

cidacs
Centro de Integração de Dados
e Conhecimentos para Saúde



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

Instituto Gonçalo Moniz

The Center for Data and Knowledge Integration for Health (CIDACS)

CIDACS Birth Cohort

SAIL User Forum, 17th March 2022

Bethânia Almeida on behalf of the Cidacs team

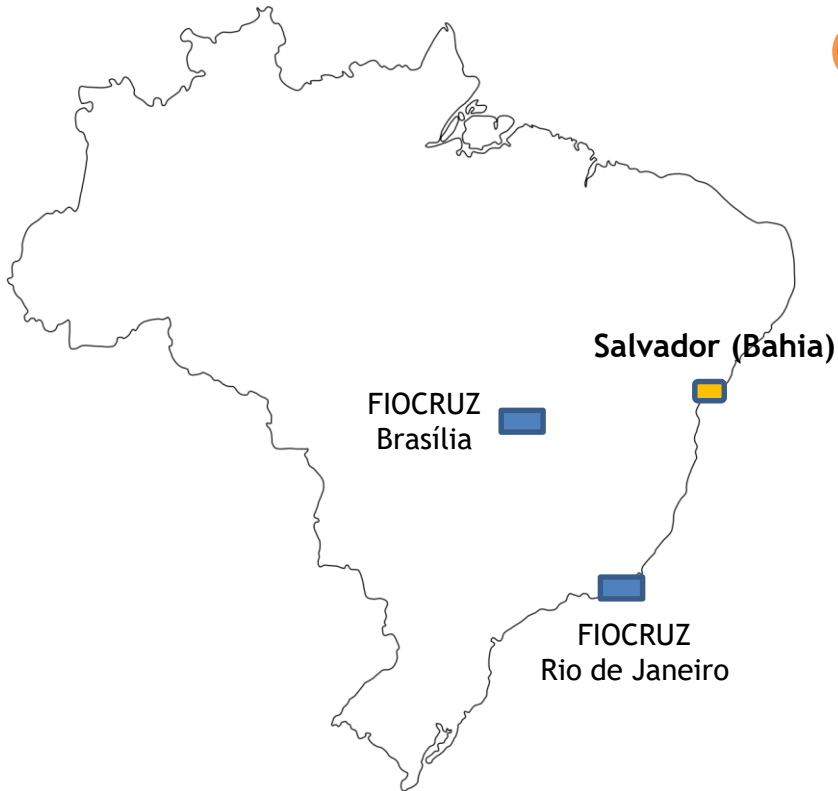
Context



Cidacs occupies 700 m² in this Bahia Technology Park building

- CIDACS was created in December 2016 in the city of Salvador (Bahia-Brazil) with the objective of conducting interdisciplinary research on population health determinants using integrated Brazilian (national) datasets to generate scientific knowledge and provide evidence to support public policymaking.
- In Brazil, very few studies have employed data linked on an individual level. The CIDACS data centre was initially constructed to house the 100 Million Brazilians Cohort.

CIDACS - The Center for Data and Knowledge Integration for Health



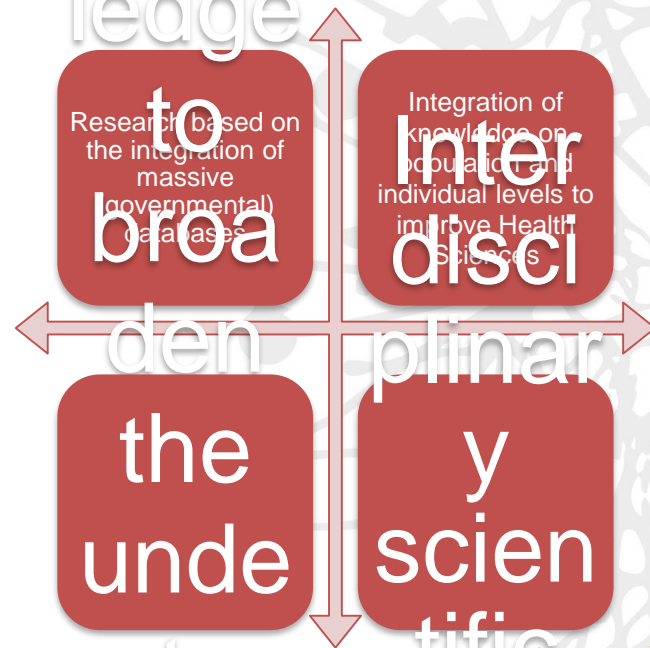
<https://cidacs.bahia.fiocruz.br/>



Managed by FIOCRUZ Bahia



Focus



CIDACS Profile

International Journal of Population Data Science (2019) 4:2:04

International Journal of Population Data Science

Journal Website: www.ijpds.org



The Centre for Data and Knowledge Integration for Health (CIDACS): Linking Health and Social Data in Brazil

Barreto, ML^{1,2*}, Ichihara, MY^{1,2}, Almeida, BA¹, Barreto, ME^{1,3}, Cabral, L¹, Fiaccone, RL^{1,4}, Carreiro, RP¹, Teles, CAS¹, Pitta, R¹, Penna, GO^{1,5,6}, Barral-Netto, M¹, Ali, MS^{1,7,8}, Barbosa, G¹, Denaxas, S⁹, Rodrigues, LC^{1,8}, and Smeeth, L^{1,8}

Submission History	
Submitted:	31/08/2019
Accepted:	21/09/2019
Published:	20/11/2019

¹Centre for Data and Knowledge Integration for Health (CIDACS), Gonçalo Moniz Institute, Oswaldo Cruz Foundation (FIOCRUZ), Salvador, Brazil.

²Institute of Collective Health, Federal University of Bahia (UFBA), Salvador, Brazil.

³Computer Science Department, Federal University of Bahia (UFBA), Salvador, Brazil.

⁴Statistics Department, Federal University of Bahia (UFBA), Brazil.

⁵Tropical Medicine Centre, University of Brasilia (UnB), Brazil.

⁶Escola Fiocruz de Governo, FIOCRUZ Brasilia, Brazil.

⁷Center for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK.

⁸Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, United Kingdom.

⁹Institute of Health Informatics, University College London, United Kingdom.

Abstract

The Centre for Data and Knowledge Integration for Health (CIDACS) was created in 2016 in Salvador, Bahia-Brazil with the objective of integrating data and knowledge aiming to answer scientific questions related to the health of the Brazilian population. This article details our experiences in the establishment and operations of CIDACS, as well as efforts made to obtain high-quality linked data while adhering to security, ethical use and privacy issues. Every effort has been made to conduct operations while implementing appropriate structures, procedures, processes and controls over the original and integrated databases in order to provide adequate datasets to answer relevant research questions. Looking forward, CIDACS is expected to be an important resource for researchers and policymakers interested in enhancing the evidence base pertaining to different aspects of health, in particular when investigating, from a nation-wide perspective, the role of social determinants of health and the effects of social and environmental policies on different health outcomes.

IJPDS International Journal of Population Data Science

[IJPDS](#) [News](#) [Issues](#) [About](#) [Conferences & Events](#) [Webinars & Videos](#) [Contact](#)

[Register](#) [Login](#)

[Home](#) / [Archives](#) / Vol. 4 No. 2 (2019): IJPDS Special Issue: Data Centre Profiles

Vol. 4 No. 2 (2019): IJPDS Special Issue: Data Centre Profiles

Published: November 20, 2019

IJPDS Special Issue:
Data Centre Profiles

Issue 2, Vol 4

IJPDS is delighted to present this special issue of population data centre profiles. A population data centre is defined in this context as an information environment providing access to linked datasets at person-level from multiple routine data sources, with explicit approaches to how this linkage is done, and how data are managed and safeguarded for the purposes of service-delivery, research or both.

With articles from population data centres across the world, this special issue represents a unique resource, bringing together into one place a wealth of information. Modes of operation vary between centres, but in general, the information includes: rationales for the establishment of the centres; population settings; architecture and IT, relevant legislation and governance; data linkage models; data access models; a description of the datasets that can be made available; and some noteworthy outputs.

We hope readers will enjoy this special issue and see it as a valuable addition to the literature base. The special issue will remain open to new submissions so that further articles can be added in the future to continue to build this resource.



Ministerio da Saúde

FIOCRUZ
Fundação Oswaldo Cruz
Instituto Gonçalo Moniz

CIDACS Profile

- Administrative social and health data is at the core of the activities conducted by CIDACS.
- Upon authorization, administrative data generated by government agencies are received and used to create specific cohorts, which may later support research initiatives in a variety of areas related to public health.

100 Million Brazilian Cohort



International Journal of Epidemiology, 2021, 1–12

<https://doi.org/10.1093/ije/dyab213>

Cohort Profile



The 100 Million Brazilian Cohort (N=131,697,800) was initially conceived to investigate the impact of social protection policies on social determinants of health in low-income populations throughout Brazil.

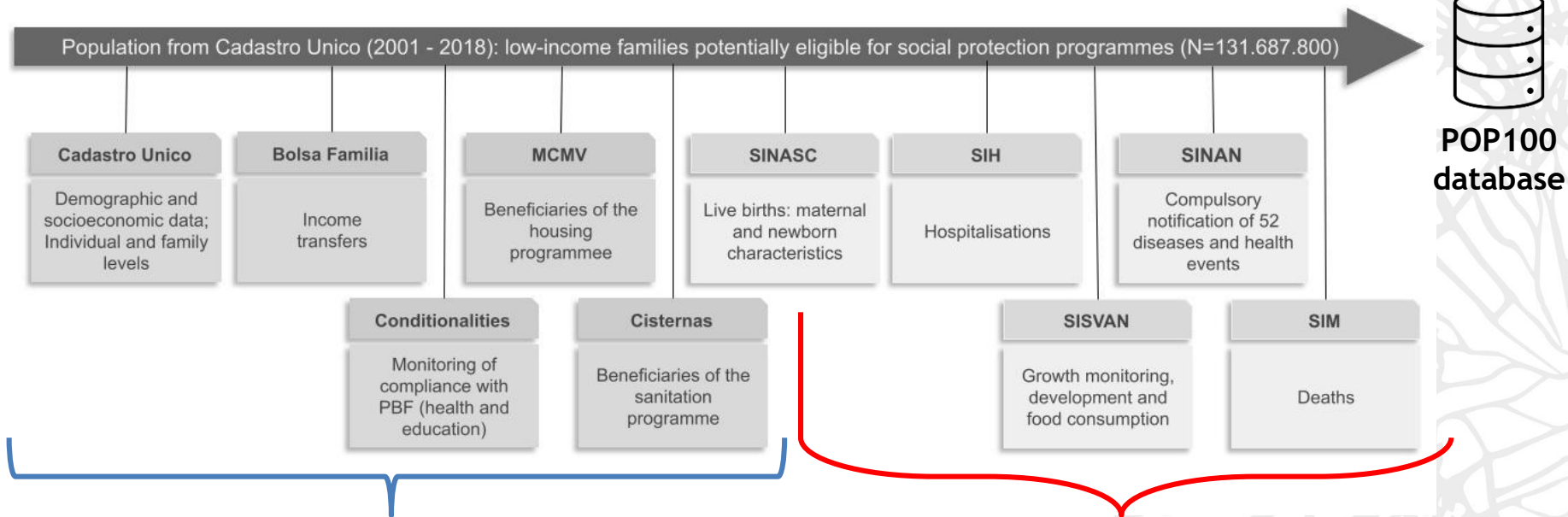
Cohort Profile

Cohort profile: The 100 Million Brazilian Cohort

The 100 Million Brazilian Cohort

- The cohort consists of individuals who have applied for social welfare assistance since 2001, and are thusly registered in a federal government database called CADUNICO.
- Individuals are eligible and registered if they receive an **income** of up to half the Brazilian minimum monthly wage (approximately USD125 in 2020) or their **total family income** does not exceed the equivalent of 3 minimum wages (approximately USD750 in 2020).
- Applicants answer a detailed form that collects demographic, economic, and social information on each member of their family, as well as family and household characteristics; this must be updated **every two years**.

The 100 Million Brazilian Cohort



Deterministic linkage (based on NIS - social ID number)

Non-deterministic linkage (name, mother's name, birthdate, gender, municipality)

CIDACS has developed different tools adapted to the type of data found in the social and health administrative databases and large-sized datasets

346

IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 22, NO. 2, MARCH 2018



On the Accuracy and Scalability of Probabilistic Data Linkage Over the Brazilian 114 Million Cohort

Robespierre Pita¹, Clécia Pinto, Samila Sena, Rosemeire Fiaccone, Leila Amorim, Sandra Reis, Mauricio L. Barreto², Spiros Denaxas, and Marcos Ennes Barreto

Abstract—Data linkage refers to the process of identifying and linking records that refer to the same entity across multiple heterogeneous data sources. This method has been widely utilized across scientific domains, including public health where records from clinical, administrative, and other surveillance databases are aggregated and used for research, decision making, and assessment of public policies. When a common set of unique identifiers does not exist across sources, probabilistic linkage approaches are used to link records using a combination of attributes. These methods require a careful choice of comparison attributes as well as similarity metrics and cutoff values to decide if a given pair of records matches or not and for assessing the accuracy of the results. In large, complex datasets, linking and assessing accuracy can be challenging due to the volume and complexity of the data, the absence of a gold standard, and the challenges associated with manually reviewing a very large number of record matches. In this paper, we present Atulmo, a hybrid probabilistic

linkage tool optimized for high accuracy and scalability in massive data sets. We describe the implementation details around anonymization, blocking, deterministic and probabilistic linkage, and accuracy assessment. We present results from linking a large population-based cohort of 114 million individuals in Brazil to public health and administrative databases for research. In controlled and real scenarios, we observed high accuracy of results: 93%–97% true matches. In terms of scalability, we present Atulmo's ability to link the entire cohort in less than nine days using Spark and scaling up to 20 million records in less than 12s over heterogeneous (CPU+GPU) architectures.

Index Terms—Data linkage, accuracy assessment, cohort study.

Barbosa et al. *BMC Med Inform Decis Mak* (2020) 20:289
<https://doi.org/10.1186/s12911-020-01285-w>

BMC Medical Informatics and
Decision Making

RESEARCH ARTICLE

Open Access



CIDACS-RL: a novel indexing search and scoring-based record linkage system for huge datasets with high accuracy and scalability

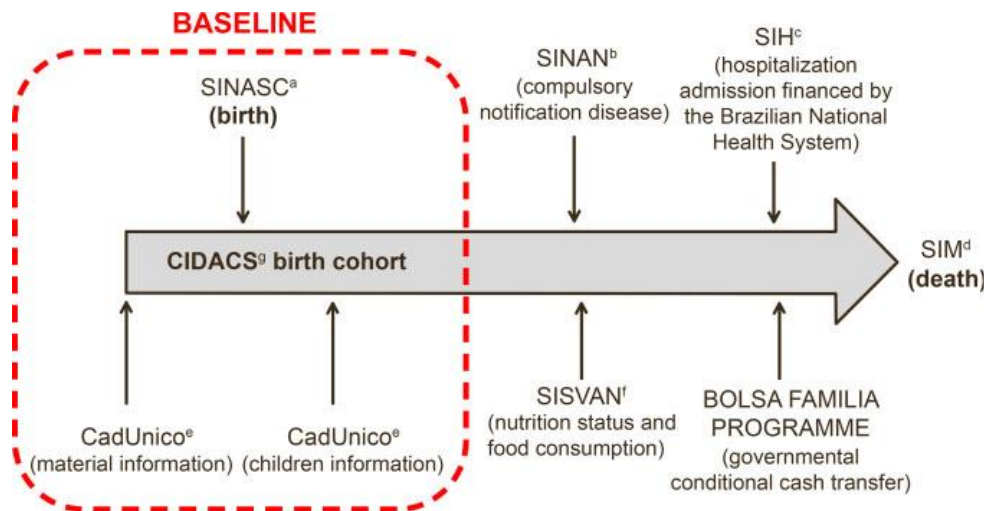
George C. G. Barbosa^{1*}, M. Sanni Ali^{1,2,3}, Bruno Araujo¹, Sandra Reis¹, Samila Sena¹, Maria Y. T. Ichihara¹, Julia Pescarini¹, Rosemeire L. Fiaccone^{1,4}, Leila D. Amorim^{1,4}, Robespierre Pita¹, Marcos E. Barreto^{1,6,7}, Liam Smeeth² and Mauricio L. Barreto^{1,5}

CIDACS Birth Cohort

- Around 3 million children are born every year in Brazil.
- The overall objective of the cohort is to research the effects of obstetric and prenatal conditions, congenital infections and other potential social and environmental determinants, as well as the impact of social policies on birth, growth, morbidity and survival, both in the overall cohort and in subgroups of interest.
- Information on each child's mother and her obstetric history [whether she previously had a stillbirth or miscarriage, a previous caesarean section or vaginal delivery], her pregnancy (length of gestation, type of delivery, fetal presentation), the newborn (birthweight, presence of congenital anomalies) and prenatal care (number of visits and when care started).

CIDACS Birth Cohort

24,695,617 live births
(1-Jan-2001 and 31-Dec-2015)
(\cong 55% of all births registered by SINASC)
were linked with the baseline of the 100 Million Brazilian
Cohort using common maternal information.



Cohort Profile

Cohort Profile: Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS) Birth Cohort

^aSINASC (Sistema de Informação sobre Nascidos Vivos / Live Birth Information System)

^bSINAN (Sistema de Informação sobre Agravos de Notificação / Information System for Notifiable Diseases)

^cSIH (Sistema de Informações Hospitalares / Hospitalization Information System)

^dSIM (Sistema de Informação sobre Mortalidade / Information System of Mortality)

^eCADUnico (Cadastro Único para Programas Sociais / Unified Register for Social Programmes)

^fSISVAN (Sistema de Vigilância Alimentar e Nutricional/ Food and Nutrition Surveillance System)

^gCIDACS (Centre for Data and Knowledge Integration for Health)

Sample of studies on newborn, child and maternal health based on CICACS cohorts

RESEARCH ARTICLE

Associations between cesarean delivery and child mortality: A national record linkage longitudinal study of 17.8 million births in Brazil

Enny S. Paixao^{1,2,*}, Christian Bottomley³, Julia M. Pescarini^{1,2}, Kerry L. M. Wong⁴, Luciana L. Cardim⁵, Rita de Cássia Ribeiro Silva^{2,3}, Elizabeth B. Brickley⁶, Laura C. Rodrigues², Flávia José Oliveira Alves^{2,3}, Maria do Carmo Leal⁴, Maria da Conceição N. Costa^{2,3}, Maria Gloria Teixeira^{2,3}, Maria Yury Ichihara^{2,3}, Liam Smeeth⁷, Maurício L. Barreto^{2,3}, Oona M. R. Campbell⁸

1 Infectious Disease Department, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom, **2** Center for Data and Knowledge Integration for Health, Oswaldo Cruz Foundation, Salvador, Brazil, **3** Nutrition School, Federal University of Bahia, Salvador, Brazil, **4** Sergio Araujo National School of Public Health, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil, **5** Collective Health Institute, Federal University of Bahia, Salvador, Brazil

* enny.paixao@lshtm.ac.uk

BMC Medicine

Rasella et al. BMC Medicine (2021) 19:127
<https://doi.org/10.1186/s12916-021-01994-7>

RESEARCH ARTICLE

Open Access

Long-term impact of a conditional cash transfer programme on maternal mortality: a nationwide analysis of Brazilian longitudinal data

Daivide Rasella^{1†}, Flávia José Oliveira Alves^{2,3†}, Poliana Rebouças^{2,3}, Gabriela Santos de Jesus⁴, Maurício L. Barreto^{2,3}, Tereza Campello^{1,5} and Enny S. Paixao^{1,6}

† School of Nutrition, Federal University of

The Lancet Regional Health - Americas 3 (2021) 10045



The Lancet Regional Health - Americas

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

Research paper

Risk of mortality for small newborns in Brazil, 2011–2018: A national birth cohort study of 17.6 million records from routine register-based linked data

Enny S. Paixao, Ph.D.^{1,2,*}, Hannah Blencowe, MD, Ph.D.², Ilia Rocha Falcao, Ph.D.^{1,3}, Eric O. Ohuma, Ph.D.², Aline dos Santos Rocha^{1,3}, Flávia José Oliveira Alves^{1,4}, Maria da Conceição N. Costa, MD, Ph.D.^{1,4}, Lorena Suárez-Idueta⁵, Naiá Ortelan, Ph.D.¹, Liam Smeeth, MD, Ph.D.², Laura C. Rodrigues, MD, Ph.D.^{1,2}, Joy E. Lawn, MB BS, Ph.D.², Marcia Furquim de Almeida, MD, Ph.D.⁶, Maria Yury Ichihara, MD, Ph.D.⁶, Rita de Cássia Ribeiro Silva, Ph.D.^{1,3}, Maria Gloria Teixeira, MD, Ph.D.^{1,4}, Maurício L. Barreto, MD, Ph.D.^{2,3}

1 Center for Data and Knowledge Integration for Health (CIDACS), Gonçalo Moniz Institute, Oswaldo Cruz Foundation, Salvador, Bahia, Brazil
2 Maternal, Adolescent, Reproductive & Child Health (MARCH) Center, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK
3 Escola de Nutrição, Universidade Federal do Bahia, Salvador, Brazil
4 Instituto de Saúde Coletiva, Federal University of Bahia, Salvador, Bahia, Brazil
5 Mexican Society of Public Health, Herschel 100, Anzures, Miguel Alemán, 11500, Mexico City
6 Faculdade de Saúde Pública, Universidade de São Paulo, Brazil

Received: 31 August 2021 | Revised: 4 November 2021 | Accepted: 11 November 2021
DOI: 10.1002/lrjh.14053

CLINICAL ARTICLE
Obstetrics

Recurrence of preterm births: A population-based linkage with 3.5 million live births from the CIDACS Birth Cohort

Aline S. Rocha^{1,2}, Rita de Cássia Ribeiro-Silva^{1,2}, Enny S. Paixao^{2,3}, Ilia R. Falção^{1,2}, Flávia José. O. Alves^{2,4}, Naiá Ortelan², Marcia F. de Almeida⁵, Rosemeire L. Fiaccone^{2,6}, Laura C. Rodrigues³, Maria Yury Ichihara², Maurício L. Barreto^{2,4}

† School of Nutrition, Federal University of

Falção et al. BMC Pregnancy and Childbirth (2020) 20:336
<https://doi.org/10.1186/s12884-020-03236-x>

BMC Pregnancy and Childbirth

RESEARCH ARTICLE

Open Access

Factors associated with low birth weight at term: a population-based linkage study of the 100 Million Brazilian cohort

Ilia R. Falção^{1,2*}, Rita de Cássia Ribeiro Silva^{1,2}, Marcia F. de Almeida⁵, Rosemeire L. Fiaccone^{2,6}, Aline dos S. Rocha^{1,2}, Naiá Ortelan², Marcia F. de Almeida⁵, Enny S. Paixao^{2,3}, Maria Yury Ichihara^{2,3}, Laura C. Rodrigues^{2,3} and Maurício L. Barreto^{2,6}

Abstract

Background: Factors associated with low birth weight at term (LBWT), a proxy for intrauterine growth restriction (IUGR), are not well elucidated in socioeconomically vulnerable populations. This study aimed to identify the factors associated with LBWT in impoverished Brazilian women.

Methods: Records in the 100 Million Brazilian Cohort database were linked to those in the National System of Information on Live Births (SIMAS) to obtain obstetric, maternal, birth and socioeconomic data between 2001 and 2015. Multivariate logistic regression was performed to investigate associations between variables of exposure and LBWT.

Results: Of 10,768,930 term live births analyzed, 3.7% presented LBWT. The highest odds of LBWT were associated with female newborns (OR: 1.49; 95% CI: 1.47–1.50), whose mothers were black (OR: 1.28; 95% CI: 1.18–1.22), had a low educational level (OR: 1.57; 95% CI: 1.53–1.62), were aged ≥35 years (OR: 1.44; 95% CI: 1.43–1.46), had a low number of prenatal care visits (OR: 2.48; 95% CI: 2.42–2.54) and were pimaquim (OR: 1.62; 95% CI: 1.60–1.64). Lower odds of LBWT were found among infants whose mothers lived in the North, Northeast and Center-West regions of Brazil compared to those in the South.

Conclusion: Multiple aspects were associated with LBWT, highlighting the need to comprehensively examine the mechanisms underlying these factors, especially in more vulnerable Brazilian populations, in order to contribute to the elaboration of health policies and promote better conditions of life for poor and extremely poor mothers and children.

Keywords: Low birth weight, Term birth, Poor populations, Cohort, Linkage

PLOS MEDICINE

RESEARCH ARTICLE

Conditional cash transfer program and child mortality: A cross-sectional analysis nested within the 100 Million Brazilian Cohort

Dandara Ramos^{1,2,3,*}, Nives B. da Silva^{2,4}, Maria Yury Ichihara^{1,2}, Rosemeire L. Fiaccone^{1,2}, Daniele Almeida^{1,4}, Samila Sena¹, Poliana Rebouças^{1,2}, Eizo Pereira Pinto Júnior¹, Enny S. Paixão^{1,2}, Sanni Algrati¹, Laura C. Rodrigues^{1,5}, Maurício L. Barreto¹

1 Center for Data and Knowledge Integration for Health (CIDACS), Fundação Oswaldo Cruz, Salvador, Bahia, Brazil, **2** Institute of Collective Health, Federal University of Bahia, Salvador, Bahia, Brazil, **3** Statistics Department, Institute of Mathematics and Statistics, Federal University of Bahia, Salvador, Bahia, Brazil, **4** Computer Science Department, Institute of Mathematics and Statistics, Federal University of Bahia, Salvador, Bahia, Brazil, **5** Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom

* dandara.ramos@ufba.br

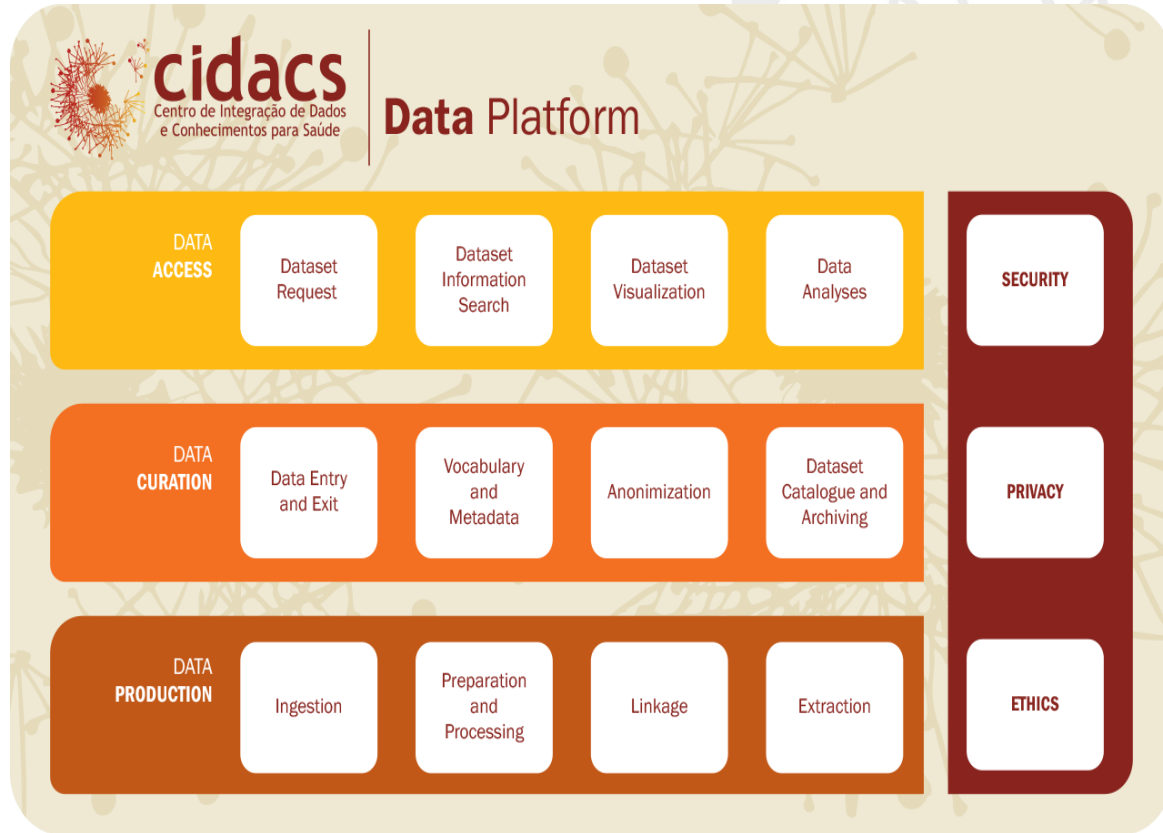


Ministério da Saúde

The centre employs a complex data management system incorporating information security, ethical principles, and privacy protection.

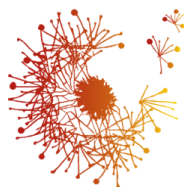
- CIDACS has been developing and consolidating its data centre operational procedures and governance.

- The usage of administrative data containing personal information for research purposes is restricted in many countries, including Brazil. Stringent requirements for access include appropriate security arrangements, exclusive use for a previously specified purpose, appropriate credentials from the requesting researcher/institution and the sound ethical basis of the research in question.

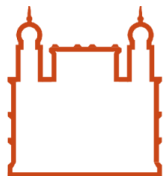


Our data centre's current focus and prospects for collaboration...

- Access, use and reuse of administrative data for public health research purposes
- Acquisition of new databases, including environmental and climate data
- Improving Data management and data governance practices
- Building Trusted Research Environments
- Sustainability of the CIDACS Population Data Center Model
- Scaling of Data Safe Haven (Data Analysis) Computational Environments
- Expanding Computing Infrastructure for Administrative Data Linkage
- Improvements in Robust Security Information and Data Anonymization Procedures
- Promotion of Public Engagement and Data literacy initiatives with different stakeholders



cidacs
Centro de Integração de Dados
e Conhecimentos para Saúde



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz
Instituto Gonçalo Moniz

Obrigada!
Diolch!
Thank you!

bethania.almeida@fiocruz.br
baraujo2010@gmail.com

COLABORADORES CIENTÍFICOS:



APOIADORES:



Parque Tecnológico da Bahia
Rua Mundo, 121, Trobogy
Salvador - Ba, CEP 41745-715



www.cidacs.bahia.fiocruz.br
cidacs@bahia.fiocruz.br

[fiocruzbahia.cidacs](https://www.facebook.com/fiocruzbahia.cidacs)
 [cidacs_fiocruz](https://twitter.com/cidacs_fiocruz)
 [Cidacs Fiocruz](https://www.youtube.com/CidacsFiocruz)

