https://lthj.qut.edu.au/ LAW, TECHNOLOGY AND HUMANS
Volume **8** (1) 2026 https://doi.org/10.5204/lthj.4005

Ethical, Legal and Social Issues Surrounding Knowledge Production in Health Using Linked Data: A Brazilian Experience

Mariana Ramos Pitta Lima and Bethânia de Araújo Almeida
The Centre for Data and Knowledge Integration for Health (CIDACS)/Oswaldo Cruz Foundation, Brazil

Abstract

The use of data derived from Electronic Health Records and Real-World Data is central to epidemiological research, particularly in population health studies. Administrative data—information collected during routine citizen-government interactions or the delivery of services—are digitally structured, falling under the broader concept of 'digital health'. Linking health data with social data holds immense potential for investigating social determinants of health. This article aims to elucidate the process of enabling the transformation of linked health and social data to support scientific research and knowledge production on social determinants of health within an epidemiological context, while focusing on relevant ethical, legal and sociotechnical issues (ELSI). The analysis is reflexive, grounded by a case study of a Brazilian initiative, the Centre for Data and Knowledge Integration for Health (CIDACS), which generates population health knowledge supported by large volumes of linked administrative data. Drawing on Science, Technology and Society (STS) Studies and Critical Data Studies, this article attempts to situate the knowledge produced by CIDACS, recognising that its data production, data infrastructure operations and data usage are intertwined and contextually embedded within sociohistorical and disciplinary frameworks. Our study concludes that concrete experiences of data practices reveal nuanced insights, underscoring the role of the Global South in advancing alternative and critical epistemologies.

Keywords*:* Data studies; linkage of health and social data; digital health; population health data; ELSI; public health.

1. Introduction

> *Our epidemiology has been directed towards constructing theoretical frameworks that integrate health-disease-care processes as components of societal organization. This pursuit is not merely an intellectual endeavour, as it reaffirms historical commitments through the denunciation of and opposition to all forms of social or health inequalities, aiming towards the construction of equity-based health policies.*[1]

The Centre for Data and Knowledge Integration for Health (CIDACS) was established in December 2016 as part of the Oswaldo Cruz Foundation (FIOCRUZ) in the city of Salvador, Bahia-Brazil. Linking social and health administrative data lies at the core of CIDACS's epidemiological studies on social determinants of health.

It is important to note that record linkage techniques require individual-level data, yet the use of secondary data containing personal information for research purposes is restricted in many countries, including Brazil. Consequently, CIDACS's data governance policies and protocols are modelled on organisational structures and procedures that incorporate technical, ethical and legal aspects to support data acquisition, management, and use specifically for public health research purposes. The

---

[1] Almeida Filho, "Panorama, Desafios e Perspectivas para uma Epidemiologia Brasileira," 687-691.

Centre's activities align with principles and guidelines for safe data linkage, as well as the Brazilian General Data Protection Law (*Lei Geral de Proteção de Dados Pessoais*, or LGPD).[2]

Administrative data are originally collected in the context of routine governmental operations and require substantial effort to be 'transformed' into datasets suitable for scientific research. This transformation involves mitigating inaccuracies, accounting for missing data, harmonising datasets and applying appropriate linkage methodologies.[3]

CIDACS has faced numerous sociotechnical challenges in producing scientific knowledge and evidence supported by linked data. To overcome these, in-house solutions have been developed, including record linkage algorithms tailored to Brazilian health and social administrative databases.[4]

Currently, CIDACS is developing additional methodological approaches to quantify linkage errors and mitigate bias, particularly among underrepresented populations, such as groups of ethnic origin and traditional communities dependent on natural resources. Additionally, CIDACS has initiated discussions and actions aimed at addressing issues related to data ethics and public engagement to enhance transparency, promote inclusion, and foster public trust and cooperation in its data practices.[5] Administrative data constitute a complex information source generated for purposes other than research.[6] Upon examining usage in social science research, Connelly[7] argued that despite being categorised as a type of big data, administrative data have been largely overlooked in broader discussions on this topic. Big data encompasses various categories, each presenting distinct opportunities and challenges across fields of knowledge. This necessitates a nuanced understanding of their respective characteristics and specific methodological approaches to ensure robust analysis.[8]

Broadly, the term 'big data' in health science and healthcare has become incorporated into the wider conceptualisation of digital technologies related to health and medicine, known as 'digital health'. Deborah Lupton,[9] a pioneer in the emerging field of 'critical digital health', defines digital health as "a term frequently adopted to encompass a wide range of technologies related to health and medicine [...] from those directed at individuals to those used at the population level."[10] Digital health involves "sociocultural products located within pre-established circuits of discourse and meaning."[11]

Although this definition highlights sociocultural and context-specific dimensions, there remains ample space in the scientific literature to explore empirical experiences illuminating these aspects. Drawing on a Brazilian case study, our analysis contributes empirical insights in an effort to advance the wider understanding of digital health. Despite the increasing prominence of big data, social research examining its implications for medicine and public health remains limited,[12] particularly from the perspective of the Global South.

The scientific literature, from a Social Studies of Science and Technology (STS) viewpoint, emphasises that data are *always situated.*[13] Authors such as Hoyer[14] and Leonelli and Tempini[15] argue that both data and data practices are multiple and contingent. Indeed, the situated nature of knowledge production is among the most well-established tenets in STS.

However, with respect to health data, the literature has primarily developed studies analysing contexts in high-income countries, often categorised as the Global North, such as Denmark and the United Kingdom.[16] The so-called 'Scandinavian model,' known

---

[2] Almeida, "CIDACS' Efforts Towards an Inclusive and Dialogic Data Governance in Brazil," 1-8; Brazil, *"Law on Treatment and Protection of Personal Data."*
[3] Barreto, "The Centre for Data and Knowledge Integration for Health (CIDACS)."
[4] Pita, "On the Accuracy and Scalability of Probabilistic Data Linkage Over the Brazilian 114 Million Cohort," 346-353; Barbosa, "CIDACS-RL." Altymo (Pita et al., 2018) and CIDACS-RL (Barbosa et al., 2020) are open-source linkage algorithms developed by the CIDACS team to overcome data-linkage limitations, making it possible to link records on a specific individual across disparate databases.
[5] Almeida, "CIDACS' Efforts Towards an Inclusive and Dialogic Data Governance in Brazil," 1-8.
[6] Connelly, "The Role of Administrative Data in the Big Data Revolution in Social Science Research," 1-12.
[7] Connelly, "The Role of Administrative Data in the Big Data Revolution in Social Science Research."
[8] Connelly, "The Role of Administrative Data in the Big Data Revolution in Social Science Research," 1-12.
[9] Lupton, Digital Sociology.
[10] Lupton, "Critical Perspectives on Digital Health Technologies," 1345.
[11] Lupton, "Critical Perspectives on Digital Health Technologies," 1349.
[12] Lupton, "Critical Perspectives on Digital Health Technologies."
[13] Kitchin, "Big Data, New Epistemologies and Paradigm Shifts," 5; Valente, "Critical Data Studies with Latin America;" Grohmann, "Latin American Critical Data Studies."
[14] Hoeyer, "Data Paradoxes."
[15] Leonelli and Tempini, "Data Journeys in the Sciences."
[16] Hoeyer, "Data Paradoxes."

for its comprehensive, unified population records and intensive health data production infrastructure,[17] is frequently cited as an example for other countries to follow. This belies local specificities, needs and practices, which vary widely and have significant conceptual implications.

Critical scholarship has extensively examined the potential risks posed by digital health, although these vary depending on the types of data, technologies, stakeholders, and purposes involved. For instance, a study by Geiger[18] exploring genomic data flows and their relation to the biomedical industry highlights the potential risks of data commodification. Other concerns include the rise of techno-solutionism,[19] and heightened individual risk, both of which exacerbate existing inequalities. These examples demonstrate how specific empirical case studies can help identify particular risks related to the use of health data.

In epidemiological studies employing linked administrative data, risks primarily concern individuals' rights to privacy and confidentiality, alongside possible harms and bias that escalate with the integration of large volumes of individual data from diverse sources.

Data practices in Latin America are contributing to this debate through examples of data activism, presenting epistemologies distinct from those focusing solely on the risks of big data. The present analysis differs by showcasing data uses for scientific research in the Brazilian public health context from a perspective aligned with the interdisciplinary field of critical data studies, which advances crucial questions regarding a more nuanced approach to data studies.[20] Scholars such as Milan and Treré[21] emphasise the importance of recognising diverse data practices (i.e., the "de-Westernisation of critical data studies"),[22] and argue that the narrative of 'data universalism' does not uniformly apply across the Global North and South.

This article describes and analyses an empirical experience from the Global South that sheds light on the process of transforming data for use in scientific research and knowledge production, specifically in the context of social epidemiological studies conducted in Brazil. It additionally focuses on relevant legal, ethical and sociotechnical factors shaping data infrastructures, access and practices.

2. Methods

The present study adopted a reflexive approach,[23] drawing on the authors' experiences while working at CIDACS, together with public information and open-access literature to provide context and examples of knowledge production and data utilisation at CIDACS, which permitted an analysis within the frameworks of Social Studies of Science and Technology (STS) and Critical Data Studies. To explore alternative and critical perspectives on CIDACS' data practices, our work presents and analyses challenges and practicalities involved in establishing the technical and organisational infrastructure necessary for linking administrative data to support epidemiological studies on social determinants of health. Additionally, ethical, legal, and societal issues surrounding linked data processing and use are examined.

Subsequently, the historical context of Brazil is briefly outlined to explain the origin of health inequalities and identify groups disproportionately affected with respect to socioeconomic indicators. The country's universal health system (SUS) is also described, as health is a constitutional right of all Brazilians, to help illustrate how social and health aspects are embedded in the linked administrative data produced to support knowledge generation.

Lastly, to exemplify data usage, four studies supported by CIDACS' linked data are presented, whose findings have been published and are freely accessible in scientific publications. This documentation and description align with standard methods of tracing data flows and *unpacking* practices that often remain invisible in research findings—or, to use a key STS metaphor, remain hidden inside a 'black box.' This article strives to situate the knowledge produced at CIDACS, recognising that data production, data infrastructure operations, and data usage are intricately intertwined, as well as contextually dependent on sociohistorical and disciplinary characteristics.

---

[17] Bauer "From Administrative Infrastructure to Biomedical Resource."
[18] Geiger, "A Tidal Wave of Inevitable Data?"
[19] Wamsley, "COVID-19, Digital Health Technology and the Politics of the Unprecedented."
[20] Valente, "Critical Data Studies with Latin America;" Grohmann, "Latin American Critical Data Studies."
[21] Milan, "Big Data from the South(s)."
[22] Milan, "Big Data from the South(s)," 321.
[23] Fook, "Developing Critical Reflection as a Research Method."

2.1 CIDACS as a Case Study

The '100 Million Brazilians Cohort' research project, which was conceived to integrate social and health data to investigate the impact of social protection policies on health outcomes among low-income populations in Brazil, served as the main motivating factor for establishing CIDACS.[24]

The cohort consists of all individuals who have applied for any type of governmental social welfare assistance since 2001 and have been thusly registered in the federal Unified Registry for Social Programs database (*Cadastro Único - CadUnico*). Applicants provide detailed demographic, economic, and social information on each of their family members, as well as household characteristics.[25]

CadUnico is a unique administrative database, characterised by high population coverage, including specific social and traditional subgroups throughout Brazil. When linked with administrative health databases and other data sources, it represents an unprecedented resource for investigating social determinants of health and assessing the effects of social protection policies on specific health outcomes across diverse groups.[26]

To transform the aforementioned cohort into a viable resource, it was necessary to obtain, process and integrate data from national health and social protection information systems, as well as to ensure secure access to linked data. The CIDACS founding group, largely comprised of epidemiologists, recognised the need to design and construct physical and computational infrastructure to provide a safe data environment, alongside developing robust methodologies for integrating large volumes of data with high-quality and accuracy.

This endeavour was informed by scientific literature and concrete international experiences. Its implementation required securing national and international research funding as well as forming interdisciplinary collaborations. These collaborative efforts facilitated the development of processes, workflows, methodologies and strategies for integrating extensive datasets to support research projects while adhering to ethical and legal regulations governing research involving humans.

2.2 CIDACS Data Platform

The CIDACS data platform integrates physical and computational infrastructure with dedicated personnel to receive, process and provide access to high-quality integrated datasets for population-based studies. Information security and the privacy of data subjects are prioritised throughout all data handling procedures.

Requests for access to linked databases are always based on updates to the 100 Million Brazilians Cohort and tailored to the needs of each individual research project. All requests prompt formal communication between a data controller and the soliciting researchers, who must provide appropriate justification for data usage in the proposed study context and submit ethical approval documentation from an institutional review board belonging to the CEP/Conep system.[27] Upon approval, CIDACS follows established workflows that vary according to each controller. After fulfilling all documentation requirements, dataset copies are prepared and securely transferred to CIDACS. Upon validating consistency, data are catalogued and transferred to a secure storage location.

All processing throughout the lifecycle of original and integrated data occurs exclusively within a secure, controlled environment in adherence with best practices for data integration to address research inquiries. Processes such as record linkage and data analysis are separated to enhance security and data integrity. This environment, known as a 'data safe haven' or 'trusted research environment', operates under the guidance of the Five Safes Framework.[28]

---

[24] Barreto, "The Centre for Data and Knowledge Integration for Health (CIDACS)."

[25] Barreto, "Cohort Profile;" Barreto, "The Centre for Data and Knowledge Integration for Health (CIDACS)."

[26] Almeida, "CIDACS' Efforts Towards an Inclusive and Dialogic Data Governance in Brazil," 1-8.

[27] The CEP/CONEP System evaluates research protocols involving human beings in Brazil with respect to ethical considerations. The system is formed by the National Research Ethics Commission (CONEP) and Institutional Review Boards (CEP).

[28] Almeida, "CIDACS' Efforts Towards an Inclusive and Dialogic Data Governance in Brazil," 1-8; Barreto, "The Centre for Data and Knowledge Integration for Health (CIDACS);" Hubbard, "Trusted Research Environments (TRE) Green Paper."

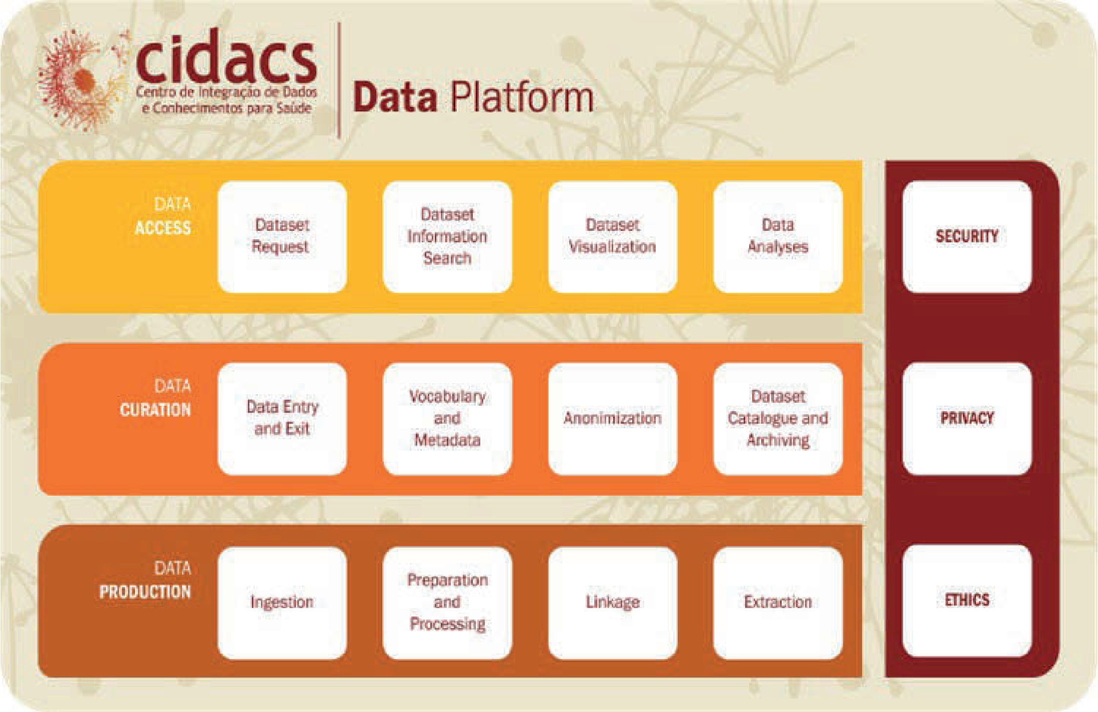

Figure 1. CIDACS Data Platform
Image credit: Barreto et al. "The Center for Data and Knowledge Integration for Health (CIDACS): An Experience of Linking Health and Social Data in Brazil." *International Journal of Population Data Science, 2019.*

The integration of administrative data and transformation into longitudinal structures suitable for addressing research questions occurs within the Secure Room environment, employing methodological strategies for selecting, cleaning, standardising and harmonising variables. CIDACS employs two linkage scenarios: Deterministic, in which pairs containing unique identifiers are linked to each other, and Non-deterministic, which is applied when common unique identifiers are unavailable, relying instead on other individual variables such as name, sex, date of birth, municipality of residence, mother's maiden name, etc. It is important to note that CIDACS has developed a novel tool to efficiently integrate large-scale administrative datasets, an algorithm denominated 'CIDACS Record Linkage'.[29]

The integrity and variable availability of the obtained databases are verified upon receipt. Metadata and data catalogues are then created, in addition to documentation of data provenance. Intensive data pre-processing and harmonisation are conducted on each original database to prepare variables for integration, while epidemiologists and statisticians help to ensure the resulting data formats are optimised for specific research applications. Following quality and accuracy assessments of linked data, anonymised or pseudonymised integrated datasets are transferred to the Data Access Environment. These datasets contain only those variables essential to answering the questions under study. Access is restricted to authorised researchers who undergo a formal accreditation process.

The *access environment* provides a virtual desktop equipped with pre-installed, audited analytical tools for data analysis. Even anonymised data cannot be downloaded. Results can be requested in tabular, graphical or script formats, and are released only after performing risk assessment to prevent the re-identification of data subjects.[30]

CIDACS has developed in-house solutions and strategies to protect personal and sensitive data during processing, linkage and access, which are employed to transform administrative data into reliable sources of information that enable the conduct of longitudinal epidemiological studies aimed at supporting knowledge production.

---

[29] Barbosa, "CIDACS-RL."
[30] Barreto, "The Centre for Data and Knowledge Integration for Health (CIDACS)."

A systematic review and thematic analysis of the literature from 2012 to 2021 identified 38 publications describing characteristics that enabled the transformation of administrative data into research-ready datasets. These publications span countries including Australia, Brazil, Canada, China, New Zealand, Taiwan, the United States and the United Kingdom. Notably, the search returned only one Brazilian publication, relating to CIDACS, in which efforts towards administrative data integration for research purposes were described.[31]

In addition to infrastructure and data linkage methods, administrative data access and use for research purposes must comply with national legislation. Beyond regulatory adherence, administrative data linkage also poses specific ethical and societal challenges.

3. Ethical, Legal and Social Issues (ELSI) Surrounding Linked Data to Support Knowledge Production at and Beyond CIDACS

CIDACS was inaugurated in December 2016, prior to the enactment of the Brazilian Data Protection Law (LGPD), which was approved by Brazil's legislature in 2018 and came into effect in 2020.

The development of CIDACS's guiding principles and frameworks was deeply influenced by the European Union's General Data Protection Regulation (GDPR)[32] and legal consultancy provided by a data protection specialist. While contextual factors shape data processing and usage, both international and local regulations play integral roles in CIDACS' ongoing efforts to achieve its specific objectives.

Following discussions surrounding the draft legislation (PL N.4.060/2012), which provided the basis for the text of the *LGPD,* CIDACS members observed prevailing public concerns focusing on the commodification and monetisation of personal data by large-scale technology companies—so-called 'big tech'. While these concerns are legitimate, they overlook the nuanced purposes underlying data processing for the purposes of conducting scientific research.[33] Embracing the concept that personal data can be transformed into scientific knowledge to understand specific problems and provide evidence for effective policymaking, the CIDACS team mobilised the Brazilian scientific community to engage in discussions on data protection that were then taking place in the country.

On 22 May 2018, the Centre's coordinator participated in a public hearing (organised by the Special Committee on PL N.4.060/12), discussing data protection in the domains of Science and Technology, Communication and Informatics. He advocated for the responsible use of personal data in scientific research aimed at knowledge production to support evidence-based public policies, while also emphasising the safeguarding of privacy and data subject rights.[34]

The *LGPD* (Law No. 13,709/2018),[35] approved on 14 August 2018 and in effect since 18 September 2020, marked a legislative milestone in Brazil by establishing rules for the processing of personal data, whether physical or digital, by natural or legal persons, in both the public and private sectors. The law explicitly recognises academic research and public health as specific contexts for processing personal and sensitive data, prescribing recommendations and establishing safeguards that must be observed according to sector-specific regulations.

Importantly, the *LGPD* includes a dedicated article (Article 13) addressing the topic of public health research, for which we believe that advocacy by CIDACS based on its experience was crucial:

> research organizations may have access to databases containing personal information, which will be handled exclusively by the institution, strictly for the purpose of carrying out studies and research; such data will be maintained in a controlled and secure environment in accordance with security practices provided for by specific regulation, including, whenever possible, the anonymisation or pseudoanonymisation of data, as well as the observance of relevant ethical standards pertaining to studies and research.[36]

In adherence to ethical standards guiding the conduct of studies and research, CIDACS/Fiocruz Bahia contributed to public consultations overseen by the National Research Ethics Committee (Conep) in September 2023 on the use of databases in

---

[31] McGrath-Lone, "What makes Administrative Data 'Research-Ready'?," 5-6.
[32] See: https://gdpr-info.eu/.
[33] Segura, "Between Data Capitalism and Data Citizenship."
[34] Available at: https://www.youtube.com/watch?v=wZHgfHZruHY&t=7346s.
[35] Brazil, "*Law on Treatment and Protection of Personal Data: Law 13.709.*"
[36] Brazil, *Law on Treatment and Protection of Personal Data: Law 13.709.*

human research.[37] Subsequently, following publication of Resolution no. 738 by the National Health Council, outlining guidelines for scientific research utilising databases containing information on humans, CIDACS requested a public hearing with Conep to clarify issues surrounding the use of secondary data containing personal information for scientific research purposes.

Through advocating for secure access to and responsible use of administrative data in alignment with legal, ethical, and privacy requirements, CIDACS has made significant contributions to national debates on the use of personal data for public health research purposes.

Kitchin (2024) insightfully notes that:

> Data are the product of discursively framed and technically mediated processes, shaped by protocols, organisational processes, measurement scales, categories, and standards that are designed, negotiated and debated. Similarly, the entirety of the data lifecycle (generation, handling, processing, storage, sharing, analysis, interpretation, and deletion) is socio-technically mediated and saturated with politics. So, too, are the production and operation of data infrastructures, and the many ways in which data are used.[38]

In an attempt to identify and address societal challenges related to administrative data linkage, CIDACS conducted an exploratory study in 2020 to examine the perceptions and attitudes of diverse societal segments toward the sharing and linking of personal data in public and private domains, with a particular focus on government administrative data.[39] This seminal study led CIDACS' leadership to realise that ethical considerations should extend beyond safeguards, control mechanisms and legal requirements, prompting the development of initiatives aimed at incorporating data ethics issues and public engagement strategies to enhance public trust and foster cooperation:

> A core factor relevant to our situation is that CIDACS acts as a custodian of a database containing data on the lowest-income individuals and families in Brazil; the Centre is also attempting to link information on additional subgroups to the existing 100 Million Brazilians Cohort. Accordingly, it is important to consider specific groups' rights, interests and concerns, as well as their cultural specificities and demographic differences in the data governance decision-making that informs scientific research and public policy.[40]

Currently, CIDACS is conducting a study on administrative data access and use for health research and policy evaluation in Brazil and Latin America, in an effort to elucidate both national and regional landscapes and specificities.

The following section provides a brief overview of Brazil's history to contextualise its structural inequalities, followed by a presentation of the country's universal health system. This information is provided to facilitate the reader's understanding of how socio-historical and health-related factors are intertwined with CIDACS' production of linked administrative data to support research efforts.

4. History of Inequalities and Public Health in Brazil (in a Nutshell)

Brazil, the largest country in Latin America, has an estimated population size of approximately 212 million inhabitants.[41] In 2022, around 92.1 million people (45.3% of the population) self-identified as *parda,*[42] 20.6 million (10.2%) as Black,[43] 1.7 million (0.8%) as Indigenous, and 88.2 million (43.5%) as White.[44] Brazil was a Portuguese colony from 1500 to 1822. Slavery was abolished in 1888, followed by the country's political independence as a republic in 1889. Throughout its history, Brazil underwent rapid industrialisation together with periods of military and authoritarian rule; its longest democracy lasted roughly

---

[37] Available at: https://www.gov.br/conselho-nacional-de-saude/pt-br/assuntos/noticias/2023/setembro/conep-consulta-publica-sobre-uso-de-bancos-de-dados-em-pesquisas-com-seres-humanos.

[38] Kitchin, "Performing Critical Data Studies from the Inside."

[39] Almeida, "Perceptions and Experiences on Data Sharing and Linkage for Research and the Evaluation of Public Health Policy."

[40] Almeida, "CIDACS' Efforts Towards an Inclusive and Dialogic Data Governance in Brazil."

[41] IBGE, "Censo Demográfico 2022."

[42] 'The category '*parda'* is defined as admixture between Whites and Indigenous; Whites and Blacks; Blacks and Indigenous; or Blacks and persons of other colours or races.' (Rebouças, 2024, 4).

[43] Some research on race, racism and health in Brazil tends to analyse its impact on Black and *parda* populations together, drawing attention to the fact that together this population represents a majority in the country and that both subgroups are affected, albeit in different ways (Góes, Ferreira and Ramos, 2022).

[44] IBGE, "Censo Demográfico 2022".

25 years.[45] Colonisation and the exploitation of Black and Indigenous populations have left a legacy of structural racism, with profound consequences for the health of Brazil's ethnic-racial groups, as well as their living conditions, civil rights, and access to a range of services.[46]

The Brazilian health system was developed in parallel with the nation's struggle for re-democratisation after a period of military dictatorship. In 1988, Brazil established a universal health system, the *Sistema Único de Saúde* (SUS), whereby health became a fundamental right and state duty enshrined in the Brazilian Constitution (1988). The key principles of Brazilian health policy are universality, integrality, decentralisation, and community participation.[47] These principles guarantee the universal right to health care across multiple levels of complexity, with decentralisation assigning responsibilities across federal, state and municipal levels of government, and social participation being integral to formulating and monitoring the implementation of health policy.[48] Notably, the Brazilian Movement for Sanitary Reform played a critical role in advocating for a more democratic and universal health system, which created the political conditions for the establishment of SUS. The movement's original slogan, "health is democracy and democracy is health," continues to resonate today.

The scientific literature highlights how the existence of a public health system influences the production and quality of administrative data suitable for research purposes.[49] In the Brazilian context, vital data emerge from SUS.[50] The country maintains diverse national information systems that support public administration, such as the Ministry of Health's Information System on Live Births (Sinasc), Mortality Information System (SIM), and Notifiable Diseases Information System (Sinan).

Social policies further complement health initiatives in Brazil. The first conditional cash transfer program implemented in the mid-1990s began as a series of local anti-poverty interventions, later expanding to the federal level in 2001. In the 2000s, discussions surrounding social protection policies intensified, becoming a priority alongside the reduction of poverty and inequality. In 2003, combating hunger and poverty was a centrepiece of the government's agenda, leading to several initiatives, including the *Bolsa Família Program* (BFP), the largest conditional cash transfer scheme ever attempted worldwide.

The objective of BFP is:

> to reduce inequalities through the conditional transfer of income to families living in poverty and extreme poverty, in order to promote emancipation and overcome their situation of vulnerability [...] it also aims to improve education and health levels through conditionalities.[51]

The use of administrative data from government social and health departments for public health research is well-established in Brazil, similar to other nations including Norway, Finland, Sweden, Canada, the United Kingdom and the United States.[52] Health knowledge production based on data is a longstanding concept. Yet advances in computational capabilities now enable the linking and organising of large and disparate databases into longitudinal structures — a significant innovation for epidemiology:

> the availability of high-quality but dispersed social and health databases in Brazil and the need to integrate data and evaluate the impact of policies aiming to improve the social determinants of health (e.g. social protection policies) on health outcomes, overall and in subgroups in a dynamic Brazilian cohort in a middle-income country with high inequalities.[53]

The 100 Million Brazilians Cohort emanated from broader societal changes, including the formulation and deployment of social protection policies targeting poverty reduction and the mitigation of social inequalities. Its implementation spanned several years and serves as a proof of concept for investigation into the potential effects of poverty alleviation on health outcomes.[54]

---

[45] Paim, "The Brazilian Health System."

[46] Rebouças, "Ethno-Racial Inequalities on Adverse Birth and Neonatal Outcomes."

[47] OECD, "OECD Reviews of Health Systems."

[48] Ortega, "Brazil's Unified Health System."

[49] Connelly, "The Role of Administrative Data in the Big Data Revolution in Social Science Research," 1-12.

[50] This a brief characterisation aimed to contextualise the CIDACS data practices. For more about SUS nuances, complexities, and contradictions, see for example: Castro, 2019; Paim, 2011; Massuda, 2018; Ortega and Pele, 2023.

[51] Menicucci, "Políticas Sociais," 147.

[52] Connelly, "The Role of Administrative Data in the Big Data Revolution in Social Science Research," 1-12.

[53] Barreto, "Cohort Profile," 2.

[54] Barreto, "Cohort Profile."

5. Data in Action: Producing Knowledge on Health Inequalities Using Large Volumes of Linked Data

The interconnected relationships between social inequalities, justice and respective impacts on populational health have been widely studied. Scholarly works examining how *the social* manifests in health have discussed the use of varied research methodologies—both quantitative and qualitative—to measure these effects.[55]

The CIDACS experience offers an opportunity to contribute to this discourse. Concerning methods for investigating complex relationships within the social-health nexus, scholars argue that quantitative epidemiological studies based on large datasets must avoid simplistic, reductive explanations of '*the social,*' rather being framed within comprehensive theoretical frameworks and situated within relevant social contexts.

Epidemiology, defined as the study of the distribution and determinants of health-related states or events within populations, can play a significant role in the prevention and control of health issues.[56] Since its inception, the discipline has relied on numerical data to elucidate health and disease-related phenomena, as well as to identify links with environmental factors.[57] Epidemiology maintains a dialogic relationship with social sciences and statistics in its pursuit to understand how social determinants influence health outcomes.

The relationships between inequity and health are complex, with nuances often obscured within administrative database variables and categories, which may in some way reflect official state policy perspectives. Nonetheless, measurable evidence retains strong explanatory power. As such, measuring and comparing health status across social markers sheds light on risk factors that may more acutely affect specific population groups whose vulnerabilities merit explanation beyond descriptive measurements.

The 100 Million Brazilians Cohort supports diverse public health research initiatives spanning infant and maternal health,[58] social inequalities in health,[59] infectious diseases[60] and mental health,[61] among other topics.

To illustrate *data in action*, i.e., the active contribution of linked data in unveiling how social inequalities affect health, our work summarises findings from four CIDACS-supported epidemiological studies employing linked data involving the 100 Million Brazilians Cohort. Three of these investigate ethno-racial inequalities in childhood health, while the fourth evaluates the impact of a social protection program on child mortality.

One study investigated ethno-racial inequalities in infant mortality from 2012 to 2018.[62] Analyzing mortality rates according to race and maternal skin colour within a cohort of 19 million Brazilian children, the study found that children born to Indigenous mothers had a sixteen-fold higher risk of death from malnutrition and a fourteen-fold higher risk of death from diarrhoea compared to children of White mothers. These results are contextualised within structural racism in Brazilian society, underscoring the need for targeted public policies to address deep-rooted inequalities. The examined indicators—childhood growth, mortality and birth outcomes—are sensitive to social inequalities, as these are preventable conditions closely linked to poverty and living conditions.

Another study encompassing 4,090,271 children born between 2008 and 2017 assessed ethno-racial disparities in child growth trajectories and nutritional status. Findings demonstrated that children born to Indigenous, Brown, and Black mothers exhibited unfavorable growth patterns compared to those of White mothers. Among Indigenous children, stature was on average 3.3 cm less than their White counterparts, indicating persistent chronic malnutrition over time.[63]

---

[55] Adams, "Re-Imagining Global Health Through Social Medicine.".

[56] Vlajinac, "Epidemiology, Aims and Scope," 350–352; see also: Porta, "A Dictionary of Epidemiology" 95.

[57] See: https://www.sciencemuseum.org.uk/objects-and-stories/medicine/epidemiology-public-health-science.

[58] Ramos, "Conditional Cash Transfer Program and Child Mortality;" Alves, "Association of Conditional Cash Transfers with Maternal Mortality Using the 100 Million Brazilian Cohort."

[59] Rebouças, "Ethnoracial Inequalities and Child Mortality in Brazil;" Rebouças, "Ethno-Racial Inequalities on Adverse Birth and Neonatal Outcomes;" Benes, "Ethnoracial Disparities in Childhood Growth Trajectories in Brazil."

[60] Pescarini, "Conditional Cash Transfer Program and Leprosy Incidence;" Paixão, "Socioeconomic Risk Markers of Congenital Zika Syndrome;" Silva, "Risk of Death Following Chikungunya Virus Disease in the 100 Million Brazilian Cohort, 2015-18."

[61] Machado, "The Impact of Social Drivers;" Toledo, "Risk of Psychiatric Hospitalization in Low-Income Youth."

[62] Rebouças "Ethnoracial Inequalities and Child Mortality in Brazil."

[63] Benes, "Ethnoracial Disparities in Childhood Growth Trajectories in Brazil," 8.

A third study[64] examined racial inequalities as persistent barriers to maternal and child health in Brazil, with adverse effects disproportionately impacting Black and Indigenous women and their offspring. Analyzing over 21 million live births from 2012 to 2019, the study estimated that premature births, low birth weight, small for gestational age and preterm deaths among neonates could have been reduced by 1.7%, 7.2%, 10.8%, and 11.8%, respectively, had these inequalities been eliminated in the country.

The final study[65] analyzed 6,309,366 children aged five years or less (2006–2015) to evaluate the Brazilian Bolsa Família Program's (BFP) effect on child mortality. Stratified by maternal education, race/skin colour, gestational age, and municipal wealth and cash transfer management quality, the findings revealed that the BFP benefit was associated with reduced mortality among children aged 1 to 4 years. This association was stronger among preterm children, children born to Black mothers, and those residing in poorer municipalities with effective program management.

These studies collectively demonstrate how structural inequalities manifest in childhood health outcomes and how poverty alleviation programs can positively impact child mortality. They further exemplify the materialisation of social determinants of health in quantitative data amenable to longitudinal monitoring. Lupton[66] conceptualises numbers as 'socio-technical devices' inseparable from the practices that generate them, as data possess agency in constituting phenomena, conferring existence and meaning. In this light, the data itself shapes the narratives constructed around social realities.

6. Final Considerations

Linking administrative social and health data to enable longitudinal epidemiological studies investigating social determinants of health and the effects of social protection policies on diverse outcomes across various groups involves complex technical, ethical, legal and social issues.

Building and sustaining physical and computational infrastructure requires substantial investment of resources and time. Developing methodologies for integrating large volumes of heterogeneous databases, and recruiting dedicated personnel for data ingestion, processing, and controlled access, demands collaboration among an interdisciplinary team of specialists.

Legal and ethical frameworks critically shape data access, acquisition, processing and usage. In observance of these frameworks and the objectives of constructing and maintaining the 100 Million Brazilians Cohort, CIDACS actively contributes to national debates on the use of personal data for scientific and public health research purposes.

Beyond physical and virtual controls and legal-ethical safeguards, considerations related to the rights of data subjects—particularly the groups represented in analyses—have emerged, especially as scientific inquiry is focused on a populational perspective, while ethics and privacy provisions often prioritise individual protections.

This article aims to provide empirical evidence to demonstrate that experiences from countries in the Global South, such as that of CIDACS in Brazil, offer alternative strategies and epistemologies for data processing and utilisation within the complex and evolving field of digital health. Critical Data Studies and STS perspectives can enrich understanding by revealing context-specific dimensions of data practices, including the intellectual projects, purposes and interests driving data use.

Funding statement
This work was supported by the Wellcome Trust [grant number 224494/Z/21/Z] and the Global Health Research Unit of the National Institute for Health and Care Research on Social and Environmental Determinants of Health Inequalities [grant number NIHR134801]. The funders had no role in the study design, analysis, decision to publish, or preparation of the manuscript.

Acknowledgements
The authors would like to thank Andris K. Walter for assistance with text revision and manuscript copyediting.

Conflict of interest
Despite being employed by the Centre for Data and Knowledge Integration for Health, the authors declare no existing conflicts of interest with regard to using CIDACS as a case study.

---

[64] Rebouças, "Ethno-Racial Inequalities on Adverse Birth and Neonatal Outcomes."
[65] Ramos, "Conditional Cash Transfer Program and Child Mortality."
[66] Lupton, Digital Sociology, 101.

Bibliography

Adams, Vincanne, Dominique Behague, Carlo Caduff, Ilana Löwy and Francisco Ortega. "Re-Imagining Global Health through Social Medicine." *Global Public Health* 14, no 10 (2019): 1383–1400. https://doi.org/10.1080/17441692.2019.1587639.

Almeida Filho, Naomar and Mauricio Barreto. "Panorama, Desafios e Perspectivas para uma Epidemiologia Brasileira." In *Epidemiologia e Saúde: Fundamentos, Métodos e Aplicações,* 687-691. Rio de Janeiro: Guanabara Koogan, 2011.

Almeida, Bethania de Araujo, Roberto Carreiro, Maìra Souza, and Mauricio Barreto. "CIDACS' Efforts towards an Inclusive and Dialogic Data Governance in Brazil: A Focused Literature Review." *International Journal of Population Data Science* 9, no 1 (2024). https://doi.org/10.23889/ijpds.v9i1.2163.

Almeida, Bethania de Araújo, Pimenta, and Denise Moraes. "Perceptions and Experiences on Data Sharing and Linkage for Research and the Evaluation of Public Health Policy." Salvador: Cidacs-Fiocruz, 2021. https://cidacs.bahia.fiocruz.br/wp-content/uploads/2021/11/Executive-Summary-English-Version.pdf

Alves, Flávia Jôse O., Dandara Ramos, Enny S. Paixão, Ila R. Falcão, Rita De Cássia Ribeiro-Silva, Rosemeire Fiaccone, Davide Rasella, et al. 2023. "Association of Conditional Cash Transfers with Maternal Mortality Using the 100 Million Brazilian Cohort.' *JAMA Network Open* 6 (2): e230070. https://doi.org/10.1001/jamanetworkopen.2023.0070.

Barbosa, George C. G., M. Sanni Ali, Bruno Araujo, Sandra Reis, Samila Sena, Maria Y. T. Ichihara, Julia Pescarini, et al. "CIDACS-RL: A Novel Indexing Search and Scoring-Based Record Linkage System for Huge Datasets with High Accuracy and Scalability." *BMC Medical Informatics and Decision Making* 20, no 1 (2020): 289. https://doi.org/10.1186/s12911-020-01285-w.

Barreto, Mauricio Lima, Maria Yuri Ichihara, Bethania De Araujo Almeida, Marcos Ennes Barreto, Liliana Cabral, Rosemeire Fiaccone, Roberto P Carreiro, et al. "The Centre for Data and Knowledge Integration for Health (CIDACS): An Experience of Linking Health and Social Data in Brazil." *International Journal of Population Data Science* 4, no 2 (2019). https://doi.org/10.23889/ijpds.v4i2.1140.

Barreto, Mauricio L., Maria Yury Ichihara, Julia M. Pescarini, M. Sanni Ali, Gabriela L. Borges, Rosemeire L. Fiaccone, Rita De Cássia Ribeiro-Silva, et al. "Cohort Profile: The 100 Million Brazilian Cohort." *International Journal of Epidemiology* 51, no 2 (2022): e27–38. https://doi.org/10.1093/ije/dyab213.

Bauer, Susanne. "From Administrative Infrastructure to Biomedical Resource: Danish Population Registries, the 'Scandinavian Laboratory,' and the 'Epidemiologist's Dream.'" *Science in Context* 27, no 2 (2014): 187–213. https://doi.org/10.1017/S0269889714000040.

Brazil. *Law on Treatment and Protection of Personal Data: Law 13.709.* From 14 August 2018, Modified by Law 13.853, from 8 July 2019. https://www.gov.br/anpd/pt-br/centrais-de-conteudo/outros-documentos-e-publicacoes-institucionais/lgpd-en-lei-no-13-709-capa.pdf.

Benes, Matos Da Silva, Helena, Rita De Cássia Ribeiro-Silva, Juliana Freitas De Mello E. Silva, Irina Chis Ster, Poliana Rebouças, Emanuelle Goes, Maria Yury Ichihara, et al. "Ethnoracial Disparities in Childhood Growth Trajectories in Brazil: A Longitudinal Nationwide Study of Four Million Children." *BMC Pediatrics* 24, no 1 (2024): 103. https://doi.org/10.1186/s12887-024-04550-3.

Castro, Marcia C., Adriano Massuda, Gisele Almeida, Naercio Aquino Menezes-Filho, Monica Viegas Andrade, Kenya Valéria Micaela De Souza Noronha, Rudi Rocha, et al. 2019. "Brazil's Unified Health System: The First 30 Years and Prospects for the Future." *The Lancet* 394 (10195): 345–56. https://doi.org/10.1016/S0140-6736(19)31243-7.

Cerqueira-Silva, Thiago, Julia M. Pescarini, Luciana L. Cardim, Clémence Leyrat, Heather Whitaker, Carlos Alexandre Antunes De Brito, Elizabeth B. Brickley, et al. 2024. "Risk of Death Following Chikungunya Virus Disease in the 100 Million Brazilian Cohort, 2015–18: A Matched Cohort Study and Self-Controlled Case Series." *The Lancet Infectious Diseases* 24 (5): 504–13. https://doi.org/10.1016/S1473-3099(23)00739-9.

Connelly, Roxanne, Christopher J. Playford, Vernon Gayle and Chris Dibben. "The Role of Administrative Data in the Big Data Revolution in Social Science Research." *Social Science Research* 59, September (2016): 1–12. https://doi.org/10.1016/j.ssresearch.2016.04.015.

Geiger, Susi and Nicole Gross. "A Tidal Wave of Inevitable Data? Assetization in the Consumer Genomics Testing Industry." *Business and Society* 60, no 3 (2021): 614–49. https://doi.org/10.1177/0007650319826307.

Góes, Emanuelle Freitas, Andrêa J. F. Ferreira and Dandara Ramos. "Anti-Black Racism and Maternal Death from COVID-19: What Have We Seen in the Pandemic?' *Ciência & Saúde Coletiva* 28, no (2023): 2501–10. https://doi.org/10.1590/1413-81232023289.08412022en.

Grohmann, Rafael. "Latin American Critical Data Studies". *Big Data and Society* 12, no 2 (2025): 20539517251330160. https://doi.org/10.1177/20539517251330160.

Fook, Jan. "Developing Critical Reflection as a Research Method". In *Creative Spaces for Qualitative Researching*, edited by Joy Higgs, Angie Titchen, Debbie Horsfall and Donna Bridges, 55–64. Rotterdam: SensePublishers, 2011. https://doi.org/10.1007/978-94-6091-761-5_6.

Hoeyer, Klaus. *Data Paradoxes: The Politics of Intensified Data Sourcing in Contemporary Healthcare*. Infrastructures. Cambridge: The MIT Press, 2023.

Hoeyer, Klaus, Susanne Bauer, and Martyn Pickersgill. "Datafication and Accountability in Public Health: Introduction to a Special Issue." *Social Studies of Science* 49, no 4 (2019): 459–75. https://doi.org/10.1177/0306312719860202.

Hubbard, Tim, Gerry Reilly, Susheel Varmae and David Seymour. "Trusted Research Environments (TRE) Green Paper". Zenodo, July 21, (2020). https://doi.org/10.5281/zenodo.4594704.

IBGE. "Demographic Census 2022: Ethnic-Racial Identification of the Population by Sex and Age: Results of the Universe." Rio de Janeiro, 2023. https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=73105

Kitchin, Rob. "Performing Critical Data Studies from the Inside. Working with Government to Change Data Regimes." In *Collaborative Research in the Datafied Society: Methods and Practices for Investigation and Intervention*, edited by Schäfer, Mirko Tobias, Es, Karin and Tracey Laurialt. Amsterdam University Press, 2024. https://library.oapen.org/bitstream/handle/20.500.12657/92631/1/9789048555925.pdf.

Kitchin, Rob. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data and Society* 1, no 1 (2014): 2053951714528481. https://doi.org/10.1177/2053951714528481.

Leonelli, Sabina and Niccolò Tempini, eds. *Data Journeys in the Sciences*. Cham: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-37177-7.

Lupton, Deborah. *Digital Sociology*. London: Routledge, 2015. https://doi.org/10.4324/9781315776880.

Lupton, Deborah. "Critical Perspectives on Digital Health Technologies." *Sociology Compass* 8, no 12 (2014): 1344–59. https://doi.org/10.1111/soc4.12226.

Machado, Daiane Borges, Jacyra Azevedo Paiva De Araujo, Flávia Jôse Oliveira Alves, Luis Fernando Silva Castro-de-Araujo, Elisângela Da Silva Rodrigues, Erika Fialho Morais Xavier, Rodrigo Lins Rodrigues, et al. "The Impact of Social Drivers, Conditional Cash Transfers and Their Mechanisms on the Mental Health of the Young; an Integrated Retrospective and Forecasting Approach Using the 100 Million Brazilian Cohort: A Study Protocol." Edited by Vanessa Carels. *PLOS ONE* 17, no 10 (2022): e0272481. https://doi.org/10.1371/journal.pone.0272481.

McGrath-Lone, Louise, Matthew A. Jay, Ruth Blackburn, et al. "What Makes Administrative Data Research-Ready?: A Systematic Review and Thematic Analysis of Published Literature." *International Journal of Population Data Science* 7, no 1 (2022). https://doi.org/10.23889/ijpds.v7i1.1718.

Massuda, Adriano, Thomas Hone, Fernando Antonio Gomes Leles, Marcia C. De Castro and Rifat Atun. "The Brazilian Health System at Crossroads: Progress, Crisis and Resilience." *BMJ Global Health* 3, no 4 (2018): e000829. https://doi.org/10.1136/bmjgh-2018-000829.

Menicucci, Telma and Sandra Gomes. *Políticas sociais: conceitos, trajetórias e a experiência brasileira.* Rio de Janeiro: Editora Fiocruz, 2018.

Milan, Stefania, and Emiliano Treré. "Big Data from the South(s): Beyond Data Universalism." *Television and New Media* 20, no 4 (2019): 319-335.

Milan, Stefania and Emiliano Treré. "Against Decolonial Reductionism: The Impact of Latin American Thinking on the Data Decolonization Project". *Big Data and Society* 11, no 4 (2024): 20539517241270694. https://doi.org/10.1177/20539517241270694.

Organisation for Economic Cooperation and Development (OECD). "OECD Reviews of Health Systems: Brazil 2021." OECD Reviews of Health Systems, Paris: OECD Publishing, 2021. https://doi.org/10.1787/146d0dea-en.

Ortega, Francisco, and Antonio Pele. "Brazil's Unified Health System: 35 Years and Future Challenges." *The Lancet Regional Health - Americas* 28, December (2023): 100631. https://doi.org/10.1016/j.lana.2023.100631.

Paim, Jairnilson, Claudia Travassos, Celia Almeida, Ligia Bahia and James Macinko. "The Brazilian Health System: History, Advances, and Challenges." *The Lancet* 377, no 9779 (2011): 1778–97. https://doi.org/10.1016/S0140-6736(11)60054-8.

Paixao, Enny S., Luciana L. Cardim, Ila Rocha Falcao, Naiá Ortelan, Natanael De Jesus Silva, Aline Dos Santos Rocha, Samila Sena, et al. "Cohort Profile: Centro de Integração de Dados e Conhecimentos Para Saúde (CIDACS) Birth Cohort." *International Journal of Epidemiology* 50, no 1 (2021): 37–38. https://doi.org/10.1093/ije/dyaa255.

Paixão, Enny S., Qeren Hapuk, R. Ferreira Fernandes, Luciana L. Cardim, Julia M. Pescarini, Maria Conceicao N. Costa, Ila R Falcão, Elizabeth B. Brickley, et al. "Socioeconomic Risk Markers of Congenital Zika Syndrome: A Nationwide, Registry-Based Study in Brazil." *BMJ Global Health* 7, no 9 (2022): e009600. https://doi.org/10.1136/bmjgh-2022-009600.

Pescarini, Julia M., Elizabeth Williamson, Maria Y. Ichihara, Rosemeire L. Fiaccone, Laura Forastiere, Anna Ramond, Joilda Silva Nery, et al. "Conditional Cash Transfer Program and Leprosy Incidence: Analysis of 12.9 Million Families From the 100 Million Brazilian Cohort." *American Journal of Epidemiology* 189, no 12 (2020): 1547–58. https://doi.org/10.1093/aje/kwaa127.

Pita, Robespierre, Clícia Pinto, Samila Sena, Rosemeire Fiaccone, Leila Amorim, Sandra Reis, Mauricio L. Barreto, Spiros Denaxas, and Marcos Ennes Barreto. "On the Accuracy and Scalability of Probabilistic Data Linkage Over the Brazilian 114 Million Cohort." *IEEE Journal of Biomedical and Health Informatics* 22, no 2 (2018): 346–53. https://doi.org/10.1109/JBHI.2018.2796941.

Ramos, Dandara, Nívea B. Da Silva, Maria Yury Ichihara, Rosemeire L. Fiaccone, Daniela Almeida, Samila Sena, Poliana Rebouças, et al. "Conditional Cash Transfer Program and Child Mortality: A Cross-Sectional Analysis Nested within the 100 Million Brazilian Cohort," edited by Zulfiqar A. Bhutta. *PLOS Medicine* 18, no 9 (2021): e1003509. https://doi.org/10.1371/journal.pmed.1003509.

Rebouças, Poliana, Emanuelle Goes, Julia Pescarini, Dandara Ramos, Maria Yury Ichihara, Samila Sena, Rafael Veiga, Laura C. Rodrigues, Maurício L. Barreto, and Enny S. Paixão. "Ethno-Racial Inequalities and Child Mortality in Brazil: A Nationwide Longitudinal Study of 19 Million Newborn Babies." *The Lancet Global Health* 10, no 10 (2022): e1453–62. https://doi.org/10.1016/S2214-109X(22)00333-3.

Rebouças, Poliana, Enny S. Paixão, Dandara Ramos, Julia Pescarini, Elzo Pereira Pinto-Junior, Ila R. Falcão, Maria Yury Ichihara, et al. "Ethno-Racial Inequalities on Adverse Birth and Neonatal Outcomes: A Nationwide, Retrospective Cohort Study of 21 Million Brazilian Newborns." *The Lancet Regional Health - Americas* 37, September (2024):100833. https://doi.org/10.1016/j.lana.2024.100833.

Segura, Maria Soledad, and Silvio Waisbord. "Between Data Capitalism and Data Citizenship." *Television and New Media* 20, no 4 (2019): 412-419.

Storeng, Katerini Tagmatarchi, Sakiko Fukuda-Parr, Manjari Mahajan, and Sridhar Venkatapuram. 2021. "Digital Technology and the Political Determinants of Health Inequities: Special Issue Introduction." *Global Policy* 12 (S6): 5–11. https://doi.org/10.1111/1758-5899.13001.

Toledo, Lidiane, Rodrigo Rodrigues, Flávia Alves, Fillipe Guedes, Jacyra Azevedo Paiva De Araújo, John A. Naslund, Maurício L. Barreto, Vikram Patel and Daiane Borges Machado. "Risk of Psychiatric Hospitalization in Low-Income Youth: Longitudinal Findings from the 100 Million Brazilian Cohort." *International Journal of Epidemiology* 53, no 6 (2024): dyae153. https://doi.org/10.1093/ije/dyae153.

Valente, Jonas C. L., and Rafael Grohmann. "Critical Data Studies with Latin America: Theorizing beyond Data Colonialism." *Big Data and Society* 11, no 1 (2024): 20539517241227875. https://doi.org/10.1177/20539517241227875.

Vlajinac, Hristina. 2008. "Epidemiology, Aims and Scope". In *Encyclopedia of Public Health*, edited by W. Kirch. Dordrecht: Springer, 2008. https://doi.org/10.1007/978-1-4020-5614-7_1006.

Wamsley, Dillon and Benjamin Chin-Yee. "COVID-19, Digital Health Technology and the Politics of the Unprecedented." *Big Data and Society* 8, no 1 (2021): 20539517211019441. https://doi.org/10.1177/20539517211019441.