








STUDY PROTOCOL

**REVISED** A nationwide longitudinal investigation on the role of prenatal exposure to infectious diseases on the onset of chronic conditions in children and adolescents in Brazil**[version 2; peer review: 2 approved]**

Enny S. Paixao <sup>1,2</sup>, Thiago Cerqueira-Silva <sup>1</sup>, Pilar T.V. Florentino<sup>2</sup>,  
 Orlagh Carroll <sup>1</sup>, Nuria Sanchez Clemente<sup>3</sup>, Deborah A. Lawlor <sup>4</sup>,  
 Rita de Cássia Ribeiro Silva<sup>2</sup>, Laura Cunha Rodrigues<sup>1</sup>, Liam Smeeth <sup>1</sup>,  
 Mauricio L. Barreto<sup>2</sup>

<sup>1</sup>Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK<sup>2</sup>Centre for Data and Knowledge Integration for Health, Salvador, Brazil<sup>3</sup>Centre for Neonatal and Paediatric Infection, St George's University, Saint George's, Saint George, UK<sup>4</sup>MRC Integrative Epidemiology Unit at the University of Bristol, United Kingdom; Population Health Science, Bristol Medical School, University of Bristol, Bristol, England, UK

**v2** First published: 19 Jun 2024, 9:320  
<https://doi.org/10.12688/wellcomeopenres.22430.1>  
 Latest published: 17 Oct 2024, 9:320  
<https://doi.org/10.12688/wellcomeopenres.22430.2>





**Abstract****Background**


In utero exposure to infections might set the stage for a chain of events leading to a wide spectrum of long-term health outcomes observed in children and adolescents. This proposal aims to investigate whether syphilis, zika, dengue and chikungunya during pregnancy can increase the risk of the offspring developing a non-infectious chronic condition during childhood and adolescence.

**Objectives**

1) Estimate the risk of non-infectious chronic conditions associated to syphilis, zika, dengue and chikungunya during pregnancy and when appropriate, explore if the risk varies by timing during pregnancy when the infection is acquired (first, second or third trimester) and severity (such as severe or mild dengue); 2) Investigate whether in uterus exposure to maternal infection affects the growth pattern of children and adolescents; 3) Examine the extent to which the

**Open Peer Review****Approval Status**  

	1	2
<b>version 2</b> (revision) 17 Oct 2024		 <a href="#">view</a>
<b>version 1</b> 19 Jun 2024	 <a href="#">view</a>	  <a href="#">view</a>

1. **Aja Louise Murray**, The University of Edinburgh, Edinburgh, UK
2. **Anna Suleri** , Erasmus University Medical Center, Rotterdam, The Netherlands

Any reports and responses or comments on the article can be found at the end of the article.

relationship between maternal infection and non-infectious chronic outcomes are mediated by intrauterine growth restriction and preterm birth.

### Methods

We will compare health outcomes and growth trajectories of children and adolescents born to mothers with and without specific infections during pregnancy using conventional multivariable regression in the whole study population, in a within sibship design, using the subgroup of offspring with at least one sibling who is not exposed to the infection, and negative control outcome. Then we will decompose the direct and mediated effects (by preterm birth and small for gestational age) of maternal infection on chronic disorders.

### Results and Conclusions

The results from this study will advance our understanding of the relationship between infections during pregnancy and chronic disorders, with widespread implications enabling targeting of critical points along the path from in utero exposure to outcomes to avoid or mitigate illness and disability over the life course.

### Plain language summary

It is not clear whether when moms get an infection during pregnancy, it might affect their child's long-term health. This means that we do not know whether we could prevent some chronic conditions by screening for and treating infections during pregnancy or by monitoring and treating the child after their birth. In this study, we are going to look at whether infections affecting low—and middle-income countries, like syphilis, neglected tropical diseases such as dengue, zika, and chikungunya, might make children more likely to develop cancer, neurological problems, autoimmune diseases, or obesity when they grow up. We will also check if these risks are connected to things like being born too soon or small. We will use a unique data set from Brazil that follows people from birth (the CIDACS birth cohort). We'll compare the health and growth of children from infancy to teenage years whose mothers had certain infections during pregnancy with those whose mothers did not. The findings from this study will help us learn more about how infections during pregnancy impact the health of children. This knowledge could help us find ways to prevent or lessen the effect of illness and disability throughout a person's life, starting from before they are even born.

### Keywords

maternal infection, syphilis, zika, dengue and chikungunya, in uterine exposure to infection, non-communicable conditions, children, adolescents, life course

**Corresponding author:** Enny S. Paixao ([enny.cruz@lshtm.ac.uk](mailto:enny.cruz@lshtm.ac.uk))

**Author roles:** **Paixao ES:** Conceptualization, Funding Acquisition, Methodology, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Cerqueira-Silva T:** Methodology, Writing – Review & Editing; **Florentino PTV:** Visualization, Writing – Review & Editing; **Carroll O:** Writing – Review & Editing; **Sanchez Clemente N:** Writing – Review & Editing; **Lawlor DA:** Methodology, Writing – Review & Editing; **Ribeiro Silva RdC:** Writing – Review & Editing; **Rodrigues LC:** Methodology, Writing – Review & Editing; **Smeeth L:** Writing – Review & Editing; **Barreto ML:** Funding Acquisition, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by Wellcome [225925].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2024 Paixao ES *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Paixao ES, Cerqueira-Silva T, Florentino PTV *et al.* **A nationwide longitudinal investigation on the role of prenatal exposure to infectious diseases on the onset of chronic conditions in children and adolescents in Brazil [version 2; peer review: 2 approved]** Wellcome Open Research 2024, 9:320 <https://doi.org/10.12688/wellcomeopenres.22430.2>

**First published:** 19 Jun 2024, 9:320 <https://doi.org/10.12688/wellcomeopenres.22430.1>

**REVISED Amendments from Version 1**

1. The revisions to this manuscript were primarily minor, focusing on addressing the reviewers' comments. The key adjustments involved clarifying that this is a general study protocol encompassing multiple research questions concerning exposure to infectious diseases during pregnancy and the subsequent development of non-infectious chronic conditions in offspring. Each research question will lead to specific analyses and multiple planned publications. For each, a tailored analysis plan will be provided, detailing the methods proposed for investigation.
2. Another notable revision was the inclusion of a comprehensive table that outlines the case definitions for each disease, following the Brazilian Ministry of Health's guidelines. This table also includes relevant information on laboratory tests, treatment protocols, and criteria for defining appropriate treatment. We have additionally provided the operational definitions specific to our study.
3. Furthermore, we incorporated several methodological refinements. We introduced inverse probability of treatment weighting (IPTW) as a potential method to control for confounding and selection bias. We also revisited the proposed growth curve models, noting in this version that we will test various models and select the one that best fits our data, rather than specifying a particular model in advance.

**Any further responses from the reviewers can be found at the end of the article**

## Introduction

The prenatal period is highly sensitive for early development of the brain and of other organs<sup>1</sup> and maternal infection during pregnancy can cause disturbances that might contribute to the causal chain of events leading to subsequent development of chronic illness<sup>2</sup>. Discussed mechanisms include the direct effects of the infectious organism on the developing fetus and exposure to maternal inflammation due to the secretion and circulation of immune-mediated components, such as cytokines, crossing the placental barrier and enter the fetal environment<sup>3</sup>. These processes may involve alterations in gene expression via epigenetic modifications<sup>2,4</sup>. Additionally, such infections might cause a transient depletion of nutrients, e.g. iron, limiting the availability of essential elements necessary for normal fetal development and growth<sup>2</sup>. Moreover, these mechanisms might result in intrauterine growth restriction and preterm birth, which have been associated with non-infectious chronic condition later in life<sup>5,6</sup>.

Empirical data on the long-term repercussions of maternal infections during pregnancy on offspring health has accumulated over the last decades but uncertainty remains. Most of the evidence stems from studies conducted in high income countries focusing their research on specific types of infection such as cytomegalovirus, influenza and herpes virus or general systematic infections (e.g. urinary tract infection)<sup>7</sup>. Another constraint of some studies is the small sample sizes<sup>8-10</sup>, and/or limitations in accurately assessing infection during pregnancy<sup>7</sup>. They commonly rely on maternal self-reported

disease or hospital records<sup>11-13</sup>. While self-reporting can capture a broader range of infections, it is susceptible to recall bias. Conversely, hospital records are biased towards severe diseases. Furthermore, associations between maternal infections during pregnancy and subsequent offspring health could be fully or partially explained by unknown or unmeasured family level confounding factors, such as socioeconomic position and genetic variation. In addition to these methodological limitations, the multifaceted nature of maternal infections underscores the necessity to independently explore the impact of various factors, including the type and timing of infection occurrence during pregnancy, the severity of the infection, and treatment availability (e.g., administration of antibiotics for syphilis treatment), given their potential differential effects on the risk of adverse childhood and adolescent outcomes.

In this study, we will investigate the relationship between infections and offspring health, focusing on four infections that impose a high burden in low- and middle-income countries, which have been overlooked by the current literature. Specifically, we will examine infections known to cause severe congenital complications, such as syphilis and Zika, as well as viral infections with documented neurotropism and/or linked to adverse birth outcomes. The latter category includes dengue and chikungunya, for which we have available data. In the case of syphilis during pregnancy, the current literature primarily focuses on congenital syphilis, and the long-term consequences have not been appropriately studied<sup>14</sup>. Arthropod-borne viral diseases have attracted interest in recent years due to their increasing incidence and expanding geographical area as a direct impact of climate change<sup>15</sup>. The recent Zika epidemic highlighted the potential teratogenic effects of this virus, including severe central nervous system anomalies that constitute Congenital Zika Syndrome. Any long-term consequences in offspring prenatally exposed without symptoms at birth are just now being described, with some<sup>16,17</sup> but not all studies reporting developmental delays in children<sup>18</sup>. Studies of the effects of prenatal exposure to chikungunya infection have suggested a neurodevelopmental risk for children infected during the intrapartum period<sup>19</sup>. Research on adverse outcomes of prenatal dengue exposure has focused on birth outcomes and demonstrated an effect on stillbirth<sup>20</sup> and spontaneous abortion<sup>21</sup>.

The selected outcomes for this study are a combination of conditions with either high morbidity or mortality during childhood and adolescence, for which a potential relationship with prenatal exposure to infection has been suggested. Infection during pregnancy has repeatedly, although not consistently, increased the risk of neurological<sup>22-25</sup> and autoimmune disorders<sup>26</sup>. However, further information on candidate pathogens, timing, and clinical features of infections is lacking. An increasing body of evidence for specific cancer onset during childhood and adolescence (e.g., leukemia, lymphoma) has demonstrated that abnormal immune responses to common infections may trigger these diseases<sup>27</sup>. Although less is known concerning in utero exposure to maternal infection during pregnancy, which could lead to chromosomal or immunological

changes and potentially cause cancer<sup>28</sup>. In addition, maternal infection can cause a congenital malformation that has been associated with childhood cancer<sup>6,29</sup>. Regarding growth patterns, most of the literature has focused on maternal antibiotic use during pregnancy or labor, which a few studies have reported to be associated with an increased risk of obesity in childhood<sup>29,30</sup>. The most prominent candidate explanation for a causal link between maternal infections during pregnancy and childhood/teenage obesity proposes a causal pathway via changes in gut microbiota detrimental to the maintenance of normal metabolic processes and homeostasis<sup>31</sup>. One study attempted to disentangle the effect of antibiotic use from underlying bacterial infections and found a small but significant increased risk<sup>32</sup> due to infection.

The Study of Early-life Exposures During Developmental Stages (SEEDS) is designed to explore the relationships between prenatal and early life exposures on health outcomes throughout the life course. While SEEDS has a broader research scope, this program protocol will specifically focus on the effect of prenatal exposure to infectious diseases on offspring. This is a general research protocol, and we will also publish analysis plans for each paper with more detailed information regarding specific analyses. SEEDS adopts the 'life course' conceptual model, recognizing that risk factors experienced during pregnancy and childhood can significantly influence developmental trajectories, potentially elevating the risk of various diseases.

## Aims and objectives

This proposal aims to investigate whether syphilis, zika, chikungunya or dengue during pregnancy can increase offspring's risk of developing neurologic disorders, auto-immune conditions, specific cancers or abnormal growth patterns during childhood and adolescence.

1. Investigate whether syphilis, zika, chikungunya or dengue during pregnancy are associated to neurologic disorders, auto-immune conditions and specific cancers and explore if the risk varies by timing during pregnancy when the infection is acquired (first, second or third trimester) and severity (such as severe or mild dengue);
2. Investigate whether in utero exposure to syphilis, zika, chikungunya or dengue affects the growth pattern of children and adolescents;
3. Examine if the relationships between syphilis, zika, chikungunya or dengue during pregnancy and neurologic disorders, auto-immune conditions, specific cancers and abnormal growth patterns are mediated by intrauterine growth restriction or preterm birth.

## Proposed methods

### Study population and period

For this study, we will use the Center of Data and Knowledge Integration for Health (CIDACS) Birth Cohort, a Brazilian population-based cohort derived from linked data (SINASC-CadUnico), developed to investigate the relationships between prenatal and early life events on health-related outcomes

for infants, children, and adolescents<sup>33</sup>. The CIDACS Birth Cohort includes Brazilian live births whose families applied for any social government benefit (poorest half of the Brazilian population). This includes approximately nineteen million live births between 1 January 2007 and 31 December 2018, with further updates expected during the study period.

### Data sources

This study is going to use routine data from the Brazilian Information System, which has a long tradition of collecting administrative data; its quality is well documented and constantly improving<sup>34-36</sup>. Several linked datasets will be used: SINASC (Live Birth Information System), CadUnico (Social records); SINAN (Information System for Notifiable Diseases); SIH (Hospitalization Information System); SIM (Information System of Mortality); SISVAN (Food and Nutrition Surveillance System) (Table 1).

1. SINASC (birth registry): This system is updated using the registration of live births. This is a legal document completed by the health professional who assisted with the delivery (whether at home or facilities).
2. CadUnico (social registry): This is an extensive questionnaire with information on the household and each individual. The CadUnico database is electronic and longitudinal and has detailed information on demographic, social, environmental, and economic features on named individuals grouped into families. The information is renewed periodically as long as the person is a candidate to receive one of the several Brazilian government social protection programs.
3. SINAN (infections registry): In Brazil, compulsory notification is required for a list of diseases following Ordinance No 264/2020, including all the infections that will be studied in this proposal. The form is disease specific and suspected cases must be reported to the Epidemiological Surveillance service on a specific numerated notification form available in any local health facility. It can be filled in by any health professional who suspects the disease. The Epidemiological Surveillance Centre then investigates to confirm or discard the suspicion based on the Brazilian definition of clinical-epidemiological cases and/or laboratory results. The type of laboratory test will depend on the disease<sup>37</sup>. Dengue, for example, could be confirmed by IgM antibodies detection by enzyme-linked immunosorbent assay (ELISA), viral Ribonucleic acid (RNA) detection via Polymerase chain reaction (PCR), Nonstructural protein 1 (NS1) viral antigen detection, or positive viral culture; type and result of tests conducted are included in the record. The case must be confirmed or discarded within 60 days. Suspected cases are discarded if one of three conditions are met: 1) negative laboratory test (once confirmed that the sample was collected in the appropriate time); 2) confirmation of differential diagnoses; 3) case without laboratory confirmation and with signs and symptoms

**Table 1. Description of datasets that will be used in this proposal.**

Data source	Data type	Approximate coverage	Examples of relevant variables	Used for:
SINASC	Live birth records in Brazil	97% of live births	Newborn characteristics (sex, Apgar scores, birth weight, ICD-10 coded congenital malformation), maternal characteristics (name, age, marital status, education, race/ethnicity, place of residence), paternal characteristics (name, age), perinatal information (reproductive history: parity, abortions, stillbirths, gestational duration, delivery mode, number of fetuses, antenatal care attendance), location of birth	The intersection between these two datasets created the CIDACS Birth Cohort
CadUnico	Social records from over 130 million individuals aged ≥16 years whose families applied for social assistance in Brazil	>50% of population	Location of residence (municipality, region, urban or rural), living conditions (housing, water sources, electricity, sewage, waste, household density), socioeconomic characteristics (education, employment, income), demographics (age, sex, race/ethnicity), social protection program participation and conditionalities	
SINAN	Records of communicable diseases of interest in the country	----	It is disease specific. Onset of symptoms, date of birth, patient name, sex, address, laboratory confirmation, symptoms and treatment	Exposure
SIH-SUS	Hospitalization admission records	75% of hospitalizations in the Brazilian National Health System	Cause of hospitalization (ICD-10 code), duration, costs, and date of hospitalization, type of hospital	Outcomes
SIM	Death records in Brazil	75–95% of Brazilian deaths, with some geographic heterogeneity	Cause of death (ICD-10 code), characteristics of the deceased (dates of birth and death, name, name of parents, sex, race/ethnicity, birth weight for infants), place of death, characteristics of mother of deceased children (maternal name, age, marital status, education, occupation, race/ethnicity, number of births, place of residence, length of gestation, number of previous stillbirths or abortions, type of delivery)	Outcomes
SISVAN	Anthropometric and food consumption records	30% of children <5 years attending primary care in the Brazilian National Health System, with heterogeneity across populations	Date of birth, age, sex, race/ethnicity, anthropometrics, breastfeeding, complementary feeding, consumption of healthy and unhealthy foods	Outcomes

compatible with another pathology. After confirmation, information on the clinical progress of the case is collected and then, it is classified into clinical categories if applicable: for example, syphilis can be classified as primary, secondary, latent and tertiary syphilis. The sensitivity of the information system to detect this infectious varies by type and we will investigate the impact of potential misclassification of exposure during the study.

4. SIH (hospitalization registry): All hospitalizations' admissions financed by the Brazilian National Health System are recorded in this system (about 75% of the Brazilian population uses the universal Brazilian system) with this percentage likely higher among the poor population registered in CadUnico.

The cause of hospitalization is reported as the diagnosis that motivated the hospitalization (coded using the ICD-10 code). It is filled by the health worker that requested the admission, normally a physician, with few exceptions such as childbirth, in which a midwife can request admission.

5. SIM (death registry): Death-related information will be obtained from the Mortality Information System, which records all Death Certificates. The death certificate is a legal document that must be completed by the physician responsible for clinical care, an assistant or another practitioner from the institution who attests to the cause of death (using the ICD-10 code)<sup>38</sup>.
6. SISVAN (food and nutrition registry): will be used to obtain child and adolescent growth. Data from

this system has information on anthropometric measurements, including weight and height, food consumption, breastfeeding and complementary feeding practices. The national population coverage of SISVAN ranges between 10% and 15%, mainly among children and adolescents. For those registered in the cash transfer program Bolsa Familia, who are also enrolled in the CadUnico, the SISVAN coverage varies from 57% to 86%<sup>39</sup>

### Data linkage

To link the datasets, the CIDACS team have used the CIDACS-RL, a novel record linkage tool developed to link big administrative data based on identifiers, such as name, sex, age or date of birth, mother's name, and the municipality of residence<sup>40</sup>. The CIDACS-RL applies indexing and searching algorithms implemented in the Apache Lucene solution as the blocking strategy. The indexation strategy allows the CIDACS-RL to search for the most similar records from the indexed dataset for each record in the main dataset and submit them to the pairwise comparisons step. Candidate linking records are ordered by the similarity scores, and only the comparison pair with the highest score is retained as a potential link. All remaining candidate records were discarded. The linkage strategy is detailed described in Barbosa *et al.* (2020)<sup>40</sup>. The linkage quality has been continuously assessed, and accuracy varies according to linked datasets.

### Exposure definition

Information on exposure to syphilis, zika, dengue, chikungunya will be assessed via record linkage between the cohort and the Information System for Notifiable Diseases (SINAN). We will use clinical and laboratory-confirmed cases and conduct sensitivity analyses, including only those laboratory-confirmed, and if applicable, analyses can also be performed according to disease severity and treatment and timing of infection. The period for which data are available for each disease will vary according to data availability and how long disease has been circulating in Brazil, with data available for syphilis and dengue from 2007 and from 2015 for chikungunya and zika (Table 2).

### Outcome definition

The outcomes will be assessed by record linkage of three different information systems depending on the disorder (SIH-hospitalization data, SIM-mortality data, and SISVAN-nutrition surveillance).

1) Neurologic disorders (cerebral palsy, epilepsy, autism, motor neurone disease), auto-immune conditions (type I diabetes, membranous nephropathy, inflammatory bowel disease, autoimmune hepatitis and juvenile arthritis), and specific cancers (lymphoma, leukemia and neuroblastoma) will be assessed by linking the cohort with hospitalization and mortality data. They will be identified using the ICD-10 code. We know that not all disease episodes will lead to hospitalization; however, due to population size, we predict enough events.

2) Growth patterns will be identified by linking the cohort with nutrition registry data (nationwide coverage reaching up to 43% among children<sup>41</sup>). We will use data on weight and height to calculate body mass index (BMI) and plot these measurements on growth charts for boys and girls. To describe trajectories in child weight, height and BMI, we plan to test different classes of models including: latent class growth, mixed effect and Super Imposition by Translation And Rotation (SITAR) models<sup>42-44</sup> to assess which model fits better to our data. Children and adolescents will be classified as wasted (weight-for-height Z-score (WHZ) < -2), stunted (height-for-age Z-score < -2), overweight or obese (WHZ > 2).

### Statistical analyses

Each pair of infectious and non-infectious outcomes will be investigated separately. We will then select all potential confounders irrespective of whether or not there is data available. Then, Directed Acyclic Graphs (DAGs) will be constructed, including the *a priori* defined set of confounders and mediators. These DAGs will be used to explore residual confounders and to inform sensitivity analyses<sup>45,46</sup>. The consideration of potential confounders will be informed by the existing literature and the knowledge of the investigator and collaborators.

The results from different epidemiological approaches will be triangulated to strengthen causal inference. The idea behind triangulation is that when we compare different approaches with assumed unrelated sources of bias<sup>47</sup>, in our study, we will be able to compare the results and if they point to the same answer, this strengthens causal inference. The key sources of bias will be explicitly acknowledged for each approach when comparing results. Where results point to different answers, we plan to understand the key sources of bias to help direct the research. The methodological approach implemented will vary according to exposure characteristics. Below, we present the main epidemiological approaches planned to answer objectives one and two.

1. Multivariable analyses and propensity score-based techniques, such as inverse probability of treatment weighting (IPTW) for the entire study population: We will employ traditional multivariable time-to-event models, adjusting for a predefined set of confounders, alongside propensity score-based methods to evaluate the association between each exposure and outcome. This will involve comparing children born to mothers with and without notified infections during pregnancy. Specifically for syphilis and Zika, we will analyze two exposed groups: those with confirmed congenital infections (congenital syphilis and congenital Zika syndrome) and those exposed without symptoms at birth. Where appropriate, we will explore variations in risk according to the timing of infection during pregnancy and the severity of the disease. However, this approach may still be subject to residual confounding.
2. Multivariable analyses within sibling groups: We will include participants with at least one sibling in the

**Table 2. Case definition, treatment and laboratory test.**

Disease	Brazilian case definition of disease	Case definition in our study	Treatment	Laboratory test
Gestational Syphilis	<p><b>Gestational syphilis</b> is defined based on criteria that combine information from laboratory tests, symptoms, and treatment. Women meeting one of the following combinations during prenatal care, childbirth, or the puerperium period should be reported as having gestational syphilis: asymptomatic women with at least one positive laboratory test for syphilis without prior treatment; symptomatic women with at least one positive laboratory test for syphilis, regardless of prior treatment; or women with positive results on both non-treponemal and treponemal tests, regardless of treatment status or symptoms.</p> <p><b>Congenital syphilis</b> is defined as individuals meeting one or more of the following criteria should be reported as cases: live births from mothers with untreated or inadequately treated syphilis; children with microbiological evidence of <i>Treponema pallidum</i> in nasal discharge, skin lesion, child biopsy, or autopsy; or children younger than 13 years with at least one of the following: clinical, cerebrospinal fluid, or radiological manifestations of congenital syphilis and a positive non-treponemal test; infants (younger than 1 year) with non-treponemal test titers greater than those of the mother in at least two dilutions; children with increasing non-treponemal test titers in at least two dilutions; non-treponemal test titers remaining positive in a child older than 6 months who was adequately treated in the neonatal period; or a positive treponemal test in a child aged 18 months without a previous diagnosis of congenital syphilis.</p>	<p>Live births from the CIDACS Birth Cohort that are linked with records from SINAN indicating gestational or congenital syphilis will be considered as exposed to syphilis during pregnancy. To investigate maternal syphilis, we will exclude records with notification dates either before conception or after the postpartum period. For investigating congenital syphilis, we will exclude records with notifications made more than 2 months prior to birth, records of suspected congenital syphilis that were ruled out following routine epidemiological investigation, and records involving fetal loss. Records not linked to SINAN syphilis data will be classified as not exposed to syphilis during pregnancy.</p>	<p>The treatment for gestational syphilis should start as soon as possible with benzathine penicillin.</p> <p>Criteria for appropriate treatment in pregnant patients include:</p> <ul style="list-style-type: none"> <li>• Administration of benzathine penicillin</li> <li>• Initiation of treatment at least 30 days before delivery</li> <li>• Therapeutic regimen according to the clinical stage of syphilis: primary syphilis requires a single dose of 2,400,000 IU; secondary syphilis requires two doses of 2,400,000 IU each; tertiary syphilis requires three doses of 2,400,000 IU each.</li> <li>• Adherence to the recommended interval between doses</li> </ul>	<p>Diagnostic tests for syphilis are divided into two main categories: direct identification of <i>Treponema pallidum</i> in samples from primary or secondary lesions, and immunological tests. In Brazil, direct identification methods include dark-field microscopy, stained material microscopy, direct immunofluorescence, and nucleic acid amplification tests (NAAT).</p> <p>Immunological tests detect antibodies produced by the body in response to the infection, utilizing blood, serum, or plasma samples. These tests are classified into two types: treponemal and non-treponemal. In Brazil, treponemal tests include the Fluorescent Treponemal Antibody Absorption Test (FTA-Abs), Enzyme-Linked Immunosorbent Assay (ELISA), chemiluminescent immunoassays, hemagglutination and agglutination tests, and rapid tests, which must demonstrate sensitivity greater than 94.5% and specificity greater than 93.0%. The primary non-treponemal test used in Brazil is the Venereal Disease Research Laboratory (VDRL) test. Results should always include the dilution titers.</p> <p>Following a positive rapid test, a blood sample is typically collected for a non-treponemal test. However, for pregnant women, treatment should be initiated based on the initial positive rapid test result, even if the second test results are pending.</p>



Disease	Brazilian case definition of disease	Case definition in our study	Treatment	Laboratory test
<p>Dengue virus</p>	<p>In Brazil, individuals who live in or have travelled within the last 14 days to an area where dengue transmission is occurring or where <i>Aedes aegypti</i> is present, and who present with fever lasting between 2 and 7 days, accompanied by two or more of the following symptoms—nausea, vomiting, rash, myalgia, headache, retro-orbital pain, petechiae, a positive tourniquet test, or leukopenia—should be reported as suspected dengue cases. Suspected cases can be confirmed through laboratory testing or clinical-epidemiological criteria, the latter based on the presence of dengue symptoms occurring in the same area and time as other confirmed cases, following an epidemiological investigation. Depending on the clinical manifestations, cases can be further classified as dengue with or without warning signs, or as severe dengue. <b>Dengue with warning signs</b> includes cases with intense and continuous abdominal pain (either reported by the patient or upon palpation); persistent vomiting; fluid accumulation (such as ascites, pleural effusion, or pericardial effusion); postural hypotension and/or syncope; hepatomegaly &gt;2 cm below the costal margin; mucosal bleeding; lethargy and/or irritability; or a progressively increasing hematocrit.</p> <p><b>Severe dengue</b> is characterized by the presence of one or more of the following conditions: a. Shock or respiratory distress due to severe plasma leakage, evidenced by tachycardia, weak or undetectable pulse, cold extremities, capillary refill time &gt;2 seconds, and convergent differential blood pressure &lt;20 mmHg, indicating late-stage hypotension. b. Severe bleeding, as assessed by a physician (e.g., hematemesis, melena, profuse metrorrhagia, or bleeding within the central nervous system). c. Severe organ impairment, such as significant hepatic damage (AST/ALT &gt;1,000), central nervous system involvement (altered consciousness), cardiac involvement (myocarditis), or damage to other organs.</p>	<p>Live births from the CIDACS Birth Cohort that are linked with SINAN dengue records indicating that the mother was reported and confirmed as a dengue case specifically during pregnancy (from conception to the date of birth) will be considered exposed to dengue. Symptomatic dengue during pregnancy includes all confirmed cases based on clinical-epidemiological and clinical-laboratory data. Laboratory-confirmed dengue cases during pregnancy refers exclusively to cases confirmed through laboratory testing.</p>	<p>There is no specific treatment for dengue; case management is primarily symptomatic. In cases of severe dengue, hospitalization may be necessary.</p>	<p>Laboratory confirmation for dengue use the following tests a. Detection of reactive NS1 protein. b. Positive viral isolation. c. Detectable RT-PCR (up to the fifth day after the onset of symptoms). d. Detection of IgM antibodies by ELISA (from the sixth day after the onset of symptoms). e. A fourfold or greater increase in antibody titers in Plaque reduction neutralization test (PRNT) or Hemagglutination Inhibition (IH) tests using paired samples (acute and convalescent phases with at least a 14-day interval). Dengue (DENV) and Zika (ZIKV) viruses are flaviviruses, which can lead to cross-reactivity in serological tests, resulting in inconclusive laboratory results.</p>

Disease	Brazilian case definition of disease	Case definition in our study	Treatment	Laboratory test
Chikungunya virus	Cases of sudden onset fever accompanied by severe acute arthralgia or arthritis, unexplained by other conditions, in individuals who reside in or have traveled to endemic or epidemic areas within 14 days prior to symptom onset, or who have an epidemiological link to a confirmed imported case, should be reported as suspected cases of Chikungunya.	Live births from the CIDACS Birth Cohort linked to SINAN records confirming that the mother was diagnosed with symptomatic Chikungunya during pregnancy (from conception to birth) will be considered exposed to the virus. Symptomatic Chikungunya during pregnancy includes all confirmed cases based on clinical-epidemiological and clinical-laboratory data. Cases specifically confirmed through laboratory testing are classified as laboratory-confirmed Chikungunya cases during pregnancy.	There is no specific treatment for Chikungunya; management is primarily symptomatic, with a focus on pain relief.	The laboratory diagnosis of CHIKV infection can be performed directly through viral isolation and the detection of viral RNA in various clinical samples, or indirectly through the detection of specific antibodies. For viral RNA detection, the main molecular techniques used are RT-PCR (Reverse Transcription Polymerase Chain Reaction) and qRT-PCR (Real-Time RT-PCR). These techniques offer a rapid and sensitive diagnosis, allowing the detection of viral nucleic acid up to approximately the eighth day after symptom onset, with the period of highest viremia occurring between the first and fifth days.  For the detection of specific antibodies, the main available techniques are the Enzyme-Linked Immunosorbent Assay (ELISA) and Point-of-Care (POC) immunochromatographic tests. Serological tests enable the detection of specific antibodies, including IgM, which can be detected from the second day after symptom onset (with the optimal period for serological investigation being from the fifth day) and IgG, which can be detected from the sixth day.

Disease	Brazilian case definition of disease	Case definition in our study	Treatment	Laboratory test
Zika virus	<p>Individuals presenting with a maculopapular rash and TWO or more of the following signs and symptoms—fever, conjunctival hyperemia without discharge and with or without itching, polyarthralgia, or periarticular edema—should be reported as suspected cases of Zika.</p> <p>Suspected cases can be classified as confirmed through laboratory testing, deemed probable based on clinical criteria, or ruled out following an epidemiological investigation.</p> <p><b>Congenital zika syndrome:</b> Live-born children who meet one or more of the following criteria have their cases reported and investigated as suspected cases of congenital Zika syndrome: microcephaly, defined as a head circumference that is more than 2 SD below the mean for age and sex (according to INTERGROWTH-21st standards for infants born at &lt;37 weeks of gestation or according to World Health Organization standards for infants born at ≥37 weeks of gestation); craniofacial disproportion (microcrania in relation to the face); central nervous system changes suggestive of congenital infection as detected on neuroimaging tests (cranial computed tomography, magnetic resonance imaging of the head, or transfontanelar ultrasonography); two or more neurologic, visual, or auditory manifestations. After notification, all suspected cases are investigated by the epidemiologic surveillance teams and classified as confirmed, probable, inconclusive, or ruled-out cases. Suspected cases of congenital Zika syndrome were considered to be confirmed if there were signs and symptoms consistent with the syndrome or laboratory evidence of ZIKV infection (from molecular or serologic testing) or the mother had reported fever or rash during pregnancy. Probable cases involved clinical changes compatible with congenital Zika syndrome and negative tests for other congenital infections although specific laboratory diagnosis for ZIKV infection was not available and the mother was asymptomatic during pregnancy. Suspected cases were ruled out if there were compatible clinical symptoms that, after clinical and epidemiologic investigation, were attributed to another cause, such as microcephaly related to restricted intrauterine growth or genetic disease. Other cases were inconclusive owing to insufficient information for proper classification or remained under investigation.</p>	<p>Live births from the CIDACS Birth Cohort linked with records from SINAN for Zika or the Public Health Event Record (RESP) for congenital Zika syndrome will be considered exposed to Zika during pregnancy.</p> <p>To investigate maternal Zika, we will exclude records with notification dates either before conception or after birth. Cases of congenital Zika syndrome can be further classified based on head circumference measurements.</p>	<p>There is no specific treatment for Zika; case management is primarily symptomatic.</p>	<p>Laboratory diagnosis of Zika virus (ZIKV) can be performed using direct methods, including viral isolation and the detection of viral genome through reverse transcription followed by polymerase chain reaction (RT-PCR), as well as indirect methods, which involve identifying the presence of viral antibodies. Due to the similarity in signs and symptoms of ZIKV infection with dengue and chikungunya, it is advised that testing for Zika begin with direct methods if Zika is the initial suspicion. Urine samples can be used to confirm viral infection up to the 15th day after the onset of symptoms.</p>

cohort and compare results between siblings who are discordant for the exposure (i.e. the sub group of participants where the offspring have at least one sibling and at least one sibling was exposed to the infection in utero and the other was not). We will fit semi or parametric survival models comparing the outcomes of siblings within nuclear families. This approach will adjust only for offspring sex, mode of delivery and birth order since other factors will be stable within families. Robust standard errors will be used to account for familial clustering at the population level. These matched analyses can control for observed, unobserved family level confounding, such as socioeconomic position and genetics. It has been very powerful to examine the effects of any in utero exposure on outcomes assessed several years later since correlation within siblings is stronger for the confounders than for the exposure of interest during the pregnancy<sup>47</sup>. However, this approach has limitations. Bias due to misclassification and individual level (as opposed to family) confounding is expected to be greater in within-siblings studies than when comparing unrelated individuals. Furthermore, they can be biased by selection, as only participants with at least one sibling are included, and contamination, where the exposure or outcome in the first sibling influences the outcome or exposure in subsequent siblings, can occur.

3. Cross-context comparisons: Brazil is a remarkably diverse country, encompassing a wide range of social and environmental factors. Our aim is to investigate whether different Brazilian regions present varying confounding structures for each association under investigation. If this assumption holds true, we will conduct stratified analyses. The primary potential source of bias in this approach is the possibility of differential misclassification of infections during pregnancy and non-infectious chronic disorders across regions.
4. Outcome negative control: This approach will select a different condition as a negative control (e.g. car accident). The outcome selected should be likely influenced by the same confounders but would not plausibly be influenced by uterus exposure to infection. There might be differences in the sources of bias between the real assessed conditions and negative control outcome (car accident), and we will explore this issue by investigating the association of observed confounders with the negative control exposure. In this approach, we would expect an unbiased negative control study to be null and the real study not null if there is a true causal effect. However, the negative control study may show a weaker association (than the real study), suggesting some bias in the real study, and we can use this information as an indicator of the extent of bias in the real study<sup>47</sup>. This approach is particularly useful to detect unmeasured

confounding, which can be a limitation of this proposal.

The results provided by all these different approaches will give a complete picture of the relationship between each selected maternal infection during pregnancy and the onset of specific non-infectious chronic disorders during childhood and adolescence.

In this study, we will conduct multiple comparisons, increasing the probability of false positivity. To adjust for multiple testing, we plan to use the false discovery rate (FDR), which has recently been proposed as an alternative metric to family-wise error rate tests. The FDR has been shown to have greater power to detect true positives while still controlling the proportion of type I errors at a specified level<sup>48</sup>.

To answer objective three, we will implement causal mediation methods to explore the complex causal pathway between infection during pregnancy and the studied outcomes. There is well-established literature about the association between maternal infections during pregnancy and small for gestational age (*proxy* of intrauterine growth restriction) and preterm birth, which are the mediators of these analyses. The simplest way to estimate the direct (number of outcomes attributed to infection during pregnancy direct path) and indirect (number of outcomes mediated through small for gestational age and prematurity) effects of maternal infection on non-infectious chronic outcomes, is to estimate relative risks with and without adjustment for potential mediating variables. However, simple adjustment for the mediator can give misleading results if there are confounders of the mediator-outcome relationship. In addition to standard regression methods, we will use statistical approaches, such as inverse probability weighting (IPW) based on the counterfactual framework<sup>49</sup>, to decompose the direct and indirect effect of maternal infection during pregnancy and non-infectious chronic disorders in children and adolescents.

### Ethics and dissemination

Ethical approval to set up the CIDACS Birth Cohort has been obtained by the Federal University of Bahia's Institute of Public Health Ethics Committee (CAAE registration number: 18022319.4.0000.5030) and the London School of Hygiene & Tropical Medicine reference number 22817. We also obtained specific approval covering the objectives presented here, by the Gonçalo Moniz Research Center - CPqGM/ FIOCRUZ/ BA (CAAE registration number: 71054923.8.0000.0040) and the London School of Hygiene & Tropical Medicine reference number 29560.

All data linkage procedures have been conducted in a physically and virtually secure environment and follow strict internal information security measures to ensure data privacy and confidentiality at CIDACS. After linkage, the databases have been fully de-identified: all identifiable elements used for linkage will be removed, and all precautions will be adopted to prevent re-identification. The final linked and de-identified

datasets will receive a Digital Object Identifier (DOI), and the complete specification of how the dataset was created will be made available upon request. The analysis will be performed in the CIDACS data analysis environment, a safe virtual infrastructure that provides remote data access (for London-based researchers) and analysis tools via Virtual Private Network (VPN).

By identifying the effect of in utero exposure to each specific infection on high-morbidity or high-mortality non-infectious chronic disorders that consume health care resources, the outputs of this study can be used to estimate attributable fractions – which will give an idea of the effect of preventing these maternal infections during pregnancy would have on the burden of these non-infectious chronic outcomes. Also, the increasing awareness of the effects of maternal infection on non-infectious chronic outcomes may offer guidance for improving public health interventions such as vaccinations, antibiotics and/or antiviral therapy, and dietary adjustments. These interventions might help to minimize the incidence of

non-infectious chronic disorders beyond what has been achieved by public health recommendations for maternal–fetal health that are already in place. We expect there to be considerable attention in these findings in light of the strong interest in the effects of maternal infection during pregnancy on the health of children and adolescents.

The research results will be disseminated through peer-reviewed publications. We anticipate publishing high-impact articles and presenting them at conferences to communicate the key findings and scientific advances. Additionally, we plan to share these findings with policymakers editing policy briefings and the public through engagement with LSHTM and CIDACS media channels, such as contributions to written media, as well as through community engagement offices, including participation in science festivals.

## Data availability

No data are associated with this article.

## References

- Lim AI, McFadden T, Link VM, *et al.*: **Prenatal maternal infection promotes tissue-specific immunity and inflammation in offspring.** *Science*. 2021; **373**(6558): eabf3002. [PubMed Abstract](#) | [Publisher Full Text](#)
- Labouesse MA, Langhans W, Meyer U: **Long-term pathological consequences of prenatal infection: beyond brain disorders.** *Am J Physiol Regul Integr Comp Physiol*. 2015; **309**(1): R1–R12. [PubMed Abstract](#) | [Publisher Full Text](#)
- Zhou J, Tong J, Ru X, *et al.*: **Placental inflammatory cytokines mRNA expression and preschool children's cognitive performance: a birth cohort study in China.** *BMC Med*. 2023; **21**(1): 449. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Choi GB, Yim YS, Wong H, *et al.*: **The maternal Interleukin-17a pathway in mice promotes autism-like phenotypes in offspring.** *Science*. 2016; **351**(6276): 933–939. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Buchmayer S, Johansson S, Johansson A, *et al.*: **Can association between preterm birth and autism be explained by maternal or neonatal morbidity?** *Pediatrics*. 2009; **124**(5): e817–25. [PubMed Abstract](#) | [Publisher Full Text](#)
- Lupo PJ, Schraw JM, Desrosiers TA, *et al.*: **Association between birth defects and cancer risk among children and adolescents in a population-based assessment of 10 million live births.** *JAMA Oncol*. 2019; **5**(8): 1150–1158. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- He JR, Ramakrishnan R, Hirst JE, *et al.*: **Maternal infection in pregnancy and childhood leukemia: a systematic review and meta-analysis.** *J Pediatr*. 2020; **217**: 98–109.e8. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Roman E, Ansell P, Bull D: **Leukaemia and non-Hodgkin's lymphoma in children and young adults: are prenatal and neonatal factors important determinants of disease?** *Br J Cancer*. 1997; **76**(3): 406–415. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fine PE, Adelstein AM, Snowman J, *et al.*: **Long term effects of exposure to viral infections in utero.** *Br Med J (Clin Res Ed)*. 1985; **290**(6467): 509–11. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bluhm E, McNeil DE, Cnattingius S, *et al.*: **Prenatal and perinatal risk factors for neuroblastoma.** *Int J Cancer*. 2008; **123**(12): 2885–90. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lee BK, Magnusson C, Gardner RM, *et al.*: **Maternal hospitalization with infection during pregnancy and risk of autism spectrum disorders.** *Brain Behav Immun*. 2015; **44**: 100–5. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Neufeld MD, Frigon C, Graham AS, *et al.*: **Maternal infection and risk of Cerebral Palsy in term and preterm infants.** *J Perinatol*. 2005; **25**(2): 108–13. [PubMed Abstract](#) | [Publisher Full Text](#)
- Cocoros NM, Lash TL, Nørgaard M, *et al.*: **Hospitalized prenatal and childhood infections and obesity in Danish male conscripts.** *Ann Epidemiol*. 2013; **23**(6): 307–13. [PubMed Abstract](#) | [Publisher Full Text](#)
- Lim J, Yoon SJ, Shin JE, *et al.*: **Outcomes of infants born to pregnant women with syphilis: a nationwide study in Korea.** *BMC Pediatr*. 2021; **21**(1): 47. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Paixão ES, Teixeira MG, Rodrigues LC: **Zika, chikungunya and dengue: The causes and threats of new and re-emerging arboviral diseases.** *BMJ Glob Health*. 2018; **3**(Suppl 1): e000530. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mulkey SB, Arroyave-Wessel M, Peyton C, *et al.*: **Neurodevelopmental abnormalities in children with in utero zika virus exposure without congenital zika syndrome.** *JAMA Pediatr*. 2020; **174**(3): 269–276. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cranston JS, Tiene SF, Nielsen-Saines K, *et al.*: **Association between antenatal exposure to zika virus and anatomical and neurodevelopmental abnormalities in children.** *JAMA Netw Open*. 2020; **3**(7): e209303. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gerzson LR, de Almeida CS, da Silva JH, *et al.*: **Neurodevelopment of nonmicrocephalic children, after 18 months of life, exposed prenatally to zika virus.** *J Child Neurol*. 2020; **35**(4): 278–282. [PubMed Abstract](#) | [Publisher Full Text](#)
- Gérardin P, Barau G, Michault A, *et al.*: **Multidisciplinary prospective study of mother-to-child chikungunya virus infections on the Island of La Réunion.** *PLoS Med*. 2008; **5**(3): e60. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Paixão ES, Costa MCN, Teixeira MG, *et al.*: **Symptomatic dengue infection during pregnancy and the risk of stillbirth in Brazil, 2006–12: a matched case-control study.** *Lancet Infect Dis*. 2017; **17**(9): 957–964. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Paixão ES, Teixeira MG, da Conceição N Costa M, *et al.*: **Dengue during pregnancy and adverse fetal outcomes: a systematic review and meta-analysis.** *Lancet Infect Dis*. 2016; **16**(7): 857–865. [PubMed Abstract](#) | [Publisher Full Text](#)
- Jiang HY, Xu LL, Shao L, *et al.*: **Maternal infection during pregnancy and risk of autism spectrum disorders: a systematic review and meta-analysis.** *Brain Behav Immun*. 2016; **58**: 165–172. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bryngre M, Sjöqvist H, Gardner RM, *et al.*: **Maternal infection during pregnancy and likelihood of autism and intellectual disability in children in Sweden: a negative control and sibling comparison cohort study.** *Lancet Psychiatry*. 2022; **9**(10): 782–791. [PubMed Abstract](#) | [Publisher Full Text](#)
- Miller JE, Pedersen LH, Streja E, *et al.*: **Maternal infections during pregnancy**

- and Cerebral Palsy: a population-based cohort study.** *Paediatr Perinat Epidemiol.* 2013; **27**(6): 542–52.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Sun Y, Vestergaard M, Christensen J, et al.: **Prenatal exposure to maternal infections and epilepsy in childhood: a population-based cohort study.** *Pediatrics.* 2008; **121**(5): e1100–7.  
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Viskari H, Knip M, Tauriainen S, et al.: **Maternal enterovirus infection as a risk factor for type 1 diabetes in the exposed offspring.** *Diabetes Care.* 2012; **35**(6): 1328–32.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Greaves M: **A causal mechanism for childhood acute lymphoblastic leukaemia.** *Nat Rev Cancer.* 2018; **18**(8): 471–484.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. He JR, Hirst JE, Tikellis G, et al.: **Common maternal infections during pregnancy and childhood leukaemia in the offspring: findings from six international birth cohorts.** *Int J Epidemiol.* 2021; **51**(3): 769–777.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Mueller NT, Whyatt R, Hoepner L, et al.: **Prenatal exposure to antibiotics, cesarean section and risk of childhood obesity.** *Int J Obes (Lond).* 2015; **39**(4): 665–70.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Wang B, Liu J, Zhang Y, et al.: **Prenatal exposure to antibiotics and risk of childhood obesity in a multicenter cohort study.** *Am J Epidemiol.* 2018; **187**(10): 2159–2167.  
[PubMed Abstract](#) | [Publisher Full Text](#)
31. Li DK, Chen H, Ferber J, et al.: **Infection and antibiotic use in infancy and risk of childhood obesity: a longitudinal birth cohort study.** *Lancet Diabetes Endocrinol.* 2017; **5**(1): 18–25.  
[PubMed Abstract](#) | [Publisher Full Text](#)
32. Li DK, Chen H, Ferber J, et al.: **Maternal infection and antibiotic use in pregnancy and the risk of childhood obesity in offspring: a birth cohort study.** *Int J Obes (Lond).* 2020; **44**(4): 771–780.  
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Paixao ES, Cardim LL, Falcao JR, et al.: **Cohort profile: Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS) birth cohort.** *Int J Epidemiol.* 2021; **50**(1): 37–38.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Gabriel GP, Chiquetto L, Morcillo AM, et al.: **Evaluation of data on live birth certificates from the information system on live births (SINASC) in campinas, são paulo, 2009.** *Rev Paul Pediatr.* 2014; **32**(3): 183–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Szwarcwald CL, do Carmo Leal M, Esteves-Pereira AP, et al.: **Evaluation of data from the brazilian information system on live births (SINASC).** *Cadernos de Saude Publica.* 2019; **35**(10): e00214918.  
[PubMed Abstract](#) | [Publisher Full Text](#)
36. De Lima EEC, Queiroz BL: **Evolution of the deaths registry system in Brazil: associations with changes in the mortality profile, under-registration of death counts, and ill-defined causes of death.** *Cad Saude Publica.* 2014; **30**(8): 1721–30.  
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Brazil. Ministry of Health: **Portaria nº 264, de 17 de Fevereiro de 2020.** 2020.  
[Reference Source](#)
38. Ministério da Saúde: **SIM. Sistema de Informações sobre Mortalidade.** *Datasus.* 2017.
39. do Nascimento FA, da Silva SA, Jaime PC: **Cobertura da avaliação do consumo alimentar no sistema de vigilância alimentar e nutricional Brasileiro: 2008 a 2013.** *Rev Bras Epidemiol.* 2019; **22**: e190028.  
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Barbosa GCG, Ali MS, Araujo B, et al.: **CIDACS-RL: a novel indexing search and scoring-based record linkage system for huge datasets with high accuracy and scalability.** *BMC Med Inform Decis Mak.* 2020; **20**(1): 289.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
41. Silva NJ, Silva JFME, Carrilho TRB, et al.: **Quality of child anthropometric data from SISVAN, Brazil, 2008–2017.** *Rev Saude Publica.* 2023; **57**(1): 62.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
42. Cole TJ, Donaldson MDC, Ben-shlomo Y: **SITAR—a useful instrument for growth curve analysis.** *Int J Epidemiol.* 2010; **39**(6): 1558–66.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. Anderson C, Hafen R, Sofrygin O, et al.: **Comparing predictive abilities of longitudinal child growth models.** In: *Stat Med.* 2019; **38**(19): 3555–3570.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. van Buuren S: **Broken stick model for irregular longitudinal data.** *J Stat Softw.* 2023; **106**(7): 1–51.  
[Publisher Full Text](#)
45. Greenland S, Pearl J, Robins JM: **Causal diagrams for epidemiologic research.** *Epidemiology.* 1999; **10**(1): 37–48.  
[PubMed Abstract](#)
46. Pearce N, Lawlor DA: **Causal inference—so much more than statistics.** *Int J Epidemiol.* 2016; **45**(6): 1895–1903.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Lawlor DA, Tilling K, Smith GD: **Triangulation in aetiological epidemiology.** *Int J Epidemiol.* 2016; **45**(6): 1866–1886.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Korthauer K, Kimes PK, Duvallet C, et al.: **A practical guide to methods controlling false discoveries in computational biology.** *Genome Biol.* 2019; **20**(1): 118.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. Richiardi L, Bellocco R, Zugna D: **Mediation analysis in epidemiology: methods, interpretation and bias.** *Int J Epidemiol.* 2013; **42**(5): 1511–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status:  

---

## Version 2

Reviewer Report 18 October 2024

<https://doi.org/10.21956/wellcomeopenres.25534.r105818>

© 2024 Suleri A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Anna Suleri** 

Erasmus University Medical Center, Rotterdam, The Netherlands

I thank the authors for their response and revisions. I think it has improved the analysis plan and look forward to seeing the results.

**Competing Interests:** No competing interests were disclosed.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

## Version 1

Reviewer Report 12 August 2024

<https://doi.org/10.21956/wellcomeopenres.24711.r87458>

© 2024 Suleri A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Anna Suleri** 

Erasmus University Medical Center, Rotterdam, The Netherlands

In this study by Paixao et al, the authors aim to explore the role of prenatal exposure to infectious diseases and the onset of chronic conditions in children and adolescents. I believe this is an important topic and that this is a comprehensive study. The protocol is well written, provides many details and I compliment the authors for their pre-registration. I have several comments and questions:

1. Please provide more information on the exposure definition:
  - Please provide details on the clinical definition
  - Please provide details laboratory definition (e.g., which assay)
  - Please provide information what sensitivity analysis will be conducted and what the rationale is
  - Please provide information how disease severity will be coded – as I see that treatment is mentioned separately
  - Please provide information how the effect of timing of infection will be measured (e.g. in individual models, G-computation models or mutually adjusted models? And how often is it measured, e.g. once per trimester?
  - It may be interesting to also conduct sensitivity analysis for those exposed to multiple viruses to examine if there is a dose-response relationship between number of viruses prenatally exposed and offspring outcomes
2. Will the outcomes be measured individually, or will a composite score be made? Is there also a way to get information from the GP or subclinical traits of e.g. autism?
3. Will clusters be made for the outcomes based on growth models and will these be placed within the regression models? If there is repeated measure data for the outcome, it may be worth considering performing mixed-effects models that allows to adjust for intra-individual confounding variables.
4. How will the authors deal with missing data in the exposure, covariates and outcome? Will they for example impute part of the missing data, and if so how?
5. Please provide the statistical packages and program the analyses will be conducted.
6. I compliment the authors for adjusting for multiple testing. Please provide the exact number of tests across the aims for which FDR will be applied.
7. I compliment the authors for aiming to create a DAG. Would it be possible to add this to the analysis plan with the exact confounders that will be used?
8. To adjust for selection bias to some extent, did the authors consider inverse probability weighting?
9. Please provide a power analysis based on the given sample size, what the smallest possible effect is the authors are powered to find.
10. Please provide a reference for the cut-offs for the weight-for-height Z-score. Did the authors consider calculating a SDS for BMI? And is this based on the WHO reference curves or another curve?
11. What will be the cut-off for the BIC to select the number of trajectory groups for outcome definition part 2? And what will the authors do if the model fit remains poor? And how will the authors ensure (clinical) interpretability of the number of groups based on the growth curves?
12. For the statistical analysis part 2, the within sibling group analyses, I tend to disagree with the authors that only sex and birth are the factors differential within families. Different siblings may for example have a different gestational age at birth, mode of delivery or birth complications which are important confounders in this specific study.
13. How will the matching take place in statistical analysis step 2? Will this be based on propensity scores for example?
14. Objective 3 – statistical analysis: which statistical package in which program will be used to apply the causal mediation analysis and how will the underlying assumptions be tested? What will be done if these associations are violated? And will the mediator be measured continuously or categorically?



15. For all exposure and outcome variables, please provide information, how and when the data is measured, and how it will be coded (e.g. continuously or categorically, and if the latter in which levels)
16. Will the code be made available on GitHub after the study is finished?
17. The authors provide a wide list of outcomes. Other important outcomes may be: other neurodevelopmental disorders such as ADHD, or birth complications (pre term birth, caesarean delivery) or offspring inflammatory outcomes such as neonatal icterus, childhood infections, allergy, atopic dermatitis or asthma, but also cardiovascular development, through e.g. obesity or cardiovascular congenital disorders such as Marfan. I wonder if the authors will be powered enough to find associations with membranous nephropathy, MND and autoimmune hepatitis given that these diseases are rare. It would be informative to conduct a power analysis.

**Is the rationale for, and objectives of, the study clearly described?**

Yes

**Is the study design appropriate for the research question?**

Yes

**Are sufficient details of the methods provided to allow replication by others?**

Yes

**Are the datasets clearly presented in a useable and accessible format?**

Not applicable

**Competing Interests:** No competing interests were disclosed.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 29 Sep 2024

**Enny Paixao Cruz**

In this study by Paixao et al, the authors aim to explore the role of prenatal exposure to infectious diseases and the onset of chronic conditions in children and adolescents. I believe this is an important topic and that this is a comprehensive study. The protocol is well written, provides many details and I compliment the authors for their pre-registration. I have several comments and questions:

1. Please provide more information on the exposure definition:
  - Please provide details on the clinical definition
  - Please provide details laboratory definition (e.g., which assay)
  - Please provide information how disease severity will be coded – as I see that treatment is mentioned separately
    - Thank you very much we have created a table including information on clinical definition (including severity), treatment and laboratory

- information used by the Brazilian Ministry of Health, and the definition used in our study.
- ○ Please provide information what sensitivity analysis will be conducted and what the rationale is
    - We will conduct multiple sensitivity analyses, tailored to the specific exposure and outcome of interest. For instance, regarding arbovirus exposure, we plan to analyze only laboratory-confirmed cases, while for syphilis, analyses can be stratified by VDRL titers. Additionally, we are considering the inclusion of historical control groups for arboviruses introduced later in the country, such as Zika and Chikungunya. This study protocol outlines the general features; however, the analysis plan for each individual study will be developed separately and published on the CIDACS website, under the "SEEDS" section, including all relevant details specific to each exposure and outcome. We have made this clear in the protocol.
  - ○ Please provide information how the effect of timing of infection will be measured (e.g. in individual models, G-computation models or mutually adjusted models? And how often is it measured, e.g. once per trimester?)
    - The data are derived from administrative databases, and in this context, the time of infection will be determined by the recorded date of symptom onset for arboviruses. The timing of infection will be treated as a time-varying variable. The selection of the model will depend on the estimation of interest. Potential models include, but are not limited to, the Cox proportional hazards model with time-varying covariates, sequential trials, marginal structural models, and G-computation. Landmark models may also be considered where appropriate.
    - For syphilis as there is universal screening and the date of symptoms onset may not be available (it might not have symptoms at all), we will take different approach and timing of infection will not be considered.
  - 1. ○ It may be interesting to also conduct sensitivity analysis for those exposed to multiple viruses to examine if there is a dose-response relationship between number of viruses prenatally exposed and offspring outcomes.
    - This is a nice idea, but it is unlikely that a woman would acquire multiple arboviruses during a single gestational period. However, we are considering including both prenatal and early-life exposures to assess whether a synergistic effect may increase the risk of these conditions in the offspring.
  - 2. Will the outcomes be measured individually, or will a composite score be made? Is there also a way to get information from the GP or subclinical traits of e.g. autism?
    - The outcome will be measured using ICD-10 codes for the cause of hospital admission and deaths. Unfortunately, we do not have GP information, so only more severe cases will be included in the outcome definition.
  - 3. Will clusters be made for the outcomes based on growth models and will these be placed within the regression models? If there is repeated measure data for the outcome, it may be worth considering performing mixed-effects models that allows to adjust for intra-individual confounding variables.
    - No clusters will be formed. The outcomes evaluated will be repeated

measurements of height and weight values standardized based on WHO references. Standardized measurements will be classified into nutritional indicators, such as severe underweight, underweight, adequate, or overweight, among others. The Broken-stick model is recommended for assessing irregular individual trajectories and standardized z-score data, providing easily interpretable estimates of childhood growth trajectories (Anderson, Hafen et al. 2019). However, the Broken-stick model does not allow for the inclusion of covariates, meaning factors that may influence growth trajectories were not considered. Given these limitations, we welcome the reviewer's suggestion to model growth trajectories using mixed-effects models, which allow for adjustment for confounding variables. We have now added to the protocol.

4. How will the authors deal with missing data in the exposure, covariates and outcome? Will they for example impute part of the missing data, and if so how?

- Exposure and outcome variables are assessed via record linkage. As a result, the absence of linkage is interpreted as the absence of exposure or outcome, meaning there will be no missing data for these variables. For covariates, the mechanisms behind missing data will be carefully considered in each analysis. If the data is believed to be missing completely at random (MCAR), or if the missingness in the exposure or covariates is not driven by the outcome, complete-case analysis may be appropriate. In other scenarios, depending on the analysis, methods such as multiple imputation using chained equations (MICE), pattern mixture models, or growth trajectory models will be considered. The choice of methods for handling missing data, along with any additional techniques such as bootstrapping for calculating 95% confidence intervals, will need to be weighed against the computational demands of working with large datasets. Please provide the statistical packages and program the analyses will be conducted.
- Potential packages that we are planning to use include, but are not limited to, Mice, tableone, survival, data.table, coxme, lme4, tidyverse, ggplot, dagitty and MatchIt.

5. I compliment the authors for adjusting for multiple testing. Please provide the exact number of tests across the aims for which FDR will be applied.

- The exact number of tests that will be conducted is not yet known. For each exposure we will compare the different subgroups considered of clinical relevance. For example, for the effect of syphilis exposure on epilepsy, we will compare congenital infection versus those not exposed, maternal (but no congenital) infection versus those not exposed and any infection exposure versus those not exposed. A total of three exposure comparisons. For the example of the impact of syphilis exposure on epilepsy. The analysis confidence intervals will be adjusted to and FDR will account for the three exposure hypotheses we are evaluating.

6. I compliment the authors for aiming to create a DAG. Would it be possible to add this to the analysis plan with the exact confounders that will be used?

- Yes, we will include an individual DAG for each plan of analyses with the exact confounders for that relationship, the ones that we are able to control for and even the unmeasured ones. These materials will be produced later during the

study's development. In this study protocol, we are giving the readers an overview of the general study. Each individual plan of analyses including DAGs will be publicized in due course in the CIDACS website, under the "SEEDS" section.

7. To adjust for selection bias to some extent, did the authors consider inverse probability weighting?
  - Yes, we will consider the IPW as a method to adjust for confounding and also selection bias. We will include this in the study protocol.
8. Please provide a power analysis based on the given sample size, what the smallest possible effect is the authors are powered to find.
  - We deliberately chose not to include a sample size calculation because we agree with the argument presented by Hernan (2022) that the importance of the research question justifies proceeding with imprecise estimate, but rather encouraging the conduct of many observational analyses. Since this study involves an observational analysis of preexisting data, concerns about underpowering should not preclude us from conducting the analysis. This approach allows us to study rare outcomes using valuable, already available data. If the study is ultimately underpowered, we will acknowledge this limitation in the discussion section of the paper.
9. Please provide a reference for the cut-offs for the weight-for-height Z-score. Did the authors consider calculating a SDS for BMI? And is this based on the WHO reference curves or another curve?
  - Standard deviations can be computed, but standardized values will be based on WHO reference. Children and adolescents will be classified as wasted (weight-for-height Z-score (WHZ) < -2), stunted (height-for-age Z-score < -2), overweight or obese (WHZ > 2) (de Onis, Onyango et al. 2007). BMI (kg/m<sup>2</sup>) will also be calculated. We derived age- and sex-standardized z-scores for BMI using WHO growth reference for under 5 years and 5–10 years. For children under 5 years, overweight and obesity was defined as weight for height z-scores above two and three standard deviations respectively. The respective cut-offs for children aged 5–10 years were one and two standard deviations.
1. What will be the cut-off for the BIC to select the number of trajectory groups for outcome definition part 2? And what will the authors do if the model fit remains poor? And how will the authors ensure (clinical) interpretability of the number of groups based on the growth curves?
  - The BIC (as well as the AIC) are model comparison measures. They do not indicate whether a model is well-fitted; they simply identify the best model among those evaluated. The general approach to assessing model fit is through residual analysis. However, certain characteristics of the evaluated data (large dataset, varied number of measures, and highly irregular measurement times) complicate this analysis. So we are planning to analyse the residual in some random samples of the data and as an alternative approach to assess the quality of the fit using the training and testing method.
  - If the model is poorly fitted, adjustments can be made by modifying the variance component, incorporating a spatial component, or exploring alternative models. The mixed effects model does not stratify the population

into distinct groups; instead, it estimates the trajectory of growth and weight gain for specific groups, such as mothers with and without particular infections during pregnancy.

2. For the statistical analysis part 2, the within sibling group analyses, I tend to disagree with the authors that only sex and birth are the factors differential within families. Different siblings may for example have a different gestational age at birth, mode of delivery or birth complications which are important confounders in this specific study.
  - You are completely right, we have added this to the manuscript now.
3. How will the matching take place in statistical analysis step 2? Will this be based on propensity scores for example?
  - Given the large size of our dataset, we sometimes use exact matching. However, when a high percentage of individuals remain unmatched, we resort to propensity score matching as an alternative.
4. Objective 3 – statistical analysis: which statistical package in which program will be used to apply the causal mediation analysis and how will the underlying assumptions be tested? What will be done if these assumptions are violated? And will the mediator be measured continuously or categorically?
  - Analyses will be conducted in R, utilizing packages such as "mediation" and "medflex." However, these packages do not fully support all types of categorical variables for the exposure, mediator, and outcome. In such cases, custom analysis can still be implemented in R using base functions to manually code the necessary procedures.
  - Mediation analysis relies on strong underlying assumptions about the available information and the relationships between variables. While these assumptions cannot be directly tested, sensitivity analyses can be performed to assess the potential impact of assumption violations.
  - The theoretical model, which serves as the basis for mediation analysis, will be developed in consultation with subject matter experts. Additionally, sensitivity analyses will be used to evaluate the impact of breaking assumptions.
  - A mediator variable, such as gestational age at birth, can be measured as a continuous variable and alternatively categorized to represent preterm birth (yes/no). We will analyze the variable in both forms to assess how these different approaches impact our results.
5. For all exposure and outcome variables, please provide information, how and when the data is measured, and how it will be coded (e.g. continuously or categorically, and if the latter in which levels).
  - The exposures and outcomes will be categorical variables assessed through record linkage. Live births from the CIDACS Birth Cohort, linked with SINAN data from conception to birth, will be classified as exposed to infectious diseases during pregnancy. Dengue exposure can be further categorized into mild dengue, dengue with warning signs, and severe dengue, with all three arboviruses classified based on laboratory testing. Live births exposed to syphilis and Zika can also be further classified as cases of congenital infection. For more details, refer to Table 2, newly added to this protocol. Live births linked to death or hospital admission records containing a relevant ICD-10 code for the specified outcome will be classified as "yes" for the outcome of

interest.

6. Will the code be made available on GitHub after the study is finished?
  - Yes, it will be published on the CIDACS website, under the "SEEDS" section.
7. The authors provide a wide list of outcomes. Other important outcomes may be: other neurodevelopmental disorders such as ADHD, or birth complications (pre term birth, caesarean delivery) or offspring inflammatory outcomes such as neonatal icterus, childhood infections, allergy, atopic dermatitis or asthma, but also cardiovascular development, through e.g. obesity or cardiovascular congenital disorders such as Marfan. I wonder if the authors will be powered enough to find associations with membranous nephropathy, MND and autoimmune hepatitis given that these diseases are rare. It would be informative to conduct a power analysis.
  - We fully agree and plan to expand the list of outcomes and exposures investigated by the SEEDS project. This was discussed with our collaborators, and the current list is only preliminary. Other disorders, such as cardiovascular conditions, are of significant interest and will be included in future analyses.

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 05 August 2024

<https://doi.org/10.21956/wellcomeopenres.24711.r87456>

© 2024 Murray A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Aja Louise Murray**

The University of Edinburgh, Edinburgh, Scotland, UK

The paper addresses an important topic with significant public health implications. Whilst there were some minor language errors/typos and it could benefit from further proof reading, it was also very well-written. The approach of triangulating across different methodologies was a strength. Some suggestions for improvement:

- The arguments for focusing on the specific exposures are well-made and convincing. Can the authors cite any prevalence information for these exposures to further give a sense of the scale of the issue (and further support that the numbers will be sufficient to support their proposed statistical analyses)
- In the exposure definition it is mentioned that timing of infection will be examined in a sensitivity analysis; however, earlier the effect of timing is mentioned as a research question, therefore, should there not be a primary analysis related to this rather than just a sensitivity analysis? Unless timing refers to something else, e.g., month of year. Either way this could benefit from clarification. Otherwise, the research questions mapped well to the proposed analyses.
- Whilst another publication is referenced, it would be helpful to have more details on how

the linkage accuracy is assessed and what the levels of accuracies are (even a range or indicative examples would be helpful additions)

- For the mediation analyses, have the authors considered using causal mediation approaches e.g.: Imai, Kosuke et al. (2010<sup>1</sup>)

### References

1. Imai K, Keele L, Tingley D: A general approach to causal mediation analysis. *Psychol Methods*. 2010; **15** (4): 309-34 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the rationale for, and objectives of, the study clearly described?**

Yes

**Is the study design appropriate for the research question?**

Yes

**Are sufficient details of the methods provided to allow replication by others?**

Yes

**Are the datasets clearly presented in a useable and accessible format?**

Not applicable

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Child development

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---