

United Nations Expert Group Meeting on "Methodology and lessons learned to evaluate the completeness and quality of vital statistics data from civil registration"
New York, 3-4 November 2016

Experience with record linkage validation studies in Brazil at national and subnational levels

Eliane de Freitas Drumond

A. A tentative chronology of civil registration in Brazil

When the Republic of Brazil was proclaimed in 1889, the state took over the civil registration of individuals and recording births, deaths, and marriages was no longer a function of the Catholic Church. These statistics were collected instead by the General Statistics Board and later the Demographic Statistical Office, Morality and Politics, as part of the Ministry of Justice. Over several periods of Brazilian history, civil registration went through numerous modifications and was coordinated by a variety of bodies. The current civil registration system resulted from an Act of 1975 which transferred the responsibility for the processing of vital statistics from the Ministry of Justice to the Ministry of Planning, through the Brazilian Institute of Geography and Statistics (IBGE, acronym in Portuguese). Civil Registry used to charge for birth and death certificates until 1997, which led to sub registry and delayed registration, especially among the poorest people living on the poorest areas of the country.

In what concerns births and deaths records, the initial official procedure required only the presence of two witnesses. Medical information about the causes of death was not available on those records. Data collection was dispersed and lacked standardization, and information on births and deaths was incomplete and untimely, all of which constituted obstacles for effective planning. In 1973, the Brazilian Ministry of Health introduced a standardized Death Declaration (DO, acronym in Portuguese) to be filled in by doctors throughout the country, following the World Health Organization (WHO) recommendations. The DO contains socio-demographic and health variables, including the causes of death attested by the doctor and the mortality information therein contained serves as the basis for the Mortality Information System (SIM, acronym in Portuguese). Pursuant to Brazilian laws, there can be no burial or cremation before a Death Certificate is issued by a notary, who must have access to the Death Declaration signed, whenever possible, by a doctor. This means that the Death Declaration is a requirement for the Death Certificate, which has a legal function (Mello-Jorge and others, 2002).

The SIM was computerized in 1979 and became the first health information system with nationwide coverage in Brazil. Its main purpose is to provide faster and better mortality information from the medical information collected, processed, and analyzed by health professionals under the general coordination of the Brazilian Ministry of Health. The creation of a standard document containing medical records as input for a national electronic health database was a crucial step to improve the vital statistics in the country. The SIM allows health policy managers to make epidemiological analyses that contribute to improvements in efficiency and effectiveness of the public health services (Mello-Jorge and others (2002)). It also became a starting point for scientific publications using data linkage in health systems.

B. The Brazilian National Health Information Systems

With the advances in Information Technology and its greater availability in Brazil, and considering the successful experience of the SIM, the path for new health information systems was cleared. Unfortunately, these new systems were conceived based solely on each specific purpose. This means that the technical staff responsible for creating them was unconcerned with the possibility of interchanging information between the systems and so there is no common identifying variable that allows record linkage.

The National Information System on Live Births (SINASC, acronym in Portuguese) was established in these circumstances in the early 1990s to gather information on births throughout the country. The data source is the

Declaration of Live Birth (DN, acronym in Portuguese), standardized by the Ministry of Health, which contains 41 variables, such as: the duration of pregnancy, newborn weight, mother's age, place of occurrence, and form of delivery (Mello-Jorge and others, 1993).

The Notification Diseases Information System (SINAN, acronym in Portuguese) also was developed in the early 90s. Its purpose is to collect and process data related to diseases whose notification is compulsory in the country, including infectious diseases like meningitis, dengue, tuberculosis, and HIV/AIDS, as well as serious injuries and deaths related to work accidents and domestic violence against women. By providing information for analyses on morbidity and mortality profiles, it contributes to the decision-making processes at the local, state, and federal levels.

An additional two national health systems were launched in Brazil around the same time. Their original purpose was to compile records with information regarding actual hospital and staff procedures and expenditures in the service of outpatients and inpatients and serve as basis for payments made to these health facilities by the Brazilian Public Health System (SUS, acronym in Portuguese). The Information System of Hospital Admissions in the Public Hospital System (SIH acronym in Portuguese) contains details from all hospital admissions financed by the public system, including certain patient sociodemographic and medical variables. Similarly, the Outpatient Information System (SIA, acronym in Portuguese) compiles data for payments in outpatient clinics.

A wide range of health information systems has been implemented in Brazil after this initial period. For the purpose of this paper, however, only studies using as sources SIM, SINASC, SINAN, SIH, and SIA will be considered. These systems (SINASC, SIM, SINAN, SIH, and SIA) constitute, therefore, the main source information on births, deaths, notifiable diseases, hospital admissions and/or outpatient consultations. At this point, it is important to clarify that while SIM, SINASC and SINAN are population-based information systems, SIH and SIA include only data from the public health system, which has coverage of around 75% of the Brazilian population.

At the time of their establishment, health information systems were managed centrally, mostly at the national level. The strategy of producing and disseminating information at the federal level was adopted in order to standardize and facilitate the electronic processing of data when the computerization process was only starting as the countrywide phenomenon. As the access to information technology became more widespread in the early 2000s, especially with the use of microcomputers, information systems could be decentralized. This decentralization was spurred by a belief that increased proximity with the health occurrences (births, diseases, hospitalizations, visits to outpatient clinics, etc.) would foster the introduction of local routines to improve data collection and processing. Currently, every eligible municipality is able to report these health occurrences to the state and federal levels, regardless of each municipality's success at implementing the decentralization plan originally devised by the Ministry of Health (Mello-Jorge and others, 2007).

There are more than 5,000 independent Brazilian municipalities, ranging from São Paulo, with over 15 million inhabitants and the highest socio-economic development, to Serra da Sauda de, with only 825 inhabitants and low socio-economic development. These socio-economic inequalities affect the structure of the municipal health systems, the training of human resources, and the infrastructure needed to input and process the data at the local level, all of which impact the quality of information produced. It is often then case that the teams dedicated to data processing are composed of professionals with secondary education only or, in a few cases, university level degrees (mainly nurses), and have no support from epidemiologists or statisticians. Moreover, because data processing and information centers are associated to policy-making instances, political changes that arise from local and state elections can cause a high turnover of professionals in this critical area.

The lack of stability in the workplace, coupled with the unspecialized knowledge of the greater part of the personnel involved, means that the extensive training in concepts and procedures, necessary for the appropriate inputting of data into the health information systems, is often lost as these employees are redirected elsewhere. Furthermore, unfortunately, suitable training is not available everywhere, and those responsible for feeding the system, maintaining the computational side of the databases or processing and analyzing the resulting data can lack the appropriate understanding to deal with the complexities of the health information systems. Additionally, a significant number of computers available to local staff are obsolete and incompatible with the newest versions of some databases (Laurenti and others, 2005).

On one hand, in places where these problems compound, the resulting system can show missing information (name, mother's name, sex, date of birth or death), incorrect naming of variables (miscarriage, abortion, stillbirths or live births, for example), typos, coding errors (procedures, and ICD-10 causes of death) and others. The usefulness of the health information provided in these contexts is thus limited, giving policy makers and researchers an unreliable view of the population health status, despite the high cost of collecting, processing, and analyzing the data.

On the other hand, especially in the capitals of the southern and southeastern states, the two historically better-developed regions of the country, the quality of the information systems is very satisfactory. Studies have shown progressive improvements in coverage, completeness, validity, and reliability of information on live births and mortality, particularly. Thus, the national databases have been used in scientific studies of cross-sectional, case-control, and cohort designs. The first fruit of linking records from Brazilian health information systems is the knowledge they gave us about the quality of those systems.

C. The uses of data linkage in Brazil: a look at the scientific literature

A bibliographic research was carried out in July-September 2016 to reveal the state of art of studies using national health database linking in Brazil. The source used was the Virtual Health Library, a regional bibliographic database resulting from the technical cooperation in information in health sciences conducted by BIREME/PAHO/WHO. The first research step included the use of the following descriptors: (linkage AND Brazil) OR (record linkage AND Brazil) OR (record linkage AND health information). Articles in Portuguese, Spanish, French, and English could be included and no time filter was used. A second step was added to retrieve the largest number of articles by following the references in each original article for possible inclusion. Only articles indexed in the BVS were included in this research and thus the number of articles obtained does not reflect the total production of Brazilian manuscripts using record linkage.

D. Available methods used to evaluate the completeness and quality of births and/or death records

The appropriateness of the methods used to evaluate the completeness and quality of births and/or death records in Brazil depends on factors: timeframe, place, and access to knowledge. Time becomes a factor given that, between 1970 and 1990, in order to check coverage, completeness and quality of the death registries, researchers have had to compare health mortality data with primary sources such as interviews with families and medical records. A total of ten articles was found in which researchers did manual record linkage of data to assess the cause of death (Fonseca and Laurenti, 1974; Guimarães and others, 1979; Nobre and others, 1989), magnitude of underreporting of infant deaths and the quality of the information about underlying causes of infant death (Barros and others, 1985; Carvalho and others, 1990; Duchiate and others, 1989; Niobey and others, 1990; Strozzi and others, 1985), cancer (Schnitman, 1990) and trauma (Mello-Jorge, 1982). Of the 25,616 deaths studied, 5774 (22.5%) occurred in children under the age of one. An upward trend in the rate of homicides was reported from a reliability analysis of the causes of 14,169 deaths. With the exception of one article, all researches were carried out in the most developed regions of the country (South and Southeast). The problems more frequently mentioned were the high underreporting of perinatal deaths, especially of fetal deaths, and incorrect registration of cause of death among infants. These results also point to possible failures in the health services in charge of children. Incorrect recording of causes of death from trauma, cancer, and mental illness were also observed. The authors concluded that the quality of the information system (only the SIM existed back then) needed improvement but also pointed for its potential as a source of information to orienting public health policies.

The 1990s was when SINASC, SIH and SIA were implemented and consolidated. The large amount of health data available in these national systems meant that the information therein contained needed to be assessed for quality, so as to determine each system's strengths and weaknesses. From 1991 to 2000 we found 18 articles on BVS bibliographic database, most of them using primary data sources as basis of comparison. Two studies (Almeida and Mello-Jorge, 1998; Mello-Jorge and others, 1997; Mello-Jorge and others, 1993) explored the quality of information on live births in 19,401 SINASC registries in six municipalities of the South and Southeast regions. They concluded that SINASC has excellent coverage and reliability. Two studies (Mathias and Soboll, 1998; Mendes and others, 2000) explored the reliability of the main diagnosis given at hospitalizations registered in SIH and concluded that measures to improve data quality were needed. The low

reliability of cause of death among infants as registered in SIM (Carvalho and Lynn, 1995; Mendonça and others, 1994), especially in post-neonate period, was observed in 652 deaths. The even larger underreporting of stillbirths was observed in a study of 125 deaths in a southern city (Bohland and Mello-Jorge, 1999). The worst quality of cause of post-neonatal deaths was observed again in 195 registries of infant death in a large city in the Southeast. Another study found a low concordance (32%) between the underlying causes of death among 452 neonates was obtained from data found in the SIM and the medical records in the hospital (Mendonça and others, 1994). Similarly, a low reliability among reports of underlying causes of death by cancer (Monteiro and others, 1997a; Monteiro and others, 1997b) and accidents (Drumond Jr. and others, 1999; Ladeira and Guimarães, 1998) was also observed. A single study considered the methodological challenges in the capture-recapture approach (Dunn and Andreoli, 1994).

High levels of child mortality was (and still is) one of the major public health problems in Brazil. The very first Brazilian scientific article using deterministic record linkage between SIM and SINASC was published in 1996 (Almeida and Mello-Jorge, 1996). After cleaning and blocking procedures (by date of birth and sex), records from both systems were linked using the mother's name. This study described a cohort of newborns in a city in the state of São Paulo and identified the quality of information regarding births and neonatal deaths according to: pregnancy characteristics, type of delivery, the mother's socio-demographic information, and on the newborn's health. Among the 3,225 live births, the 55 neonatal deaths had a lower coverage in SINASC. This underreporting was probably due to misunderstanding about the need register the live birth and then register the neonatal death. The second cohort study was conducted in the municipality of Goiânia, in the Center-West region (Neto and Barros, 2000). It focused on risk factors for neonatal and post-neonatal mortality by automated deterministic linking followed by manual revision of matched live births and infant death records. A total of 20,981 live births and 342 infant deaths constitute the retrospective cohort. The study showed that prematurity, low birth weight, delivery in state hospital and the mother's low level of education increase the risk of infant mortality in the neonatal and post-neonatal periods and suggests that these risk groups should be prioritized by health services. The third cohort study deterministically linking SINASC (1,187 live birth) and SIM (61 deaths) aimed at calculating infant mortality rate and the risk factors related to deaths. Almeida and others (2011) studied the quality of information in records of stillbirths, and Nakamura-Pereira and others (2013) studied hospitalization records for sentinel events of maternal near-miss. Important results of these studies on child mortality were: a) they evaluated the quality of information in each of these information systems and proposed the measures to improve them; b) they demonstrated that information missing from a system could be recovered from the other; c) they pointed to the risk factors of infant death and failures in the health service.

The article "Reclink: an application for database linkage implementing the probabilistic record linkage method", published in 2000 opened new possibilities for Brazilian research about probabilistic linkage. Reclink was the first software developed by Brazilian researchers who were interested in having it tested and thus made it available free of charge (Camargo Jr. and Coeli, 2000). Reclink went through a number of iterations and reached a stable level in version 3, which became widely used in Brazil. In 2015 the same authors published the article "Going open source: some lessons learned from the development of OpenRecLink". This article's purpose was to (1) report that Reclink was now open source, (2) implement a cross-platform version and (3) support internationalization. Another article aimed at explaining the capture-recapture method gave examples of its use and application in clinical epidemiology (Camargo Jr and Coeli, 2015).

In summary, we can say that data linkage until the early 2000s was done mostly between primary sources and one of the national health databases, and enabled improvements in information quality for studying diseases or vital statuses in a target population. Information on methodology, sensitivity, specificity and positive predictive value of the results by linkage are quite scarce in this period. It is worth noting that the use of primary sources in quality assessment of records gave way to assessments by data linkage. The good results obtained in researches of this period pointed to the requirements and the potential of deterministic and probabilistic methods of data linkage in Brazil. They also motivated researchers to then carry out methodological and evaluative studies on data relationship, as can be seen in the years 2001-2016.

This new wave of studies focusing on methodology and evaluation comprises 20 papers aimed at improving the data linkage process in Brazil. Four of these used record linkage to assess quality of data. The first validated the chain of information about hospital admissions from 2,288 household interviews and 130 records of admission from SIH (Coeli and others, 2003). A probabilistic linkage in multiple steps of 293,123 records from SIH and

223,443 records from SINASC using the mother's name and the child's birth date showed an agreement of only 68% (Coeli and Camargo Jr., 2002; Coutinho and others, 2008; Drumond and Machado, 2008). The low level of agreement obtained was related to incomplete records in SIH and problems in the linkage process, mainly (Portela and others, 1997). A higher concentration of unpaired records was observed among populations of greater social vulnerability. The low sensitivity of the probabilistic method (60.9%) used to relate 92 records of live births (from SINASC) with information obtained from interviews was assigned to a few common fields available (Coutinho and others, 2008). The capture-recapture method was successfully applied to a third study using data on diabetes mellitus surveillance in three information systems (SIM, SIH and SIA) in another study (Coeli and others, 2004). The most efficient linkage strategy was revealed by a fourth article (Coeli and Camargo Jr., 2002) to be a multistep approach with a comparison of different blocking strategies used in the pairing of 59,065 deaths (SIM) and 531 hospital admissions (SIH).

Six articles can be grouped because they use record linkage to evaluate the reusability of secondary databases for purposes others than those for which they were created (Cherchiglia and others, 2007; Coutinho and Coeli, 2006; Fonseca and others, 2010; Migowski and others, 2011; Pacheco and others, 2008; Sousa and others, 2008). The probabilistic record linkage between SIM (119,325 entries) and SIH (1,002,890 entries) using name, sex and date of birth was blindly tested twice to determine cause of death and gave a 93% agreement. The authors considered the variations in the structures of these databases and the long processing time important limitations. Another study evaluated the accuracy of probabilistic record linkage using a five-step strategy to ascertain death in survival studies with a sensitivity of 85.5%, specificity 99.4%, positive predictive value of 98.1% and positive predictive value of 94.9%. The third study described a method that combined the stages of deterministic and probabilistic linkage of patients undergoing renal replacement therapy. A total of 176,773 records were matched, reconstructing the patients' trajectories at the primary, secondary, and tertiary levels of the Brazilian public health system. A fourth study included a comparison of two probabilistic relationship strategies (in three or more steps) followed by manual review to explore the severe morbidity (near miss) and maternal (Harak and others, 2005) mortality in the capitals of all Brazilian states in 2002 and concluded that the strategies yield very similar results. Another study attempted to validate a hierarchical deterministic record linkage algorithm using data from 2,666 patients infected with HIV, 15,000 co-infected with HIV/TB and 835,066 deaths in Brazil. The study's sensibility was of almost 90% and specificity 100%. The sixth study (Fonseca and others, 2010) evaluated the accuracy of a probabilistic record linkage in identifying deaths among 559,442 cases reported to the SINAN and obtained a high sensibility (87.6%), specificity (99.6%) as well as a positive predictive value (99.2%). The authors observed a lower sensitivity for patients younger than 13 years. During this period, literature reviews on the probabilistic and deterministic (Frias and others, 2014; Nascimento and others, 2008) linkage procedures used by Brazilian researchers were also published (Coeli and others, 2015; Machado, 2004; Silveira and Artmann, 2009).

The release of the Millennium Development Goals (MDG) in 2000 focused global attention on some critical health problems. As Brazilian authorities, healthcare staff, and the scientific community all became engaged to achieve these goals raising the profiles of these health concerns in our society. Locally managed databases can speed up decision-making for faster and more effective health actions. Good quality of information, however, remains an essential and necessary condition for effectiveness. An often-encountered problem is that there is no regular evaluation plan effective nationwide to monitor the quality of data from the various health information systems. The MDG increased the need for reliable information so that healthcare professionals and researchers have intensified their efforts to evaluate and improve the quality of the records. Several actions were launched to train the sources of information, to purchase equipment and train staff. As a consequence, there was a significant increase in the number and the spectrum of scientific studies published after 2000.

Indeed, reducing child and maternal mortality is a top priority in the social health policies of the Brazilian government. Among the 17 studies related to infant mortality (Almeida and others, 2006; Almeida and others, 2002; Ferreira and others, 2011; Frias and others, 2013; Giglio and others, 2005; Guerra and others, 2008; Maia and others, 2012, 2015; Martins and Velásquez-Meléndez, 2004; Mello-Jorge and Gotlieb, 2001; Mendes and others, 2012; Nascimento and others, 2008; Pereira and others, 2013; Ribeiro and others, 2009; Santa Helena and others, 2005; Sarinho and others, 2001; Silva Antônio Augusto Moura and others, 2001; Silva and others, 2009; Silva and others, 2006; Silva and others, 2014; Theme Filha and others, 2004), 7 used deterministic data linkage to assess the coverage, completeness and validity of information on live births and deaths; 10 studied

risk factors for death and conducted a literature review on the use of linkage in the study of infant deaths. SINASC and SIM were deterministically linked in these studies to analyze a total of 424,783 live births and 4,005 infant deaths. The linkage method used proved to be very useful, enabling adequate investigation of the factors most strongly related to mortality among infants. However, there is no report on the linking parameters used in these 10 studies or the results of the classification of links or manual review of matched and unmatched pairs. It is possible that the exposure and/or the outcome variables were misclassified, thus introducing a bias in measures of association. Critical analysis of the results of various Brazilian studies using deterministic linkage concluded that their specificity is high, but that other steps would be required to increase sensitivity.

Using different linkage methods showed new possibilities, weaknesses and strengths of information systems on births and deaths; led to more reliable numbers of infant mortality rates by each component; pointed out weaknesses in prenatal care, delivery, and newborn care that allowed improved health policies to be developed and evaluated. SIM coverage was proven to be consistently better than coverage by civil registration, for the first time in 2005 (Paes, 2005).

We can also find 6 studies about maternal mortality. For many decades now, a correction factor on the number of maternal deaths has been used in Brazil to adjust the lack of information on the real number of these deaths by direct or indirect causes. A study sought to evaluate the underreporting of maternal deaths in SIM by deterministically linking these records with data from home interviews (Laurenti and others, 2004; Luizaga and others, 2010; Sousa and others, 2007). Underreporting of deaths related to maternal causes in the Brazilian capitals was found to be at (Almeida and others, 2006) 21.4% and thus a new correction factor was proposed. Maternal diseases affecting pregnancy and/or child are studied by Krauss and others (2014), and Cavalcante and others (2005).

The quality of cause of death information also continued to be evaluated. Manual data linkage between SIM and police reports of 3,643 deaths showed low concordance in cases assigned to transport accidents on land. Low agreement between SIM, SINAN, and SIH was also observed in manual data linkage of deaths by accidents with poisonous animals (4,418 snakes, 2,266 spiders and 5,689 scorpions) (Bochner and Struchiner, 2002). Underreporting of deaths in work accident in SIM was reported by three studies in capitals of the Southeast (Bordoni and others, 2016; Correa and Assunção, 2003; Silva and Drumond, 2013). Coverage and quality of information were also assessed in studies of land transportation accidents (Andrade and Mello-Jorge, 2001; Paixão and others, 2015) as well in other sources of trauma (Gawryszewski and others, 2004; Melione and Mello-Jorge, 2008; Mello-Jorge and others, 2012). The capture-recapture method was the most frequently used in studies linking SIM, SINAN, and police reports. The increase observed in deaths occurring on the way to work is one of the serious consequences of violence in urban traffic and increased motorcycle use in Brazil. Occupation-related homicides (as in the case of watchmen) were the most underreported.

Adequate surveillance of notifiable diseases remains a major public health concern in Brazil given their high incident and prevalence. The health information system designed to be source of information for control and monitoring actions, SINAN, observes a very complex workflow that often hinders the very actions it is supposed to support. SINAN records information on multiple stages in the process of notifiable diseases, processing the moment of notification, of monitoring, and of closure, thus requiring continued input for every case. Due to the high number of annual notifications, data treatment (such as evaluations of completeness, timeliness and duplications) at local or national levels is neither quick nor easy. The workflow is particularly complex for diseases that evolve chronically and that lead patients to search for care continually and often at primary, secondary and tertiary levels. In this context, studies evaluating the underreporting of notifiable diseases and the closing of cases as deaths in SINAN are frequent. Probabilistic record linkage of 16,567 cases in SINAN and 1,146 deaths in SIM demonstrated there was undercounting of tuberculosis cases in SINAN, besides the occurrence of late diagnosis and failure to comply with protocol treatment. Other two similar studies (with a total 548,199 cases and 7,077 deaths) observed almost 40% of underreporting in SINAN, higher among serious cases of tuberculosis. Both highlighted the need to improve surveillance and monitoring of cases. Probabilistic linkage of SIM and SINAN and SIH using name, date of birth and address showed a significant underreporting of TB in SINAN among older patients or children aged less than 15 years (Bartholomay and others, 2014; Façanha, 2005; Lindoso and others, 2008; Oliveira and others, 2016; Oliveira and others, 2012; Selig and others, 2004; Selig and others, 2010; Sousa and Pinheiro, 2011). The aforementioned operational difficulties intrinsic to SINAN meant that records related to tuberculosis, in particular, were found to have many

duplications, very incomplete filling of variables and underreporting, as shown by four different studies evaluating a total of 32,162 deaths. The authors advocate that probabilistic linkage of SIM and SINAN be used routinely in TB surveillance and HIV/AIDS co-infection (Ferreira and others, 2000; Fonseca and others, 2010; Lemos and Valente, 2001; Miranda and others, 2009; Prado and others, 2011). A methodological study was carried out to evaluate the use of deterministic linkage with a merge process using data on HIV/AIDS from SINAN and SIM, and found that there was a 25% underreporting of deaths in SINAN and that in 12% of deaths in SIM, HIV/AIDS was left out as a cause of death, thus pointing to need to improve data sources and surveillance of HIV/AIDS (Pacheco and others, 2011). Other diseases studied include Hansen's, leptospirosis, and dengue (Brum and Kupek, 2005; Duarte and França, 2006; Façanha and others, 2006).

The publications using linkage showed a very strong increase over time in Brazil and there were important differences in the number of studies each region contributed and the database used. The Southeast region contributed with 80% all studies and majority used the mortality database SIM. Most of these articles can be found in two traditional Brazilian journals dedicated to public health (*Cadernos de Saúde Pública (Rio de Janeiro)*, edited in Rio de Janeiro, and *Revista de Saúde Pública (São Paulo)*, edited in São Paulo). The authors defined the methodology predominately as cross-sectional, followed by cohort and case-control studies. Methodological studies mainly devoted themselves to probabilistic linkage. The majority of authors work at universities but some worked at surveillance levels inside the local or state health service institutions. Medical records (MR) were crucially used whenever a new information system was first put into operation or whenever its performance in a particular setting was put into doubt. The results found were not (and still are not, in some cases) doubted, and many true pairs failed to be registered. This failure to find true pairs can result in severe bias in situations involving vulnerable populations, for which pairing errors can be less prejudicial than the exclusion of true pairs left unmatched. The 2000s saw an improvement in the circumstances for dissemination and consolidation of the available epidemiological and administrative databases, matched by a drive to link data across them. In every study found, authors acknowledged the validity of linkage as a tool for better and more widespread use of health information systems, as well as to develop grounded criticism of their quality. These papers show that results obtained from linked data have epidemiological consistency and are theoretically plausible.

E. The lessons learned from the experiences of using linkage methods

Given that information is sometimes missing from individual records in any one system, the possibility of filling these gaps with information from other health databases through linkage was very well received by health managers and researchers, who use local databases as easily available sources to respond fast and consistently to the demands of the health policy makers and the public. This work, however, is complicated by the absence of a common variable that could make deterministic linkage more accurate, as well as the use of different formats for displaying information, purposes, administrators and timeframes. There is no information on the extent to which deterministic record linkage is used in Brazil for surveillance purposes.

It is therefore crucial that researchers and those involved in surveillance have the appropriate knowledge to reflect and evaluate critically the results they obtain after linkage, regardless of the method chosen for the pairings. The possibility of procedural errors that affect sensitivity, specificity and predictive values of the linkage performed must be considered and include both inappropriate pairings (when siblings are matched as the same individual, that pair must be broken down during manual debugging) and unpairings (when names are misspelt or events misclassified). These errors are more likely to occur among socially vulnerable populations, such as afrodescendants and inhabitants of shanty towns or "favelas" because common fields are often left blank or incomplete, and many authors have pointed out that the bad quality of information is, most of the time, related to bad quality of care.

First names pose a particular challenge: in addition to typos, certain "foreign" names are more frequent among the lower social classes and present a variety of possible spellings. The English name "Michael" can be rendered as "Maicou", "Maicol", "Maicon", "Maykon", and so on. Given that the deterministic relationship strategy more often used in Brazil as a unique identifier is the patient's name, the incorrectness of the information compromises the relationship process. Exact matching has only two outcomes: true or false pair. In most of

cases, the deterministic relationship was "all or nothing", i.e. absolute agreement of all digits or characters of the identifier (key match). These true pair matchings have a high specificity, but leaves out partial agreement. In order to increase sensitivity, the researcher must conduct prior data cleaning to reduce the heterogeneity or proceed with cleaning and reattempt linkage or adopt some form of probabilistic matching. Additional steps to increase sensitivity, however, were not reported in the majority of studies published in Brazil. While simple to understand and implement, the deterministic strategy can become laborious and time consuming when it involves subjective decisions made at each step. Moreover, there is precious little written on deterministic linkage methodologies in Portuguese, which can account for its uncritical use in Brazil. Official papers by the Ministry of Health are of great value in the training of public health professionals and having easy and free access to publications describing methodologies could help these professionals think more critically about their results and adopt measures that can increase the sensitivity of the deterministic relationships they perform.

The health information systems available in Brazil vary greatly because of the differing administration purposes for which they were created and thus while some are patient-oriented, other are procedure-oriented, some are population-based, others are limited to specific populations, for example, those accessing the public health system only, and there is no unique identifier across them. Information systems such as SIH and SIA, whose main purpose is management and financing, use procedure logs from in-patient and outpatient clinics so that any given individual can have multiple records in the database, even for the same hospitalization. Making these systems patient-centered would be an important step in revealing the links between the various stages on that individual's life, from birth to hospitalizations and finally death. It would also be crucial to reveal the burden of health disparities in different socioeconomic groups as well as to focus on the surveillance of clinical subgroups. If linkage could be done routinely, assessments of the health systems become a possibility, for example, using different outcomes such as different infant mortality rates by public and private facility.

Furthermore, data sources are often managed by different sectors, which then claim ownership of data. It means that the process of link may result difficult and compromised. The channels of communication between these different sectors may not always be open and identifying common goals that would make the sharing of knowledge, as well as of data, possible. Organizational obstacles include fear of criticism about the quality of the information system, which can block efforts to merge data sources. Certain sectors are also more involved with the processes needed for improving data quality, including data production, analysis, and assessment. Others, however, are only responsible for collecting data and, unsurprisingly, those are often the sectors less willing to share their databases. There are many reasons to justify reluctance to share their data; commonly used linkage variables are personal identifying information such as name, mother's name, date of birth, sex, and address. Just as in other countries, the matter of privacy is also being discussed in Brazil. Concerns related to confidentiality and the patients' right to privacy are behind certain requests such as the need to obtain consent from each individual whose personal data are included – something that is operationally impossible when dealing with large databases. If, on the one hand, too strict an individual consent policy can mean that data that is crucial for improving the entire health system may become inaccessible and thus hamper affect the collectivity, on the other there must be mechanisms that guarantee rigorous ethical, technological and methodological control for data protection. The benefits of linkage must be carefully weighed against these risks and the potential scientific and social worth of such use clearly demonstrated. Users of the health system have been increasingly concerned with the possibility of confidentiality breaching and exposure of their personal information. While proposals such as the creation of data linkage units, cryptography and others are still far from our reality, confidentiality and protection of health-related data must be more intensively discussed by all of those involved in health systems, be it as managers, as staff, as healthcare providers, or as users, so that a consensus can be reached.

A serviceable, secure, and air gapped computing platform is essential, especially for research using large datasets. The team and expertise required for linkage may vary according the complexity of the research project; more complex projects require more experience and expertise from the team. Ideally, a research and data support team will already be in place before the proposal is processed. In addition to the quality of the database used, the linkage process involves decisions that can affect its validity and whether or not its conclusions can be generalized. For those reasons, the methodology must be clear and accuracy measures explicit. The biases introduced when linkages are incomplete or when results are inaccurate may influence the willingness of sectors and institutions to adopt routine linkage.

Publications with methodological proposals and critical evaluations of probabilistic are more abundant and available in Portuguese. The possibility of using the Reclink software without costs boosted the use of the methodology in Brazil. Most work undertaken in Brazil using probabilistic linkage is based on the model proposed by Fellegi and Sinter in 1969 and follow steps such as standardization, blocking, pairing, decision rules, and manual review. The rules that govern the linkage and the accuracy measures are more clearly stated in articles that employ probabilistic relationships than in those that use deterministic linkage.

In conclusion, while the use of record linkage has considerably increased both in academic and service contexts, more widespread recognition of this methodology could certainly improve health surveillance, especially in relation to perinatal and maternal outcomes, and TB. We also noticed a trend toward probabilistic methodologies in academic works, which are, regardless of approach, more easily quantified in terms of bibliography, in opposition to the more common use of deterministic approaches in the management instances of the public health system in Brazil. When dealing with large datasets such as those of SIM and SINASC, probabilistic linkage could be a better option, but that would require adequate support, including that which can be offered by cooperation with other institutions and academics. For many researchers, the quality of the information contained in the database remains the culprit for low accuracy, and there is relatively little reflection on the linkage process itself. The use of a sample of links whose status is determined by manual review could also be used in an attempt to determine a gold standard. Adequate use of linkage methodologies can be disseminated with professional training, inclusion of epidemiologists and statisticians to health teams, and investments in IT.

References

- Almeida, M. F., G. P. Alencar, I. França, H. M. D. Novaes, A. A. F. Siqueira, D. Schoeps, L. C. Rodrigues (2006). Validation of birth certificates based on data from a case-control study. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 22, No. 3, pp. 643.
- Almeida, M. F., G. P. Alencar, D. Schoeps, E. G. Minuci, Z. P. Silva, L. P. Ortiz, P. C. Santos (2011). Qualidade das informações registradas nas declarações de óbito fetal em São Paulo, SP. *Revista de Saúde Pública* (São Paulo), vol. 45, No. 5, pp. 845-853.
- Almeida, M. F. & M. H. P. Mello-Jorge (1996) The use of the linkage technique of information systems in cohort studies on neonatal mortality. *Revista de Saúde Pública* (São Paulo), vol. 30, No. 2, pp. 141.
- Almeida, M. F. & M. H. P. Mello-Jorge (1998). Pequenos para idade gestacional: fator de risco para mortalidade neonatal. *Revista de Saúde Pública* (São Paulo), vol. 32, No. 3, pp. 217-224.
- Almeida, M. F., H. M. D. Novaes, G. P. Alencar, L. C. Rodrigues (2002). Neonatal mortality: socio-economic, health services risk factors and birth weight in the City of São Paulo. *Revista Brasileira de Epidemiologia* (São Paulo), vol. 5, No. 1, pp. 93-107.
- Andrade, S. M. & M. H. P. Mello-Jorge (2001). Traffic accidents in a city in Southern Brazil: an evaluation of coverage and quality of data. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 17, No. 6, pp. 1449-1456.
- Barros, F. C., C. G. Victoria, A. M. B. Teixeira, M. Puerto Filho (1985). Mortalidade perinatal e infantil em Pelotas, Rio Grande do Sul: nossas estatísticas são confiáveis? *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 1, No. 3, pp. 348-358.
- Bartholomay, P., G. P. Oliveira, R. S. Pinheiro, A. M. N. Vasconcelos (2014). Melhoria da qualidade das informações sobre tuberculose a partir do relacionamento entre bases de dados. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 30, No. 11, pp. 2459-2470.
- Bochner, R., & C. J. Struchiner (2002). Recording of venomous bites and stings by National Information Systems in Brazil. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 18, No. 3, pp. 735-746.
- Bohland, A. K. & M. H. P. Mello-Jorge (1999). Mortality among children of less than one year of age in a region of Brazil. *Revista de Saúde Pública* (São Paulo), vol. 33, No. 4, pp. 366-373.

- Bordoni, P. H. C., L. S. Bordoni, J. M. Silva, E. F. Drumond (2016). Utilização do método de captura-recaptura de casos para a melhoria do registro dos acidentes de trabalho fatais em Belo Horizonte, Minas Gerais, 2011. *Epidemiologia e Serviço de Saúde*(Brasília), vol. 25, No. 1, pp. 1-10.
- Brum, L., & E. Kupek (2005). Record linkage and capture-recapture estimates for underreporting of human leptospirosis in a Brazilian health district. *The Brazilian journal of infectious diseases: an official publication of the Brazilian Society of Infectious Diseases*(São Paulo), vol. 9, No. 6, pp. 515.
- Camargo Jr, K. R., & C. M. Coeli (2015). Going open source: some lessons learned from the development of OpenRecLink. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 31, No. 2, pp. 257.
- Camargo Jr., K. R. & C. M. Coeli (2000). Reclink: an application for database linkage implementing the probabilistic record linkage method. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 16, No., pp. 439-447.
- Carvalho, M. L., & S. D. Lynn (1995). Reliability of reported underlying causes of neonatal death: implications for the study of preventable mortality. *Revista de Saúde Pública* (São Paulo), vol. 29, No. 5, pp. 342-348.
- Carvalho, M. L., F. M. L. Niobey, N. N. Miranda, P. C. Sabroza (1990). Agreement as to the determination of the basic cause of death among children of under one year of age in Metropolitan Region of Rio de Janeiro, Brazil, 1986. *Revista de Saúde Pública* (São Paulo), vol. 24, No. 1, pp. 20-27.
- Cavalcante, M. S., A. N. Ramos Jr, L. R. S. K. Pontes (2005). Relacionamento de sistemas de informação em saúde: uma estratégia para otimizar a vigilância das gestantes infectadas pelo HIV. *Epidemiologia e Serviço de Saúde* (Brasília), vol. 14, No. 2.
- Cherchiglia, M. L., A. A. G. Júnior, E. I. G. I. Andrade, C. J. Machado, F. A. Acúrcio, W. M. Júnior, O. V. Queiroz (2007). A construção da base de dados nacional em Terapia Renal Substitutiva (TRS) centrada no indivíduo: aplicação do método de linkage determinístico-probabilístico. *Revista Brasileira de Estudos de População*(Rio de Janeiro), vol. 24, No. 1, pp. 163-167.
- Coeli, C. M., R. Blais, M. C. E. Costa, & L. M. Almeida. (2003). Probabilistic linkage in household survey on hospital care usage. *Revista de Saúde Pública* (São Paulo), vol. 37, No. 1, pp. 91-99.
- Coeli, C. M. & K. R. Camargo Jr (2002). Evaluation of different blocking strategies in probabilistic record linkage. *Revista Brasileira de Epidemiologia*(São Paulo), vol. 5, No. 2, pp. 185-196.
- Coeli, C. M., E. S. F. Coutinho, R. P. Veras (2004). The challenge of applying capture-recapture methodology to diabetes mellitus surveillance in the elderly: lessons from a Brazilian experience. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 20, No. 6, pp. 1709-1720.
- Coeli, C. M., R. S. Pinheiro, K. R. Camargo Jr (2015). Conquistas e desafios para o emprego das técnicas de record linkage na pesquisa e avaliação em saúde no Brasil. *Epidemiologia e Serviço de Saúde*(Brasília), vol. 24, No. 4, pp. 795-802.
- Correa, P. R. L. & A. A. Assunção (2003). A subnotificação de mortes por acidentes de trabalho: estudo de três bancos de dados. *Epidemiologia e Serviço de Saúde* (Brasília), vol. 12, No. 4.
- Coutinho, E. S. F. & C. M. Coeli (2006). Accuracy of the probabilistic record linkage methodology to ascertain deaths in survival studies. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 22, No. 10, pp. 2249.
- Coutinho, R. G. M., C. M. Coeli, E. Faerstein, D. Chor (2008). Sensitivity of probabilistic record linkage for reported birth identification: Pró-Saúde Study. *Revista de Saúde Pública* (São Paulo), vol. 42, No. 6, pp. 1097.
- Drumond, E. F. & C. J. Machado (2008). Linkage entre registros do Sihsus e SINASC: possíveis vieses decorrentes do não-pareamento. *Revista Brasileira Estudos da Populacao*, vol. 25, No. 1.
- Drumond Jr., M., M. M. T. A. Lira, M. Freitas, T. M. V. Nitrini, K. Shibaó (1999). Evaluation of the quality of mortality information by unspecified accidents and events of undetermined intent in a metropolis of Southeastern Brazil. *Revista de Saúde Pública* (São Paulo), vol. 33, No. 3, pp. 273-280.

- Duarte, H. H. P. & E. B. França (2006). Qualidade dos dados da vigilância epidemiológica da dengue em Belo Horizonte, MG. *Revista de Saúde Pública* (São Paulo), vol. 40, No. 1, pp. 134-142.
- Duchiade, M. P., M. L. Carvalho, M. C. Leal (1989). As mortes em domicílio de menores de um ano na região metropolitana do Rio de Janeiro em 1986: um evento-sentinela na avaliação dos serviços de saúde. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 5, No. 3, pp. 251-263.
- Dunn, J. & S. B. Andreoli (1994). Capture-recapture: a new methodology for epidemiological research. *Revista de Saúde Pública* (São Paulo), vol. 28, No. 6, pp. 449-453.
- Façanha, M. C. (2005). Tuberculosis: underreporting of tuberculosis cases that resulted in death in Fortaleza-CE. *Revista Brasileira de Epidemiologia*(São Paulo), vol. 8, No. 1, pp. 25-30.
- Façanha, M. C., A. C. Pinheiro, J. R. C. Lima, M. L. L. T. Ferreira, G. F. D. Teixeira, M. Z. Rouquayrol (2006). Leprosy: underreported cases in Fortaleza - Ceará, Brazil. *Anais Brasileiros de Dermatologia*(São Paulo), vol. 81, No. 4, pp. 329-333.
- Ferreira, J. S. A., M. B. R. Vilela, P. S. Aragão, R. A. Oliveira, R. F. Tiné (2011). A Evaluation of the quality of information: linkage between SIM and SINASC in Jaboatão dos Guararapes, Pernambuco State. *Ciência & Saúde Coletiva*(Rio de Janeiro), vol. 16, No., pp. 1241-1246.
- Ferreira, V. M. B., M. C. Portela, M. T. I. Vasconcellos (2000). Variables associated with underreporting of AIDS patients, Brazil, 1996. *Revista de Saúde Pública* (São Paulo), vol. 34, No. 2, pp. 170-177.
- Fonseca, L. A. M. & R. Laurenti. (1974). The quality of medical certification for the cause of death in S. Paulo, Brazil. *Revista de Saúde Pública* (São Paulo), vol. 8, No. 1, pp. 21-29.
- Fonseca, M. G. P., C. M. Coeli, F. F. A. Lucena, V. G. Veloso, M. S. Carvalho (2010). Accuracy of a probabilistic record linkage strategy applied to identify deaths among cases reported to the Brazilian AIDS surveillance database *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 26, No. 7, pp. 1431-1438.
- Frias, P. G., C. L. Szwarcwald, P. R. B. S. Junior, W. S. Almeida, P. I. C. Lira (2013). Correção de informacoes vitais: estimacão da mortalidade infantil, Brasil, 2000-2009. *Revista de Saúde Pública* (São Paulo), vol. 47, No. 6, pp. 1048-1058.
- Frias, P. G., C. L. Szwarcwald, P. I. C. Lira (2014). Avaliação dos sistemas de informações sobre nascidos vivos e óbitos no Brasil na década de 2000. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 30, No. 10, pp. 2068-2280.
- Gawryszewski, V. P., M. S. Koizumi, M. H. P. Mello-Jorge (2004). Morbidity and mortality from external causes in Brazil, 2000. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 20, No. 4, pp. 995-1003.
- Giglio, M. R. P., J. A. Lamounier, O. L. M. Neto, C. C. César (2005). Low birth weight in a cohort of newborns in Goiânia-Brazil in 2000. *Revista Brasileira de Ginecologia e Obstetrícia*(Rio de Janeiro), vol. 27, No. 3, pp. 130-136.
- Guerra, F. A. R., J. C. Lerena Jr, S. G. N. Gama, C. B. Cunha, M. M. Theme Filha (2008). Birth defects in Rio de Janeiro, Brazil: an evaluation through birth certificates (2000-2004). *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 24, No. 1, pp. 140-149.
- Guimarães, C., J. M. P. Souza, M. H. P. Mello-Jorge, R. Laurenti, S. L. D. Gotlieb, A. H. Santo, .L. S. C. Júnior (1979). Mortalidade de adultos de 15 a 74 anos de idade em São Paulo, Botucatu e São Manuel (Brasil), 1974/1975. *Revista de Saúde Pública* (São Paulo), vol. 13, No., pp. 1-73.
- Harak, C. A. C., S. L. D. Gotlieb, R. Laurenti. (2005). Accuracy of the Mortality Data System in a city in the south of the State of São Paulo, Brazil. *Revista Brasileira de Epidemiologia*(São Paulo), vol. 8, No. 1, pp. 19-24.
- Krauss, P. P., C. L. Vieira, J. F. C. Santos, L. A. Lima, L. F. Legay, G. M. Lovisi (2014). Avaliação de desfechos perinatais/ infantis em partos de pacientes com transtornos mentais maiores de um hospital psiquiátrico do Rio de Janeiro, Brasil. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 30, No. 8, pp. 1654-1666.
- Ladeira, R. M. & M. D. C. Guimarães (1998). Analysis of agreement between coding of underlying cause of death by road accidents. *Revista de Saúde Pública* (São Paulo), vol. 32, No. 2, pp. 133-137.
- Laurenti, R., M. H. P. Mello-Jorge, S. L. D. Gotlieb(2004). Maternal mortality in Brazilian State Capitals: some characteristics and estimates for an adjustment factor. *Revista Brasileira de Epidemiologia*(São Paulo), vol. 7, No. 4, pp. 449-460.

- Laurenti, R., M. H. P. Mello-Jorge, M. L. Lebrão, S. L. D. Gotlieb, M. F. Almeida (2005). Vital statistics: counting births and deaths. *Revista Brasileira de Epidemiologia*(São Paulo), vol. 8, No. 2, pp. 108-110.
- Lemos, K. R. V. & J. G. Valente (2001). Death certificates as a marker for under-recording of AIDS cases. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 17, No. 3, pp. 617-626.
- Lindoso, A. A. B. P., E. A. Waldman, N. K. Komatsu, S. M. Figueiredo, M. Taniguchi, L. C. Rodrigues (2008). Profile of tuberculosis patients progressing to death, city of São Paulo, Brazil, 2002. *Revista de Saúde Pública* (São Paulo), vol. 42, No. 5, pp. 805-812.
- Luizaga, C. T. M., S. L. D. Gotlieb, M. H. P. d. M. Jorge, R. Laurenti (2010). Maternal deaths: reviewing the adjustment factor for official data. *Epidemiologia e Serviços de Saúde*(Brasília), vol. 19, No., pp. 8-15.
- Machado, C. J. (2004). A literature review of record linkage procedures focusing on infant health outcomes *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 20, No. 2, pp. 362-371.
- Maia, L. T. S., W. V. Souza, A. C. G. Mendes (2012). Differences in risk factors for infant mortality in five Brazilian cities: a case-control study based on the Mortality Information System and Information System on Live Births. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 28, No. 11, pp. 2163-2176.
- Maia, L. T. S., W. V. Souza, A. C. G. Mendes (2015). A contribuicao do linkage entre o SIM e SINASC para a melhoria das informacoes da mortalidade infantil em cinco cidades brasileiras. *Revista Brasileira de Saude Materno Infantil*(Recife), vol. 15, No. 1, pp. 57-66.
- Martins, E. F. & G. Velásquez-Meléndez (2004). Determinants of neonatal mortality in a cohort of born alive infants, Montes Claros, Minas Gerais, 1997-1999. *Revista Brasileira de Saude Materno Infantil*(Recife), vol. 4, No. 4, pp. 405-412.
- Mathias, T. A. F. & M. L. M. S. Soboll (1998). Reliability of diagnoses on authorization forms for hospital admittance. *Revista de Saúde Pública* (São Paulo), vol. 32, No. 6, pp. 526-532.
- Melione, L. P. R. & M. H. P. Mello-Jorge (2008). Data reliability of hospital admissions due to external causes in a public hospital in São José dos Campos, São Paulo, Brazil. *Revista Brasileira de Epidemiologia*(São Paulo), vol. 11, No. 3, pp. 379-392.
- Mello-Jorge, M. H. P. (1982). Mortality from violent causes in the city of S. Paulo, Brazil: IV - The situation in 1980. *Revista de Saúde Pública* (São Paulo), vol. 16, No. 1, pp. 19-41.
- Mello-Jorge, M. H. P., A. M. Cascão, A. C. Reis, R. Laurenti (2012). Em busca de melhores informações sobre a causa básica do óbito por meio de linkage: um recorte sobre as causas externas em idosos - Estado do Rio de Janeiro, Brasil, 2006. *Epidemiologia e Servico de Saúde*(Brasília), vol. 21, No. 3, pp. 407-418.
- Mello-Jorge, M. H. P., S. L. D. Gotlieb, S. M. Andrade(1997). Analysis of legal registration of live births in an urban area of Southern Brazil. *Revista de Saúde Pública* (São Paulo), vol. 31, No. 1, pp. 78-89.
- Mello-Jorge, M. H. P., S. L. D. Gotlieb, R. Laurenti (2002). The national mortality information system: problems and proposals for solving them I - Deaths by natural causes. *Revista Brasileira de Epidemiologia*(São Paulo), vol. 5, No. 2, pp. 197-211.
- Mello-Jorge, M. H. P., S. L. D. Gotlieb, M. L. Soboll, M. F. Almeida, M. R. D. O. Latorre. (1993). Avaliação do sistema de informação sobre nascidos vivos e o uso de seus dados em epidemiologia e estatísticas de saúde. *Revista de Saúde Pública* (São Paulo), vol. 27, No. suppl, pp. 1-46.
- Mello-Jorge, M. H. P., R. Laurenti, S. L. D. Gotlieb (2007). Quality analysis of Brazilian vital statistics: the experience of implementing the SIM and SINASC systems. *Ciência & Saúde Coletiva*(Rio de Janeiro), vol. 12, No. 3, pp. 643-654.
- Mello-Jorge, M. H. P. d., S. L. D. Gotlieb (2001). O Sistema de Informação de Atenção Básica como fonte de dados para os Sistemas de Informações sobre Mortalidade e sobre Nascidos Vivos. *Informe Epidemiológico do SUS*(Brasília), vol. 10, No. 1, pp. 7-18.
- Mendes, A. C. G., M. M. Lima, D. A. Sá, L. C. S. Oliveira, L. T. S. Maia (2012). The use of the interrelation of data bases to improve information on child mortality in municipalities in the Brazilian State of Pernambuco. *Revista Brasileira de Saude Materno Infantil*(Recife), vol. 12, No. 3, pp. 243-249.

- Mendes, A. C. G., J. B. Silva Junior, K. R. Medeiros, T. M. Lyra, D. A. Melo Filho, D. A. Sá (2000). Evaluation of the hospital information system (SIH/SUS) as a complementary information source for surveillance and monitoring of notifiable diseases. *Informe Epidemiológico do SUS*(Brasília), vol. 9, No. 2, pp. 67-86.
- Mendonça, E. F., E. M. A. Goulart, J. A. D. Machado (1994). Reliability of the medical certificates of underlying cause of infant deaths in a metropolitan region of southeastern Brazil. *Revista de Saúde Pública* (São Paulo), vol. 28, No. 5, pp. 385-391.
- Migowski, A., R. B. M. Chaves, C. M. Coeli, A. L. P. Ribeiro, B. R. Tura, M. C. C. Kuschnir, R. M. A. Xavier (2011). Accuracy of probabilistic record linkage in the assessment of high-complexity cardiology procedures. *Revista de Saúde Pública* (São Paulo), vol., No., pp. 0-0.
- Miranda, A. E., J. E. Golub, F. F. Lucena, E. N. Maciel, M. F. Gurgel, R. Dietze (2009). Tuberculosis and AIDS co-morbidity in Brazil: linkage of the tuberculosis and AIDS databases. *The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases*(Salvador), vol. 13, No. 2, pp. 137.
- Monteiro, G. T. R., R. J. Koifman, S. Koifman (1997a). Confiabilidade e validade dos atestados de óbito por neoplasias. I. Confiabilidade da codificação para o conjunto das neoplasias no Estado do Rio de Janeiro. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 13, No. suppl.1, pp. 39-52.
- Monteiro, G. T. R., R. J. Koifman, S. Koifman (1997b). Reliability and accuracy of reported causes of death from cancer. II. Accuracy of stomach cancer reported in the municipality of Rio de Janeiro County, Brazil. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 13, No., pp. S53-S65.
- Nakamura-Pereira, M., W. Mendes-Silva, M. A. B. Dias, M. E. Reichenheim, G. Lobato (2013). Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH-SUS): uma avaliação do seu desempenho para a identificação do near miss materno. *Cadernos de Saúde Pública*(Rio de Janeiro), vol. 29, No. 7, pp. 1333-1345.
- Nascimento, E. M. R., M. C. N. Costa, E. L. A. Mota, J. S. Paim (2008). Investigation of risk factors for infant mortality by linking health databases. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 24, No. 11, pp. 2593-2602.
- Neto, O. L. M. & M. B. A. Barros (2000). Risk factors for neonatal and post- neonatal mortality in the Central-West region of Brazil: linked use of life birth and infant death records. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 16, No. 2, pp. 477-485.
- Niobey, F. M. L., A. M. Cascão, M. P. Duchiate, P. C. Sabroza (1990). The quality of the filling - in of death certificates of children below one year of age in the metropolitan region of Rio de Janeiro. *Revista de Saúde Pública* (São Paulo), vol. 24, No. 4, pp. 311-318.
- Nobre, L. C., C. G. Victora, F. C. Barros, C. Lombardi, A. M. B. Teixeira, S. C. Fuchs (1989). Avaliação da qualidade da informação sobre a causa básica de óbitos infantis no Rio Grande do Sul (Brasil). *Revista de Saúde Pública* (São Paulo), vol. 23, No. 3, pp. 207-213.
- Oliveira, G. P., A. L. S. Bierrenbach, K. R. Camargo, C. M. Coeli, R. S. Pinheiro (2016). Accuracy of probabilistic and deterministic record linkage: the case of tuberculosis. *Revista de Saúde Pública* (São Paulo), vol. 50, No., pp. 49.
- Oliveira, G. P., R. S. Pinheiro, C. M. Coeli, D. Barreira, S. B. Codenotti (2012). Mortality information system for identifying underreported cases of tuberculosis in Brazil. *Revista Brasileira de Epidemiologia*(São Paulo), vol. 15, No. 3, pp. 468-477.
- Pacheco, A. G., V. Saraceni, S. H. Tuboi, L. H. Moulton, R. E. Chaisson, S. C. Cavalcante, L. H. Harrison (2008). Validation of a hierarchical deterministic record-linkage algorithm using data from 2 different cohorts of human immunodeficiency virus-infected persons and mortality databases in Brazil. *American Journal of Epidemiology*(Baltimore), vol. 168, No. 11, pp. 1326.
- Pacheco, A. G., V. Saraceni, S. H. Tuboi, L. M. Lauria, L. H. Moulton, J. C. Faulhaber, M. Schechter (2011). Estimating the extent of underreporting of mortality among HIV-infected individuals in Rio de Janeiro, Brazil. *AIDS research and human retroviruses*, vol. 27, No. 1, pp. 25.
- Paes, N. A. (2005). Assessment of completeness of death reporting in Brazilian states for the year 2000. *Revista de Saúde Pública* (São Paulo), vol. 39, No. 6, pp. 882-890.

- Paixão, L. M. M. M., E. D. Gontijo, S. A. Mingoti, D. A. d. S. Costa, A. A. d. L. Friche, W. T. Caiaffa. (2015). Urban road traffic deaths: data linkage and identification of high-risk population sub-groups. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 31, No. suppl. 1, pp. 92-106.
- Pereira, C. C. B., S. A. Vidal, P. I. Carvalho, P. G. Frias. (2013). Evaluation of the Live Birth Information System (SINASC) in the Brazilian State of Pernambuco. *Revista Brasileira de Saude Materno Infantil* (Recife), vol. 13, No. 1, pp. 39-49.
- Portela, M. C., J. M. A. Schramm, V. L. E. Pepe, M. F. Noronha, P. C. A. M., M. P. Cianieli (1997). Algoritmo para a composição de dados por internação a partir do sistema de informações hospitalares do sistema único de saúde (SIH/SUS) - Composição de dados por internação a partir do SIH/SUS Algorithm for establishing hospital admittance data based on the hospital information system in the Brazilian Unified Health System (SIH/SUS). *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 13, No. 4, pp. 771-774.
- Prado, T. N., A. L. Caus, M. Marques, E. L. Maciel, J. E. Golub, A. E. Miranda (2011). Epidemiological profile of adult patients with tuberculosis and AIDS in the state of Espírito Santo, Brazil: cross-referencing tuberculosis and AIDS databases. *Jornal Brasileiro de Pneumologia*(Brasília), vol. 37, No. 1, pp. 93-99.
- Ribeiro, A. M., M. J. Guimarães, M. C. Lima, S. W. Sarinho, S. B. Coutinho (2009). Risk factors for neonatal mortality among children with low birth weight. *Revista de Saúde Pública* (São Paulo), vol. 43, No. 2, pp. 246-255.
- Santa Helena, E. T., C. A. Sousa, C. A. Silva (2005). Risk factors for neonatal mortality in Blumenau, Santa Catarina: linkage between database. *Revista Brasileira de Saude Materno Infantil*(Recife), vol. 5, No. 2, pp. 209-217.
- Sarinho, S. W., D. A. M. Filho, G. A. P. Silva, M. C. Lima. (2001). Risk factors for neonatal death in Recife: a case-control study. *Jornal de Pediatria*(Rio de Janeiro), vol. 77, No. 4, pp. 294-298.
- Schnitman, A. (1990). Análise da fidedignidade da declaração da causa básica de morte por câncer em Salvador, Brasil. *Revista de Saúde Pública* (São Paulo), vol. 24, No. 6.
- Selig, L., M. Belo, A. J. L. A. Cunha, E. G. Teixeira, R. Brito, A. L. Luna, A. Trajman (2004). Deaths attributed to tuberculosis in the State of Rio de Janeiro. *Jornal Brasileiro de Pneumologia*(Brasília), vol. 30, No. 4, pp. 335-342.
- Selig, L., A. L. Kritski, A. M. Cascão, J. U. Braga, A. Trajman, R. M. G. Carvalho (2010). Proposta de vigilância de óbitos por tuberculose em sistemas de informação. *Revista de Saúde Pública* (São Paulo), vol. 44, No. 6, pp. 1072-1078.
- Silva, A. A. M., Ribeiro, V. S., F. Borba Júnior, Antônio, C. C. Liberata, R. A. Silva (2001). Avaliação da qualidade dos dados do Sistema de Informações sobre Nascidos Vivos em 1997-1998. *Revista de Saúde Pública* (São Paulo), vol. 35, No. 6, pp. 508-514.
- Silva, C. F., A. J. M. Leite, N. M. G. S. Almeida (2009). Linkage between live birth and infant death records in a municipality in Northeast Brazil: quality of the information system. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 25, No. 7, pp. 1552-1558.
- Silva, C. F., A. J. M. Leite, N. M. G. S. Almeida, R. C. Gondim (2006). Risk factors for infant mortality in a city of northeastern Brazil: linkage between live birth and infant death databases - 2000 to 2002. *Revista Brasileira de Epidemiologia*(São Paulo), vol. 9, No. 1, pp. 69-80.
- Silva, J. M. & E. F. Drumond (2013). Evaluation of the strategy for identification and measurement of the fatal accidents at work. *Ciência & Saúde Coletiva*(Rio de Janeiro), vol. 18, No. 5, pp. 1361-1365.
- Silva, L. P., C. M. M. Moreira, M. H. C. Amorim, D. S. Castro, E. Zandonade (2014). Avaliação da qualidade dos dados do Sistema de Informações sobre Nascidos Vivos e do Sistema de Informações sobre Mortalidade no período neonatal, Espírito Santo, Brasil, de 2007 a 2009. *Ciência & Saúde Coletiva*(Rio de Janeiro), vol. 19, No. 7, pp. 2011-2020.
- Silveira, D. P. & E. Artmann (2009). Accuracy of probabilistic record linkage applied to health databases: systematic review. *Revista de Saúde Pública* (São Paulo), vol. 43, No. 5, pp. 875-882.
- Sousa, L. M. O. & R. S. Pinheiro (2011). Unnotified deaths and hospital admissions for tuberculosis in the municipality of Rio de Janeiro. *Revista de Saúde Pública* (São Paulo), vol. 45, No. 1, pp. 31-39.

- Sousa, M. H., J. G. Cecatti, E. Hardy, S. J. Serruya (2008). Probabilistic record linkage: an application to severe maternal morbidity (*near miss*) and maternal mortality. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 24, No. 3, pp. 653-662.
- Sousa, M. H., J. G. Cecatti, E. E. Hardy, S. J. Serruya(2007). Declared maternal death and the linkage between health information systems. *Revista de Saúde Pública* (São Paulo), vol. 41, No. 2, pp. 181-189.
- Strozzi, G. M., J. B. Strozzi, M. L. Souza, M. D. Schutel(1985). A study of the basic cause of death in children under 15, during 1982, in Children Hospital (Florianópolis, Santa Catarina, Brasil). *Revista de Saúde Pública* (São Paulo), vol. 19, No. 2, pp. 123-132.
- Theme Filha, M. M., S. G. N. Gama, C. B. Cunha, M. C. Leal (2004). Confiabilidade do Sistema de Informações sobre Nascidos Vivos Hospitalares no Município do Rio de Janeiro, 1999-2001. *Cadernos de Saúde Pública* (Rio de Janeiro), vol. 20, No. suppl.1, pp. 83-91.