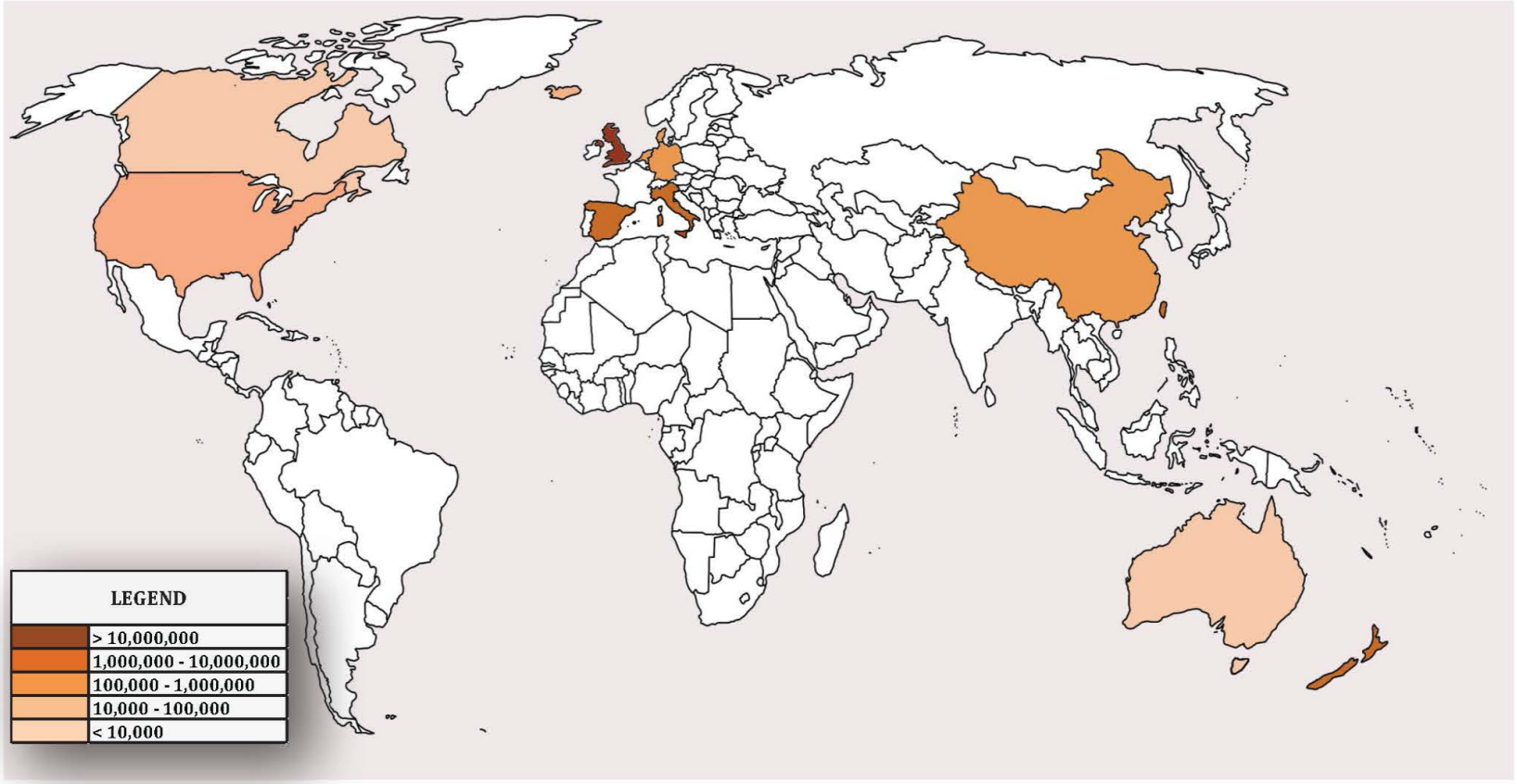
Population based pediatric health care data

Identify healthcare databases comprising paediatric data (2012)



Published: Ferrajolo C. https://repub.eur.nl/pub/77131/
Chapter 6

Identify healthcare databases comprising paediatric data (2012)



64 responses out of 125, 34 willing to participate in GRIP:
Health care data on more than 50 million children

Published: Ferrajolo C. https://repub.eur.nl/pub/77131/
Chapter 6

Conclusion on availability of paediatric data for use, effectiveness and safety studies

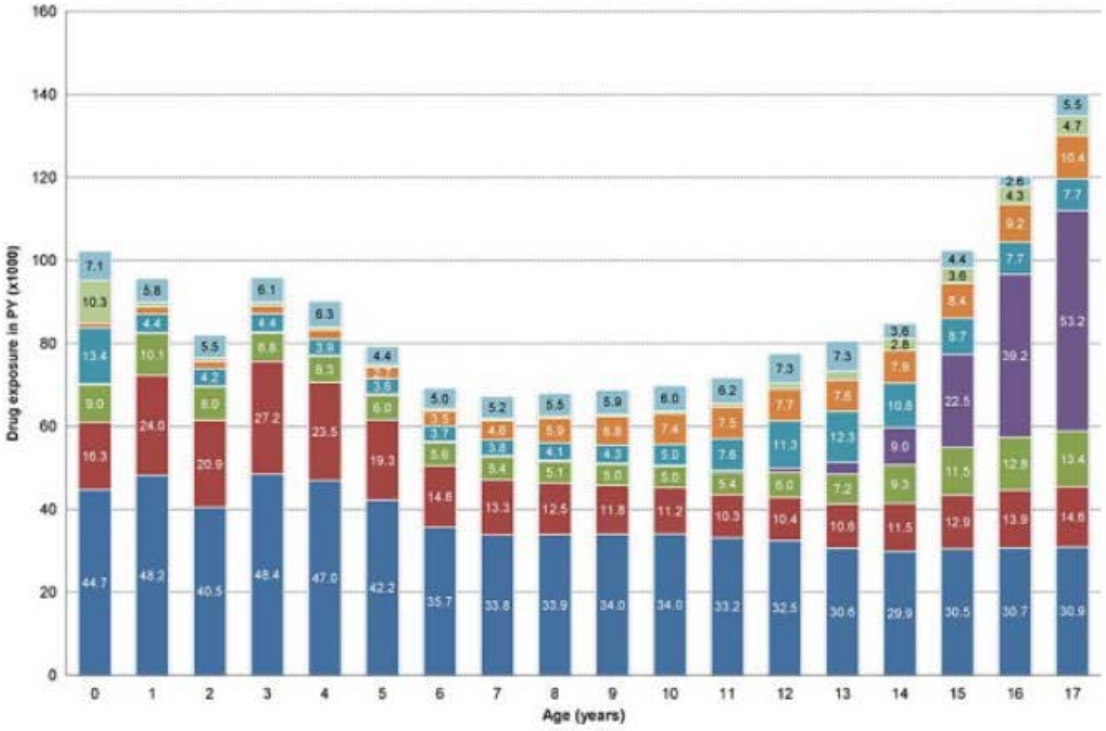
- **Spontaneus reports**: millions of reports on pediatrics are publicly available

- **Health care records**: Data on many children available around the world, databases with >50 million children willing to collaborate

Global Research in Paediatrics

We need to pool and combine to increase ability to detect in pediatrics

EU-ADR network: 8 databases, 5 million children, 2170 different drugs, 25 million PY follow-up



"The 1.6 million PYs of exposure were distributed over 2170 individual drugs, compared with 2289 for the overall population (all ages) in the database network. Of these, only 18 represented 50% and 158 drugs represented 90% of the total drug exposure time.

Drug exposure in person-years by age. Note: Drug exposure is aggregated on the first ATC level (anatomical main group). 'Other' represents all other drug groups with a total exposure of <5000 PYs. ■ respiratory, ■ anti-Infectives, ■ dermatological, ■ genitourinary, ■ alimentary, ■ neurologic, ■ blood, ■ other (<5 000 PYs)

De Bie et al. Br J Clin Pharmacol. 2015 Aug; 80(2): 304–314.

Methods in electronic health care databases: size & power for paediatric studies within EU-ADR databases with 5 million children

Table 1

Amount of required drug exposure to identify potentially drug-induced adverse events

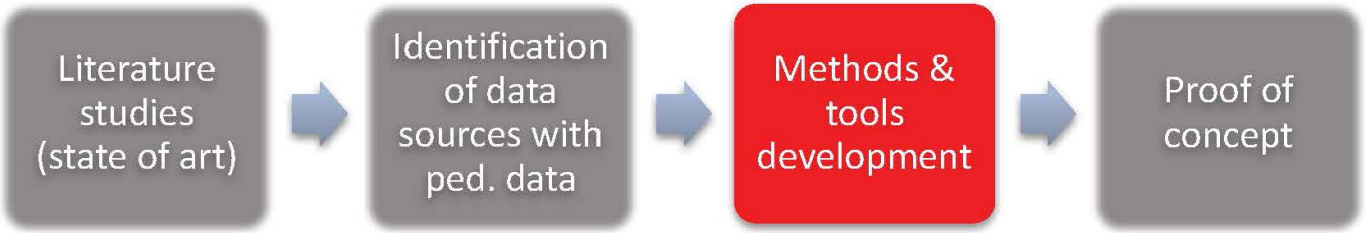
		Weak association (RR ≥2)			Moderate association (RR ≥4)			Strong association (RR ≥6)		
Event type	IR per 100 000 PY	Required exposure (PY)	Drugs *n*	% of Exp	Required exposure (PY)	Drugs *n*	% of Exp	Required exposure (PY)	Drugs *n*	% of Exp
Hip fracture	15.3	52 501	6	29.5	8039	42	67.8	3589	81	80.4
Upper GI bleeding	14.4	55 725	5	26.2	8532	39	66.3	3810	79	79.9
Neutropenia	8.1	99 259	2	13.0	15 198	25	56.9	6786	48	70.5
Acute liver injury	4.0	202 733	0	0	31 041	9	37.3	13 860	26	57.8
Pancytopenia	3.7	215 469	0	0	32 991	9	37.3	14 730	25	56.9
Bullous eruption	3.6	224 394	0	0	34 358	9	37.3	15 341	24	56.0
Anaphylactic shock	3.2	248 526	0	0	38 053	8	35.0	16 990	20	52.1
Cardiac valve fibrosis	2.9	275 840	0	0	42 235	8	35.0	18 858	15	46.6
Acute renal failure	1.6	517 050	0	0	79 168	3	17.9	35 348	9	37.3
Acute pancreatitis	1.6	519 664	0	0	79 568	3	17.9	35 527	9	37.3

Drugs (*n*): Number of drugs at fifth ATC, chemical substance level that have enough PY of exposure to detect a potential signal (total 2170). % of Exp: Proportion of PYs of exposure of the drugs with enough exposure compared with the total PYs of exposure for all drugs. IR, incidence rate; PY, person years; RR, relative risk; upper GI bleeding, upper gastrointestinal bleeding

Br J Clin Pharmacol / **80:2** / 307

Global collaboration is needed

De Bie et al. Br. J Clin Pharmacol 2015: 304-314



Methods & tools to mine big health data in pediatrics

Global Research in Paediatrics

methods for paedatric signal detection in spontaneous reporting databases

Goal

- What are the best methods to mine for safety signals in children in spontaneous reporting databases?

- Comparison of performance of different data mining methods in spontaneous reporting databases

GRIP workflow

- Creation of reference sets
- Cleaning & completion of VAERS, FAERS and EUDRAVIGILANCE sets (machinelearning)
- Terminology mapping
- Running and analysis

Drug Saf (2015) 38:207–217
DOI 10.1007/s40264-015-0265-0

ORIGINAL RESEARCH ARTICLE

Pediatric Drug Safety Signal Detection: A New Drug–Event Reference Set for Performance Testing of Data-Mining Methods and Systems

Osemeke U. Osokogu · Federica Fregonese · Carmen Ferrajolo · Katia Verhamme · Sandra de Bie · Geert 't Jong · Mariana Catapano · Daniel Weibel · Florentia Kaguelidou · Wichor M. Bramer · Yingfen Hsia · Ian C. K. Wong · Madlen Gazarian · Jan Bonhoeffer · Miriam Sturkenboom

Reference set drugs

Table 2 Classification of each drug–event pair as positive control (green: PC1 or PC2) or negative control (red: NC2)

Selected Drugs	Bullous eruption	Aplastic anemia	Agranulocytosis	Thrombocytopenia	Psychosis	Suicide	Vent. arrhythmia	Sudden death	QT prolongation	Venous thromboembolism	Anaphylaxis	Seizure	Acute kidney injury	Acute liver injury	Sepsis	SIDS
flucloxacillin																
clarithromycin																
doxycycline																
lopinavir																
isoniazid																
praziquantel																
mebendazole																
quinine																
fluticasone																
montelukast																
isotretinoin																
loperamide																
domperidone																
methylphenidate																
ibuprofen																
cyproterone/eth.est.																

Abbreviations: Vent. - ventricular; SIDS - Sudden Infant Death Syndrome; eth.est.- ethinylestradiol

GRIP-Reference set vaccine ADRs

ELSEVIER

Vaccine

Available online 20 October 2015

In Press, Corrected Proof — Note to users

Reference set for performance testing of pediatric vaccine safety signal detection methods and systems

Yolanda Brauchli Pernus[a, 1], Cassandra Nan[b, 1], Thomas Verstraeten[b], Mariia Pedenko[c], Osemeke U. Osokogu[c], Daniel Weibel[c], Miriam Sturkenboom[c], Jan Bonhoeffer[a, d], on behalf of the GRIP consortium

[a] Brighton Collaboration Foundation, Switzerland
[b] P95 Pharmacovigilance and Epidemiology Services, Leuven, Belgium
[c] Erasmus University Medical Center, The Netherlands
[d] University Children's Hospital Basel, University Basel, Switzerland

Received 1 October 2015, Accepted 5 October 2015, Available online 20 October 2015

Show less

doi:10.1016/j.vaccine.2015.10.013

Cleaning spontaneous reporting datasets: enhancing quality of big data

Cleaning steps

- Creation of GRIP spontaneous reporting common data model
- Deduplication of records (within and between FAERS/VAERS/Eudravigilance)
- Coding of events (MEDDRA) and drugs (ATC)
  - Mapping tool (Machine learning) for coding of drug names
- Transfer of FAERS, VAERS and EUDRAVIGILANCE in common data model

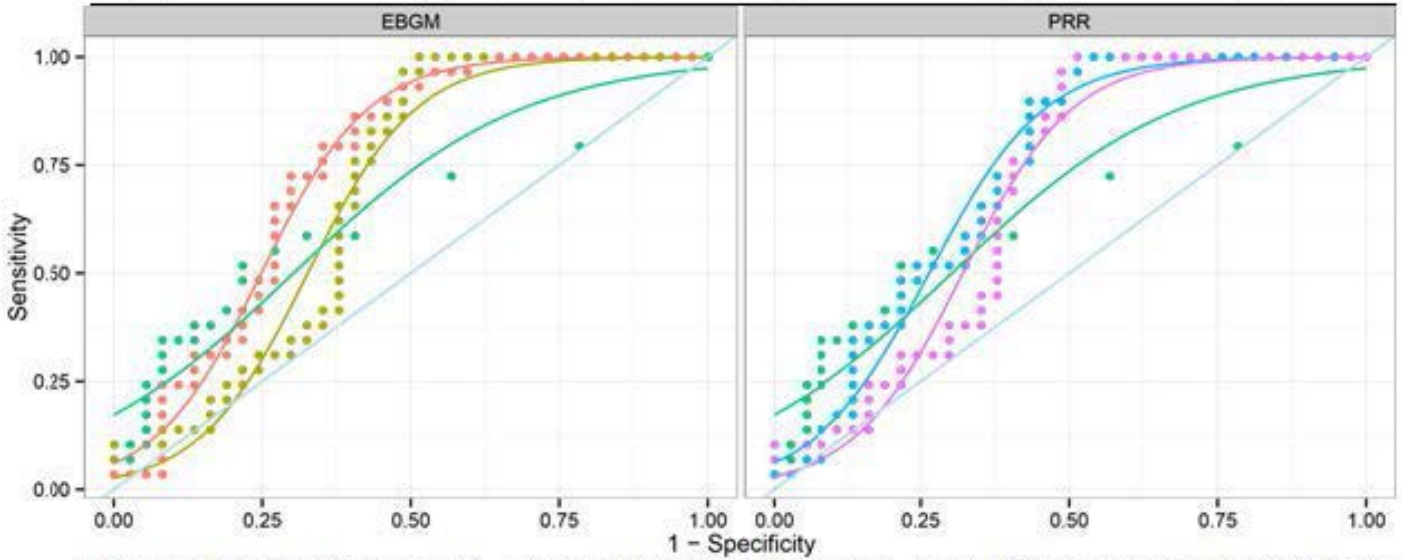
testing automated signal identification methods (example)

FAERS (FDA adverse events)
EUDRAVIGILANCE (EMA)
VAERS (CDC adverse events vaccines)

No difference in performance of methods on FAERS, adjustment for age worsens performance

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

SDA	Sensitivity	Specificity	PPV	NPV	AUC
Number of reports	58.62	67.57	58.62	67.57	0.634
PRR	64.71	63.27	37.93	83.78	0.731
EBGM	62.50	58.62	17.24	91.89	0.745
After age adjustment[a]					
PRR	66.67	62.75	34.48	86.49	0.688
EBGM	50.00	56.90	13.79	89.19	0.683



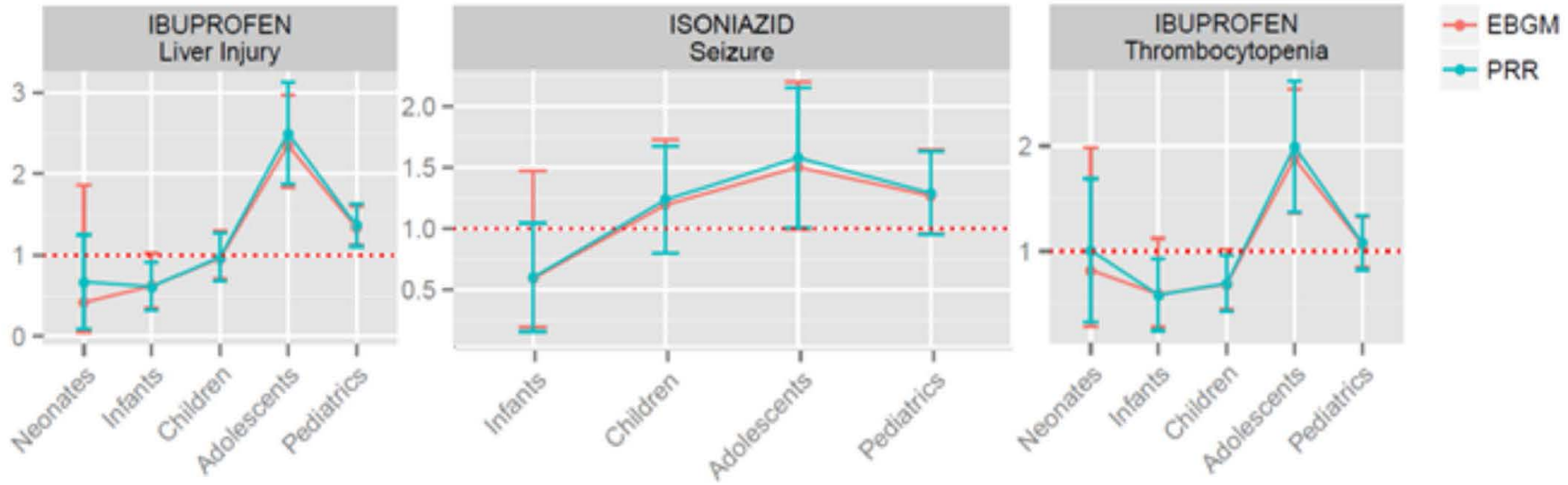
SDA-signal detection algorithm; PRR= Proportional reporting ratio; EBGM= Empirical Bayes Geometric Mean; AUC=area under the curve; PPV=positive predictive value; NPV-negative predictive value

[a] adjusted PRR/ROR values calculated by combining the individual estimates from each age stratum into one measure according to the Mantel-Haenszel approach.

Osokogu OU, etal . Drug Safety Monitoring in Children: Performance of Signal Detection Algorithms andImpact of Age Stratification. Drug Saf. 2016 Sep;39(9):873-81.

Impact of age stratification, some signals unmasked

Fig. 2 Variation of PRR and EBGM estimates across pediatric specific strata –selected examples



Osokogu OU, etal . Drug Safety Monitoring in Children: Performance of Signal Detection Algorithms andImpact of Age Stratification. Drug Saf. 2016 Sep;39(9):873-81.

Recommendations for pediatric signal detection on FAERS

- The Signal detection algorithms 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.

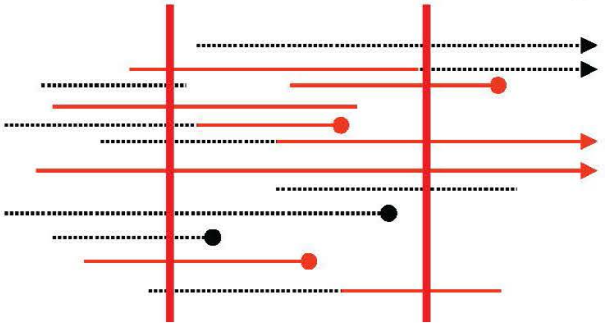
Global Research in Paediatrics

methods for conducting studies in big health care databases

methods for conducting studies in electronic health care databases: how to estimate incidence & prevalence in children given dynamic populations

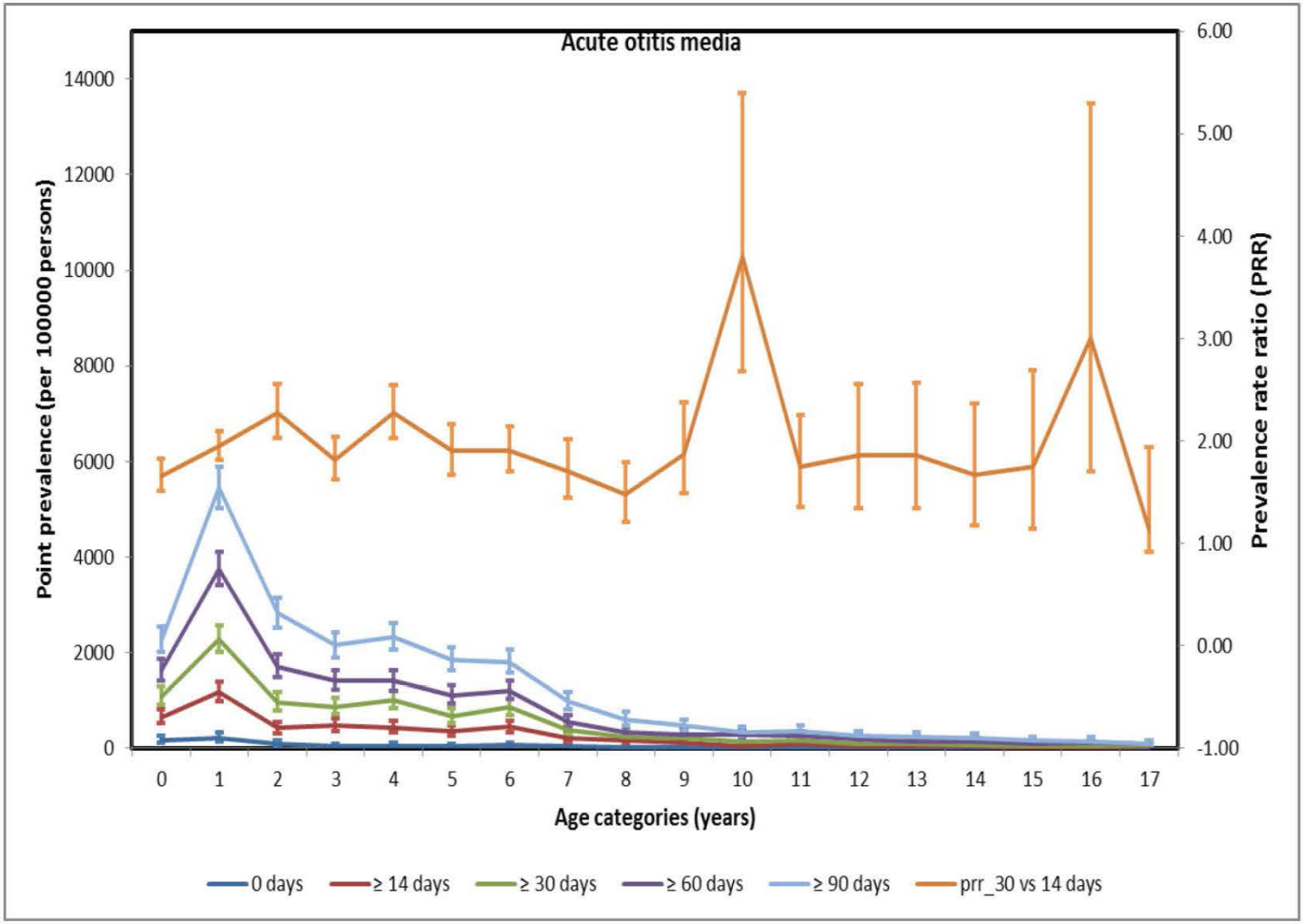
Explanation of issue

- Health care databases comprise
  - Population file
  - All drugs prescribed/dispensed
  - Events (primary care, hospitalization)
- On registered population
- However population is dynamic and we only see a fraction of the 'life'



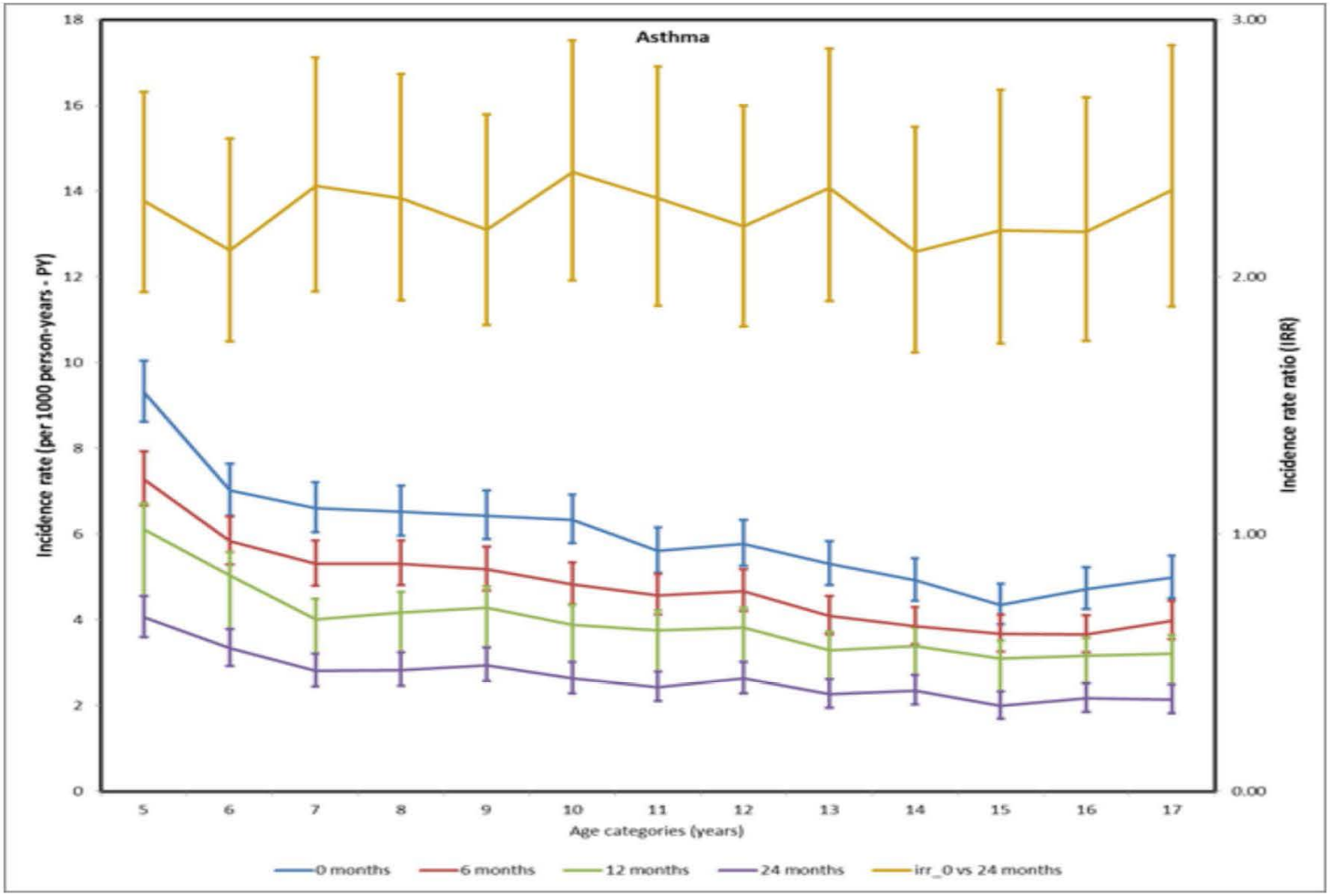
Osokogu: https://repub.eur.nl/pub/95504/

Methods in electronic health care databases: estimation of incidence & prevalence: impact of episode duration on incidence estimation



Osokogu: https://repub.eur.nl/pub/95504/

Methods in electronic health care databases: estimation of incidence & prevalence: impact of naïve period on incidence estimation



Osokogu: https://repub.eur.nl/pub/95504/

Methods in electronic health care databases: estimation of incidence & prevalence: **recommendations for studies**

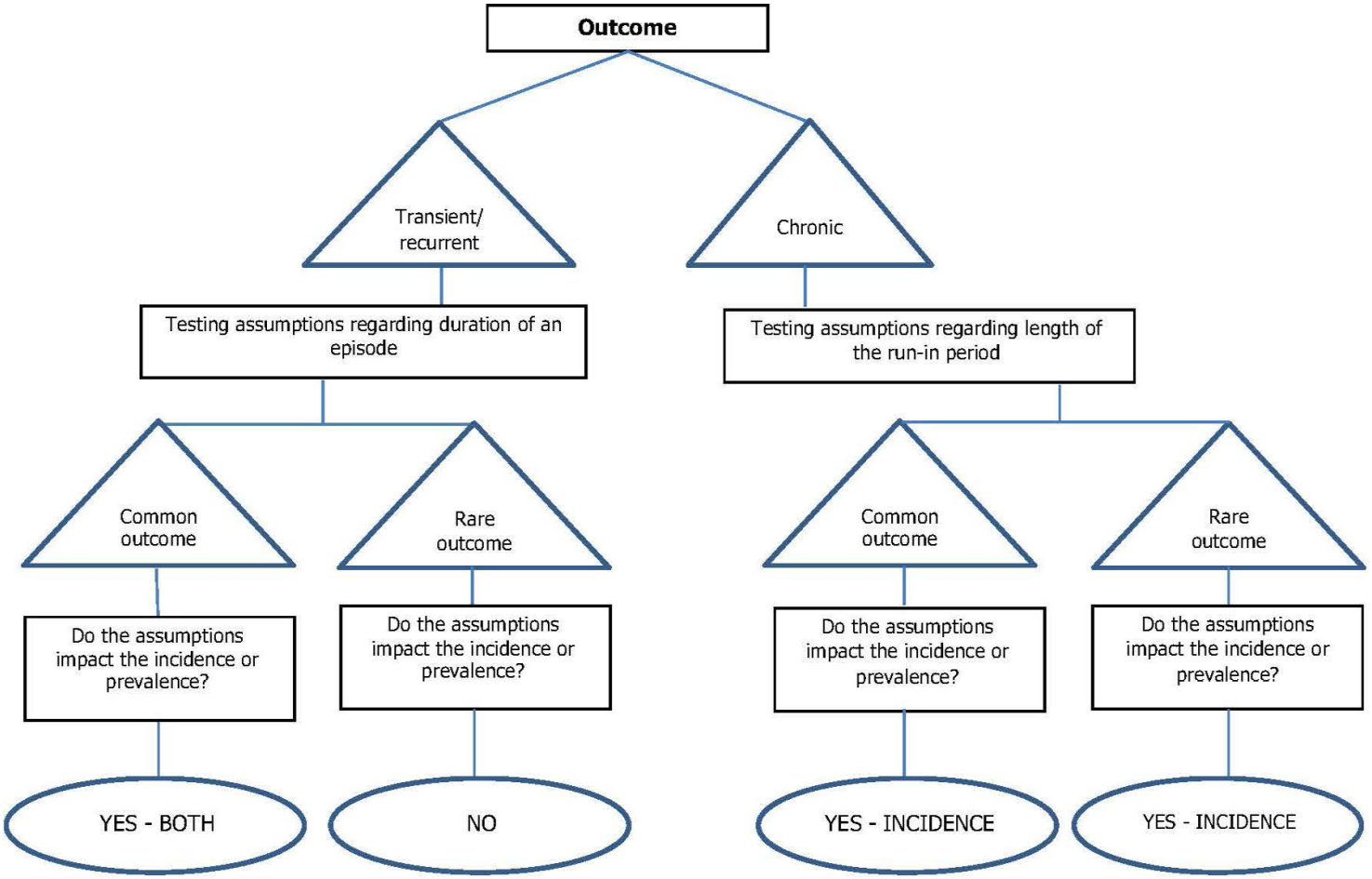


Figure 4: Summary of the impact of assumptions on the investigated outcomes

Osokogu: https://repub.eur.nl/pub/95504/

methods for conducting studies in big health care databases:

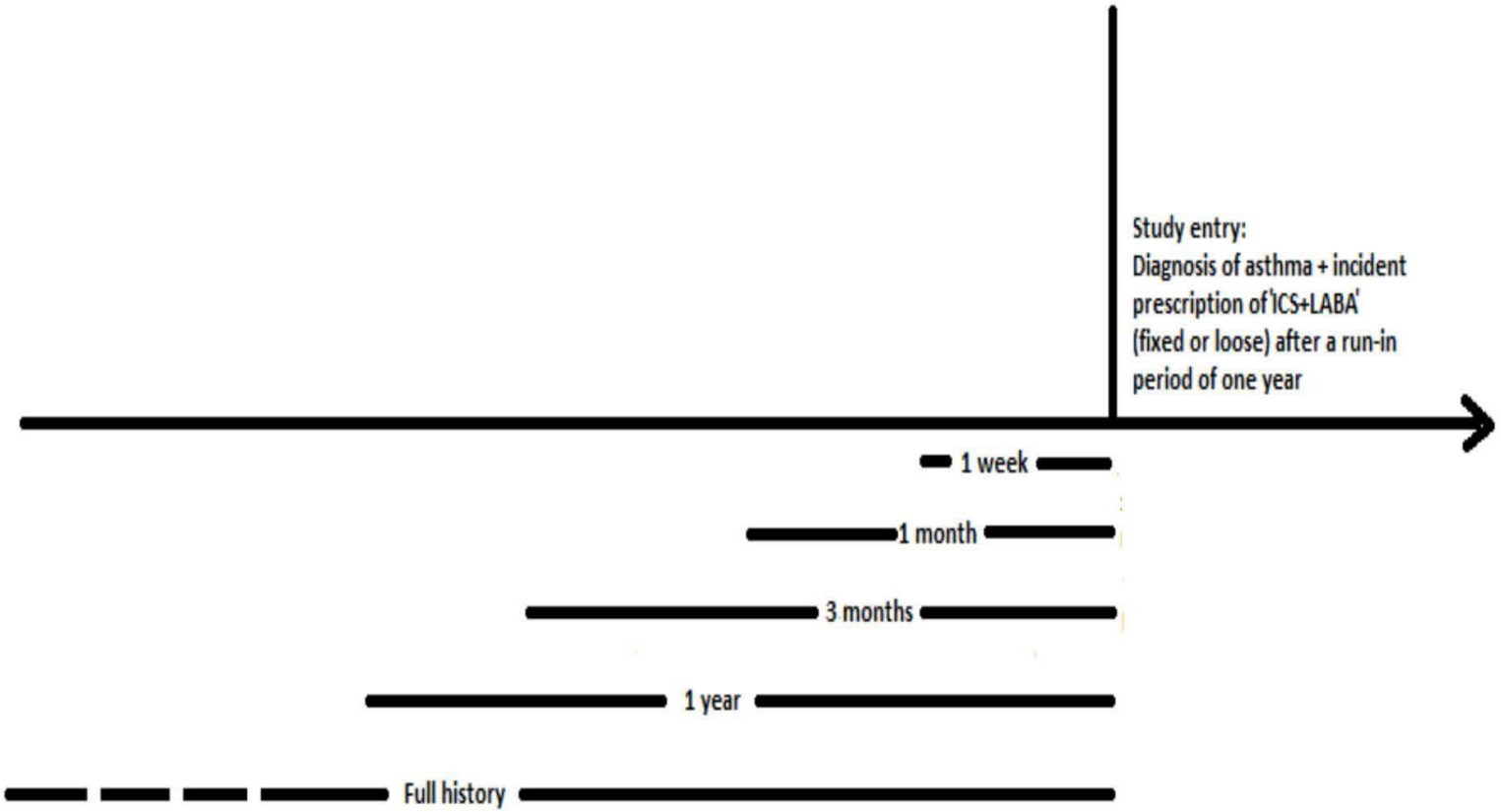
how to best adjust for confounding in pediatric observational studies?

Propensity scores

- Propensity score: statistical model that predicts the 'assignment of treatment based on covariates'
- Allows for matching on this score to create balance between the different treatment groups
- Not much used in pediatrics and not clear what the 'look back period' should be (co-morbidity may be acute)
- Used an example regarding effectiveness of asthma medication re exacerbations in IPCI

Osokogu: https://repub.eur.nl/pub/95504/

Design of methods study on look back period for construct of propensity score



Impact of look back period and 'adjustment method'

Time	Analysis	HazardRatio	95 % Hazard Ratio Confidence Limits
	Crude	0.366	(0.202 ; 0.664)
1 Week	Multivariate	0.406	(0.221 ; 0.745)
	PS-matching	0.480	(0.222 ; 1.306)
	PS-IPTW	0.381	(0.196 ; 0.742)
	PS-adjustment	0.433	(0.231 ; 0.811)
1 Month	Multivariate	0.468	(0.250 ; 0.877)
	PS-matching	0.601	(0.262 ; 1.375)
	PS-IPTW	0.408	(0.207 ; 0.806)
	PS-adjustment	0.565	(0.287 ; 1.112)
3 Months	Multivariate	0.533	(0.282 ; 1.008)
	PS-matching	0.699	(0.311 ; 1.571)
	PS-IPTW	0.428	(0.209 ; 0.878)
	PS-adjustment	0.564	(0.290 ; 1.097)
1 Year	Multivariate	0.552	(0.296 ; 1.030)
	PS-matching	0.561	(0.253 ; 1.244)
	PS-IPTW	0.452	(0.207 ; 0.988)
	PS-adjustment	0.608	(0.315 ; 1.174)
Full History	Multivariate	0.477	(0.259 ; 0.880)
	PS-matching	0.575	(0.242 ; 1.364)
	PS-IPTW	0.384	(0.166 ; 0.887)
	PS-adjustment	0.525	(0.282 ; 0.978)

The impact of different look back periods and the choice of the way to implement the PS are important. The results on a matched analysis are comparable to clinical trial data on the comparison between fixed and loose ICS+LABA combinations in preventing worsening of asthma.

Osokogu: https://repub.eur.nl/pub/95504/

Summary

- 'Big data' is a great & challenging opportunity also in pediatrics, many data sources are available
- Many applications can be found/exist where big data analysis may assist clinical and regulatory decision making
- Computing and data facilities for distributed systems need to be improved but great developments are on the way, in pediatrics global collaboration is needed!
- Collaboration needed between Data Scientists, pharmacologists, regulators, epidemiologists, pediatricians to improve the field
- Machine learning methods can help, but the human mind will remain necessary for interpretation & generalization


Thank you