1. Title Page

Title	TARGET-EU: Risk of adverse birth and neurodevelopmental outcomes in children born alive to fathers exposed to valproate versus levetiracetam for generalised epilepsy
Research question & Objectives	What is the cumulative incidence of death, major congenital malformations, autism and ADHD in offspring born alive to fathers diagnosed with generalised epilepsy exposed to valproate compared to levetiracetam?
Protocol version	V1.0
Last update date	06 March 2026
Contributors	**Primary investigator contact information:** Carlos González Poses (c.gonzalezposes-2@umcutrecht.nl) Dr. Linda Nab (l.nab@umcutrecht.nl) **Contributor names:** Dr. Constanza Andaur Navarro Dr. Oisín Ryan Prof. Ian Douglas Dr. Daniala Weir Clara Rodríguez-Bernal Gabriel Sanfélix Gimeno Aníbal García Sempere Fran Lopis-Cardona
Study registration	**Site:** https://catalogues.ema.europa.eu/node/4440/administrative-details **Identifier:** EUPAS1000000539
Sponsor	**Organization:** EU PE&PV research network **Contact:** eupepv@uu.nl
Conflict of interest	None

Table of contents

2. Abstract

Background

Multiple studies have suggested that the intake of valproate for pregnant women during pregnancy has a negative effect on pregnancy outcomes (including congenital anomalies and neurodevelopmental disorders (Liu et al., 2023: Tomson et al., 2019: Bromley et al., 2023). The evidence regarding the effect of paternal valproate exposure is much less clear. A study conducted by IQVIA concluded there was an increased risk of neurodevelopmental disorders (IQVIA, 2021). A recent study, using partially the same data, concluded there were no increased risks of neurodevelopmental disorders or congenital malformations (Christiansen at al., 2024: 2025). The UK Medicines Agency has recommended further studies about the effect of valproate in the spermatogenic cycle, as well as long-term effects (Medicines & Healthcare products Regulatory Agency, 2023).

Objectives

The primary objective is to estimate the cumulative incidence of death, major congenital malformations, ADHD and autism in offspring born alive to fathers diagnosed with generalised epilepsy exposed to valproate compared to levetiracetam.

Methods

We will conduct an observational cohort study following the target trial emulation approach in combination with the estimands framework using linked electronic health records from the Valencian Health Integrated Database (García-Sempere et al, 2020), a comprehensive database of longitudinal electronic health records of the Valencia region in Spain. Eligible individuals are males of reproductive age (≥ 18 years) diagnosed with generalised epilepsy exposed to valproate or levetiracetam. In the primary analysis, the hypothetical strategy is used for treatment switch and the treatment policy strategy is used for treatment discontinuation; a principal stratum strategy is used where the principal stratum consists of those who father live births. A Poisson regression is used to estimate relative risks, weighted by the propensity score.

3. Amendments and updates

Version date	Version number	Section of protocol	Amendment or update	Reason
06 March 2026	1.0			

4. Milestones

Table 1. Milestones

Milestone	Date
Study protocol for RWD study	08 August 2025
Preliminary results RWD study	April 2026
Final Study report	10 June 2026

5. Rationale and background

What is known about the condition: Epilepsy is a neurological condition characterised by recurrent, unprovoked seizures, typically treated with anti-seizure medication, including valproate and levetiracetam.

What is known about the exposure of interest: Multiple studies suggest that the intake of valproate for pregnant women during pregnancy has a negative effect on pregnancy outcomes (including congenital anomalies and neurodevelopmental disorders, Liu et al., 2023: Tomson et al., 2019: Bromley et al.,2023). The evidence regarding the effect of paternal valproate exposure is much less clear. A study conducted by IQVIA concluded there was an increased risk of neurodevelopmental disorders (IQVIA, 2021). A recent study, using partially the same data, concluded there were no increased risks of neurodevelopmental disorders or congenital malformations (Christiansen at al., 2024: 2025). The UK Medicines Agency has recommended further study about the effect of valproate in the spermatogenic cycle, as well as long-term effects (Medicines & Healthcare products Regulatory Agency, 2023).

Gaps in knowledge: It remains uncertain whether paternal exposure to valproate contributes to adverse pregnancy outcomes or neurodevelopmental disorders in offspring

What is the expected contribution of this study? We address this question using cutting-edge causal inference methods, namely the target trial emulation and estimands framework. This work seeks to address some of the limitations of previous studies, including an imprecisely defined time zero and retrospective assessment of exposure and potential confounding by indication.

6. Research questions and objectives

The overall aim is to assess the risk of adverse birth and neurodevelopmental outcomes in children born alive to fathers exposed to valproate versus levetiracetam for generalised epilepsy.

6.1. Primary Estimand 1

Research question targeted by the primary estimand: What is the cumulative incidence ratio of death, major congenital malformations, autism and ADHD in the offspring of men with generalised epilepsy who would father live births irrespective of treatment with valproate or levetiracetam, when exposed to valproate before conception compared to levetiracetam before conception, regardless of conceiving within 3 months of treatment initiation and treatment discontinuation, and in the absence of treatment switch?

Table 1. Core Emulation Table: Primary estimand (Estimand 1)

Attribute	Target Trial	Target Trial Emulation	Comment
Population	Adult males diagnosed with generalised epilepsy who would conceive under either valproate or levetiracetam	Adult males diagnosed with generalised epilepsy who would father live births under either valproate or levetiracetam	No other population is identifiable in the data source to address the research question. For a discussion on why the population has been restricted to the principal stratum of fathers of live births, see the rows titled 'Endpoint' and 'Intercurrent events and strategies to handle them' below. There is potential for misclassification of generalised epilepsy, as it can be challenging to distinguish between generalised and focal epilepsy based on diagnostic codes, due to the limited

			specificity of code usage when patients present to clinicians.
Treatment Conditions	Intervention group: valproate as monotherapy Active comparator group: levetiracetam as monotherapy	Intervention group: valproate as monotherapy Active comparator group: levetiracetam as monotherapy	Treatment strategies are inferred from electronic health records based on dispensing information.
Endpoint	First occurrence of: adverse pregnancy outcomes, death of offspring after birth, and diagnosis of autism or ADHD in offspring by the age of 12,	First occurrence of: death of offspring after birth, diagnosis of major congenital malformations in live birth, and diagnosis of autism or ADHD in offspring, by the age of 8	Fathers can only be linked to live births in our data source. We therefore do not have information on all adverse pregnancy outcomes. In this study, we assume that the fathers linked to children in the data are their biological fathers. We make explicit that in the trial emulation, we focus on major congenital malformations because these are less subjective than minor malformations, which may be inconsistently diagnosed or reported in electronic health records. There is potential for outcome misclassification, as it is assumed that absence of records on these health outcomes reflects no occurrence of them.
Summary Measure	Cumulative incidence ratio	Cumulative incidence ratio	
Intercurrent events and strategies to handle them	**Same for both treatment strategies**	**Same for both treatment strategies**	The principal stratum of our study is defined by “fathering a live birth”,

	No conception: Principal stratum Conception within 3 months post-treatment initiation: treatment policy Treatment switch prior to conception: hypothetical strategy Treatment discontinuation prior to conception: treatment policy Still birth, spontaneous abortion or congenital malformation: composite strategy Death of offspring after birth: composite strategy	Non-live birth: principal stratum Conception within 3 months post-treatment initiation: treatment policy Treatment switch prior to conception: hypothetical strategy Treatment discontinuation prior to conception: treatment policy Major congenital malformation identified in live births: composite strategy Death of offspring after birth: composite strategy	different from “conceiving” in the target trial. This deviation is necessary because non-live births (miscarriages, stillbirths) as well as congenital malformations in non-live births are not linkable to the father in our data source. These events are addressed using a principal stratum strategy, as they all do not result in live births. Major congenital malformations in live births can be identified and will be handled with the composite strategy, in accordance with the target trial. Ascertainment of intercurrent events likely to be less complete in the emulation and is prone to uncertainty regarding timing of treatment discontinuation or switch.

Rationale for why selected strategies to handle intercurrent events are chosen

Non-live births: principal stratum.

Rationale: In the target trial, we focus on the principal stratum of men who would conceive (through their partners). For more information on the rationale for choosing this principal stratum, see the target trial protocol. In the trial emulation, we further restrict this to the principal stratum of live births, as non–live births cannot be linked to fathers in our data source. The stratum of men whose partners would conceive follows naturally from the fact that the endpoint of the target trial is defined only for males with partners who conceive. In contrast, the narrower focus on live births in the emulation is determined by data constraints.

In consequence, in the emulation, the principal stratum of interest is men who would father live births, irrespective of treatment with valproate or levetiracetam. This should not be confused with the observed subgroup of men who father live births under their actual treatment, as we cannot know whether they would have done so under the alternative treatment.

Conception within 3 months post-treatment initiation: treatment policy.
Rationale: In the target trial, patients would be instructed to remain on treatment for at least 3 months before attempting conception, ensuring that a full spermatogenesis cycle has been completed. However, partners might become pregnant before the recommended three-month period has elapsed. This intercurrent event is more likely to occur in real-world settings. Understanding the associated risks in these cases remains clinically relevant. In addition, the timing and extent to which valproate or levetiracetam may affect adverse birth and neurodevelopmental outcomes remain uncertain. Therefore, conceptions occurring within three months after treatment initiation are still considered informative regarding the treatment effect and accordingly this intercurrent event is handled with a treatment policy strategy.

Treatment switch prior to conception: hypothetical strategy.
Rationale: Because starting therapy with another treatment may confound the risk of the outcome (i.e., the occurrence of the outcome may be attributable to the initiation of the alternative treatment rather than the initial treatment), we will consider the treatment effect in a hypothetical world where patients would not initiate another therapy prior to conception. Treatment switch also includes switching to other treatments for epilepsy (e.g., lamotrigine), or combinations of treatments for epilepsy – which are not allowed in the target trial, but might occur anyway.

Treatment discontinuation prior to conception: treatment policy.
Rationale: Exposure to valproate or levetiracetam before conception might affect outcomes irrespective of discontinuation. The period before conception during which treatment exposure may influence adverse health outcomes in offspring is uncertain. Therefore, in the primary estimand, we consider any exposure to valproate or levetiracetam before conception relevant regardless of treatment discontinuation, hence the choice of the treatment policy strategy.

Major congenital malformations in live births: composite strategy.
Rationale: Because these malformations may be related to the use of valproate we consider them part of the outcome.
Death of offspring after birth: composite strategy.
Rationale: We consider death of offspring after birth as part of the outcome because it is an important adverse health outcome, particularly in neonates (birth up to 28 days of age).

6.2. Supplementary Estimand 2

Same as for the primary estimand (see Table 2), except that treatment discontinuation is categorised as more than 3 months or less than 3 months before conception, with the former handled with a hypothetical strategy and the latter handled with a treatment policy strategy. The estimand targets the effect of treatment during the spermatogenesis cycle.

Research question targeted by the estimand: What is the cumulative incidence ratio of death, major congenital malformations, autism and ADHD in the offspring of men with generalised epilepsy who would father live births irrespective of treatment with valproate or levetiracetam, when exposed to valproate before conception compared to levetiracetam before conception, regardless of conceiving within 3 months of treatment initiation and treatment discontinuation less than 3 months before conception, and in the absence of treatment discontinuation more than 3 months before conception and of treatment switch?

Rationale for why selected strategies to handle intercurrent events are chosen

Treatment discontinuation more than 3 months before conception: hypothetical strategy.
Rationale: One hypothesis is that valproate affects outcomes during the spermatogenesis period (~3 months). We aim to estimate the effect of valproate on the outcome in a hypothetical scenario where no treatment discontinuation occurs more than three months before conception.

Treatment discontinuation less than 3 months before conception: treatment policy.
Rationale: These individuals took valproate during the critical spermatogenesis window, so any potential effect on sperm has already occurred. Therefore, the fact that patients discontinue treatment shortly before conception is not relevant.

7. Research methods

7.1. Study design

Research design (e.g. cohort, case-control, etc.): We will use the *sequential trial emulation* approach (Hernán and Robins, 2024, p. 309-311: Hernán et al., 2008). This consists of creating multiple overlapping cohorts, one per each calendar month from January 2010 to March 2016. In each calendar month, a new cohort is created where eligible individuals are classified according to their observed treatment strategy (i.e., 'starting, continuing or switching to valproate' or 'starting, continuing or switching to levetiracetam'). Individuals are artificially censored if they cannot be linked to a live birth within 3 years and 9 months from time zero (see below for rationale). This artificial time window is chosen because treatment trajectories in individuals diagnosed with generalised epilepsy are often complex, and a live birth may not occur until up to several decades after time zero. Therefore, we explicitly emulate a pragmatic feature of the design of the target trial within each of the sequential (monthly) trials of this non-

interventional study, and apply a follow-up period of 3 years up to conception. The same individual can be part of multiple cohorts, and not necessarily in the same treatment arm in each cohort. Variance estimation will be adjusted via bootstrapping to reflect this design choice.

Rationale for study design choice: In the target trial, participants are randomised to one of the treatments upon trial entry. Depending on their treatment history, this could correspond to treatment continuation (i.e., individuals already on valproate remain on valproate, and analogously with levetiracetam), a treatment switch (i.e., individuals on levetiracetam are randomised to valproate, and vice versa), or treatment initiation (for treatment-naïve individuals).

In contrast, observational data lack randomisation, and multiple points in time after an individual becomes eligible for the study could, in principle, serve as a proxy for treatment assignment (continuing, switching or starting treatment). To justify the sequential trial emulation, we considered alternative designs and discuss why they were deemed less appropriate:

- A natural alternative would be to define time zero as treatment initiation. However, for generalised epilepsy, treatment is often initiated before puberty, well before individuals meet the study's eligibility criterion of being aged ≥18 years (see Section 7.3.2).

- Another possible approach would be to start follow-up when individuals turn 18 and continue until a live birth occurs. Yet, treatment trajectories in epilepsy are complex, and live births may occur decades later. By that time, many individuals will no longer be on their initial medication and may have switched treatments multiple times before fathering a live birth. This creates challenges to ascertain the exposure of interest and leave limited information for inference, particularly since treatment switch is handled under the hypothetical strategy.

- Starting follow-up at age 18 and restricting it to three years and nine months would not solve this problem, as most individuals would not father a live birth by age 21, making the design inefficient.

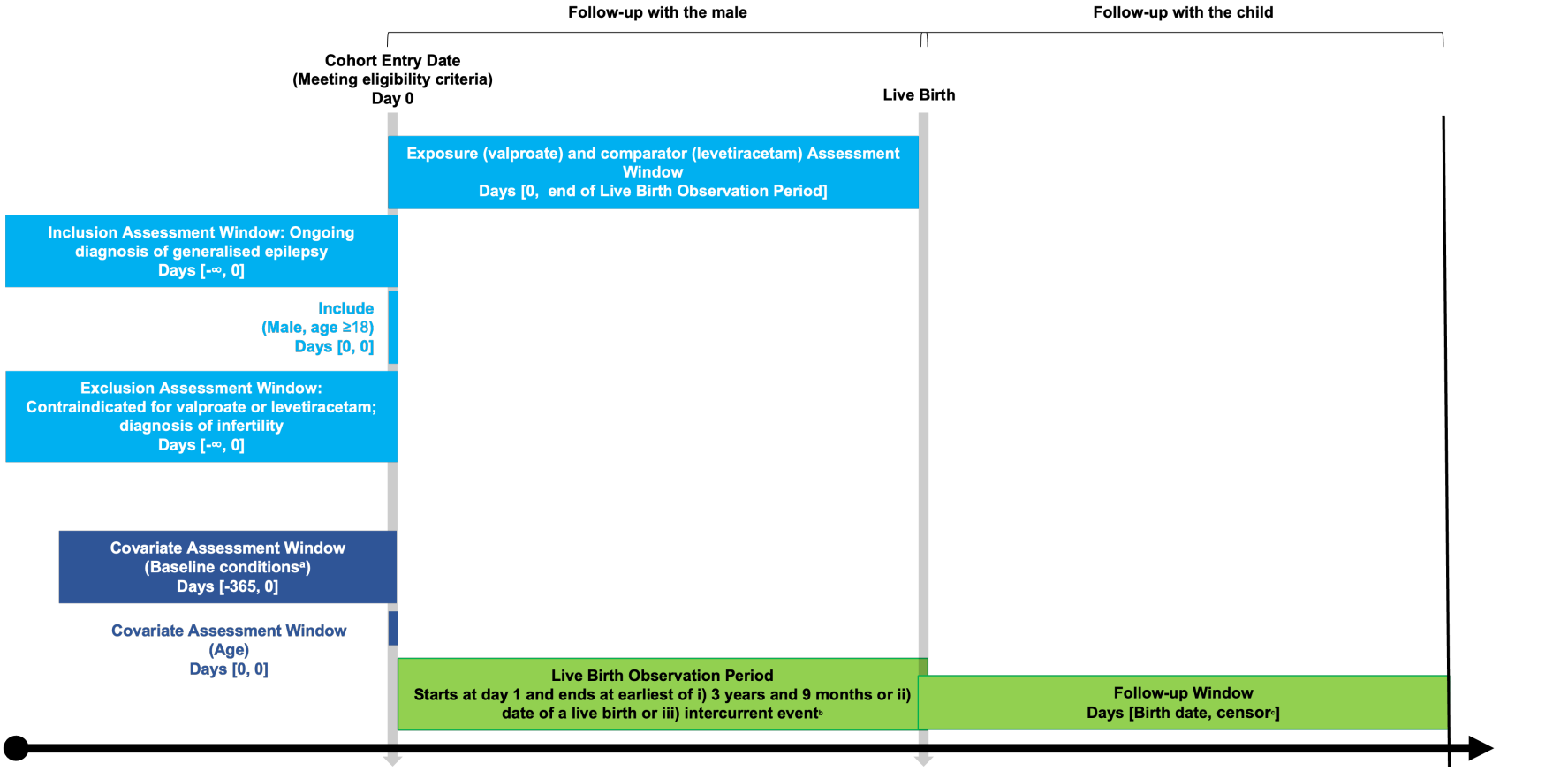
- Defining time zero retrospectively based on the timing of a live birth may introduce immortal time and selection bias.

To address these challenges, which prevent the emulation of a realistic target trial, we apply monthly sequential trial emulations, allowing us to efficiently capture relevant exposure and outcome information over time. In addition, by applying a follow-up period of 3 years up to conception (as in the target trial), this design ensures that a live birth can occur no more than 3 years and nine months after time zero.

The last cohort is defined 8 years and 9 months (March 2016) prior to the administrative end of the database (December 2024), ensuring that only live births occurring at least 5 years before the database cutoff are included. This approach is used to ensure a minimum of 5 years of follow-up for all included live births.

7.2. Study design diagram

The study design is illustrated in Figure 1.



a. Full list and code algorithms provided in protocol
b. Primary estimand: censor at treatment switch. Secondary estimand: censor at switch or discontinuation >3 months pre-conception.
c. Censored at earliest of death, major congenital malformation, autism, ADHD, end of the study period (8 years after birth), deregistration or end of administrative follow up (31/12/2024).

Figure 1. Cohort description and look-back periods

7.3. Setting

7.3.1. Definition of time 0 (and other primary time anchors) for entry to the study population

Time zero (index date) is defined as the date on which an individual meets the eligibility criteria. In the target trial, treatment assignment occurs randomly at time zero. In observational data, however, treatment assignment is not observed. Consequently, time zero is not uniquely defined and may correspond to any point during which an eligible individual from the study population is using one of the study treatments. To address this, we use a sequential design that allows multiple time zeros per individual (one per month). A detailed rationale for this approach, including why alternative definitions of time zero are not viable, is provided in Section 7.1.

Table 2. Operational Definition of Time 0 (index date) and other primary time anchors

Study population name(s)	Time Anchor Description (e.g. time 0)	Number of entries	Type of entry	Washout window	Care Setting[1]	Code Type[2]	Diagnosis position[3]	Incident with respect to…	Measurement characteristics/ validation	Source of algorithm
Exposure: valproate	Date of dispensing of valproate (time 0)	*Multiple entries (one individual can be part of multiple cohorts)*	The type of entry can be either incident or prevalent. See Section 7.1 for an explanation why treatment initiation is not always chosen as time 0 in our study.	No washout period	OP	ATC/MPID		N/A	N/A	N/A
Active comparator: levetiracetam	Date of dispensing of	*Multiple entries (one individual can be part*	The type of entry can be either incident or	No washout period	OP	ATC/MPID		N/A	N/A	N/A

	levetiracetam (time 0)	*of multiple cohorts)*	prevalent. See Section 7.1 for an explanation why treatment initiation is not always chosen as time 0 in our study.							

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

[2]See appendix for listing of clinical codes for each study parameter

[3] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter) (indicated with "primary") or can be in any position (indicated with "any")

7.3.2. Study inclusion criteria:

Individuals must be older than 18 years so that they are in reproductive age. Individuals must be males as the paternal exposure to valproate is the question of interest. Individuals must be diagnosed with generalised epilepsy to reduce confounding by indication across arms.

Table 3. Operational Definitions of Inclusion Criteria

Criterion	Details	Order of application	Assessment window	Care Settings[1]	Code Type[2]	Diagnosis position[3]	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
Age >=18 and <=50	Time 0 – date of birth.	Before	[0,0]	OT	N/A		Males	N/A	N/A
Males		Before	[0,0]	OT	N/A		Males	N/A	N/A
Ongoing generalised epilepsy	Ongoing disease would be identified as diagnosis of	Before	[-∞, 0]	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Males	N/A	N/A

	'generalised epilepsy' without an end date								

N/A

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

[2] See appendix for listing of clinical codes for each study parameter

[3] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter) (indicated with "primary") or can be in any position (indicated with "any")

7.3.3. Study exclusion criteria

Participants must not be contraindicated of valproate or levetiracetam, because that is the main exposure of interest and the comparator. Patients must not be diagnosed with a condition that prevents fertility, as the interest is in conditions of the offspring.

Table 4. Operational Definitions of Exclusion Criteria

Exclusion criterium: contraindication for valproate or leviteracetam

Rationale:

Criterion	Details	Order of application	Assessment window	Care Settings [1]	Code Type[2]	Diagnosis position[3]	Applied to study populations:	Measurement characteristics / validation	Source for algorithm
Contraindication for valproate/levetiracetam		Before	[-∞, 0]	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Males	N/A	N/A
Infertile		Before	[-∞, 0]	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Males	N/A	N/A

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

[2] See appendix for listing of clinical codes for each study parameter

[3] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter) (indicated with "primary") or can be in any position (indicated with "any")

7.4. Variables

7.4.1. Exposure(s) of interest

Valproate is the drug of interest. The exposure group consists of patients who use valproate at any dose, formulation or regimen.

Levetiracetam was chosen as an active comparator because, in the context of generalised epilepsy, the use of an active comparator is the most appropriate approach since it aligns with real-world clinical practice, under which epileptic patients are unlikely left untreated, and helps to address confounding by indication. The comparator group of interest consists of patients who use levetiracetam at any dose, formulation or regimen.

Algorithm to define duration of exposure effect:

Duration of exposure will be measured via dispensing information. For the construction of treatment episodes based on dispensing records, overlapping days between dispensing are handled by carrying forward any unused supply. Specifically, if a refill occurs before the end of the previous dispensed prescription's calculated days' supply, the overlapping days are added to the end of the new dispensed prescription's duration. We assume that each dispensed prescription has a lasting effect of up to 30 days. Therefore, a gap of up to 30 days between the end of a dispensed prescription's days' supply and the subsequent refill is allowed, without considering the patient as having discontinued treatment. Additionally, for the final dispensed prescription in a treatment episode, we extend exposure by 30 days beyond the calculated end date to account for any residual pharmacological effect or continued use.

Table 5. Operational Definitions of Exposure

Exposure group name(s)	Details	Washout window	Assessment Window	Care Setting[1]	Code Type[2]	Diagnosis position[3]	Applied to study populations:	Incident with respect to...	Measurement characteristics/ validation	Source of algorithm
Exposure	valproate	None	[0, live birth observation period end]	OP	ATC/PMID		Males	The type of entry can be either incident or prevalent. See Section 7.1 for an explanation why treatment initiation is	N/A	N/A

								not always chosen as time 0 in our study.		
Comparator	levetiracetam	None	[0, live birth observation period end]	OP	ATC/PMID		Males	The type of entry can be either incident or prevalent. See Section 7.1 for an explanation why treatment initiation is not always chosen as time 0 in our study.	N/A	N/A

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] See appendix for listing of clinical codes for each study parameter
[3] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter) (indicated with "primary") or can be in any position (indicated with "any")

7.4.2. Outcome(s) of interest

Outcomes in this study are assessed in live births, and not in their fathers. The composite outcome of interest is the first of: death of offspring after birth, diagnosis of major congenital malformation in live births, and diagnosis of autism or ADHD in offspring by the age of 8.

The main source for identifying major congenital malformations in VID (our data source) is the Congenital Anomalies Registry of the Valencia region, which is based on the EUROCAT rules, capturing predominantly major congenital malformations. Minor congenital malformations are only registered if they are associated with major congenital malformations.

Table 6. Operational Definitions of Outcome

Outcome name	Details	Primary outcome?	Type of outcome	Washout window	Care Settings[1]	Code Type[2]	Diagnosis Position[3]	Applied to study populations:	Measurement characteristics/ validation	Source of algorithm
Offspring death after live birth	Measured up to the age of 8	Yes	Binary	None	OT	N/A	Any	Offspring	N/A	N/A
Major congenital malformations in live births	Measured up to the age of 8	Yes	Binary	None	IP, OT	ICD-10CM	Any	Offspring	N/A	N/A
Diagnosis of autism	Measured up to the age of 8	Yes	Binary	None	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Offspring	N/A	N/A
Diagnosis of ADHD	Measured up to the age of 8	Yes	Binary	None	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Offspring	N/A	N/A

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] See appendix for listing of clinical codes for each study parameter
[3] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter) (indicated with "primary") or can be in any position (indicated with "any")

7.4.3. Follow up

Follow-up in this study is divided into two parts: males are followed from time zero until the end of the live birth observation period (for reference, see first study design diagram), and live births are followed from the time of birth until the end of follow up (for reference, see second study design diagram). Details on the end of follow-up for these two distinct periods are provided below.

Table 7A. Operational Definitions of Paternal Follow Up

Follow up start	Day 1	
Follow up end[1]	**Select all that apply**	**Specify**
Date of outcome	No	Outcomes not measured in fathers
Date of death	No	Children linked to fathers who died before completion of the child's follow-up are retained in the analysis as much as possible (in specific cases where the father dies shortly after the child's birth, linkage may not be possible). Paternal survival throughout the child's lifetime is not required.
End of observation in data	No	Children linked to fathers who are deregistered before completion of the child's follow-up are retained in the analysis as much as possible (in specific cases where the father deregisters shortly after the child's birth, linkage may not be possible). Continuous paternal registration throughout the child's lifetime is not required.
Day X following index date *(specify day)*	Yes	Follow-up will end 1370 days (3 years and 9 months) after index date for individuals not fathering a live birth (administrative censoring)
End of study period (specify date)	No	No fathers will have incomplete follow-up due to the study ending, as the study period ends in December 2024 and the last cohort is defined 8 years and 9 months before this cutoff (March 2016), while maximum follow-up is 3 years and 9 months.
End of exposure *(specify operational details, e.g. stockpiling algorithm, grace period)*	Yes	Individuals are artificially censored at treatment discontinuation >3 months pre-conception (estimand 2). Details are provided in text below Table 2.
Date of add to/switch from exposure *(specify algorithm)*	Yes	Individuals are artificially censored at treatment switch (estimand 1 and 2). Details are provided in text below Table 2.

Other date *(specify)*	Yes	Live birth

[1] Follow up ends at the first occurrence of any of the selected criteria that end follow up.

Table 88B. Operational Definitions of Child Follow Up

Follow up start	Live Birth	
Follow up end[1]	**Select all that apply**	**Specify**
Date of outcome	Yes	Major congenital malformation, diagnosis of autism or ADHD
Date of death	Yes	
End of observation in data	Yes	Censor child at date of deregistration, or database end (31/12/2024) (non-administrative censoring)
Day X following index date *(specify day)*	Yes	Follow-up will end 8 years after birth (administrative censoring)
End of study period (specify date)	Yes	31/12/2024
End of exposure *(specify operational details, e.g. stockpiling algorithm, grace period)*	No	Exposure is not measured in children
Date of add to/switch from exposure *(specify algorithm)*	No	Exposure not measured in children
Other date *(specify)*	No	

[1] Follow up ends at the first occurrence of any of the selected criteria that end follow up.

7.4.4. Covariates (confounding variables and effect modifiers, e.g. risk factors, comorbidities, comedications)

We include as confounders the following characteristics of the father: age, treatment start's calendar year and month, history of psychiatric disorders of males, seizure types, BMI, liver dysfunction, previous exposure to anti-seizure medication, ethnicity, use of known teratogenic medication, alcohol abuse and smoking. These will be included in a propensity score model.

- Paternal age is included because it may influence the choice of medication (e.g., younger males might be more likely prescribed the newer drug levetiracetam, and clinician prescribing patterns might also vary by patient age) and advanced paternal age is linked to de novo mutations in sperm, increasing risk of neurodevelopmental disorders in offspring.
- Calendar time is included because Levetiracetam has increased its use over time (patients closer to 2024 are more likely to receive levetiracetam), while standard of care as well as neurodevelopmental disorders might have a calendar trend.
- History of psychiatric disorders is included because levetiracetam has psychiatric side-effects so that it is less likely given to certain types of patients, mostly patients with generalized epilepsy, and valproate is also indicated for patients with bipolar disorder. Psychiatric conditions in the father are expected to be a risk factor for neurodevelopmental disorders in the offspring, given heritability.
- Seizure type is included because it determines treatment choice, and some types of epilepsy share genetic pathways with neurodevelopmental disorders.
- BMI is included because valproate can cause weights problems and is therefore less likely to be given to obese patients, while obesity can be a risk factor for neurodevelopmental disorders.
- Liver dysfunction is included because males with liver dysfunction are less likely to be prescribed valproate due to hepatoxicity risk and liver dysfunction independently affects reproductive outcomes (i.e., sperm quality and hormone balance which can impact gamete development).
- Previous exposure to anti-seizure medication is included because prior use of valproate increases the likelihood of being on valproate at day 0 (and similarly for levetiracetam), and prior exposure may also influence the risk of neurodevelopmental disorders and major congenital malformations.We note that previous exposure to anti-seizure medication is included as a covariate (categorised as valproate, levetiracetam, none and other, details in Table 9). Beyond serving as a confounder, this also serves the purpose of emulating stratified randomisation (treatment assignment is assumed to be random conditional on covariates). These will be included in the propensity score model and outcome model, so that effect modification by previous exposure can be assessed.
- Ethnicity, use of known teratogenic medication, alcohol abuse and smoking are considered weaker confounding factors. They are risk factors for the outcomes, but the extent to which they influence treatment choice is less clear. However, even small univariable associations between the exposure and these risk factors may become more impactful in multivariable settings and could introduce confounding. We have therefore decided to treat them as potential confounding factors.

Table 9 includes paternal confounders and maternal risk factors.

Maternal risk factors are, a priori, not included in the set for confounding adjustment. The reason is both theoretical and pragmatic. Theoretically, a confounder is a variable that determines treatment and outcome. A risk factor in the partner is unlikely to determine whether a patient that has

generalised epilepsy takes valproate or levetiracetam. Pragmatically, in the emulation we do not know if patients have a partner (and who their partner is) at time-zero. We expect risk factors on the mother level equally distributed among both treatment groups. The distribution of maternal risk factors is examined and, where appropriate, incorporated into the sensitivity analyses described in Section 7.6.2.

Table 9.9 Operational Definitions of Covariates

Characteristic	Details	Type of variable	Assessment window	Care Settings [1]	Code Type[2]	Diagnosis Position [3]	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
Paternal confounders									
Age		Continuous	[0,0]	OT	N/A	Any	Males	N/A	N/A
Calendar time		Categorical	[0,0]	N/A	N/A	Any	Males	N/A	N/A
History of psychiatric disorders		Categorical	[-365, 0]	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Males	N/A	N/A
Seizure type	It is currently unknown how exhaustively seizures are recorded in our data source	Categorical	[-365 0]	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Males	N/A	N/A
Obesity	Using coded BMI information and measured BMI values Obesity (BMI ≥30 kg/m2)	Categorical	[-365,0]	OP	ICD-9CM/ICD-10CM Local values	Any	Males	N/A	N/A
Liver dysfunction	Ongoing disease would be identified as diagnosis of 'liver dysfunction' without an end date	Categorical	[-365,0]	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Males	N/A	N/A
Previous exposure to anti-seizure medication	Categorised as: valproate, levetiracetam, none, other by counting the number of dispensed	Categorical	[-365,0]	IP, OP, ED	ATC/PMID	Any	Males	N/A	N/A

Characteristic	Details	Type of variable	Assessment window	Care Settings [1]	Code Type[2]	Diagnosis Position [3]	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
	prescriptions in the past year and assigning each patient to the medication with the highest number of dispensed prescriptions. If two or more medications had the same number of dispensed prescriptions, the category was determined based on the most recent dispensed prescription within the year.								
Ethnicity	Reflecting structural, environmental, and health-system factors	Categorical	[0, 0]	OT	N/A	Any	Males	N/A	N/A
Use of known teratogenic medication		Categorical	[-365, 0]	IP, OP, ED	ATC/PMID	Any	Males	N/A	N/A
Alcohol abuse		Categorical	[-365, 0]	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Males	N/A	N/A
Smoking	Using information on tobacco use	Categorical	[-365, 0]	OT	ICD-9CM/ICD-10CM	Any	Males	N/A	N/A
Maternal covariates (risk factors for neurodevelopmental disorders and major congenital malformations)									
Age		Continuous	[Live birth – 9 months, Live birth – 9 months]	OT	N/A	Any	Females	N/A	N/A
BMI	Using coded BMI information and measured BMI values	Categorical	[-365, Live birth – 9 months]	OP	ICD-9CM/ICD-10CM	Any	Females	N/A	N/A

Characteristic	Details	Type of variable	Assessment window	Care Settings [1]	Code Type[2]	Diagnosis Position [3]	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
Diabetes	Ongoing disease would be identified as diagnosis of 'diabetes' without an end date	Categorical	[-365, Live birth – 9 months]	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Females	N/A	N/A
Gestational diabetes		Categorical	[-9 months, Live birth]	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Females	N/A	N/A
Smoking	Using information on tobacco use	Categorical	[-9 months, Live birth]	OT	ICD-9CM/ICD-10CM	Any	Females	N/A	N/A
Use of known teratogenic medication		Categorical	[-9 months, Live birth period]	IP, OP, ED	ATC/PMID	Any	Females	N/A	N/A
Maternal covariates (risk factors for neurodevelopmental disorders)									
Mental disorders	Ongoing disease would be identified as diagnosis of 'mental disorder' without an end date	Categorical	[-365, Live birth – 9 months]	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Females	N/A	N/A
Concomitant medications associated with valproate-indicated psychiatric conditions or neuropsychiatric adverse effects		Categorical	[-365, Live birth]	IP, OP, ED	ATC/PMID	Any	Females	N/A	N/A
Maternal covariates (risk factors for major congenital malformations)									
Infections during pregnancy	TORCHES VLAP: Toxoplasmosis, rubella, cytomegalovirus, herpes simplex, enteroviruses	Categorical	[-9 months, Live birth]	IP, OP, ED	ICD-9CM/ICD-10CM	Any	Females	N/A	In progress

Characteristic	Details	Type of variable	Assessment window	Care Settings [1]	Code Type[2]	Diagnosis Position [3]	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
	and syphilis, varicella, Lyme disease, AIDS, and parvoviruses.								

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] See appendix for listing of clinical codes for each study parameter
[3] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter) (indicated with "primary") or can be in any position (indicated with "any")

7.5. Core Emulation Table - Design Summary

Table 1010. Core Emulation Table: Comparison of Target Trial and Proposed Target Trial Emulation Design Elements

	Target Trial	Target Trial Emulation	Comment
Inclusion criteria	1. Participants must be 18 years or older. 2. Participants must be males. 3. Participants must have a female partner with which they intend to conceive. 4. Participants must have a documented clinical history of generalised epilepsy. 5. Participants must be in disposition to give and understand informed consent.	1. Participants must be in reproductive age, i.e., between 18 and 50 years old. 2. Same. 3. Partner status and intention to conceive are not captured in electronic health records. 4. Same, ascertained from electronic health record coding. 5. Not applicable, as informed consent is not requested.	1. Reproductive age is used in the emulation because criterion 3 ("must have a female partner with whom they intend to conceive") cannot be emulated. 2. The individual must be recorded as male in their electronic health records; this may include individuals who identify as male, depending on how sex and gender are documented in the system. 3. It is not possible to (reliably) derive information on partner status and planning to conceive. This entails

			including much more patients (essentially almost any patient with generalized epilepsy) in the emulation. 4. Relies on accuracy of coded records. There is potential for misclassification of generalised epilepsy, as it can be challenging to distinguish between generalised and focal epilepsy based on diagnostic codes, due to the limited specificity of code usage when patients present to clinicians.
Exclusion criteria	1. Participants must not have any known contraindication for either valproate or levetiracetam use. 2. Participants partner must not be diagnosed with generalised epilepsy. 3. Participants must not have any medical condition that permanently prevents them from conception (i.e., infertility or any condition that makes a future conception impossible).	1. Same. 2. Partner status is not captured in electronic health records. 3. Same.	1. Relies on accuracy of coded record 2. It will not be possible to emulate exclusion criteria 2, as partner status is not known at time zero. We assume the likelihood of partners having generalised epilepsy is the same in both treatment groups. 3. We note that fertility ascertainment with electronic health records is likely of lower quality than in the trial. In trials, fertility can be assessed directly through specific questions or tests, whereas electronic records typically capture only certain causes of infertility or previously diagnosed cases.
Setting	Multicenter	Routine care data sources in the Valencia region in Spain (e.g., primary care and/or administrative databases) capturing	Real-world data captures care as delivered

		prescriptions, dispensing information, outcomes, and covariates.	
Time *(when follow up begins and ends):*	Follow-up start: 1. Begins at fulfilment of eligibility criteria and randomisation. Ends at first of: event of interest, children's reaching 12 years of age, deviation from treatment strategy (treatment switch or treatment discontinuation), loss-to-follow up, or administrative censoring.	Follow-up starts in fathers and outcomes are measured in their offspring. Follow-up start in fathers: Begins at fulfilment of eligibility criteria and use of study treatment (i.e. continuation of valproate/leviteracetam, switch to valproate or leviteracetam, or start of valproate/leviteracetam). Follow-up end in fathers: at the earliest of i) 3 years and 9 months; ii) date of live birth; iii) intercurrent event (treatment switch or treatment discontinuation > 3 months pre-conception (estimand 2)). Follow-up start in offspring: Birth. Follow-up end in offspring: at the earliest of: death, major congenital malformation, diagnosis of autism or ADHD, children's reaching 8 years of age, end of data availability (31/12/2024) or disenrollment from database	Follow-up of individuals will stop if they do not father a live birth 3 years and 9 months after time zero, mimicking the target trial (see section 7.1 for an explanation). Children linked to fathers can be followed between 2010–2024 in our data source. Therefore, only offspring born between 2010 and 2013 have complete follow-up. To increase the proportion of offspring with complete follow-up in our study, we will shorten the follow-up period to 8 years. Given that the mean age at diagnosis is 5 years for autism and 8 years for ADHD, this shortened follow-up is expected to capture the majority of diagnoses for both conditions (see Appendix for an overview of available evidence of timing of diagnosis).
Study treatment conditions	Intervention group: valproate as monotherapy Active comparator group: levetiracetam as monotherapy	Intervention group: valproate as monotherapy Active comparator group: levetiracetam as monotherapy Ascertained from prescribing records	Treatment strategies are inferred from the database based on dispensing information

Method of Assignment to Trial Intervention	Stratified randomisation, in four strata: current (at the time of randomization) valproate users, current levetiracetam users, current users of any other antiepilectic drug, and individual who are not current users of any drug	Randomisation cannot be directly emulated. We will adjust for baseline differences via a propensity score model, assuming treatment is administered as if random conditional on confounders. Strata will be included to the propensity score model, thus randomisation would be emulated within each strata (because the treatment is assumed to be assigned as if random conditional on covariates and strata). The strata will also be part of the outcome model.	Treatment allocation will be emulated based on the occurrence of study treatment dispensing. Because individuals receive the study treatments chronically (and may not be eligible for the study at treatment initiation), there is no single, uniquely defined time zero for many participants (See Section 7.1 for an explanation of why treatment initiation cannot be used as time zero and other considerations).
Outcome (including operational definition)	Composite of any of: diagnosis of autism or ADHD in offspring, adverse pregnancy outcomes, and death of offspring after birth	Composite of any of: diagnosis of major congenital malformation in live birth, death of offspring after birth, and diagnosis of autism or ADHD in offspring	Fathers can only be linked to live births: the outcome in the trial emulation is therefore partly different (it does not include adverse pregnancy outcomes that are not observed in live births). We focus on major congenital malformations in offspring because these outcomes are less subjective, more clinically distinct, and more reliably captured in electronic health records compared with minor malformations. Outcomes will be identified via codelists.
Intercurrent Events and strategies to handle them	**Same for both treatment strategies** 1. No conception: Principal stratum. 2. Conception within 3 months post-treatment initiation: treatment policy.	1. Non-live birth: principal stratum. 2. Conception within 3 months post-treatment initiation: treatment policy.	1. Handled via subsetting analyisis to individuals who father live births within 3 years and 9 months from treatment initiation. This assumes men who fathered a live birth under one treatment

	3. Treatment switch prior to conception: hypothetical strategy. 4. Treatment discontinuation prior to conception: treatment policy. 5. Still birth, spontaneous abortion or congenital malformation: composite strategy. 6. Death of offspring after birth: composite strategy.	3. Treatment switch prior to conception: hypothetical strategy. 4. Treatment discontinuation prior to conception: treatment policy. For estimand 2, intercurrent event 4 is subdivided into: 4A. Treatment discontinuation > 3 months prior to conception: hypothetical strategy. 4B. Treatment discontinuation < 3 months prior to conception: treatment policy. 5. Major congenital malformations in live births: composite strategy. 6. Death of offspring after birth: composite strategy.	would have also fathered a live birth under the other treatment. A sensitivity analysis will be performed to assess the sensitivity of the results to the assumption made in the primary analysis to identify the principal stratum. For the principal stratum we explicitly apply a time limit of 3 years and 9 months for a live birth from study entry, identical in both the target trial and the trial emulation. To implement the **treatment policy** strategy, explicit identification of IE2, IE4, IE4B is not strictly required; to implement the **hypothetical policy** strategy, explicit identification of IE3 and IE4A is required as individuals will be censored at IE3 and IE4A. Timing of conception in observational data will be identified by going back 9 months from live birth. 3. Dispensing information will be used to define treatment switch. 4/4A/4B. Dispensing information will be used to define treatment discontinuation. Of note, for the construction of treatment episodes based on dispensing records and defining treatment switch (IE3) and treatment discontinuation (IE4/4A/4B),

			overlapping days between dispensing are handled by carrying forward any unused supply. Specifically, if a refill occurs before the end of the previous dispensed prescription's calculated days' supply, the overlapping days are added to the end of the new dispensed prescription's duration. See *Section 7.4.1* for details on how exposure will be assessed. To implement the composite strategy, IE5 and IE6 can be retrieved using: 5. Diagnostic codes for major congenital malformations 6. Documented death in offspring
Loss to follow up	Individuals are considered lost to follow-up if, after multiple contact attempts, there is no response by the individual and health status or other relevant study information is hence unknown.	Outcomes in individuals whose offspring are censored or lost to follow-up before the age of 8 will not be observed and are therefore missing. Children linked to fathers who are deregistered or died before completion of the child's follow-up are retained in the analysis as much as possible (in specific cases where the father deregisters or dies shortly after the child's birth, linkage may not be possible).	In rare cases, the father or child may move outside the Valencian region but remain registered in the regional census for a period of time. Because this is so uncommon, we assume that individuals who are not deregistered remain under follow-up.

7.6. Data analysis

7.6.1. Analysis plan

Overview
For the primary estimand, the main estimand to support decision making, the cumulative incidence ratio will be estimated in the principal stratum of interest using a weighted Poisson regression model, using inverse probability of treament weights. To identify the principal stratum, we will assume that men who fathered a child in their assigned treatment would have also done so had they been assigned to the other treatment. Sensitivity analyses will be performed to assess the robustness of the results to this and other assumptions made in the primary analysis.

The same model and assumptions will be used in estimand 2.

Supplementary analyses including diagnostic and descriptive assessments to support the main analysis will be displayed. To contextualise data and results from the primary and sensitivity analyses. These will include number of men whose child experienced the outcome of interest per treatment group, crude incidence rates, weight and propensity score distributions, covariate balance before and after weighting, number of individuals with missing outcome, number of individuals who experience the identified intercurrent events as well as model diagnostics.

7.6.2. Primary Estimand (1) Analysis

i. Objective

To estimate the cumulative incidence of death, major congenital malformations, autism and ADHD in offspring born alive to fathers diagnosed with generalised epilepsy exposed to valproate compared to levetiracetam

ii. Exposure contrast

Valproate vs levetiracetam

iii. Outcome

Composite of death after birth, major congenital malformations in livebirths, and diagnosis of autism or ADHD in offspring by age of 8

iv. Analytic software

R

v. Handling of intercurrent events

- Non-live birth (IE1)
  Strategy used: Principal Stratum
  Handling: The analysis is restricted to the subpopulation of individuals observed to father live offspring under their allocated treatment. The assumptions under which this analysis would correspond to the principal stratum of interest are detailed below.
- Conception within 3 months post-treatment initiation (IE2)
  Strategy used: Treatment Policy
  Handling: Data after the occurence of the IE will be included in the analysis as per the treatment policy strategy.
- Treatment switch prior to conception (IE3)
  Strategy used: Hypothetical
  Handling: Non-administrative censored at the time of switching treatment. Outcomes are treated as missing. Missing data assumptions are described in the 'missing data methods' section of this table.
- Treatment discontinuation prior to conception (IE4)
  Strategy used: Treatment Policy
  Handling: Data after the occurrence of the IE will be included in the analysis as per the treatment policy strategy.
- Major congenital malformations (IE5)
  Strategy used: Composite endpoint with neurodevelopmental disorders.
  Handling: Considered as part of the outcome.
- Death of offspring after birth (IE6)
  Strategy used: Composite endpoint with neurodevelopmental disorders.
  Handling: Considered as part of the outcome.

Diagnostics for IE handling:
As a descriptive measure, we will tabulate for both treatment arms how often each IE occurs. For the IEs leading to censoring of follow-up (IE3), we report time to IE by treatment arm (median and IQR).

Handling of IE of non-live birth with the principal stratum strategy:
We consider the trial population to be divided into four strata depending on whether participants would father a live birth under both treatments (a); would father a live birth under valproate but not levetiracetam (b); would father a live birth under levetiracetam but not valproate (c); would not father

a live birth under either treatment (d) (for reference see Table 13 in Section 7.6.2). None of these strata are identifiable from the observed data, since they refer to cross-world counterfactual quantities.

The stratum of interest is stratum (a) as these are the individuals who would father a live birth under both treatments. In the primary analysis, we assume that the observed subgroup of men who father a live birth in their assigned treatment belong to stratum (a). In other words, we assume strata (b) and (c) are empty. This assumption is relaxed in a sensitivity analysis.

vi. Outcome Model

We estimate the relative risk parameterised as the ratio of the 8-year cumulative incidence between exposure groups.

The cumulative incidence ratio will be estimated using a weighted Poisson regression model, including terms for treatment and the stratification factor of previous medication use (categorised as valproate, levetiracetam, none or other, see Table 6 for details). Inverse probability of treatment weights (IPTW) will be used (see below).

The data from all sequential monthly trials are analysed together. A dataset will be constructed in which each individual contributes multiple records, one for each possible trial start month (referred to as "stacked" dataset). To account for the correlation induced, subject-specific random effects will be included in the Poisson model.

Variance estimation

Bootstrap will be used to estimate variance. 1000 bootstrap resamples will be performed. In each bootstrap sample, each unique individual is resampled with replacement, in order to conserve the original sample size. In each bootstrap sample the whole analytical pipeline consisting of 1) the creation of multiple cohorts defined by the calendar month index and its stacking, and 2) the estimation of different models based on the stacked dataset (propensity score model, and Poisson regression) will be repeated. The limits of the 95% confidence interval will be the 2.5% and 97.5% percentiles of the bootstrap distribution for the treatment effect estimate. Bootstrapping is performed so that the estimator's variance takes into account the correlation induced by the fact that single individuals are included multiple times in the analysis with overlapping intervals of follow-up.

Assumptions

We assume that the distribution of the outcome follows a Poisson distribution. Although the outcome is binary and a Binomial likelihood would fit, we do not use a Binomial model here because, when the sample size is large and the event is rare (i.e., the probability of the outcome is small), the Binomial distribution is well approximated by a Poisson distribution. We formally assess the sensitivity of the results to this assumption in a sensitivity analysis in which we model the outcome using a Binomial likelihood instead of the Poisson.

vii. Confounding adjustment

Propensity score weighting

The propensity score is defined as the probability of belonging to the valproate treatment arm, conditional on confounders detailed in section 7.4.4. The rationale for confounder selection is explained there. The idea of inverse probability of treatment weights is to reweight both treatment groups so that they have the same covariate distribution after weighting, therefore achieving exchangeability.

The propensity score is modelled using a logistic regression model with treatment as the outcome and paternal age categorised in groups (grouped into 10-year bins, based on the range of ages in the dataset), calendar time, history of psychiatric disorders, seizure types, BMI, liver dysfunction and previous exposure to anti-seizure medication (categorised in four strata) as covariates.

Stabilised weights will be calculated by dividing the probability of receiving the observed treatment conditional on previous exposure to anti-seizure medication by each individual's estimated propensity score (i.e., the conditional probability of receiving their observed treatment). To limit the influence of extreme values, weights will be truncated at the 1st and 99th percentiles.

Weight truncation reduces the influence of individuals with highly improbable treatment assignments but does not resolve propensity score non-overlap. Therefore, if regions of the propensity score distribution show insufficient overlap, we plan to restrict analyses to the overlapping region (trimming) or apply overlap weights.

Assumptions Underlying IPTW

- **No unmeasured confounding** (all relevant baseline confounders are included in the propensity score model).
- **Positivity** (each individual has a non-zero probability of receiving either treatment, given their covariates).
- **Correct model specification** (the propensity score model is correctly specified [functional form, covariate inclusion]).
- **Consistency** (each individual's potential outcome under the observed treatment equals their actual outcome).

Diagnostics

- **Covariate balance:** Check that baseline characteristics are balanced across treatment groups after weighting.
    - Evaluate standardized mean differences (SMDs): SMDs < 0.1 will be considered acceptable.
    - Maternal risk factors (Table 9) are also assessed for balance across treatment groups. A sensitivity analysis will be performed if any imbalance is identified.
- **Positivity check:** Describe overlap in propensity score distributions between treatment groups to support estimation (graphically).
    - Figures representing the distribution of IPTW for each treatment arm (to be conducted before and after truncation)

viii. Missing data methods

Missing Exposure Data

We assume that the absence of dispensed prescription records for valproate or levetiracetam reflect true treatment discontinuation, and not incomplete data capture or dispensed prescriptions issued outside the database.

Missing Outcome Data

Outcomes are observed in children. For males who belong to the principal stratum (i.e., they father a live birth) but experience IE3 (treatment switch, handled using the hypothetical strategy), the outcomes of their children will be treated as missing. In addition, children who do not have complete follow-up and have not yet experienced the outcome before censoring will also be treated as having missing outcomes. A substantial number of children will not have the full 8 years of follow-up, since children born between December 2016 and December 2019 cannot accumulate 8 years of observation; therefore, only children born before December 2016 have the potential for complete follow-up.

Given the considerable number of males with missing child outcomes, missingness will be addressed using multiple imputation by chained equations (MICE) under a Missing at Random (MAR) assumption (conditional on the covariates included in the outcome model and IPTW model). This assumption will be relaxed in a sensitivity analysis.

Before performing imputation, we will examine the extent and patterns of missingness to evaluate whether imputation is appropriate. Specifically, we will:

- Quantify the percentage of missing outcomes overall and by treatment group.
- To describe observation time, we will also report time to censoring for children who are deregistered or do not have complete follow-up and for whom outcomes are imputed (overall and by treatment group). Time to IE3 in fathers who are censored will be reported under the diagnostic for IE handling.

Missing Covariate Data

Missing data in covariates is handled using a combination of multiple imputation or variable omission strategies depending on the degree of missingness.

Note that, for variables that are assessed using diagnostic codes or records of prescription/dispensation of a medicine, we assume complete coverage, and absence of a record will be taken to reflect a true absence of the corresponding event/medicine prescription. As such, missingness in this context largely concerns lifestyle factors such as BMI, smoking and alcohol abuse. We will first quantify the extent and pattern of missingness to evaluate which missing data strategy is most appropriate

- Missing values for lifestyle factors will in general be imputed using multiple imputation with chained equations (MICE) under the Missing at Random (MAR) assumption.
- If a covariate has more than 40% missing data, the variable will be excluded. Thresholds of 40% have been cited because effect estimates begin

to be less reliable as the level of missingness increases beyond this threshold [Jakobsen et al. (2017), "When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts"]

Imputation Model

The MICE procedure will include all covariates used in the outcome and propensity score model, outcomes of interest will also be included.

Full Conditional Distributions

MICE will use variable-specific conditional models:

- Logistic regression for binary variables
- Multinomial logistic regression for categorical variables with >2 categories.
- Predictive mean matching for continuous variables

Number of Imputations and Diagnostics

We will generate at least 10 imputed datasets (to ensure stable estimates given the level of missingness) and pool results across imputations using Rubin's rules. Diagnostics will include:

- Checking whether imputed values are plausible and consistent with observed distributions.
- Evaluating convergence of the chained equations.
- Assessing stability and consistency of results across imputed datasets, by plotting the estimates across imputations and visually inspecting their variability

Effect Estimation Under Multiple Imputation

Since bootstrap is used to obtain confidence intervals, we need to consider how to combine multiple imputation with bootstrap. Following Schomaker & Heumann (2017), *"Bootstrap inference when using multiple imputation"*, we adopt the **MI BOOT** approach because it has been shown to be valid and is less computationally demanding than the **BOOT MI** alternative. Please find the formulas used to calculate standard errors and point estimates below (copied from Schomaker & Heumann (2017)).

- **Method 2, MI Boot:** Multiple imputation is utilized for the data set $D = \{D^{obs}, D^{mis}\}$. For each of the M imputed data sets D_m,B bootstrap samples are drawn, which yields $M \times B$ data sets $D^*_{m,b}$; $b = 1, \dots, B$; $m = 1, \dots, M$. The bootstrap samples are used to estimate the standard error of (each scalar component of) $\hat{\theta}_m$ in each imputed data set respectively, i.e., $\widehat{\mathrm{Var}}(\hat{\theta}_m) = (B-1)^{-1} \sum_b (\hat{\theta}_{m,b} - \hat{\bar{\theta}}_m)^2$ with $\hat{\bar{\theta}}_m = B^{-1} \sum_b \hat{\theta}_{m,b}$. This results in M point estimates (calculated from the imputed, but not yet bootstrapped data), and M standard errors (calculated from the respective bootstrap samples). One can thus calculate standard multiple imputation confidence intervals, possibly based on a t_R- distribution, as explained above.

These estimates will be combined across the imputed datasets using Rubin's rules (not based on a t_R distribution), which account for both within-imputation variance (the average estimation error within each imputed dataset) and between-imputation variance (the variability in estimates across imputations). The total variance therefore reflects uncertainty from both the imputation process and the estimation, producing valid confidence intervals.

The number of imputations (**M**) is set to 10. The number of bootstrap samples is 500 if sufficient computational power is available; otherwise, we will use 100 bootstrap samples.

ix. Subgroup analyses

Not applicable.

7.6.3. Supplemental Estimand (2) Analysis

The analysis for estimand 2 is kept the same with one difference: how the intercurrent event of treatment discontinuation is handled (IE4). In Estimand 1 this is dealt with via treatment policy. In Estimand 2, this is dealt with via hypothetical strategy 3 months before conception (which implies missing outcome data) (IE4A), and via treatment policy in the 3 months before conception (same as in estimand 1) (IE4B). Like with treatment switch handled with the hypothetical strategy in estimand 1, if these males fall within the principal stratum, the outcomes in their children are treated as missing. Like for missing outcomes in estimand 1, these will be imputed using MICE, following the same assumptions and models as described for estimand 1 above.

For the diagnostics, time to IE4A will also be reported for Estimand 2, summarised by treatment group using the median and IQR.

7.6.4. Sensitivity Analyses

(1a, 2a) Principal stratum effect: principal score weighting

In the main analysis of estimands (1) and (2) we assume the males observed to father a live birth are all in stratum (a) (see Table 13) under the assumption that they would father a live birth under both treatments.

In this sensitivity analysis we relax this assumption for both estimand (1) and (2), using a monotonicity assumption under which men who fathered a live birth under valproate would have also fathered a live birth under levetiracetam and a principal score weighting technique, which assumes potential outcomes are assumed independent of principal strata membership conditional on covariates .

Table 11. Overview of the four stratums in our population. A denotes treatment and S(A) denotes potential principal stratum membership under treatment A.

		Takes valproate (A = 0)	
		Fathers a live birth [S(0) = 1]	Does not father a livebirth [S(0) = 0]
Takes levetiracetam (A = 1)	Fathers a live birth [S(1) = 1]	Always-father (a)	Father-if-levetiracetam only (c)
	Does not father a live birth [S(1) = 0]	Father-if-valproate only (b)	Never-father (d)

Analysis methods:
Principal score weighting under a monotonicity assumption.

Non-live birth

- Strategy Used: Principal Stratum
- Handling:

We consider the population divided into the four strata shown in Table 11.The principal stratum of interest is stratum (a) as these are the individuals who would father a live birth under both treatments. We assume stratum (b) is empty under the monotonicity assumption, under which those who father a live birth under valproate would also have done so had they been exposed to levetiracetam. Consequently, in the group exposed to valproate, we directly observe membership of stratum (a): these fathers would also have fathered live births had they been exposed to levetiracetam.

Members of the group exposed to levetiracetam who father a live birth, may belong to strata (a) or (c). In order to estimate the principal stratum effect, we assume that, in the leviteracetam arm, membership to stratum (a) is independent of potential outcomes conditional on observed covariates. Under this assumption we can estimate the effect in the principal stratum using a principal score weighting technique [Jo & Stuart 2010 "On the Use of Propensity Scores in Principal Causal Effect Estimation"]. In this approach, we first fit a model predicting the probability of membership to the principal stratum in the valproate arm, where principal stratum membership is known under the monotonicity assumption. The predictors in this model are the baseline confounders (Table 9).

In the second step we use this estimated model to predict probabilities of stratum (a) membership (the principal score) amongst the group exposed to levetiracetam who fathered a live birth. These levetiracetam-exposed group members receive a weight of *PrincipalScore/(1- PrincipalScore)*.

In summary, the analysis will include valproate males who fathered a livebirth (with a weight of 1) and levetiracetaram fathers who fathered a livebirth (with a weight equal to PrincipalScore/(1-PrincipalScore). These weights will be combined with the propensity score weighs via multiplication. Covariate balance at baseline after this multiplication will also be assessed. Valproate males who did not father a livebirth will be used only to estimate principal score probabilities. Levetiracetam fathers who did not father a livebirth will not be included in the analysis; however, the proportion of males under levetiracetam (and valproate) who fathered a livebirth will be reported in both arms.

Assumptions:

- Principal ignorability: Principal stratum membership is assumed to be independent of potential outcomes, given the observed covariates.
- Transportability: The relationship between covariates and principal stratum membership is transportable across treatment arms, allowing the principal score model estimated under valproate to be applied under levetiracetam.
- Monotonicity: Those who father a live birth under valproate would also have done so had they been exposed to levetiracetam;
- Membership to the principal stratum (a) is independent of treatment assignment conditional on observed covariates;
- Correct specification of the principal score model: Principal stratum membership will be predicted using baseline confounders described in Table 9. A limitation of this approach is that the probability of fathering a live birth under either treatment also depends on characteristics of the partner. We do not include maternal factors in the model because partner information is missing for males who do not father a live birth.

(1b, 2b) Robustness of results to departures from the MAR assumption for missing outcomes: best/worst case scenario under the censoring not at random assumption.

In estimands (1) and (2) we assume outcomes are MAR given treatment group, previous use of antiepileptic medication and additional baseline covariates included in the MICE procedure.

In this sensitivity analysis we explore the robustness of the results to departures of this assumption. The goal is to assess the impact on the estimated treatment effect of selected missing non-at-random assumptions. The assumptions chosen represent the 4 extremes of a tipping point sensitivity analyses.

Analysis methods:
For males with a missing outcome in their children (either because of occurrence of IE or incomplete follow-up of children), repeat the analysis under four scenarios, assuming a) all missing outcomes are imputed as occurrence of the outcome; b) all missing outcomes are imputed as non-occurrence of the outcome, respectively for each treatment group. This equates to the following scenarios:

Table 12. Overview of best/worst case scenarios

Scenario	valproate	levetiracetam	Interpretation
1	Best case (lowest event rate)	Worst case (highest event rate)	Maximally favours valproate
2	Worst case (highest event rate)	Best case (lowest event rate)	Maximally favours levetiracetam
3	Best case	Best case	Optimistic for both groups
4	Worst case	Worst case	Pessimistic for both groups

Assumptions:

- Missing not at random assumptions are made whereby a extreme increase/decrease in the probability of the occurrence of the outcome is assumed in men whose live birth has missing outcome status, leading to all or none of them experiencing the outcome of interest.

(1c, 2c) Relaxing the MAR assumption made in the primary analysis for missing outcomes using additional information

In estimands (1) and (2) we assume outcomes are MAR conditional on baseline covariates included in the MICE procedure.

Analysis methods:

The imputation model is the same as that used in the primary analysis, with the addition of the following covariates:

- Time-varying paternal risk factors at the time of censoring: age and calendar time.

The imputation models are the same as those employed in the primary analysis.

Assumptions:

- Missing outcomes are MAR conditional on the baseline and time-varying covariates included in the imputation model.
- Imputation models are correctly specified.

(1d, 2d) Outcome model: log-binomial

In estimands (1) and (2) we use a Poisson regression to estimate relative risks under the assumption that a Poisson distribution approximates a Binomial distribution when sample size is large and the outcome is rare.

In this sensitivity analysis, we change this assumption by modelling the outcome using a Binomial likelihood instead of a Poisson likelihood.

Analysis methods:

The analysis will follow the primary analysis plan, with the exception that a log-binomial regression is used to model the outcome. Log-binomial regression can experience convergence issues in certain circumstances. If such problems occur, the corresponding results will not be reported.

Assumptions:

The outcome model is correctly specified.

(1e, 2e) Balancing maternal risk factors

In estimands (1) and (2) we assume maternal risk factors (Table 9) are balanced across weighted treatment arms.

In this sensitivity analysis, we relax this assumption by including imbalanced maternal risk factors in the IPTW model. Imbalance is defined as a standardised mean difference of 0.1 or more.

Analysis methods:
The analysis will follow the primary analysis plan, with the exception that any imbalanced maternal risk factors will be added to the IPTW model.

Assumptions:
The IPTW model is correctly specified.

Table 13. Sensitivity analyses – rationale, strengths and limitations

	What is being varied? How?	Why? (What do you expect to learn?)	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
(1a, 2a) Principal stratum **(principal risk score)**	We relax the assumption that stratums (b) and (c) (for reference, see Table 13) are empty by instead assuming only stratum (b) is empty. Levetiracetam-exposed group members receive a weight of PrincipalScore/(1- PrincipalScore)	Males who father a live birth under levetiracetam might not do so under valproate.	We relax the assumption that males who father a live birth would have done so under both treatments.	We can estimate the effect in the principal stratum using a principal score weighting technique by assuming membership of stratum (a) is independent of the potential outcomes in the levetiracetam arm conditional on observed covariates. This assumption might not hold because we do not include maternal covariates in the model for estimating probabilities of stratum membership, or generally because we might not observe all relevant covariates that would render principal stratum independent of potential outcomes in the levetiracetam arm. One limitation of using principal score weighting as proposed by Jo and Stuart (2011) is that their

				approach was developed in a different context, and it is unclear whether it can be directly applied in our setting. In their paper, all individuals in the control arm (in our case, all individuals in the levetiracetam arm, regardless of whether they father a live birth) are included in the analysis. In our application, however, we must deviate from this approach. By definition, outcomes (which are assessed in children) cannot be observed for individuals in the levetiracetam arm who do not father a live birth. These individuals therefore receive a weight of zero in our implementation of the method. It remains unclear whether this deviation introduces methodological concerns.
(1b,2b) MAR assumption **(best/worst case scenario under MNAR)**	The assumption that outcomes are MAR conditional on covariates included in the outcome model and the MICE procedure is varied.	To assess the robustness of the treatment effect estimate to violations of the MAR assumption made in the primary analysis for missing outcomes. If results are stable across scenarios, confidence increases that findings are not driven by bias (due to missing outcomes).	Explores alternative assumptions about missingness in outcomes. Sets bounds on the extent of maximum possible bias due missing not at random outcomes.	The best/worst case scenarios considered are the most extreme assumptions of a tipping point sensitivity analysis. Values in between the two extremes are not explored, which limits the ability to explore the variability of the treatment effect under a wide spectrum of MNAR assumptions.
(1c, 2c) MAR assumption	The assumption that outcomes are MAR conditional on covariates	To assess the robustness of the treatment effect estimate to a different MAR	Additional time-varying variables are used in the imputation model.	Variance of estimates might be increased.

(MAR using additional information)	included in the outcome model and the MICE procedure is varied.	assumption for missing outcomes. If results are similar to those in the primary analysis, confidence increases that findings are not driven by bias (due to missing outcomes).		Additional covariates might add noise instead of improving imputation.
(1d, 2d) Outcome model **(log binomial)**	The outcome model to estimate relative risks	To evaluate whether the treatment effect estimate is sensitive to modelling assumptions about the distribution of the outcome.	Assumes outcome follows a Binomial distribution	Log-binomial regression can experience convergence issues in certain circumstances; if such problems occur, the results will not be reported.
(1e, 2e) Imbalanced maternal risk factors	The assumption that maternal risk factors are balanced across weighted treatment arms is relaxed.	Maternal risk factors (Table 9) may be unevenly distributed across weighted treatment arms due to chance. These maternal risk factors are established risk factors for the outcomes and could therefore confound the relation under study if imbalance is detected.	Assumes exchangeability conditional on the covariates present in the primary analysis plus the additional maternal factors.	Variance of estimates might be increased. Improving balance for these factors might imbalance other factors. Model misspecification of the IPTW model.

7.6.5. Other Supplemental Analyses

No additional supplemental analyses will be conducted beyond those indicated in Section 7.6.1.

7.6.6. Core Emulation Table – Estimation Summary

Table 1114. Core Emulation Table: Estimation Summary

	Target Trial	Target Trial Emulation	Comment
Analysis Method	Poisson Regression	Weighted Poisson Regression. Weights are calculated using the Inverse Probability of Treatment, computed based on logistic regression.	IPTW used to emulate randomization in observational data Trimming of observations based on PS distribution represents a departure from the original target trial but is considered best practice when using propensity score methods in emulation. By removing patients in regions of non-overlap, the analysis is restricted to a population where treatment assignment is more comparable across groups. As a result, the estimated effect no longer applies to the entire eligible population but to this more comparable subset. If some patients have an extremely low probability of receiving one of the treatments, valid causal contrasts cannot be identified for them. Without trimming, effect estimates in these regions rely on unsupported extrapolation, making the results unstable and potentially biased.
Missing Data Assumptions and Methods to Handle	Missing data occurs at random given treatment and stratum of previous antiepileptic medication.	Individual who experience intercurrent events handled using the hypothetical strategy, that is with non-administrative censoring, will have missing outcomes. We assume MAR conditional on treatment assignment, stratum of prior	Losses to follow-up are likely to be underestimated: unless and individual actually disenrolls from the healthcare provider, this will not be observed. However, sometimes individuals change healthcare providers without informing it.

		treatment and confounders (see Table 9). Missing data in covariates is handled using a combination of multiple imputation or variable omission strategies depending on the degree of missingness.	Sensitivity analyses under alternative missing data assumptions will be performed.
Statistical Model Assumptions	Model assumes that the outcome follows a Poisson distribution conditional on the covariates. It further assumes: - Missingness occurs at random given the variables in the model. - Correct model specification (i.e., correct functional form and inclusion of relevant covariates) - Exchangeability between treatment groups due to randomisation	The assumptions are the same with the difference that: - Exchangeability (even though not strictly a modelling assumption) is assumed conditional on covariates used in propensity score models (assuming propensity score models are correctly specified). - The use of a MICE model to impute the outcome allows us to assume a weaker form of non-informative censoring.	Exchangeability (even though not strictly a modelling assumption) is achieved through randomisation in the trial and through adjustment (weighting) in the emulation. Missing outcomes are imputed using a MICE model in the trial emulation.
Sensitivity Analyses	Sensitivity analysis for principal stratum analyses.	Several sensitivity analyses will be conducted to assess the robustness of the primary analysis: a. **Principal stratum:** In the emulation, a sensitivity analysis for the assumption made in the primary analysis to identify the principal stratum is	The principal stratum has changed from males whose partners would go on to conceive under both treatments (target trial) to males who father live births under both treatments (trial emulation). Sensitivity analyses are described in more detail in section 7.6.2.

		implemented using a principal score analysis based on the probability of principal stratum membership. b. **Robustness of results to departures of the MAR assumption for outcomes:** To test the assumption that missing outcomes are Missing at Random (MAR), we will conduct analyses under the MNAR assumption using best-/worst-case scenarios from a tipping-point analyses. c. **Relaxing the MAR assumption using additional information:** We will perform a further sensitivity analysis that relaxes the MAR assumption by incorporating time-varying covariates into the imputation framework. d. **Varying the outcome model:** To assess sensitivity to the assumed outcome model, we will re-estimate the effect using a log-binomial regression model in place of the primary Poisson model. e. **Balancing maternal**	

		risk factors If maternal risk factors are imbalanced across weighted treatment arms, they will be added to the IPTW models in a sensitivity analysis.	

7.7. Data sources

7.7.1. Data sources

Rationale for selection and feasibility: The Valencia Integrated Database was selected because it provides most relevant information to study this question with observational data, including: prescription and dispensing information, diagnostic codes, adverse pregnacy outcomes, father-child linkage, and most relevant covariates.

Strengths of data source(s): The database provides coverage of 98% of the population of a largely inhabited region, 5 million people. A key strength is that database is one of the few in Europe that provide father-linkage information. Other strengths include the detection of adverse pregnancy outcomes.

Limitations with potential impact in the study results: Father-child linkage is deterministic, but its reliability has not been studied systematically. It is based on an algorithm combining address of father and child, coincidence of last names of father and child. Another limitation is that only health encounters with the public health system are recorded – so events recorded in the private sector may go unrecorded. A last limitation is that father-child linkage is only available for around 60% of births, lowering precision of the study. Although children linked to their fathers can be followed from 2010 to 2025 (allowing for a maximum of 15 years of follow-up) the analysis restricts fathers' follow-up to the period from 2010 to 2017. This ensures that all children included have a minimum of 5 years of follow-up.

Data source provenance/curation: VID data is managed by FISABIO in collaboration with the Valencian Publich Health System. More information on the data provencance can be found here: https://catalogues.ema.europa.eu/node/1077/administrative-details . We refer to Step 1 of the feasibility assessment, copied in Appendix 1.

Table 1215. Metadata about data sources and software

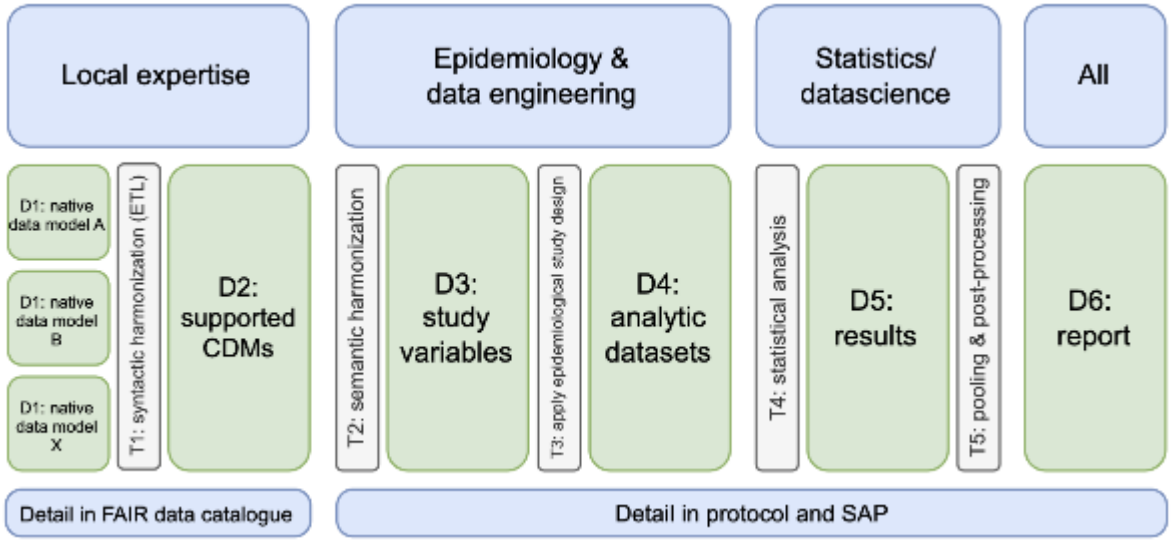
	Data 1
Data Source(s):	Valencia Integrated Database
Study Period:	2010-20245
Eligible Cohort Entry Period:	2010-20245
Data Version (or date of last update):	2025
Data sampling/extraction criteria:	To be discussed
Type(s) of data:	Primary care EHR, prescription and dispensing data.
Data linkage:	Yes, deterministic
Conversion to CDM*:	Yes, OMOP and Conception Data Model
Software for data management:	<Text>

*CDM = Common Data Model

7.8. Data Management

The study will be conducted in a distributed manner using the UMCU, ARS Toscana and VAC4EU tools, procedures, and pipeline. Figure 2 specifies the data sets (D) and transformation processes (T), programming follows this pipeline, with involvement of different types of experts.

Figure 2 Data Management from the data transformation perspective



D1: Original data can be in any native format

The RWD-RWE pipeline used by VAC4EU starts with data banks that are controlled by the Data Expert and Access Partner (DEAP) and can be in any format. Data always stays local and never leaves the secure environments of the DEAPs. The ETL (extract, transform, load, see below for more details under 'T1') design is shared in a searchable FAIR VAC4EU catalogue. The VAC4EU FAIR Molgenis data catalogue is a meta-data management tool designed to contain searchable meta-data describing organisations that can provide access to specific data sources.

T1: Syntactic harmonisation (ETL)

T1: Syntactic harmonisation is conducted through an extraction, transformation, and loading (ETL) process of native data into the ConcePTION common data model (CDM) (see section 'D2: Common data model'). To harmonise the structure of the data sets stored and maintained by each data partner, a shared syntactic foundation is used. The ETL process has various structured steps as described by Thurin et al (2021):

- DEAPs are asked to share the data dictionaries of their data banks (selected tables and variable names/structure)
- Metadata (descriptive data about the data sources and databanks) & data dictionaries, are uploaded in FAIR data catalogue (Molgenis).

D2: Common data model

For this project, the CDM (D2) is the ConcePTION common data model. The CDM version that is used is v2.2, which is available as an open-source CDM. In this CDM, data are represented in a common structure, but the values of the data remain in their original language (e.g. codes will have either ICD9/10/ICPC/SNOMED or MEDCODEID values).

T2: Semantic harmonisation

During the T2 step, many data transformations occur related to the completion of missing features in the data. Based on the relevant diagnostic medical codes and keywords, as well as other relevant concepts (e.g., medications), one or more phenotype algorithms are constructed (typically one sensitive, or broad, algorithm and one specific, or narrow, algorithm) to operationalise the identification and measurement of each event. In this step we conduct time anchoring (observation periods, look back periods), clean the data such as the dose of vaccines, sort on record level, aggregate across multiple records, and combine concepts for implantation of algorithms, and rule-based creation of study variables.

In this phase of the creation of study variables, semantic mapping is conducted. This semantic mapping across different vocabularies is conducted as part of the R-study script using different functionalities. To reconcile differences between different terminologies and native data availability, machine-readable code lists are used that comprise the terminologies that are used in the network (e.g. ICD-9, ICD10, SNOMED, ICPC and DEAP specific adaptations). This is combined with the BRIDGE metadata file that defines risk windows, look-back periods, and algorithms for each study variable (Cid Royo et al.).

D3: Study variables

D3 datasets are interim data sets with information on study variables for each study participant, the unit may be a person, a medicine, or episode of time. The design of these datasets is described in codebooks. Examples of D3 datasets are the outputs of the ConcePTION pregnancy algorithm (ttps://github.com/IMI-ConcePTION/ConceptionTools/wiki#conception-pregnancy-algorithm), and outputs of functions that define smoking. Multiple functions/packages exist within the VAC4EU, for different study variables.

T3: Application of epidemiological design

In the T3 step epidemiological designs are applied such as sampling, matching (on specific variables and/or propensity scores), and selection based on inclusion and exclusion criteria using the study variables in the D3 datasets. The designs will be implemented for the various study objectives using R-scripts, and these may use the existing functions (R-cran) or functions that have been developed in the VAC4EU community (e.g. matching).

D4: Analytical data set

D4 is an analytical dataset, and multiple D4 data sets may be produced based on the objectives of the study. The format is described initially in a code book for communication between programmers and statisticians.

T4: Statistical analysis

This step in the data transformation pipeline will produce statistical estimates such as descriptives (counts, percentages), distributions (mean, percentiles), rates (prevalence, incidence), regression coefficients, or other relevant estimates. This will be conducted using R.

D5: Results

D5 is the set of estimates, tables or aggregate data that is transferred from the DEAPs to the Digital Research Environment (DRE). The aggregated results produced by these scripts at the DEAP's site will be uploaded to the UMCU DRE for post-processing, pooling and visualisation (Figure 1). The DRE is a cloud-based, globally available research environment where data are stored and organised securely and where researchers can collaborate. The DRE is made available through UMCU. The DRE applies double authentication where researchers can collaborate using data that are stored and organised securely [ref]. UMCU is responsible for data processing and data security.

All researchers who need access to the DRE will be granted access to study-specific secure workspaces by UMCU. Access to the workspaces will be possible only after double authentication using an identification code and password together with the user's mobile phone for authentication.

Uploading files will be possible for all researchers with access to the workspace within the DRE. Downloading of files will be possible only after requesting and receiving permission from a workspace member with an "owner" role, who will be a UMCU team member.

T5: Post-processing/pooling

In this step, the result from different DEAPs is pooled and converted into tables and figures for reporting.

7.9. Quality Control

All key study documents such as the hypothetical trial protocol, target trial emulation protocol and study reports will undergo senior scientific and editorial review.

7.9.1. Data quality

For all data sources and for each data instance we will conduct *INSIGHT* level 1-2 quality checks, detailed statistical analysis plans for the indicators are available on the public repositories:

- https://github.com/UMC-Utrecht-RWE/INSIGHT-Level1 Hoxhaj, V. (2023). UMC-Utrecht-RWE/INSIGHT-Level1: https://doi.org/10.5281/zenodo.10035167
- https://github.com/UMC-Utrecht-RWE/INSIGHT-Level2 Hoxhaj, V., & van den Bor, R. (2023). UMC-Utrecht-RWE/INSIGHT-Level2: https://doi.org/10.5281/zenodo.10035169

Briefly, level 1 verifies Data Completeness and level 2 Data Consistency.

Level 1 – Data Completeness

The purpose of the level 1 check is to verify the completeness of the ETL process and the data in the variables. Examples of tests are:

- Presence of variables in each of the CDM tables in D2
- Checks for misspellings and letter case in variable names in each of the CDM tables
- Verification of vocabularies
- Check date formats
- Check conventions of values
- Missing data analysis
- Frequency tables for categorical variables

Level 2 – Data Consistency

Real data is not random but follows certain logical constraints that reflect rules governing real-world situations. Examples of indicators generated by level 2 checks are:

- Event dates before date of birth
- Event dates after date of death
- Event dates out of observation periods
- Subjects having an observation but not present in the PERSONS table
- Observations associated with a visit id and occurred before/after the visit start/end date
- Subjects younger than 12 years old reported as parents
- Age at the observation period older than 115 y old Data

7.9.2. Code Quality

These coding practices define how the TARGET programming team collaborates to write clean, reliable, and reproducible code for the VAC4EU Real-World Evidence (RWE) Analytical Pipeline. They aim to ensure clarity, consistency, and maintainability across all case studies within the project.

Coding conventions

To ensure clarity, consistency, and maintainability across the project, the following conventions will be applied to all codebases within the project:

- Consistent style: Code follows a consistent and readable style (see the tidyverse style guide for R).
- Meaningful names: Use clear, descriptive names for variables, functions, and files to convey their purpose.
- Modular code: Break down code into small, reusable functions where possible.
- No hardcoded paths: Use configuration files or relative paths to ensure portability.

Following these conventions makes the code easier to understand, test, and reuse across case studies and teams.

Documenting Code

Code documentation is used to promote good coding practices and ensure our work is understandable, maintainable, and reproducible. To achieve this, we will:

- Use descriptive comments that explain the purpose and rationale behind code sections, focusing on why something is done, not just what.
- Clearly document function inputs, outputs, and side effects, using standardized formats (e.g., roxygen2 in R) where appropriate and supported.
- Write meaningful variable and function names to make the code as self-explanatory as possible.

Version Control

We use Git and GitHub to manage version control. These tools support good coding practices by enabling collaboration, tracking changes, accessing a project's history, and ensuring code quality through review and documentation.

A dedicated GitHub organisation has been created for the project (https://github.com/target-roc19). Each case study is managed in its own repository within this organisation. Repositories are structured consistently across case studies, to reinforce modularity. Access to repositories is controlled through teams.

During development, all repositories remain private to ensure confidentiality. Once the project is finalised, relevant repositories will be made public and assigned a digital object identifier (DOI) via Zenodo to support transparency, reproducibility, and reuse by the wider research community.

To maintain code quality and clarity, we follow the git and GitHub guidelines below.

- Always use pull requests (PRs): never push directly to the main branch.
- Open an issue before creating a new branch. Ideally, one PR resolves one issue to keep changes focused and reviewable.
- Every PR must be reviewed by at least one other person before merging.
- The PR author merges the PR after it has been reviewed and approved.
- Write clear, descriptive commit messages.
- Write informative PR descriptions, including:
    - A concise title
    - Links to related issues
    - A summary of the changes

Continuous Integration

Continuous Integration (CI) is set up to automatically check code quality and run tests whenever changes are pushed to the repository or submitted through a pull request (PR). The CI workflow ensures that the package adheres to predefined style guidelines and that all automated tests pass before changes are merged.

Coding Template

Every case study follows the general coding template used across all code in the TARGET project. The folder structure is organised as follows:

```
case-study-template
|____data
| |____D2_cdm
| |____D3_study_variables
| |____D4_analytic_datasets
| |____D5_results
| |____D6_report
|____docs
|____logs
|____run
|____tests
|____transformations
| |____T2_semantic_harmonization
```

```
| |____T3_study_design
| |____T4_statistical_analysis
| |____T5_processing_results
|____CHANGELOG.md
|____LICENSE
|____README.md
```

Project Data Structure and Storage

The data folder follows the Real-World Evidence pipeline structure. Data conforming to the common data model is stored in the D2_cdm folder.

Results from transformations T2, T3, T4, and T5 are saved in the respective folders:

- D3_study_variables
- D4_analytic_datasets
- D5_results
- D6_report

Each dataset is associated with a codebook, explained in more detail below.

All data remain securely stored on the Data Expert and Access Partners (DEAPs) servers and are never transferred externally. For testing purposes, dummy datasets are created. These fall into two categories:

- Unit test data: Small, predefined input and output pairs used to test individual transformation steps. These are stored in the tests folder, not in data, and can support automated testing.
- Pipeline test data: Larger, more complex dummy datasets used to test whether the full pipeline runs as expected. These may be included in the repository only if they remain below GitHub's 100 MiB file size limit and will otherwise be shared via SharePoint.

Logging System

When the pipeline is executed, log files are saved in the logs folder. These logs are especially helpful when running the code in the DEAPs environment, as they help trace and diagnose potential errors. We recommend using the logger R package to handle logging throughout the pipeline. A sample logging setup can be found in the logger.R script located at the root of the project directory.

Executing the Analytical Pipeline

The run folder contains scripts used to execute each transformation step in the pipeline.

- A central script, run_pipeline.R, orchestrates the full pipeline from start to finish.
- Subscripts (e.g., run_T2.R or similar) are available to run individual transformation steps separately.

Typically, the run_pipeline.R script is the main entry point used by a DEAP to execute the full pipeline. Before running it in the DEAP environment, the pipeline may need to be adapted to local settings. This can be done using a configuration file that defines variables required to tailor the pipeline to a specific DEAP. Please note that configuration files should not include sensitive information.

Such a file might include variables like:

- The name of the DEAP
- The path to the local data instance
- The path to any required external resources

Testing and Quality Assurance

The tests folder contains scripts to test the analytical pipeline. Tests will be used to ensure code behaves as expected and remains stable over time. By systematically checking inputs, outputs, and edge cases, tests help catch errors early and make future changes safer. We use the testthat R package to structure and run unit tests.

Continuous integration (CI) is used to automate testing. With CI, tests are automatically run each time code is pushed to the repository (e.g., via GitHub Actions). This helps identify issues immediately, ensures that new changes do not break existing functionality, and supports better collaboration by enforcing consistent code quality across contributors.

Modular Data Transformation Workflow

The transformations folder follows the Real-World Evidence pipeline structure. It contains the source code for all transformation steps, which is typically written in R. Each subfolder corresponds to a specific step in the pipeline (e.g., T2_semantic_harmonization, T3_study_design, T4_statistical_analysis, T5_processing_results) and includes the relevant scripts and helper functions for that step.

During the T2 step, a database is usually created (e.g., using DuckDB). This database can be queried using SQL, and it is recommended that all SQL queries be saved as clearly named, standalone SQL script files to ensure readability and reusability.

The purpose of the transformations folder is to structure and modularise the processing logic, making it easier to maintain, test, and reuse across different case studies. By organising code by transformation step, teams can work in parallel, increasing efficiency.

Changelog

A changelog will be kept for all notable changes in the project. Changelogs help track the evolution of the project over time, making it easier for collaborators to understand what has changed between versions. We follow the structure and best practices outlined in Keep a Changelog.

Codebooks

Before developing code, codebooks are created to describe each dataset (D) within the pipeline. A codebook is a comprehensive document that outlines the structure, contents, and metadata of a dataset. It serves as a detailed reference guide for anyone working with the data and plays a crucial role in guiding the development of the analytical pipeline by clearly defining both the inputs and expected outputs.

All codebooks are summarized in a central index file, which provides a high-level overview of the pipeline's structure. For each codebook, the index file includes:

- A brief description of its purpose,
- A list of the scripts used to generate the corresponding dataset,
- A description of the input datasets and input parameters required.

The datasets D2, D3, D4, and D5 are typically subdivided into multiple smaller transformation steps, each detailed within their respective codebooks. These smaller transformation steps ensure that each part of the pipeline is clearly scoped and well-documented.

In addition to supporting development, codebooks help ensure quality control by making transformation logic transparent and verifiable, and they enhance reproducibility by documenting exactly how data is structured and used throughout the analytical pipeline.

Deployment

The analytical pipeline is delivered to DEAPs as a GitHub release, tagged with a version number. Versioning follows the format: vYYYYMMDD.XX, where the date indicates the release date and XX denotes the sub-version or revision number.

Any deployment issues can be reported via the GitHub repository using the issues feature, where the programming team responsible for the R code will collaborate with the local DEAP to resolve them as needed.

Reproducibility

It is recommended to locally use the renv R package to maintain the R version and version of packages for reproducibility purposes.

At this time, however, using renv reliably across different systems and environments remains challenging. For this reason, we currently recommend its use only in local development setups.

We are actively monitoring developments in the R ecosystem related to cross-platform reproducibility. As soon as a more stable and portable solution becomes available, we will revisit this guidance and promote broader adoption.

Licensing

The code will be made available under an open source license.

README Guidelines

Each case study repository includes a README that covers the following points:

- Project Overview: brief summary of the study goals and key research questions.
- Background: context and rationale for the study.
- Repository Structure: Outline of main folders and their contents.
- Data Overview: Description of data sources, formats, and data privacy considerations.
- How to Run: Instructions for running the pipeline and key scripts, plus where outputs are saved.
- Testing: How to run tests to verify code functionality.
- Contributing: Guidelines for code contributions and issue tracking.
- License: Information about the code license.
- Contact: Who to reach out to for help or questions.

7.10. Study size and feasibility

The sample size calculations from our target trial are as follows: 'We use a two-sided alpha of 0.05, a power of .80, assuming a cumulative incidence of the outcome of 0.2 in the Levetiracetam arm, a minimum clinically relevant effect size of 0.05 percentage points and a 1:1 ratio allocation in treatment vs control. For that design, a sample size of 2188 individuals is necessary. Since we expect 15% of people to not conceive, a sample size of 2574 is necessary.'

An estimate of the precision of our estimate is particularly challenging in our analysis because of the repeated cohort analysi. In practice, the estimator's variance is not derived analytically but relies on a bootstrapped approximation. The precision of the estimate will depend on the number of individuals in each treatment arm (which is not currently know), the estimated propensity scores, the proportion of individuals who go on to father a

livebirth within 3 years of index date, the number of individuals lost to follow-up, who engage in treatment switching or who engage in treatment discontinuation, as well as the number of events observed.

Here, we report estimates of the confidence interval under simplifying assumptions and scenarios. First, we keep the cumulative incidence in each arm constant and equal to the one define in the target trial's sample size calculation. Second, we vary the sample size per group. Third, we compute the standard error of a relative risk as sqrt(1/a - 1/(a+b) + 1/b - 1/(b+d)) where a is the number of events in the valproate arm, a + b is the number of individuals in the valproate arm, b is the number of events in the levetiracetam arm, and b+d is the number of individuals in the levetiracetam arm. We report the standard error and the confidence interval in the different scenarios in Table 16.

Table 136. Estimates' precision in several scenarios

Scenario	Cumulative incidence Valproate	Cumulative Incidence Levetiracetam	Sample size Valproate	Sample size Levetiracetam	Confidence interval for Risk Ratio	Standard error of Log RR
Scenario 1: required sample size in trial is achieved	0.25	0.20	1144	1144	(1.07-1.46)	0.08
Scenario 2: 50% larger sample size	0.25	0.20	1716	1716	(1.10-1.42)	0.06
Scenario 3: 100% large sample size	0.25	0.20	2288	2288	(1.12-1.39)	0.06
Scenario 4: 50% lower sample size	0.25	0.20	572	572	(1.01-1.56)	0.11
Scenario 5: 75% lower sample size	0.25	0.20	1144	286	(0.97-1.62)	0.13

Levetiracetam, 100% Valproate						

We do not have information on counts of Valproate or Levetiracetam users in the Valencia Integrated Database. However, given that it comprises 5 million individuals, and assuming a prevalence of generalized epilepsy of 0.25%, we should expect around 6.250 males diagnosed with epilepsy at any given time point. In combination with the sequantial trials approach, this should make achieving our target sample size achievable.

8. Limitation of the methods

A limitation of this study is the choice of time zero, which could not be emulated in the absence of randomisation. In the trial emulation, time zero is defined as the date on which an individual meets the eligibility criteria. In the target trial, treatment assignment occurs randomly at time zero. In observational data, however, treatment assignment is not observed. Men may become eligible at multiple ages and can initiate, discontinue, or switch the treatments of interest several times between first eligibility and fathering a live birth, over a period spanning many years. Consequently, time zero is not uniquely defined and may correspond to any point during which an eligible individual from the study population is using one of the study treatments. This inability to define time-zero leads us to use a design known in the causal inference literature as *sequential trial design*, where every eligibility period is seen as a time-zero and multiple cohorts are generated, overlapping in terms of follow-up and individuals. A detailed rationale for this approach, including why alternative definitions of time zero are not viable, is provided in Section 7.1.

A related limitation in this study is that the sequential trial does not align with the target trial cohort experimental design. This is because in experimental cohort parallel-arm designs individuals are only enrolled once and in only one arm. In this sense, the target trial parallel-arm cohort experimental design could not be emulated. However, considered in isolation, each of the subcohorts are formed in a way that aligns with the target trial design, and other design elements were tried to be followed as closely as possible (including treatment groups, intercurrent events and strategies to handle them. The statistical analyses are broadly similar, but in the observational study we add subject-specific random effects to the Poisson model to account for correlations due to repeated and overlapping follow-up within individuals across both arms.

Another deviation of this study from the target trial is the principal stratum of interest. Some intercurrent events (adverse pregnancy outcomes resulting in a non-live birth) could not be measured because adverse pregnancy outcomes cannot be linked to the father in the database. Therefore, the intercurrent event in the observational study was ‘non-live births’ instead of ‘no conception’. The primary analysis assumes that males who father a live birth would have done so under either treatment, which is a very strong assumption that may not be valid given the potential impact of valproate on fertility. This assumption is relaxed in a sensitivity analysis using a principal score approach. A limitation of the principal score model is that it comes with an additional assumption, that potential outcomes are independent of principal strata given covariates, which is difficult to verify or satisfy with our data, particularly since maternal predictors of live birth could not be included. Additionally, the performance of the method when only part of

the control group is weighted, as in our case, has not been thoroughly examined (see also the discussion in the limitation section of Table 13). Lastly, the method, particularly in combination with propensity score weights, might result in extreme weights that make the estimates unstable.An additional deviation of each sequential trial from the target trial includes absence of randomisation (mitigated to an extent using inverse probability of treatment weights).

Other limitations including possible bias common to all observational studies, such as confounding bias, measurement error, missing data and informative censoring. Specifically with respect to this case study, we do not have information on all father-child pregnancies. We may lack detailed information on some potential confounders, such as seizure type, as the granularity and availability of this information in the data source are currently unknown. We additionally lack information on the specific reasons for treatment switch which might be an important confounder. Some of these limitations have been addressed, when possible, using sensitivity analyses.

9. Protection of human subjects

This is a non-interventional study using secondary data collection and does not pose any risks for individuals. Each data source research partner will apply for an independent ethics committee review according to local regulations. Data protection and privacy regulations will be observed in collecting, forwarding, processing, and storing data from study participants.

This is a non-interventional study using secondary data collection and does not pose any risks for individuals. Each data source research partner will apply for an independent ethics committee review according to local regulations. Data protection and privacy regulations will be observed in collecting, forwarding, processing, and storing data from study participants. Patient information This study involves data that exists in an anonymized structured format and contains no patient personal
information. All parties will comply with all applicable laws, including laws regarding the implementation of organisational and technical measures to ensure the protection of patient personal data. Such measures will include omitting patient names or other directly identifiable data in any reports, publications, or other disclosures, except where required by applicable laws. Patient personal data will be stored at DAPs in encrypted electronic form and will be password protected to ensure that only authorised study staff have access. DAPs will implement appropriate technical and organisational measures to ensure that personal data can be recovered in the event of a disaster. In the event of a potential personal data breach, DAPs shall be responsible for determining whether a personal data breach has in fact occurred and, if so, providing breach notifications as required by law.

Patient consent

As this study does not involve data subject to privacy laws according to applicable legal requirements, obtaining informed consent from individuals is not required.

10. Reporting of adverse events

For studies in which the research team uses only data from automated healthcare databases, according to the International Society for Pharmacoepidemiology Guidelines for GPP. “Aggregate analysis of database studies can identify an unexpected increase in risk associated with a particular exposure. Such studies may be reportable as study reports, but typically do not require reporting of individual cases. Moreover, access to automated databases does not confer a special obligation to assess and/or report any individual events contained in the databases. Formal studies conducted using these databases should adhere to these guidelines.” For non-interventional study designs that are based on secondary use of data, such as studies based on
medical chart reviews or electronic health records, systematic reviews, or meta-analyses, reporting of adverse events/adverse drug reactions is not required. Reports of adverse events/adverse drug reactions should only be summarized in the study report, where applicable. According to the EMA Guideline on GVP, Module VI – Management and Reporting of Adverse Reactions to Medicinal Products, “All adverse events/reactions collected as part of [non-interventional postauthorization studies with a design based on secondary use of data], the submission of suspected adverse reactions in the form of [individual case safety reports] is not required. All adverse events/reactions collected for the study should be recorded and summarized in the interim safety analysis and in the final study report.” Module VIII – Post-Authorization Safety Studies echoes this approach. Legislation in the EU further states that for certain study designs such as retrospective cohort studies, particularly those involving electronic health records, it may not be feasible to make a causality assessment at the individual case level.

11. References

Bromley, Rebecca L., Philip Bullen, Ellen Campbell, et al. “Neurodevelopment of Babies Born to Mothers with Epilepsy: A Prospective Observational Cohort Study.” *Epilepsia* 64, no. 9 (2023): 2454–71. https://doi.org/10.1111/epi.17709.

Christensen, Jakob, Betina B. Trabjerg, and Julie Werenberg Dreier. “Risk of Neurodevelopmental Disorders and Paternal Use of Valproate During Spermatogenesis.” *JAMA Network Open* 8, no. 5 (2025): e2512139. https://doi.org/10.1001/jamanetworkopen.2025.12139.

Christensen, Jakob, Betina B. Trabjerg, and Julie Werenberg Dreier. “Valproate Use During Spermatogenesis and Risk to Offspring.” *JAMA Network Open* 7, no. 6 (2024): e2414709. https://doi.org/10.1001/jamanetworkopen.2024.14709.

Hernán, Miguel A., Alvaro Alonso, Roger Logan, et al. “Observational Studies Analyzed Like Randomized Experiments: An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease.” *Epidemiology*19, no. 6 (2008): 766–79. https://doi.org/10.1097/EDE.0b013e3181875e61.

Hernán, Miguel A., Issa J. Dahabreh, Barbra A. Dickerman, and Sonja A. Swanson. “The Target Trial Framework for Causal Inference From Observational Data: Why and When Is It Helpful?” *Annals of Internal Medicine*, February 18, 2025, ANNALS-24-01871. https://doi.org/10.7326/ANNALS-24-01871.

Hernan, Miguel A., and James M. Robins. *Causal Inference: What If*. First edition. Taylor and Francis, 2024.

IQVIA, and Valproate EU consortium. *A Post-Authorisation Safety Study (PASS) to Evaluate the Paternal Exposure to Valproate and the Risk of Neurodevelopmental Disorders Including Autism Spectrum Disorder as Well as Congenital Abnormalities in Offspring - a Population-Based Retrospective Study*. PASS 1.1. 2023.

Liu, Chen, Samuel Pace, Rebecca Bromley, and Ruth Dobson. "Exposure to Medication for Neurological Disease in Pregnancy – Time to Consider the Long-Term Implications?" *eClinicalMedicine* 63 (September 2023): 102157. https://doi.org/10.1016/j.eclinm.2023.102157.

Medicines and Healthcare products Regulatory Agency. *Valproate: Review of Safety Data and Expert Advice on Management of Risks*. Public Assesment Report. 2023.

Tomson, Torbjörn, Dina Battino, Rebecca Bromley, et al. "Management of Epilepsy in Pregnancy: A Report from the International League Against Epilepsy Task Force on Women and Pregnancy." *Epileptic Disorders* 21, no. 6 (2019): 497–517. https://doi.org/10.1684/epd.2019.1105.

Royo AC, Elbers Jhj R, Weibel D, et al. Real-World Evidence BRIDGE: A Tool to Connect Protocol With Code Programming. Pharmacoepidemiol Drug Saf. 2024;33(12):e70062. doi:10.1002/pds.70062

12. Appendices

Appendix 1. Shell Tables

Table 1. Standardized mean differences in covariates between Valproate and Levetiracetam users at baseline before and after reweighting, in subset of individuals who do conceive within 3 years

Figure 1. Overlap in baseline propensity scores

Table 2. Proportion of individuals who father a livebirth within 3 years out of total individuals (marginal and stratified by treatment groups and covariates).

Table 3. Proportion of individuals who engage in treatment switch and/or treatment discontinuation within subset of individuals who conceive within 3 years.

Figure 2. Standardized mean differences in covariate values by time-point, stratified by arm, among individuals who do not engage in treatment switch or treatment discontinuation, before and after reweighting

Table 4. Estimate of relative risk from Poisson models and log-binomial model, with and without weighting due to loss to follow-up, with bootstrapped adjusted variance

Figure 3. Cumulative incidence curve by treatment arm, per year from index date, computed based on IP weighted pooled logistic model standardized for calendar year.

Appendix 2. Review of age at diagnosis for autism and ADHD

Added as a separate document.