Vaccine 89 (2026) 128899

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Updating the evidence on brighton collaboration's GAIA maternal immunisation case definitions: a systematic review of performance and implementation challenges

Robert Mboizi [a,b], Kirsten Maertens [c], Kirsty Le Doare [a,d], Clare L. Cutland [e], Flor M. Muñoz [f,1], Hannah G. Davies [d,g,h,*,1]

[a] Makerere University Johns Hopkins Research Collaboration (MUJHU), Kampala, Uganda
[b] University of Antwerp, Antwerp, Belgium
[c] Centre for the Evaluation of Vaccination, Vaccine & Infectious Diseases Institute, University of Antwerp, Antwerp, Belgium
[d] Centre for Neonatal and Paediatric Infection, City St George's, University of London, London, United Kingdom
[e] University of the Witwatersrand, Johannesburg, South Africa
[f] Departments of Pediatrics and Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, United States of America
[g] London School of Hygiene & Tropical Medicine, London, United Kingdom
[h] Department for Infection and Population Health, Institute for Global Health, University College London, United Kingdom

ARTICLE INFO

Keywords:
Vaccine safety
Maternal immunisation
Neonatal infections
Brighton collaboration
GAIA

ABSTRACT

Background: The Brighton Collaboration's GAIA initiative developed harmonised case definitions (CDs) to standardise safety monitoring for maternal immunisation studies globally. These definitions use levels of diagnostic certainty (LOC) to support use across settings with differing diagnostic capacity, but their real-world performance and applicability have remained uncertain. A 2023 review found that around half of GAIA CDs had not been evaluated and that several key outcomes performed poorly with high rates of cases that could not be classified at LOC 1–3 (achieving only LOC-4).

Methods: Ovid MEDLINE, EMBASE, and Web of Science databases were searched for research studies in English-language publications from 2014 to May 5th, 2025, that assessed the performance of the GAIA case definitions in routine care or research. Two reviewers independently screened studies and extracted data, resolving any discrepancies through consensus. The review protocol was registered with PROSPERO (CRD42025620433).

Results: Database searches identified 4260 records; after deduplication and two-stage screening, 10 studies including 42,587 cases were eligible. Three definitions had not been evaluated (ectopic pregnancy, postpartum endometritis, neurodevelopmental delay), and several others have only been assessed in high-income settings. While CDs for maternal death, neonatal death, and preterm birth performed well, others had high proportions of unclassifiable cases, achieving only LOC-4. Variability was seen with some studies able to classify 100% of potential cases, whilst others were only able to classify 50% at LOC 1–3.

Conclusions: Although progress has been made in evaluating GAIA case definitions, challenges remain in understanding their applicability across settings and data sources. Variability likely reflects differences in setting-level resources, data availability and documentation quality, as well as limitations inherent to the definitions themselves, highlighting the need for targeted refinement and broader validation to support robust, standardised, and comparable maternal immunisation safety surveillance globally.

1. Introduction

The Brighton Collaboration's Global Alignment of Immunisation Safety Assessment in Pregnancy (GAIA) project was established to address the lack of standardised definitions and surveillance approaches for adverse event monitoring within maternal immunisation trials [1]. To achieve this, the GAIA group developed harmonised case definitions for key maternal, fetal, and neonatal adverse outcomes. These

* Corresponding author at: Institute for Global Health, University College London, United Kingdom.
E-mail address: Hannah.g.davies@ucl.ac.uk (H.G. Davies).
[1] Denotes equal contribution.

https://doi.org/10.1016/j.vaccine.2026.128899
Received 16 June 2026; Accepted 29 June 2026
Available online 13 August 2026

definitions were designed to be globally applicable and suitable for use across diverse clinical and research settings, thereby enabling consistent and comparable monitoring of immunisation safety during pregnancy.

Outcome selection was guided by expert consensus following recommendations from a World Health Organization–convened panel in 2014 [1]. The resulting definitions were released between 2016 and 2019 and are structured around levels of diagnostic certainty (LOC), with higher levels reflecting greater diagnostic specificity and lower levels offering increased sensitivity. Levels 1–3 are typically regarded as classifiable cases, whereas level 4 are reported cases with insufficient evidence to classify (at Levels 1–3), and level 5 non-cases (supplementary table 2). This tiered approach was intentionally designed to support applicability across settings with varying resources and diagnostic capacity.

Despite these efforts, uptake and implementation of GAIA definitions in maternal vaccine safety studies have remained limited [2,3]. Proposed explanations for poor uptake include limited implementation capacity, lack of awareness and uncertainty regarding the performance of these definitions when applied to real-world clinical data [2,4]. Several studies have evaluated the performance of GAIA case definitions in both clinical trials and observational research, using prospective and retrospective data sources [5,6].

A 2023 review synthesised the early validation evidence; however, at that time, half of GAIA-defined outcomes had not yet undergone formal evaluation [3]. The key findings from the 2023 review are summarised in Table 1. While key neonatal definitions such as neonatal death, preterm birth, bloodstream infection, and low birth weight generally performed well, others—especially stillbirth, small for gestational age, neonatal meningitis, and several maternal and intrapartum outcomes—performed poorly. Studies further demonstrated that relatively minor clarifications, such as allowing higher-level diagnostic criteria to contribute to classification at lower levels, could substantially improve performance [6].

The review also highlighted that the definitions are resource-intensive [7] and were designed primarily for prospective trials, whereas much pregnancy vaccine safety data is derived from retrospective electronic health records and claims databases, requiring adaptation of GAIA definitions to medical coding and algorithm-based approaches. Importantly, many definitions had either not been evaluated at all or had been assessed only in high-income settings, leaving substantial uncertainty about their performance and applicability in low- and middle-income countries [3]. These gaps underscored the need for further validation and broader testing to support the use of GAIA definitions in pooled global vaccine safety analyses.

The present review provides an updated and expanded synthesis of the literature, incorporating newly published studies and substantially increasing the evidence base to include data for a greater number of outcomes. This update aims to identify definitions that continue to demonstrate suboptimal performance and may require revision, as well as those that remain insufficiently validated across diverse healthcare settings.

2. Methods

This review was performed in accordance with the updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses [8] and was registered in the international prospective register of systematic reviews; PROSPERO ID: CRD42025620433.

2.1. Inclusion criteria

Studies were eligible for inclusion if they assessed the performance of GAIA case definitions using real-world data, whether derived from routine healthcare sources or research settings. Studies were included regardless of whether data were collected prospectively or retrospectively. Studies that applied GAIA definitions, for example, in vaccine

Table 1
Key findings from Davies et al. review paper assessing performance of GAIA definitions [3].

Theme	What was found	Relevance
Overall validation of GAIA definitions	13 of 26 GAIA case definitions (50%) have never been field-tested; two of the highest-priority definitions (maternal death, postpartum haemorrhage) have not been evaluated at all.	Large gaps exist in evidence for tools that are supposed to support global vaccine safety monitoring.
Geographic limitation of evidence	Five definitions had only been tested in a single study using retrospective data and only in high-income countries.	Their performance in low- and middle-income countries (LMICs), where maternal vaccine trials are expanding, is unknown.
Definitions that performed well	Neonatal death, preterm birth, low birthweight, and hypertensive disorders, generally classified well across multiple studies and designs.	These definitions are relatively robust and suitable for broader use.
Poorly performing definitions	There were large proportions of unclassifiable (level 4) cases for stillbirth, small for gestational age, and neonatal infections.	Important outcomes can not be classified at LOC1-3, reducing the accuracy of safety estimates.
Challenges with retrospective data application	Non-reassuring fetal status, dysfunctional labor, congenital microcephaly and pathways to preterm labor were unassessable in large proportions of retrospective data from high-income settings.	These case definitions may not be applicable in surveillance studies without revised definitions or improved data.
Effect of simple rule changes	Explicitly allowing higher-level diagnostic criteria to count at lower LOCs would greatly improve classification (e.g., stillbirth classification improved from 29% to 100% in one study).	Minor wording and logic changes can produce major improvements in definition performance.
Inconsistent requirements within and across definitions	Gestational age and birthweight requirements were inconsistent across neonatal outcome definitions, with some being more lenient and others more stringent.	This led to additional complexity applying the definitions.
Ambiguity in interpreting the definitions	For some outcomes (e.g. stillbirth) there was variable interpretation of the definition across studies.	This ambiguity led to heterogeneous interpretation and inconsistency in application.
Data and workload burden	Record review took 1–2 h per case; training and adjudication were also time-consuming.	This limits feasibility for large studies and routine safety monitoring, especially in LMICs.
Mismatch with real-world data sources	GAIA CDs were designed for prospective trials, but most vaccine safety data come from EHRs, claims data, and ICD coding.	Definitions must be adapted to be usable in post-marketing and observational studies.
Future directions	Ongoing refinement, expansion to additional outcomes, and broader global integration were identified as priorities.	To ensure GAIA case definitions remain relevant as new maternal vaccines and platforms are introduced.
Overall conclusion	GAIA definitions are essential but need validation, clarification to remove ambiguity, and adaptation for real-world and LMIC data.	Without these changes, global vaccine safety comparisons will remain unreliable.

trials or safety studies, but did not assess performance were excluded. Only English-language peer-reviewed publications were eligible. Reference lists of all relevant studies were also screened to identify additional eligible publications. Review papers were excluded.

2.2. Search strategy and selection criteria

Ovid MEDLINE, EMBASE, and Web of Science databases were searched for English-language publications from 2014 to May 5th, 2025. Articles were identified using a search strategy adapted from prior work by Davies et al. [3], combining MeSH terms and keywords related to four core concepts: vaccination: (e.g., "immunisation," "vaccin*"), safety: (e.g., "safety," "pharmacovigilance," "adverse event"), standardisation: (e.g., "standard* defin*," "GAIA"), and population: (e.g., "pregnant women," "matern*," "newborn"). The search strategy was peer reviewed by the London School of Hygiene & Tropical Medicine library team according to the PRESS guideline statement [9]. The full search strategy is shown in Supplementary Table 1. Conference abstracts were not systematically searched as part of the study selection process. However, one potentially relevant study was identified initially through a conference abstract, and data were extracted from the full text article, which became available during this review.

2.3. Screening

The review followed a dual-reviewer screening process at all stages. Search results were imported into Rayyan, where duplicate records were removed using a combination of automated and manual procedures. Two reviewers (RM and HGD) independently screened titles and abstracts against the predefined eligibility criteria. Records that clearly did not meet the inclusion criteria were excluded. Full texts of potentially eligible studies were retrieved and independently assessed by two reviewers (RM and HGD). Studies for which full texts could not be accessed were excluded. Reasons for exclusion were documented and reported in the PRISMA flow diagram. Discrepancies or uncertainties regarding eligibility were resolved through discussion, with consultation of a third reviewer where necessary.

2.4. Data abstraction

Data were extracted using a standardised form developed in Excel and independently checked by a second reviewer for accuracy and completeness. Extracted information included study setting (number of countries and sites), World Bank country income level, data type (research, routine), number of case definitions and cases assessed, and level of certainty assessments (LOC1–4). The results and discussion sections of included articles were also reviewed for a narrative assessment of challenges encountered when applying the definitions. Where required data were not reported in sufficient detail within the published manuscripts, study authors were contacted to request additional information.

2.5. Data synthesis and analysis

The distribution of classifiable cases by levels of certainty (LOC 1–3), as well as those classified as level 4 (reported case with insufficient evidence to classify into LOC 1–3) were summarised across studies. Variability in definition performance was assessed in two ways: first, by presenting the overall distribution of cases across LOC1–4 for each outcome across all studies; and second, by providing the proportion of classifiable cases (LOC1–3) for each outcome within individual studies. Advanced data analysis was conducted in R Version 2025.09.1 + 401. The results, discussion and supplementary materials of included articles were reviewed using a structured narrative approach to identify reported challenges in the application of GAIA definitions and suggested modifications. Relevant text was identified during full-text review and abstracted according to the specific definition to which it referred.

2.6. Risk of bias of included studies

This review examined the applicability and performance of established case definitions in published studies, rather than evaluating intervention effects. As the GRADE framework is intended to assess certainty in effect estimates from comparative studies, its use was not considered appropriate for the objectives of this review. Studies included were not assessed for, nor ranked based on, limitations in design or possible bias.

3. Results

A total of 4260 records were identified through electronic database searches. Following title and abstract screening, 28 articles were selected for full-text review. After assessment for eligibility, 19 papers were excluded, resulting in nine papers included in the review. One additional study was identified through a conference abstract, yielding a total of 10 included studies (Fig. 1). Authors of three included studies were contacted via email for additional data; all responded and provided the requested information. Across the 10 included studies, the performance of GAIA case definitions were evaluated using clinical data in a total of 42,587 cases (Table 3).

Three studies assessed GAIA definitions using routine clinical data, four used research datasets, and two used a combination of both data sources. Data were available from 15 countries, including four high-income countries, eight middle-income countries and three low-income (Table 2). Five studies evaluated definitions using retrospectively collected medical records, three recruited participants prospectively, and two used a combination of retrospective and prospective data collection (Table 2). Overall, performance had been assessed for 11 of 13 (85%) pregnancy and obstetric outcomes; postpartum endometritis and ectopic pregnancy have not yet been evaluated. Among infant outcomes, 12 of 13 (92%) definitions have been assessed, with neurodevelopmental delay the only remaining outcome not yet evaluated in any published study.

The number of cases assessed varied substantially by outcome. Preterm birth (n = 18,404), low birthweight (n = 5863), and small for gestational age (n = 6371) were the most frequently assessed outcomes. In contrast, spontaneous abortion (n = 64), neonatal seizures (n = 27) and failure to thrive (n = 334) were evaluated in single studies. Maternal death was assessed in the fewest cases (n = 13), across two studies conducted in LMIC settings.

Fig. 2 shows the distribution of cases across levels of certainty (LOC 1–4) for the maternal and fetal GAIA outcomes (Panel A) and the neonatal or infant outcomes (Panel B) assessed in the 10 included studies. Among maternal-fetal outcomes, large proportions of cases of gestational diabetes mellitus and non-reassuring fetal status were unclassifiable (LOC-4), accounting for 95.2% and 74.4% of cases, respectively. Approximately 40% of spontaneous abortion cases were also classified as LOC-4; all spontaneous abortion cases were derived from a single study [10]. In contrast, all cases of maternal death and fetal growth restriction were classifiable (LOC 1–3); although both outcomes were assessed in small numbers (supplementary table 3).

Among neonatal outcomes, high proportions of cases of neonatal meningitis (83.3%), encephalopathy (52.8%) and stillbirth (51.5%) were not classifiable (LOC-4). In contrast, most cases of failure to thrive (100%), neonatal death (99.4%) and gestational age/preterm (96.9%) were classifiable at LOC1–3 (supplementary table 4). Overall, several outcomes demonstrated high performance, with more than 80% of cases classifiable at LOC1–3, although some were based on small numbers or limited settings.

Given the diversity of study settings and data sources, Figs. 3 and 4 illustrate substantial variability in the proportion of cases classifiable at LOC1–3 across studies. Among maternal–fetal outcomes, marked

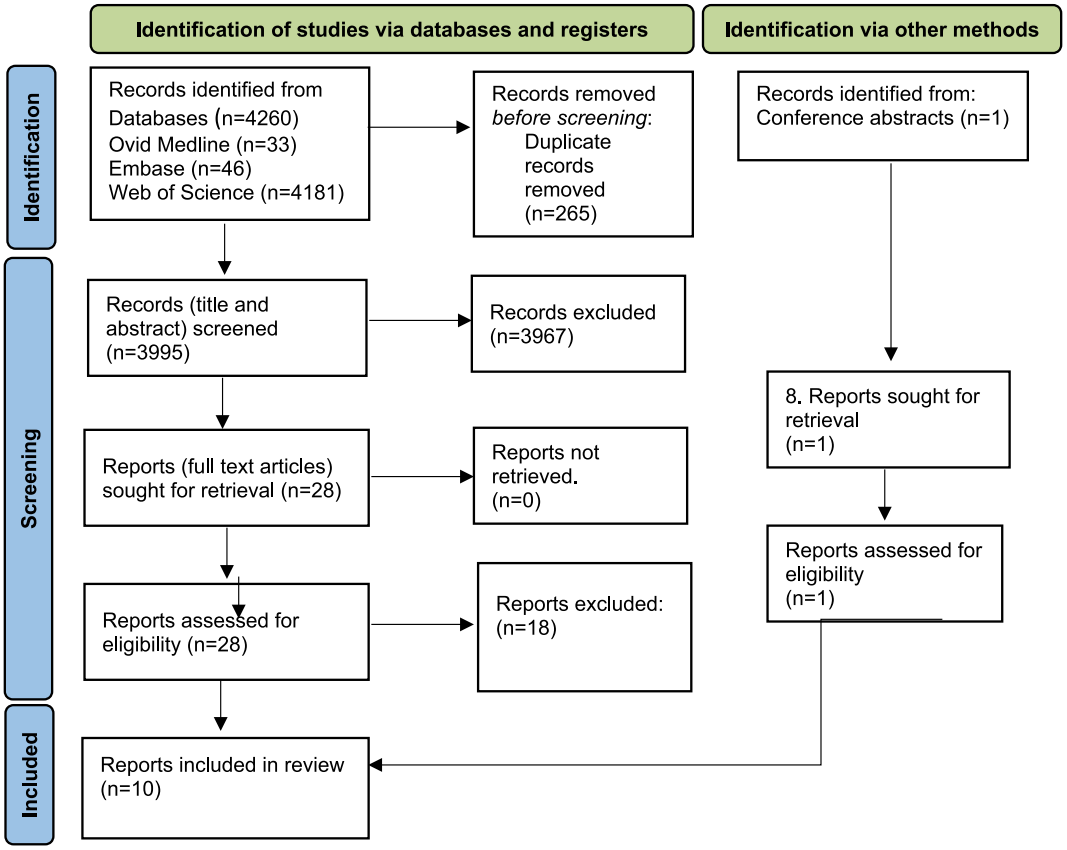

Fig. 1. PRISMA flow-chart showing results for database searches and screening.

heterogeneity was observed for the preterm labour definition, with classification rates ranging from 41.5% to 100%. Similarly, performance of the non-reassuring fetal status definition varied by setting: all cases in high-income settings were classifiable at LOC1–3, whereas none were classifiable in the Ugandan study.

Among neonatal outcomes, a high degree of variability in the proportion of cases classifiable at LOC1–3 between studies was seen for stillbirth (range 0%–78.6%). The highest proportion of classifiable cases was reported in a Kenyan study using prospectively collected research data (55/70 cases) [11], whereas none of the 271 cases assessed retrospectively using routine data from DRC were classifiable at LOC 1–3 [12]. Similarly, high variability was observed for small for gestational age (51.0–100%), microcephaly (57.2–100%) and low birthweight (69.4–100%) outcomes (Figs. 3 and 4).

4. Discussion

Since the 2023 review, which found that 13 GAIA definitions had not undergone formal evaluation [3], important progress has been made in evaluating definition performance. Approximately 30,000 additional cases have now been assessed, leaving only three definitions—ectopic pregnancy, neurodevelopmental delay, and postpartum endometritis—yet to be evaluated. Given public concern about potential neurodevelopmental effects following maternal vaccination [13], it is particularly important that this outcome is rigorously evaluated in a sufficient number of cases to enable robust assessment of definition performance and implementation challenges. Furthermore, several other outcomes of high clinical relevance, including spontaneous abortion and congenital anomalies, remain incompletely evaluated.

Overall, several maternal outcomes (maternal death, dysfunctional labour, and hypertensive disorders) and neonatal outcomes (neonatal death, preterm birth, congenital anomalies, neonatal seizures, low birthweight, and failure to thrive) demonstrated good performance, with more than 80% of assessed cases classifiable at LOC 1–3. However, these findings should be interpreted with caution, as some outcomes have been evaluated only in limited settings or with small numbers of cases—for example, maternal death and congenital anomalies were assessed in relatively few cases, while outcomes such as neonatal seizures and failure to thrive have been evaluated in a limited number of settings.

Notably, definition performance varied substantially across studies, with the proportion of potential cases classifiable at LOC1–3 ranging from approximately 50% to nearly 100%, depending on the data sources and study setting (Figs. 3 and 4). This variability did not appear to be explained solely by national income level or by whether data were collected prospectively or retrospectively, although these factors may have influenced performance for some definitions—particularly those requiring diagnostics more commonly available in higher-resource settings, or those dependent on integrating information across multiple time points or data sources. This may have contributed to the low proportion of classifiable cases observed for outcomes such as gestational diabetes and meningitis, which rely on glucose testing and lumbar puncture, respectively, although this is based on a limited number of studies.

These differences appear to be driven in part by variation in data source and case ascertainment. For example, Izulla et al. classified all low birthweight and small-for-gestational-age cases at LOC1–3 using prospectively collected research data from Kenya [11], whereas only 69.4% and 59.7%, respectively, were classifiable using prospective and retrospective routine data from the Democratic Republic of Congo [12,14]. Similarly, for hypertensive disorders of pregnancy, 100% of cases identified in retrospectively analysed research datasets were classifiable (LOC1–3) [15] compared with only 78.1% of cases identified from routine clinical data in South Africa [16].

Table 2
Summary of the included studies.

Reference	Number CDs	Study Setting	Study Setting (Countries)	Study Setting (Countries)	Sites (number)	Data type	Case definitions assessed
Stuurman et al. 2018 [20]	4	Mixed	8	Ghana, South Africa, Tanzania, Zimbabwe, Spain, India, Nepal, Iran	24	Routine	Preterm birth, neonatal death, bloodstream infections, stillbirth
Kochhar et al. 2019 [15]	4	LMICs	2	South Africa, The Gambia	2	Research Data	Preterm birth, stillbirth, hypertension, GA assessment (enabling term).
Stuurman et al. 2021 [6]	7	Mixed	7	Ghana, Tanzania, Zimbabwe, Iran, India, Nepal, Spain	21	Routine Data	Low birthweight, preterm birth, small for gestational age, stillbirth, neonatal death, neonatal infections (bloodstream, respiratory, meningitis), congenital microcephaly.
Watson et al. 2021 [7]	10	High-income	3	United States, United Kingdom, Australia	7	Both	Preterm birth, low birthweight, small for gestational age, respiratory distress, congenital microcephaly, preterm labor, fetal growth restriction, hypertensive disorders (pre-eclampsia), non-reassuring fetal status, dysfunctional labor.
Gadoth et al. 2022 [12]	6	LMIC	1	Democratic Republic of Congo (DRC)	10	Routine	Low birthweight, preterm birth, congenital microcephaly, neonatal infections (bloodstream), small for gestational age, stillbirth)*
Izulla et al., 2023 [11]	6	LMIC	1	Kenya	N/A	Routine & Research	Neonatal death (term births), small for gestational age, preterm birth, stillbirth, low birthweight, neonatal infections (bloodstream)
Arena et al. 2023 [14] (prospective)	7	LMIC	1	Democratic Republic of Congo (DRC)	10	Research Data	Stillbirth, preterm birth, low birthweight, small for gestational age, microcephaly, neonatal infections (bloodstream), neonatal deaths.
Cutland et al. 2024 [16]	11[#]	LMIC	1	South Africa	2	Routine data	Hypertensive disorders (hypertension, pre-eclampsia, pre-eclampsia with severe features), maternal deaths, gestational diabetes, pathways to preterm (PPROM), postpartum haemorrhage, antepartum bleeding (placental abruption), congenital microcephaly, low birthweight, small for gestational age, neonatal infections (bloodstream, meningitis, respiratory), neonatal death.
Arena et al. 2023 [21] (retrospective)	7	LMIC	1	Democratic Republic of Congo (DRC)	10	Routine	Stillbirth, preterm birth, low birthweight, small for gestational age, microcephaly, neonatal infections (bloodstream), neonatal deaths.
Davies et al [10]	20	LMIC	1	Uganda	1	Observational research data	Hypertensive disorders (hypertension, pre-eclampsia, pre-eclampsia with severe features), maternal death, dysfunctional labor (first stage and second), pathways to preterm, postpartum haemorrhage, antepartum bleeding, spontaneous abortions, chorioamnionitis, stillbirth, preterm birth, congenital microcephaly, low birthweight, small for gestational age, congenital anomalies, neonatal encephalopathy, neonatal infections (bloodstream, meningitis, respiratory, respiratory distress, neonatal seizures, neonatal death, failure to thrive.

*Aimed to look at neonatal deaths but insufficient follow-up to 28 days noted.
[#]Described incidence of stillbirth and spontaneous abortions but did not provide a breakdown of the level of certainty.

This likely reflects variation in how cases are identified and verified. Research datasets typically involve structured data capture and active review of provisional diagnoses, whereas routine clinical records may retain initial diagnostic labels that are not revised as additional information becomes available. Consistent with this, Davies et al. found that neonatal encephalopathy was frequently over-recorded in Ugandan clinical notes, with many cases later judged to be LOC5 (non-cases) by an expert panel [10]. Variability in how suspected cases are initially identified, and how LOC4 and LOC5 are applied across settings, further contributes to inconsistent classification. While LOC4 classifications reflect insufficient evidence to meet LOC 1 to 3, and therefore greater diagnostic uncertainty, they may still be informative for signal detection and may represent a more realistic level of classification in settings where documentation is incomplete or diagnostic capacity is limited. However, this depends on the definition functioning as intended; where definitions systematically classify true cases as LOC4, as previously observed for stillbirth, this may instead reflect limitations in the definition itself and overestimate uncertainty.

This is further illustrated by differences in case identification and classification processes across settings. For example, a prospective study from the Democratic Republic of Congo, Arena et al. found that of 132 potential cases of neonatal sepsis identified using a screening tool, 84% were ultimately classified as LOC5 (non-cases), with only 16% meeting the case definition [14]. This suggests that important differences in case identification and classification processes between settings may influence observed definition performance. Overall, the observed variability likely reflects a combination of differences in setting-level resources, data structure, case ascertainment procedures, diagnostic verification, documentation quality, and interpretation of the definitions, rather than any single factor.

In addition to these setting-related factors, limitations inherent to the GAIA definitions themselves were identified. For infection outcomes,

Table 3
Table showing published GAIA definitions with the year of publication, whether any study validation has been performed, the number of studies in which the definition has been assessed and the total number of cases assessed.

GAIA outcome	Publication Date	Field Validation?	Studies	Cases Assessed (N)*
Stillbirth [22]	2016	Yes	8	1649
Neonatal death [23]	2016	Yes	6	967
Maternal death [24]	2016	Yes	2	13
Congenital anomalies [19]	2016	Yes	2	55
Congenital microcephaly [25]	2017	Yes	7	2531
Fetal growth restriction [26]	2017	Yes	1	99
Non-reassuring fetal status [27]	2016	Yes	2	133
Antenatal bleeding [28]	2017	Yes	2	151
Dysfunctional labor [29]	2017	Yes	2	236
Gestational diabetes mellitus [30]	2017	Yes	1	124
Hypertensive disorders of pregnancy [31]	2016	Yes	3	1631
Pathways to preterm birth [32]	2016	Yes	3	1968
Spontaneous abortion [33]	2017	Yes	1	64
Ectopic Pregnancy [33]	2017	No	–	–
Postpartum haemorrhage [34]	2016	Yes	2	120
Neonatal encephalopathy [35]	2017	Yes	2	125
Failure to thrive [36]	2017	Yes	1	334
Low birthweight [37]	2017	Yes	8	5863
Preterm birth [38]	2016	Yes	10	18,404
Respiratory distress [39]	2017	Yes	2	201
Small for gestational age [40]	2017	Yes	8	6371
Neonatal infections - respiratory [41]	2016	Yes	3	309
Neonatal infections - bloodstream	2016	Yes	8	1107
Neonatal infections - meningitis	2016	Yes	3	96
Chorioamnionitis [42]	2019	Yes	1	9
Neonatal seizures [43]	2019	Yes	1	27
Neurodevelopmental delay [44]	2019	No	–	–
Postpartum endometritis [45]	2019	No	–	–
Total				42,587

some definitions permit fewer clinical signs at LOC-3 than at LOC-2, which may unintentionally restrict classification at lower certainty levels in settings where diagnostic investigations are limited (e.g., blood culture or radiography) or where clinical documentation is incomplete. This inconsistency is illustrated in Supplementary Table 4, where fewer criteria are permitted at LOC-3 than at LOC-2 for conditions such as neonatal bloodstream and respiratory infections. This limitation has been shown to reduce the ability to classify probable cases in practice, as reported by both Davies et al. [10] and Cutland et al. [16] Aligning clinical criteria across LOCs, such that the same signs are permitted at both levels, but a greater number are required to meet LOC2, would likely improve definition performance.

Issues of internal inconsistency were also identified within individual definitions. Overall, the preterm birth definition demonstrated a high proportion of classifiable cases (96.9%); however, performance varied by data source. In routine clinical data from South Africa, 5.7% of cases were assigned LOC-4 (unclassifiable), highlighting challenges when key gestational age elements were missing [16]. In contrast, all preterm birth cases were classifiable in a research setting with routine ultrasound dating [10]. However, some cases with implausible gestational age estimates were still classified at LOC1/2 because both ultrasound dating and a recorded last normal menstrual period (LNMP) were present. This highlights a limitation of the definition; whereby formal criteria can be met despite internally inconsistent data. In addition, variability was observed in how the certainty of LNMP was determined across studies: some authors accepted any recorded LNMP as evidence of certainty, whereas others required explicit confirmation from the pregnant woman. While the definition specifies how compatibility between LNMP and ultrasound findings should be assessed, it does not provide guidance on how the certainty of LNMP itself should be established, nor on how to handle implausible gestational age estimates once criteria are met. Providing explicit guidance on assessing LNMP reliability and downgrading certainty in the presence of implausible gestational age would strengthen the definition.

A related issue was identified in the small-for-gestational-age (SGA) definition, where birthweight is permitted as a criterion for gestational age assessment, introducing circular logic. Gestational age should instead be determined as a standalone parameter using a consistent approach, with SGA classification applied only once gestational age has been established. Aligning the SGA definition with the gestational age criteria used for preterm birth [38] would improve conceptual clarity and reduce the potential for inconsistent classification. More broadly, related inconsistencies across definitions have also been noted. Stuurman et al. reported that the neonatal death definition incorporates an internal assessment of maturity that differs from the gestational age criteria specified in the preterm birth definition, introducing further inconsistency across GAIA outcomes [6].

Ambiguity in the criteria and their interpretation in practice was also observed. The previous review highlighted uncertainty in the stillbirth definition regarding the number of clinical signs required for classification; this definition is already undergoing revision to address this lack of clarity and the high proportion of LOC-4 (unclassifiable) cases. Similarly, Davies et al. reported ambiguity in the non-reassuring fetal status definition, where a strict interpretation requiring the presence of all listed signs resulted in no cases (0/99) being classifiable at LOC1–3 [10]. These findings suggest that lack of clarity in the definitions, together with variation in how they are interpreted in practice, contributes to inconsistent application. Future revisions of GAIA definitions should prioritise improving clarity, with additional guidance or training resources supporting consistent interpretation in practice.

Given the limited inclusion of pregnant women in early clinical trials of new vaccines and therapeutics, post-licensure observational studies using real-world data often provide some of the earliest evidence on vaccine safety during pregnancy [4,17]. Such studies frequently rely on information captured in electronic health records [18], medical and insurance claims data, and population-based or regional registries. However, these data are typically recorded using standardised medical coding systems (e.g., ICD, MedDRA, and SNOMED), which often lack the clinical detail required to fully apply GAIA case definitions. As a result, many cases identified through these sources are classified as LOC-4 due to insufficient supporting information. Incorporating these coding systems into revised case definitions—particularly by allowing coded data to contribute to lower levels of certainty (e.g., LOC3), alongside clear guidance on validation approaches—would facilitate more complete case ascertainment and reduce the proportion of unclassifiable cases. This would improve the utility of real-world data while maintaining appropriate levels of diagnostic certainty. This need has already been recognised in the revision of the stillbirth definition and in more recently published definitions, including congenital anomaliess [19], and should be extended to other GAIA definitions.

This study has several limitations. Although GAIA definitions were originally developed for use in clinical trials in pregnant women, most included studies in this review assessed their performance using observational research or routine data sources. As a result, the observed variability in definition performance may partly reflect the constraints

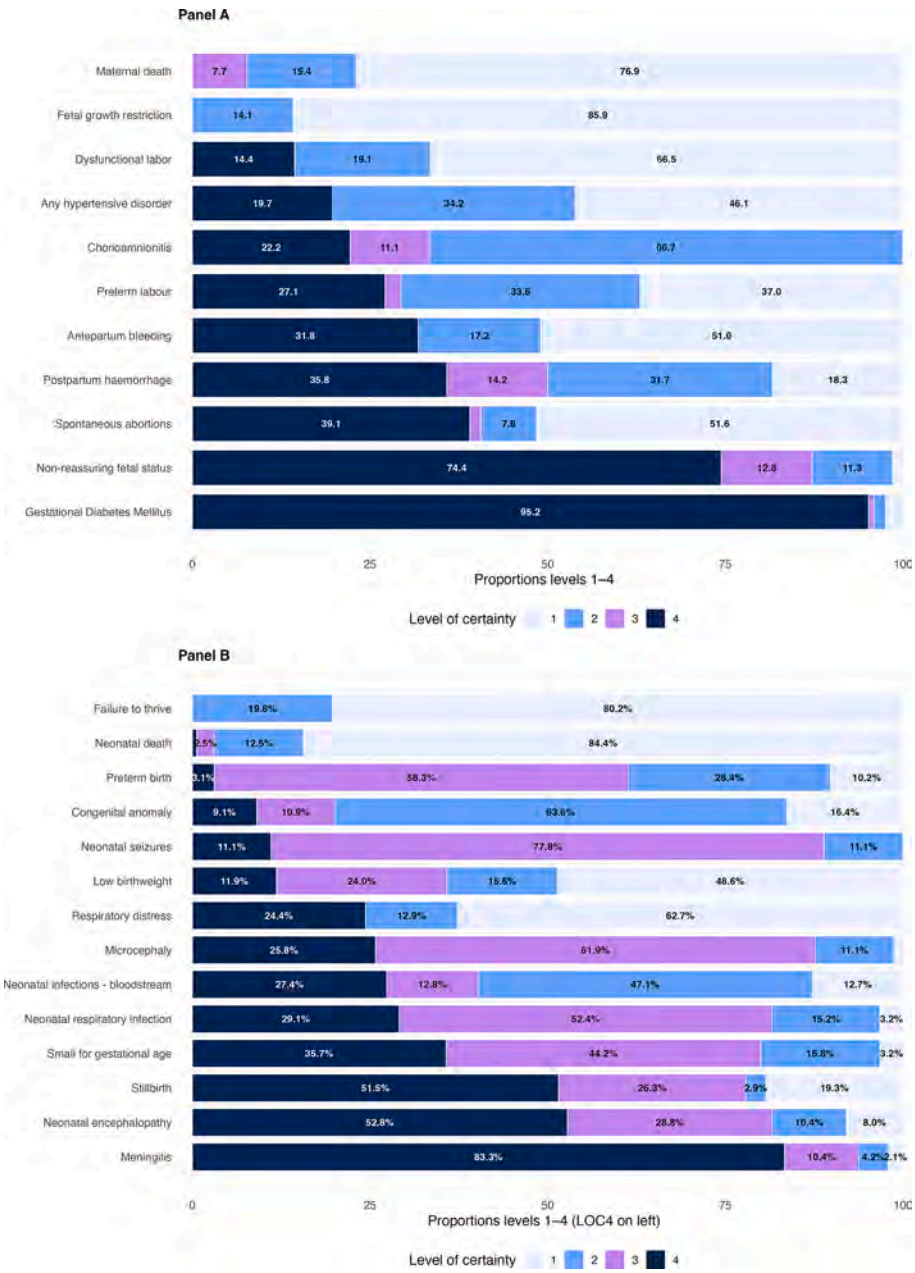

Fig. 2. Figure showing a breakdown of the level of certainty (levels 1–4) for the maternal and fetal GAIA outcomes (Panel A) and the neonatal or infant outcomes (Panel B).

of these data sources, and it remains unclear whether performance would be improved in settings with more complete and systematically collected trial data. In addition, some outcomes were evaluated in small numbers of cases or in limited settings, and variability in data quality and reporting across studies may have influenced the findings. The search was conducted in three major bibliographic databases (Web of Science, Embase, and Ovid MEDLINE); while these provide broad coverage of biomedical literature, additional sources were not searched, and some relevant studies may not have been captured. Finally, the narrative synthesis relied on information reported within included studies, which may have varied in detail and completeness.

In conclusion, important progress has been made in evaluating GAIA case definitions; however, challenges remain that limit understanding of their applicability across settings and data sources. Variability in performance is likely to reflect a combination of differences in setting-level resources, data availability, documentation quality, and limitations inherent to the definitions themselves, as illustrated by the variation observed across data sources and study settings in this review. Understanding the relative contribution of these factors is essential for targeted revision. Clarifying ambiguous criteria, ensuring internal logical consistency, providing additional training resources, and adapting definitions for use with real-world data, including explicit guidance on validation, will be essential to support robust and comparable maternal vaccine safety assessments globally.

Statement on artificial intelligence

During the preparation of this work, the authors used ChatGPT (OpenAI) to assist with language editing, grammar, sentence flow, and structural refinement of the manuscript. After using this tool, the

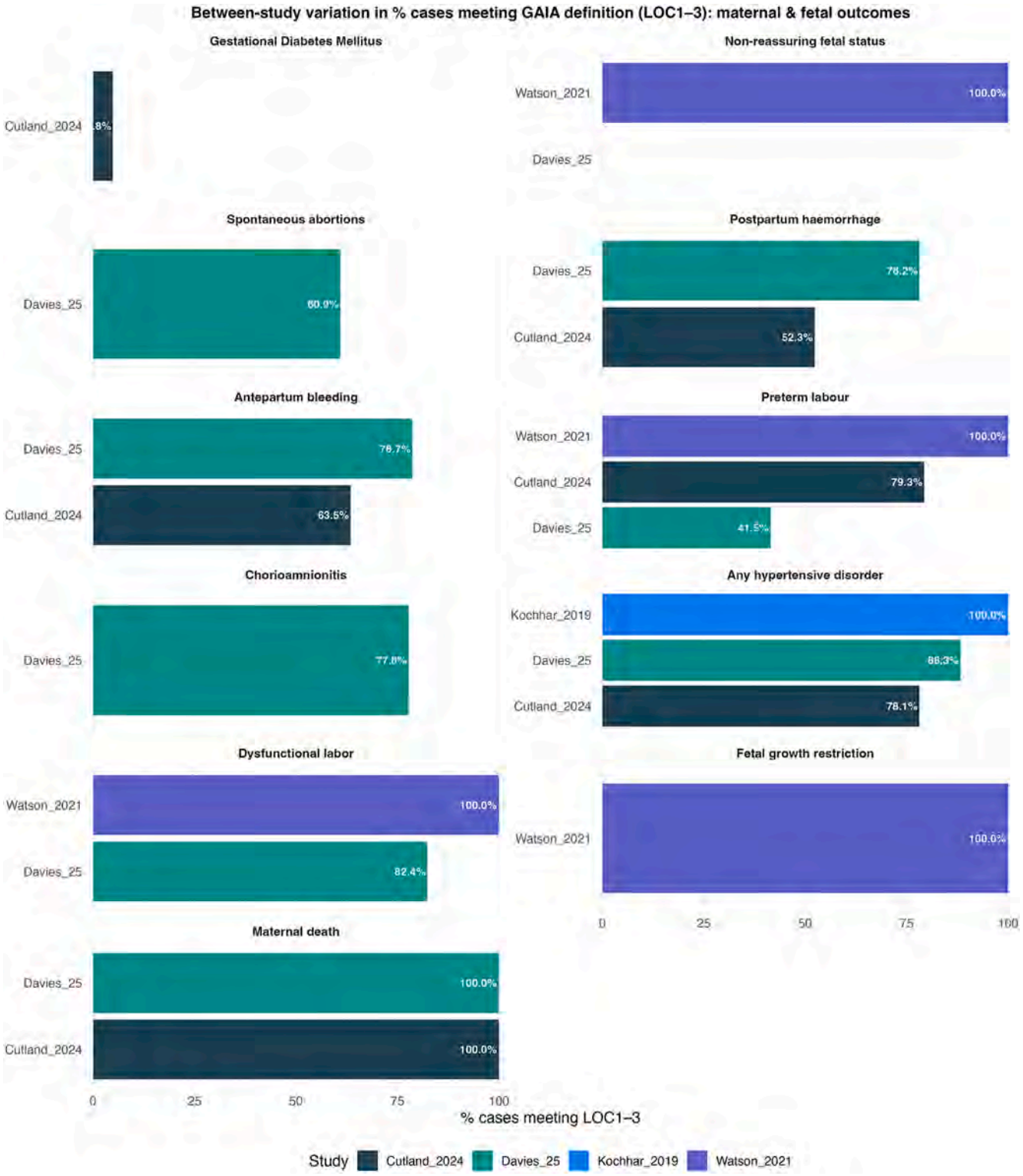

Fig. 3. Figure showing the variability between studies for the total proportion of maternal and fetal cases assessed (levels 1-4) meeting the case definition (levels 1-3).

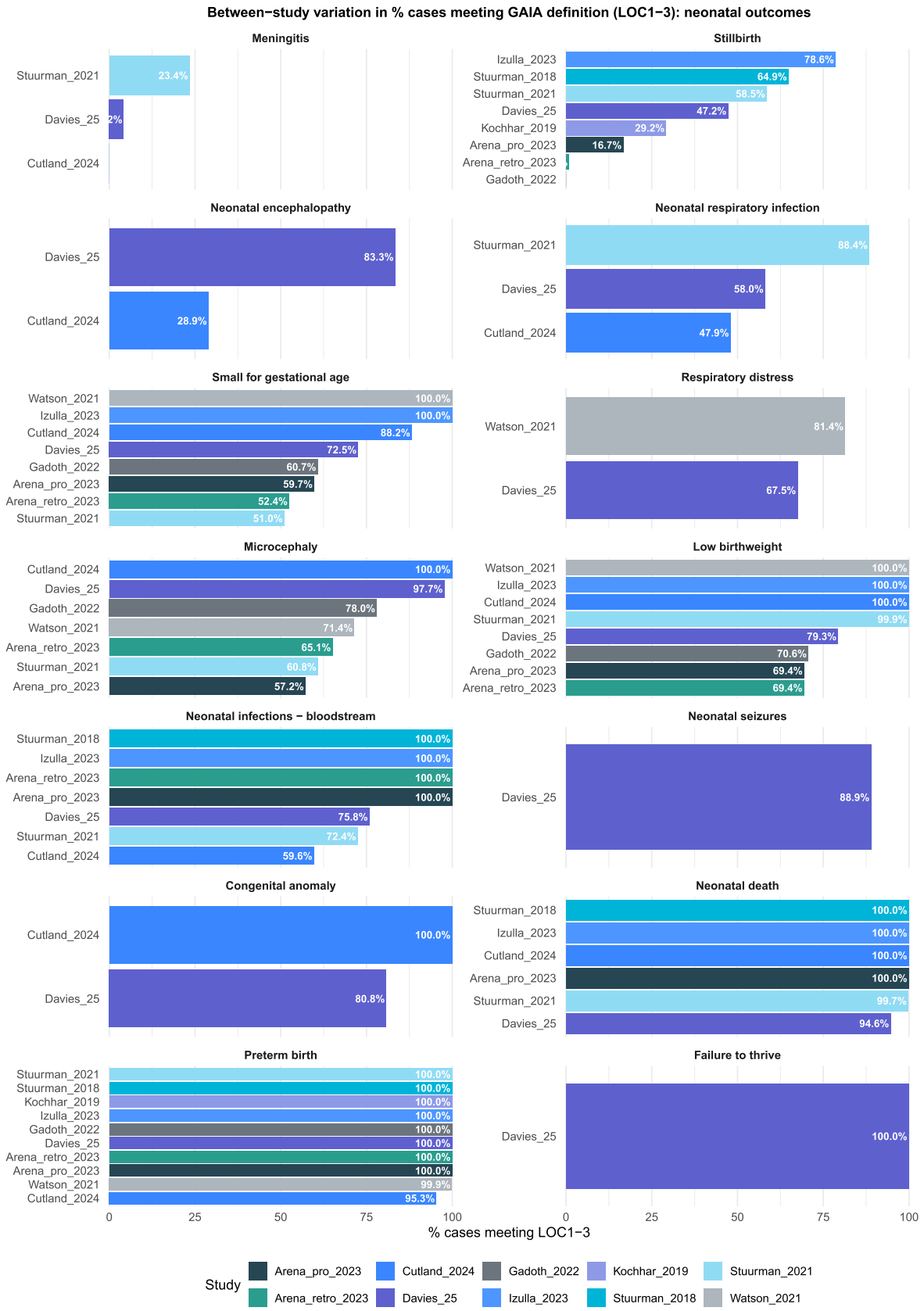

Fig. 4. Figure showing the variability between studies for the total proportion of neonatal and infant cases assessed (levels 1-4) meeting the case definition (levels 1-3).

authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Robert Mboizi: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Kirsten Maertens:** Investigation, Supervision, Writing – review & editing. **Kirsty Le Doare:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Clare L. Cutland:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Flor M. Muñoz:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing. **Hannah G. Davies:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding statement

This work was performed as part of the Safety Platform for Emergency Vaccines (SPEAC), a project of the Coalition for Epidemic Preparedness Innovations (CEPI) and Brighton Collaboration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vaccine.2026.128899.

Data availability

The authors do not have permission to share data.

References

[1] Bonhoeffer J, Kochhar S, Hirschfeld S, Heath PT, Jones CE, Bauwens J, et al. Global alignment of immunization safety assessment in pregnancy - the GAIA project. Vaccine 2016;34(49):5993–7. https://doi.org/10.1016/j.vaccine.2016.07.006.

[2] Davies HG, Thorley EV, Al-Bahadili R, Sutton N, Burt J, Hookham L, et al. Defining and reporting adverse events of special interest in comparative maternal vaccine studies: a systematic review. Vaccine X Elsevier Ltd 2024. https://doi.org/10.1016/j.jvacx.2024.100464.

[3] Davies HG, Conor Bowman, Watson G, Dodd C, Jones CE, Munoz FM, et al. Standardizing case definitions for monitoring the safety of maternal vaccines globally: GAIA definitions, a review of progress to date. Int J Gynaecol Obstet 2023;00:1–10. https://doi.org/10.1002/IJGO.14843.

[4] Regan AK, Rowe SL. Use of routinely collected "real-world" data for the postlicensure safety assessment of vaccination during pregnancy 2025;44(2):108–10. https://doi.org/10.1097/INF.0000000000004606.

[5] Kochhar S, Bonhoeffer J, Jones CE, Munoz FM, Honrado A, Bauwens J, et al. Immunization in pregnancy clinical research in low- and middle-income countries - study design, regulatory and safety considerations. Vaccine 2017;35(48 Pt A):6575–81. https://doi.org/10.1016/j.vaccine.2017.03.103.

[6] Stuurman AL, Sharan A, Jahagirdar S, Elango V, Riera-Montes M, Kashyap N, et al. WHO global vaccine safety multi-country collaboration project on safety in pregnancy: assessing the level of diagnostic certainty using standardized case definitions for perinatal and neonatal outcomes and maternal immunization. Vaccine X 2021;9:100123. https://doi.org/10.1016/j.jvacx.2021.100123.

[7] Watson G, Dodd C, Munoz FM, Eckert LO, Jones CE, Buttery JP, et al. Applicability of the GAIA maternal and neonatal outcome case definitions for the evaluation of adverse events following vaccination in pregnancy in high-income countries. Pediatric Infectious Disease Journal 2021;40(12):1127–34. https://doi.org/10.1097/inf.0000000000003261.

[8] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372. https://doi.org/10.1136/BMJ.N71 [PubMed PMID: 33782057].

[9] McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016;75:40–6. https://doi.org/10.1016/j.jclinepi.2016.01.021 [PubMed PMID: 27005575].

[10] Davies HG, Rukundo G, Komugisha C, Kipyeko S, Young A, Kisakye J, et al. Baseline adverse birth outcomes in a prospective pregnancy cohort in Uganda: Evidence to inform maternal vaccine readiness. Vaccine 2026. https://doi.org/10.1016/j.vaccine.2026.128952. In press.

[11] Izulla P, Wagai JN, Akelo V, Ombeva A, Okeri E, Onyango D, et al. Vaccine safety surveillance in Kenya using GAIA standards: a feasibility assessment of existing national and subnational research and program systems [internet]. 2023. https://doi.org/10.1016/j.vaccine.2023.07.063.

[12] Gadoth A, Mukadi Nkamba D, Arena PJ, Hoff NA, Dzogang C, Kampilu D, et al. Assessing the feasibility of passive surveillance for maternal immunization safety utilizing archival medical records in Kinshasa, Democratic Republic of the Congo. Vaccine 2022;40(26):3605–13. https://doi.org/10.1016/J.VACCINE.2022.04.073 [PubMed PMID: 35570074].

[13] Neeman S, Hemo M, Meiri G, Shmueli D, Menashe I. Maternal influenza vaccination during pregnancy and risk of autism spectrum disorder in the offspring. JCPP Advances 2025. https://doi.org/10.1002/jcv2.70085.

[14] Arena PJ, Gadoth A, Nkamba DM, Dzogang C, Hoff N, Barrall AL, et al. Capability and feasibility of the global alignment of immunisation safety assessment in pregnancy criteria for the assessment of pregnancy and birth outcomes in Kinshasa, Democratic Republic of the Congo: a prospective cohort study. BMJ Glob Health 2023;1:35. https://doi.org/10.1136/bmjph-2023-000035.

[15] Kochhar S, Clarke E, Izu A, Emmanuel Kekane-Mochwari K, Cutland CL. Immunization in pregnancy safety surveillance in low and middle-income countries- field performance and validation of novel case definitions. Vaccine 2019;37(22):2967–74. https://doi.org/10.1016/j.vaccine.2019.03.074.

[16] Cutland CL, Sawry S, Fairlie L, Barnabas S, Frajzyngier V, Le Roux J, et al. Obstetric and neonatal outcomes in South Africa [internet]. 2024. https://doi.org/10.1016/j.vaccine.2024.01.054.

[17] World Health Organization. Guidance for best practices for clinical trials. Geneva; 2024.

[18] Mesfin YM, Cheng A, Lawrie J, Buttery J. Use of routinely collected electronic healthcare data for postlicensure vaccine safety signal detection: a systematic review. BMJ Glob Health 2019;4(4):1–14. https://doi.org/10.1136/bmjgh-2018-001065.

[19] DeSilva M, Munoz FM, Mcmillan M, Kawai AT, Marshall H, Macartney KK, et al. Congenital anomalies: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2016;34(49):6015–26. https://doi.org/10.1016/j.vaccine.2016.03.047.

[20] Stuurman AL, Riera M, Lamprianou S, Perez-Vilar S, Anderson SA, Mangtani P, et al. Vaccine safety surveillance in pregnancy in low- and middle-income countries using GAIA case definitions: a feasibility assessment. Vaccine 2018;36(45):6736–43. https://doi.org/10.1016/j.vaccine.2018.09.033.

[21] Arena PJ, Dzogang C, Gadoth A, Nkamba DM, Hoff NA, Kampilu D, et al. Comparison of adverse pregnancy and birth outcomes using archival medical records before and during the first wave of the COVID-19 pandemic in Kinshasa, Democratic Republic of Congo: a facility-based, retrospective cohort study. BMC Pregnancy Childbirth 2023;23(1):1–12. https://doi.org/10.1186/s12884-022-05291-w [PubMed PMID: 36647021].

[22] Tavares Da Silva F, Gonik B, McMillan M, Keech C, Dellicour S, Bhange S, et al. Stillbirth: case definition and guidelines for data collection, analysis, and presentation of maternal immunization safety data. Vaccine 2016;34(49):6057–68. https://doi.org/10.1016/j.vaccine.2016.03.044.

[23] Pathirana J, Munoz FM, Abbing-Karahagopian V, Bhat N, Harris T, Kapoor A, et al. Neonatal death: case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2016;34(49):6027–37. https://doi.org/10.1016/j.vaccine.2016.03.040.

[24] Patwardhan M, Eckert LO, Spiegel H, Pourmalek F, Cutland C, Kochhar S, et al. Maternal death: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2016;34(49):6077–83. https://doi.org/10.1016/j.vaccine.2016.03.042.

[25] DeSilva M, Munoz FM, Sell E, Marshall H, Tse Kawai A, Kachikis A, et al. Congenital microcephaly: case definition & guidelines for data collection, analysis, and presentation of safety data after maternal immunisation. Vaccine 2017;35(48 Pt A):6472–82. https://doi.org/10.1016/j.vaccine.2017.01.044.

[26] Easter SR, Eckert LO, Boghossian N, Spencer R, Oteng-Ntim E, Ioannou C, et al. Fetal growth restriction: case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2017;35(48 Pt A):6546–54. https://doi.org/10.1016/j.vaccine.2017.01.042.

[27] Gravett C, Eckert LO, Gravett MG, Dudley DJ, Stringer EM, Mujobu TBM, et al. Non-reassuring fetal status: case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2016;34(49):6084–92. https://doi.org/10.1016/j.vaccine.2016.03.043 [PubMed PMID: 27461459].

[28] Prabhu M, Eckert LO, Belfort M, Babarinsa I, Ananth CV, Silver RM, et al. Antenatal bleeding: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2017;35(48 Pt A):6529–37. https://doi.org/10.1016/j.vaccine.2017.01.081.

[29] Boatin AA, Eckert LO, Boulvain M, Grotegut C, Fisher BM, King J, et al. Dysfunctional labor: case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2017;35(48 Pt A):6538–45. https://doi.org/10.1016/j.vaccine.2017.01.050.

[30] Kachikis A, Eckert LO, Walker C, Oteng-Ntim E, Guggilla R, Gupta M, et al. Gestational diabetes mellitus: case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2017;35(48 Pt A): 6555–62. https://doi.org/10.1016/j.vaccine.2017.01.043.
[31] Rouse CE, Eckert LO, Wylie BJ, Lyell DJ, Jeyabalan A, Kochhar S, et al. Hypertensive disorders of pregnancy: case definitions & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2016; 34(49):6069–76. https://doi.org/10.1016/j.vaccine.2016.03.038.
[32] Harrison MS, Eckert LO, Cutland C, Gravett M, Harper DM, McClure EM, et al. Pathways to preterm birth: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2016;34(49): 6093–101. https://doi.org/10.1016/j.vaccine.2016.03.054.
[33] Rouse CE, Eckert LO, Babarinsa I, Fay E, Gupta M, Harrison MS, et al. Spontaneous abortion and ectopic pregnancy: case definition & guidelines for data collection, analysis, and presentation of maternal immunization safety data. Vaccine 2017;35 (48 Pt A):6563–74. https://doi.org/10.1016/j.vaccine.2017.01.047.
[34] Kerr R, Eckert LO, Winikoff B, Durocher J, Meher S, Fawcus S, et al. Postpartum haemorrhage: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2016;34(49):6102–9. https://doi.org/10.1016/j.vaccine.2016.03.039.
[35] Sell E, Munoz FM, Soe A, Wiznitzer M, Heath PT, Clarke ED, et al. Neonatal encephalopathy: case definition & guidelines for data collection, analysis, and presentation of maternal immunisation safety data. Vaccine 2017;35(48 Pt A): 6501–5. https://doi.org/10.1016/j.vaccine.2017.01.045.
[36] Ross E, Munoz FM, Edem B, Nan C, Jehan F, Quinn J, et al. Failure to thrive: case definition & guidelines for data collection, analysis, and presentation of maternal immunisation safety data. Vaccine 2017;35(48 Pt A):6483–91. https://doi.org/10.1016/j.vaccine.2017.01.051.
[37] Cutland CL, Lackritz EM, Mallett-Moore T, Bardají A, Chandrasekaran R, Lahariya C, et al. Low birth weight: case definition & guidelines for data collection, analysis, and presentation of maternal immunization safety data. Vaccine 2017;35 (48 Part A):6492. https://doi.org/10.1016/J.VACCINE.2017.01.049 [PubMed PMID: 29150054].
[38] Quinn JA, Munoz FM, Gonik B, Frau L, Cutland C, Mallett-Moore T, et al. Preterm birth: case definition & guidelines for data collection, analysis, and presentation of immunisation safety data. Vaccine 2016;34(49):6047–56. https://doi.org/10.1016/j.vaccine.2016.03.045.
[39] Sweet LR, Keech C, Klein NP, Marshall HS, Tagbo BN, Quine D, et al. Respiratory distress in the neonate: case definition & guidelines for data collection, analysis, and presentation of maternal immunization safety data. Vaccine 2017;35(48 Pt A): 6506–17. https://doi.org/10.1016/j.vaccine.2017.01.046.
[40] Schlaudecker EP, Munoz FM, Bardaji A, Boghossian NS, Khalil A, Mousa H, et al. Small for gestational age: case definition & guidelines for data collection, analysis, and presentation of maternal immunisation safety data. Vaccine 2017;35(48 Pt A): 6518–28. https://doi.org/10.1016/j.vaccine.2017.01.040.
[41] Vergnano S, Buttery J, Cailes B, Chandrasekaran R, Chiappini E, Clark E, et al. Neonatal infections: case definition and guidelines for data collection, analysis, and presentation of immunisation safety data. Vaccine 2016;34(49):6038–46. https://doi.org/10.1016/j.vaccine.2016.03.046.
[42] Kachikis A, Eckert LO, Walker C, Bardaji A, Varricchio F, Lipkind HS, et al. Chorioamnionitis: case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2019;37(52):7610–22. https://doi.org/10.1016/j.vaccine.2019.05.030.
[43] Pellegrin S, Munoz FM, Padula M, Heath PT, Meller L, Top K, et al. Neonatal seizures: case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2019;37(52):7596. https://doi.org/10.1016/J.VACCINE.2019.05.031 [PubMed PMID: 31783981].
[44] Villagomez AN, Muñoz FM, Peterson RL, Colbert AM, Gladstone M, MacDonald B, et al. Neurodevelopmental delay: case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2019;37(52):7623. https://doi.org/10.1016/J.VACCINE.2019.05.027 [PubMed PMID: 31783983].
[45] Rouse CE, Eckert LO, Munoz FM, Stringer JSA, Kochhar S, Bartlett L, et al. Postpartum endometritis and infection following incomplete or complete abortion: case definition & guidelines for data collection, analysis, and presentation of maternal immunization safety data. Vaccine 2019;37(52):7585–95. https://doi.org/10.1016/j.vaccine.2019.09.101.