Antiseizure meDication Exposure and Pregnancy and neonaTal outcomes research (ADEPT)

Feasibility of estimating the risk of adverse pregnancy, neonatal and long-term child outcomes associated with *in utero* maternal antiseizure medications (ASMs) exposure or peri-conceptional maternal/paternal ASM exposure

Rationale and background

The safety of certain ASMs, including valproate, during pregnancy remains a key concern due to their known teratogenic and neurodevelopmental risks. Conflicting findings also suggest potential neurodevelopmental impacts from paternal ASM exposure pre-conception. Thus, assessing the feasibility of conducting causal studies on ASM exposure, especially given the complexity of father-child data linkage, was considered essential.

Research question and objectives

This study aimed to assess whether available real-world data sources can be used to investigate the effects of maternal and paternal exposure to ASMs on pregnancy, neonatal and child outcomes, and to assess their overall fitness-for-purpose. The following sub-objectives were addressed: a) to estimate the availability of relevant information for pregnant women; b) to estimate availability of relevant information for men and linkage with pregnancies; c) to estimate availability of relevant information for neonates/children, including comparisons between paternal linked and non-linked records where applicable; and d) to evaluate the overall fitness-for-purpose of the data sources for different types of studies on periconceptional and prenatal ASM exposure and associated outcomes.

Study design

A retrospective cohort study was conducted to assess the suitability of the available databases for ASM exposure studies, and to complete a fit-for-purpose evaluation using the SPIFD (Structured Process to Identify Fit-for-Purpose Data) framework. In addition, the study investigated where to obtain additional information that may not be captured in available databanks of the data source and explored the potential to enhance EHR data through supplementary sources by a survey among participating data experts and access providers (DEAPs).

Population

The source population comprised on average 11,544,388 women of childbearing age and 13,433,397 men of reproductive potential (≥12 years) per calendar year. In total, 7,986,129 pregnancies with the required lookback period of 365 days prior to last menstrual period (LMP) were identified. Paternal linkage was available for 1,245,338 pregnancies. The population of liveborn neonates recorded within four weeks after birth included 3,270,673 neonates, all of whom had linked maternal information, and 1,165,192 (35.6%) had linked paternal information.

Variables

The main exposures of interest were ASMs (ATC code N03A) and gabapentinoids (N02BF) and all benzodiazepines with antiepileptic properties. Outcome parameters were considered per sub-objective. Clinical outcomes of interest included adverse pregnancy outcomes (i.e., spontaneous pregnancy loss, induced abortion, stillbirth and preterm birth) and adverse neonatal/child outcomes (i.e., small for gestational age (SGA), congenital anomalies, Neonatal Intensive Care Unit (NICU) admission, neonatal death, infant growth failure, intellectual disability, autism spectrum disorder, ADHD and speech delay).

Data sources

The study was conducted using electronic health record data from 9 data sources in 7 European countries, covering an overall population of 65 million persons. These included BIFAP (ES), CPRD Gold (UK), EFEMERIS (FR), Finnish linked registries (FI), Norwegian linked registries (NO), PHARMO (NL), SIDIAP (ES), Val Padana LHU (IT), and VID (ES).

Data analysis

For objectives a-c, a total of 40 feasibility parameters were estimated using descriptive analyses and visualizations. These parameters informed the structured fitness-for-purpose assessment in objective d, which evaluated whether the data sources contain the information required to address two potential research questions. For all scores below 4, potential strategies for data enrichment were identified. The study-specific design elements, minimal criteria, and the scoring system were defined in collaboration with the EMA. Feasibility was assessed by evaluating the availability and applicability of participating data sources using the SPIFD framework, including the presence of key study variables and their potential use for the research objectives. In addition, a survey of participating data experts and access providers (DEAPs) was conducted to identify potentially relevant supplementary data sources, including opportunities for primary data collection. The survey leveraged local expert knowledge to map additional resources across participating countries and to assess their potential contribution to future studies. As the identification of sources relied on DEAP knowledge rather than a standardized, systematically validated source-mapping process, additional relevant data sources may exist. The feasibility of incorporating these supplementary resources in future research was also evaluated.

Main results on the EHR data

On average, the data sources covered 11.5 million women of childbearing age and 13.4 million men of reproductive potential per year. A total of 7,986,129 pregnancies with adequate lookback were identified. Among these, 5.3% (n=424,092) involved ASM use prior to pregnancy and 2.6% (n=208,143) during pregnancy. Father–pregnancy linkage was available for 1,346,334 pregnancies (16.9% of all identified pregnancies), of whom 3.0% (n=39,771) had paternal ASM exposure in the year prior to conception. The neonatal cohort comprised 3,752,232 liveborn children, with 87.2% (n=3,270,673) recorded within four weeks after birth and 37.3% (n=1,219,944) linked to paternal information. Among neonates, 1.5% (n=47,666) were exposed in utero, 3.1% (n=101,875) had periconceptional maternal exposure, and 3.2% (n=38,451) had periconceptional paternal exposure. Notable variability was observed across databases in exposure prevalence and data capture, reflecting differences in healthcare systems and recording practices. Most data sources provided good coverage of women of childbearing age, reliable pregnancy identification, and well-established mother–child linkage, supporting studies of maternal exposure and common pregnancy and early-life outcomes. Only four data sources were able to perform father–child linkage, and linkage completeness varied considerably. Only VID and NLRs currently enable deterministic linkage through unique personal identifiers, allowing identification of fathers in a high proportion of cases. In contrast, PHARMO and CPRD rely on probabilistic approaches, resulting in more limited linkage completeness. Although many data sources supported follow-up into childhood, differences in follow-up duration limited the assessment of long-term outcomes, including neurodevelopmental disorders, which were incompletely and inconsistently captured.

Overall, the strongest fitness-for-purpose profiles were observed in databases combining comprehensive healthcare coverage with deterministic family linkage and long-term follow-up capabilities, whereas the principal limitations across databases related to incomplete paternal linkage, heterogeneous child outcome ascertainment, and incomplete integration of secondary care and external registry information.

Table S1. Summary of fitness-for-purpose assessment by data source and outcome

Data source	Maternal exposure studies[i]	Paternal exposure studies[ii]	Neonatal outcomes[iii]	Long-term child outcomes[iv]	Fitness-for-purpose assessment for the current data instances
VID (Spain)	High	High	High	High	Follow-up up to 14.5 years is available (herein shown until 11 years) available; father linkage available (52.4%), with moderate proportions across paternal linkage indicators.
Val Padana LHU (Italy)	High	Low	High	Moderate	No father linkage; follow-up to age 15 years not available; neonatal

					mortality ascertainment affected by recording practices.
SIDIAP (Spain)	High	Low	High	High	No father linkage. Secondary care information limited mainly to hospital discharge diagnoses.
EFEMERIS (France)	Moderate	Low	Moderate–High	Low–Moderate	Restricted to pregnancies with outcomes; no father linkage; limited lookback period; ADHD not captured.
Finnish linked registries (FLR)	High	Low	High	High	Restricted to pregnancies and offspring from those pregnancies; no father linkage.
Norwegian linked registries (NLR)	High	High	High	High	Follow-up to age 15 years available across multiple outcomes; father linkage available (67.1%), with high proportions across paternal linkage indicators.
PHARMO (Netherlands)	High	Moderate	Moderate–High	Moderate–High	Mother–child linkage only moderate (>50%); father linkage available (25.5%); no secondary care data.
BIFAP (Spain)	Moderate–High	Low	Low	Low	No mother–child linkage; no father linkage; neonatal and child follow-up elements not captured as a consequence.
CPRD Gold (UK)	High	Moderate	Moderate	Low	Father linkage available (16.1%); secondary care outcomes not included in the current version. HES linkage ongoing.

[i]*Maternal exposure assessment categories are primarily driven by structural characteristics of the underlying data source, including the definition of the source population (broad female population vs pregnancy-restricted cohorts) and longitudinal completeness (lookback period and continuity of historical data capture). Secondary considerations include completeness of pregnancy phenotyping (pregnancy endings, gestational age, multiple pregnancies, and conception type) and richness of maternal clinical information (risk factors, comorbidities, medication use, and healthcare utilisation), while mother–child linkage is considered supportive but not determinative. Limited source population definition and/or restricted lookback period may cap the maximum achievable categorisation irrespective of pregnancy data quality.*

[ii]*Paternal exposure assessment categories are primarily based on the proportion of pregnancies with linked father–child records among live births. Thresholds were defined as: High >70%, Moderate–High 50–70%, Moderate 1–49%, Low no linkage. Final categorisation was further informed by linkage robustness (deterministic vs probabilistic), type of linkage identifier (father ID vs household proxy), and internal consistency of linkage distribution across pregnancy outcomes.*

[iii]*Neonatal outcome assessment categories are primarily determined by the availability and completeness of mother–child linkage, which is a prerequisite for neonatal follow-up. Where linkage is available, categorisation is further informed by the breadth and completeness of routinely captured neonatal outcomes, including birthweight, gestational age, small-for-gestational-age status, NICU admission, congenital anomalies, and neurodevelopmental outcomes. High completeness requires near-complete linkage and comprehensive outcome capture. Moderate–High reflects substantial but incomplete linkage and/or minor gaps in outcome availability. Moderate reflects limited outcome completeness despite linkage availability. Absence of mother–child linkage results in classification as Low, as neonatal outcomes cannot be reliably ascertained.*

[iv]*Long-term child outcome categories are primarily assessed based on (i) availability and completeness of mother–child linkage, (ii) adequacy of longitudinal follow-up across clinically relevant developmental windows (1–3, 5, 11 and 15 years), and (iii) completeness of outcome capture across key neurodevelopmental domains (intellectual disability, autism spectrum disorders and ADHD) and supporting developmental outcomes (growth failure/stunting and speech delay). High: mother–child linkage available; follow-up extending to 15 years; comprehensive capture of key neurodevelopmental outcomes (intellectual disability, ASD and ADHD) across relevant age windows; no major limitations in long-term outcome ascertainment. Moderate–High: mother–child linkage available; follow-up available at least to 11 years; capture of most key outcome domains (≥75%), with only limited gaps (e.g. missing 15-year follow-up for some outcomes or absence of selected supporting outcomes such as stunting or speech delay). Moderate: mother–child linkage available; follow-up mainly restricted to 5–11 years and/or capture of approximately 50–75% of outcome domains, including partial coverage of key neurodevelopmental outcomes. Low–Moderate: mother–child linkage available but capture of fewer than 50% of outcome domains and/or important gaps in key neurodevelopmental outcomes (intellectual disability, ASD or ADHD), limiting assessment of long-term child health. Low: mother–child linkage unavailable, or long-term child outcomes not captured, precluding reliable ascertainment of long-term child outcomes.*

Main results on the study of data availability and accessibility outside the EHR data

The survey suggests that the main obstacle is generally not the absence of data, but rather the ability to access, link, standardise, and routinely operationalise existing data banks, by the data processors. The challenge is greatest for data sources requiring new linkage infrastructure (e.g. BIFAP, PHARMO), whereas databases operating within established registry ecosystems (e.g. VID, FLR, NLR) are already better positioned to address many of the identified information gaps.

Table S2. Summary of main findings on data availability and accessibility

Data source	Main hurdle identified	How it could be overcome
VID (Spain)	No major structural barriers identified. Most relevant registries are already linked through established infrastructure. Main issue is time to obtain data	Continue expanding routine use of existing linked registries and maintain deterministic linkage capabilities.
Val Padana LHU (Italy)	Some complementary information is available outside the routine data instance but is not systematically integrated.	Establish routine linkage for relevant registries and outcome sources. Requires reasons for doing this and investment
SIDIAP (Spain)	Additional information exists but father linkage and some external sources are not routinely incorporated.	Develop father–child linkage infrastructure and integrate complementary registries. Requires reasons for doing this and investment
EFEMERIS (France)	Pregnancy-focused structure limits access to paternal, prior history and long-term child information.	Expand linkage to external child health and outcome registries where feasible. Requires reasons for doing this and investment
Finnish linked registries (FLR)	Additional long-term outcome information may exist outside the current data instance. Main issue is time to obtain approvals and data	Extend linkage to complementary registries and data collections when required for specific research questions. Lag time to be improved
Norwegian linked registries (NLR)	Some potentially relevant data banks were not included in the current data instance. Main issue is money and time to obtain approvals and data	Expand use of existing national linkage capabilities to incorporate additional information. Lag time to be improved
PHARMO (Netherlands)	School records, insurance data, tertiary referral centre data and specialised registries have not previously been linked; feasibility and administrative requirements remain uncertain.	Establish and validate new linkage procedures, obtain governance approvals, and assess the feasibility of incorporating educational, insurance and specialised clinical data sources. Requires reasons for doing this and investment
BIFAP (Spain)	Secondary care information may be available but is not routinely accessible; some information is stored in free text, standardisation approaches and linkage mechanisms are not yet established. National registries are available but not routinely linked.	Develop linkage mechanisms with national registries, establish ETL and standardisation procedures, and explore NLP or other methods to extract information from unstructured clinical text. Requires reasons for doing this and investment
CPRD Gold (UK)	HES and LSOA linkage ongoing.	Secure routine access and approvals for linked datasets (e.g. secondary care data) and incorporate them into future research instances. Requires reasons for doing this and investment.

Discussion

This multi-database feasibility study across nine European data sources demonstrates the potential of routinely collected healthcare data to study ASM safety in pregnancy and early life. The large population size and availability of adequate lookback periods provide a strong basis for studying maternal exposure and common pregnancy outcomes. However, the observed variability across databases in exposure prevalence and data capture highlights important challenges for multi-database research. These differences likely reflect variation in healthcare systems, coding practices, and data collection methods, and may affect the comparability of results across settings. This underlines the importance of harmonised definitions and analytical approaches.

The feasibility of studying paternal exposure was more limited. Father-child linkage was only available in four data sources and varied in completeness. This remains an important gap, particularly given the increasing interest in the role of paternal exposures in reproductive and child health outcomes.

For neonatal and long-term child outcomes, registry-based infrastructures generally performed best, particularly the Norwegian and Finnish linked registries. In contrast, limitations related to follow-up duration and linkage completeness reduced suitability in several EHR-based databases. Neurodevelopmental outcomes were incompletely captured in some sources, and BIFAP was unable to support neonatal and child follow-up analyses because mother–child linkage had not yet been established. t

Although many data sources supported follow-up into childhood, differences in follow-up duration, timeliness of registration limited the assessment of long-term outcomes, including neurodevelopmental disorders, which were incompletely and inconsistently captured. This suggests that studies focusing on long-term child health require careful selection of data sources and may benefit from linkage to additional data systems (e.g., education databases, specialized registries, or child health records). Overall, fitness-for-purpose varied by research question, with no single data source capturing all required elements, highlighting the need for careful database selection and methodological harmonisation in future studies.

Conclusions

This study demonstrated that the included European real-world data provide a strong foundation for investigating maternal ASM exposure during pregnancy, supported by large populations, robust pregnancy ascertainment using own or ConcePTION algorithms, and generally well-established mother–child linkage.

However, the feasibility assessment also identified important limitations for studies of paternal ASM exposure and long-term child neurodevelopmental outcomes. Only four of the nine participating data sources supported father–pregnancy linkage, and high-quality linkage was largely limited to registry-based infrastructures with deterministic linkage. In addition, ascertainment of neurodevelopmental outcomes varied substantially across databases because of differences in follow-up duration, linkage capabilities, and outcome capture.

The findings indicate that studies relying solely on primary care or claims data may be affected by incomplete paternal linkage, incomplete ascertainment of long-term child outcomes, and variable availability of key confounders, potentially limiting the validity and interpretability of results. Requirements are linkage to primary care, secondary care records and family linkage/pregnancy registers over prolonged follow-up period.

The survey of data experts showed that many of these gaps could potentially be addressed through linkage to complementary registries and other data sources already available within national data ecosystems. However, the main challenge is not the availability of additional data, but the ability, reason and investment to access, link, standardise, and routinely integrate these sources into research-ready datasets. Future studies evaluating paternal ASM exposure and neurodevelopmental outcomes should therefore prioritise data sources with established family linkage infrastructures and long-term follow-up capabilities, while considering the potential added value of complementary registries and other linked data sources. Continued investment in deterministic linkage infrastructure, data integration and reducing lag times will be essential to strengthen the generation of regulatory-grade evidence in this field.