



## **Data Science Approaches to Improve Maternal and Child Health in Brazil**

### **1. Preamble**

This joint call for proposals targeted specifically to Brazilian researchers is the result of a partnership between the Ministry of Health (MoH), the National Council for Scientific and Technological Development (CNPq), the National Council for State Funding Agencies (CONFAP), engaged State Funding Agency (FAPs) and The Bill & Melinda Gates Foundation (BMGF). This call is part of another initiative funded by the Gates Foundation in 2010 named Healthy Birth, Growth and Development Knowledge integration (*HBGDki*). The goal of this program is to use data science tools to develop a deep understanding of the risk factors contributing to poor outcomes in preterm birth, physical growth faltering and impaired neurocognitive development. Through Grand Challenges Explorations - Brazil, the above-mentioned partners share this goal and wish to build on the growing expertise in data science, epidemiology and public health in Brazil to address priority issues in maternal and child health.

### **2. THE CHALLENGE**

There remain key knowledge gaps in understanding how nutrition, prenatal and antenatal care, maternal support, and environmental and social factors contribute to an elevated risk of poor maternal and childhood health outcomes. Such an understanding is required to determine what interventions, including health policies, should be delivered to which group of individuals at what point in their lifecycle to ensure optimal outcomes.

### **3. THE OPPORTUNITY**

Developing and validating approaches to foster maternal and child health is challenging due to the complex interaction of biological, environmental and social factors. Furthermore, policy recommendations for such approaches frequently lack sufficient supporting scientific evidence, while clinical trials are expensive, time-consuming, and increasingly difficult to implement. There is now a key opportunity to accelerate research in this area by analyzing administrative and clinical outcomes databases in Brazil to formulate public health recommendations that are data-driven and cost effective.

The purpose of this call for proposals is to promote new and novel approaches to analyzing data related to social programs and public health in Brazil to produce



novel insights which can be used to improve maternal and child health in Brazil and around the world. Applicants may choose to work with large datasets available to them or to collaborate with Center for Data and Knowledge Integration for Health (CIDACS) to explore their linked anonymized dataset (100 million Brazilian Cohort), which integrates information from the Cadastro Unico (CADU) with the live births system (SINASC), the mortality system (SIM) and Bolsa Família program (BFP) - see Appendix A for more information about data access through CIDACS.

Successful applicants to this call should assume that with the appropriate agreements and ethical approvals, they will not only receive funding but will be able to access anonymized subsets of these data available from Center for Data and Knowledge Integration for Health (CIDACS) researchers. The utilization of the health and social data, as a key focus of this call, is intended to engage a broad spectrum of collaborators - including research and clinical scientists working with data scientists, bioinformaticians, statisticians, epidemiologists, engineers and computer programmers - to identify how innovative data analytic approaches can be used to develop improved solutions to tackle the burden of maternal and child health problems in Brazil.

We also welcome applicants who have access to other relevant data sets, including publicly available data, clinical research, cohort and survey studies and other large datasets that can help address the questions below, to submit their proposals under this call. In this case, researchers who submit applications involving other data sets are responsible for securing appropriate authorization and access to the data. Applicants are encouraged to describe the steps required to access the proposed dataset in their proposals.

### **3.1. WHAT WE ARE LOOKING FOR**

We seek proposals designed to answer critical scientific questions related to maternal and child health and development outcomes. Proposals should use innovative data analytics and modeling approaches that can be applied to the CIDACS linked datasets or to other relevant data sets that applicants can access. Proposals should be based on these linked datasets or existing primary data in Brazil and yield actionable results with a potential to significantly impact public health policy.

#### **3.1.1 We will give highest priority to proposals that:**



- a) Support innovative collaborations between Brazilian research scientists, healthcare experts, and data scientists;
- b) Answer critical scientific questions identified in this GCE, while building and strengthening data science capacity for Brazil;
- c) Take into account social, environmental and cultural determinants of outcomes and incorporate an understanding of the target community that includes barriers and constraints to delivery of interventions and to implementation of public health programs;
- d) Contribute to a portfolio of funded projects that addresses regional diversity and the need to provide health equity for diverse and vulnerable populations;
- e) Explain how answers will have the highest likelihood of being relevant for implementation broadly in the public health system;
- f) Describe mechanistic models for establishing the relationship between interventions and its related outcomes.

**3.1.2. Examples of what we are looking for include analytical approaches:**

- a) That apply innovative analyses or machine learning techniques to identify patterns in data from “natural experiments” (e.g., the impact of economic cycles on the quality of primary care and health outcomes);
- b) That stratify risk of adverse pregnancy outcomes, including preterm birth and low birth weight;
- c) That incorporate weight gain during pregnancy as a variable, including helping to determine the relative contributions to neonatal health outcomes of maternal diet quantity versus quality;
- d) That determine the relative contributions to infant health outcomes of diet quantity versus quality (e.g., protein quantity versus quality);
- e) That target underexplored subsets of data (e.g., rare but significant HBGD events that can be studied because of the large size and statistical power of the database);
- f) That help convert correlations to causal hypotheses (e.g., health outcomes correlated to sex differences, maternal education, birth spacing, age of first pregnancy, establishing causal impact of air pollution on fetal growth);
- g) That identify new ways to aggregate risk factors and identify vulnerable populations for adverse maternal and child health outcomes, including innovative data integration strategies and visualization tools;



- h) That specifically incorporate the roles of women – as perceived locally – from adolescence to motherhood (including pregnancy during adolescence);
- i) That evaluate programs for pre-pregnancy intervention for women and the effect of doing so on prenatal, maternal, fetal and neonatal mortality;
- j) That determine the best care for low-birth-weight babies;
- k) That help determine the window of opportunity to foster catch-up growth for preterm and low-birth-weight babies, and the most effective interventions for doing so;
- l) That help identify critical periods for intervention during pregnancy and early childhood;
- m) That stratify risk of stunting and wasting from birth through two years of age;
- n) That target root causes of maternal mortality, including caesarean section, and address the most vulnerable population groups considering age and ethnicity;
- o) That investigate the “double burden” of disease in Brazil leading to pockets of stunting and wasting in parallel with pockets with rising rates of childhood obesity.
- p) That stratify risks for child development aiming to establish national indicators for healthy development from the neonatal period to the first two years of the children addressing preferably most vulnerable population groups considering age and ethnicity
- q) That help to understand the relationship between social indicators, nutritional conditions and mortality from the prenatal period to the early childhood. We welcome applications addressing traditional and vulnerable populations.

**3.1.3. Examples of what we are NOT looking for:**

- a) Proposals submitted by applicants from outside of Brazil.
- b) Proposals for new studies to generate new data;
- c) Proposals that do not focus on health outcomes in Brazil;
- d) Proposals not related to maternal and child health challenges
- e) Development of new primary data collection tools;
- f) Approaches that do not meaningfully involve data from adolescents, mothers, or infants;



- g) Applications proposing data science algorithm development without clear relevance to answering the types of questions described in this call for proposals;
- h) Ideas without a clearly articulated and testable hypothesis together with metrics of success;
- i) Ideas for which the described indicator of success cannot be demonstrated or significantly advanced within the scope of the GCE Phase 1 award (USD \$100,000 over 18 months);
- j) Proposals that do not describe the innovation's potential effects on health policy making;
- k) Analyses that are only slight improvements over existing approaches (e.g., replication of an approach in a new geography in the absence of added innovation);

## APPENDIX

### Center for Data and Knowledge Integration for Health

The Center for Data and Knowledge Integration for Health (CIDACS) was launched in 2016 as part of Fiocruz in Bahia to conduct innovative scientific research that broadens understanding of the effects of social and environmental determinants, including public policy, on population health in Brazil. To achieve this goal, it links information from social programs and other databases of health information systems, such as mortality, birth, communicable and non-communicable infectious diseases, among other outcomes, from different information systems maintained by the Ministry of Health. A central feature of the CIDACS platform is to produce linked datasets, whose main product is the 100 million Brazilians cohort (2006-2015). This cohort is built by linking data from individuals in the Cadastro Único (CADU) to data of the same individuals in the Bolsa Familia Program (BFP), SINASC and SIM. CADU is a national registry of individuals who could be eligible to federal government social programs. BFP is a national conditional cash transfer program which transfers cash to poor households when they comply with conditions related to health and education. Individuals in the BFP database must be registered in CADU. SINASC is the registry of birth certificates. SIM is the registry of death certificates.

Previous analyses performed by the CIDACS team have demonstrated that the social protection program contributed to a decrease in childhood mortality overall and in particular for deaths attributable to poverty-related causes such as malnutrition and diarrhea.



The CIDACS platform employs the highest levels of security to ensure the privacy, integrity, and ethical use of the data. Considerable research has been required to develop the processes for data receipt, curation, and to provide authorized access for analysis. In this Grand Challenges Explorations – Brazil program, the CIDACS platform will be available for the selected applicants to conduct innovative, interdisciplinary projects in the areas of maternal and child health.

**Additional Resources**

[HBGDKI.org](http://HBGDKI.org)

[Center for Data and Knowledge Integration for Health \(CIDACS\)](#)

Other resources to follow in the coming days