

Using a deterministic matching computer routine to identify hospital episodes in a Brazilian de-identified administrative database for the analysis of obstetrics hospitalisations

Claudia Medina Coeli^{1,*}, Rosa Maria Soares Madeira Domingues², Lana Meijinhos³, Daniela Medina Coeli Bastos¹, Rejane Sobrino Pinheiro¹, Valeria Saraceni³, Marcos Augusto Bastos Dias⁴, Natália Santana Paiva¹, and Kenneth Rochel de Camargo Jr⁵

Submission History

Submitted:	17/06/2024
Accepted:	02/12/2024
Published:	03/03/2025

Abstract

Introduction

The absence of a unique patient identifier in the Brazilian hospital administrative database prevents the identification of hospital episodes with multiple hospitalisations of the same patient.

Objectives

This study aims to evaluate the information gain by using a computer routine to identify acute Obstetrics hospital episodes and its impact on assessing marks of case severity.

Methods

The data source was a de-identified Brazilian hospital administrative database from 2017 to 2020, including hospitalisations records of women of reproductive age (10 to 49 years old) for treating acute conditions (N=16,087,490). We processed this database by combining C++ and Python routines to create a hospital episodes database. From the latter, we selected obstetrics hospital episodes from 2018 to 2019 (N=4,926,877). We compared selected characteristics of the hospital episodes according to their type (multiple vs single records per episode), testing for differences using effect size measures. We compared relative differences in case severity marks when using the hospital episode as the unit of analysis to that of isolated hospitalisations (N=5,018,350).

Results

Compared to single-record episodes, multiple-records episodes had longer length of stay, higher amount reimbursed, and lower proportion of discharge alive. When comparing isolated hospitalisations to hospital episodes analysis, we observed an increase in all case severity indicators, especially for hospital deaths, with an increment of 13.15%. The computer routine decreased the hospital admissions with a reason for hospital discharge that did not indicate the outcome (hospital stay or inter-hospital transfer) from 2.29% to 0.73%.

Conclusions

The deterministic matching computer routine proved valuable for identifying records that refer to the same hospital episode, which improved the assessment of severe cases.

Keywords

hospitalization; health administrative data; obstetrics; record linkage; computer applications software

*Corresponding Author:

Email Address: coeli@iesc.ufRJ.br (Claudia Medina Coeli)

Introduction

Hospital administrative databases have been used for severe maternal morbidity (SMM) surveillance at the population level [1], following the recommendation of SMM surveillance as a complementary strategy for reducing maternal mortality [2].

The Brazilian Unified Healthcare System, also known as SUS (from the Portuguese, *Sistema Único de Saúde*), was created in 1988 and structured on the constitutional principles of universal coverage, comprehensive care and citizens' participation [3]. Under SUS, hospital care is provided by public institutions at the three levels of government (federal, state, municipal) and private institutions that deliver services under contract to the public system. Although the system is in theory universal, in 2019 28.5% of the population purchased additional coverage from the supplementary private medical system, varying according to factors such as age and state of residence [4].

Hospital admissions funded by SUS are recorded using the same information system (SUS Hospital Information System, from Portuguese *Sistema de Informação Hospitalar do SUS—SIH-SUS*) and under the same standards, yielding a National Database. The AIH (Hospitalisation Authorization, from the Portuguese acronym for *Autorização de Internação Hospitalar*) database is the core SIH-SUS table. The AIH is a claim form the hospital administrative staff fills out for each discharge based on the medical record abstract [5]. Data is continually produced, with de-identified microdata consolidated and made publicly available monthly for download [6]. Considering the high hospital delivery coverage in the country and the data availability, we proposed using SIH-SUS to monitor SMM, using the WHO (World Health Organisation) definition of potentially life-threatening conditions (PLTC) [7].

The challenges for implementing the SMM surveillance using administrative data are the criteria definition and its operationalisation in the available database [1, 8]. Using the SIH-SUS poses an additional challenge, as the AIH form does not have a unique patient identifier, preventing the identification of all hospitalisations of the same patient. Analysing the AIH forms as isolated hospitalisations may underestimate markers of case severity, such as prolonged length of stay (LOS) and the occurrence of clinical complications [9]. To overcome this limitation, we developed a deterministic matching routine to identify hospital episodes in the SIH-SUS de-identified database. We used the Manitoba Centre for Health Policy (MCHP) definition of a hospital episode: “a single, continuous stay in the hospital system, irrespective of transfers between hospitals” [10]. We excluded long-term chronic care hospitalisations according to the MCHP's definition.

This study aims to evaluate the information gain by using a computer routine to identify acute obstetrics hospital episodes and its impact on assessing marks of case severity.

Methods

Study design

We conducted a cross-sectional study using a matching deterministic computer routine to identify obstetric acute

hospital episodes with admission dates from 2018 to 2019 in the Brazilian publicly funded hospital administrative database

Data Source

We used the AIH database from SIH-SUS. Although the AIH database fields changed through time, the general table structure includes patient data (e.g. sex, date of birth, postal code, city of residence); hospital data (e.g. hospital code, municipality); hospitalisation data (e.g., AIH code, type of care, admission and discharge dates, LOS, number of days in intensive care unit [ICU], amount reimbursed, principal diagnosis, up to 12 other diagnoses, principal procedure and reason for the discharge) [11]. The AIH code is the database's primary key, identifying one, and only one, AIH record. The field type of care has two attributes: type 1, acute and type 5, long-term. The reason for the discharge has five main attributes: discharge alive, hospital stay, inter-hospital transfer, death and administrative discharge.

The SIH-SUS supports the prospective payment system adopted in Brazil to reimburse hospitals for the inpatient services rendered. The principal procedure field is coded according to the SUS Procedures, Medications, Orthoses, Prostheses and Special Materials Management System Table [12]. It defines the amount reimbursed and the expected LOS based on an estimated resource consumption for similar cases [5]. However, whenever the patient needs additional treatment (e.g. a patient admitted for an obstetric procedure requiring surgical intervention) or a longer-than-expected hospital stay due to the characteristics of the clinical presentation, the current AIH form is closed, and a new AIH form with a new code is generated. Likewise, when the patient is transferred to another hospital due to the need for a procedure not available in the original hospital or to continue the treatment in a hospital with a more complex level of care, a new AIH form with a new code is generated [13]. Multiple AIH forms with different codes are generated for the same patient in both cases. Finally, whenever it is necessary to transfer a patient to long-term care, a new AIH form is generated, receiving the same AIH code of the original event, and code 5 (long-term care) in the AIH form type of care field. One patient can have one or multiple type 5 AIH forms, depending on the time needed to stay in the hospital to complete the treatment. In the case of multiple type 5 AIH forms, the same AIH code is assigned to all of them.

We used previously downloaded and processed databases to develop a maternal health surveillance panel [14]. In summary, due to the size of the files, for each Brazilian federative unit (26 states and the federal district) and year (from 2012 to 2020), we downloaded the AIH databases using the microdatasus R package [15], running R version 4.3.1 for Linux [16], with the R Studio interface (version 2023.06.1 for Linux) [17]. Then, we selected records according to the following criteria: female sex and age between 10 and 49 years. Since the adolescent fertility rate is high in Brazil, the Ministry of Health adopts the age range 10 to 19 years in order to monitor adolescent pregnancies [18]. Next, we merged all the year's files of a specific federative unit. Finally, we selected only AIH records for acute care (type 1). We processed each of the 27 federative unit databases, creating two binary fields to identify whether the AIH record met the diagnosis

or the procedure criteria for an obstetric admission. We assigned the value 1 to a diagnosis flag field if any diagnoses were filled out with ICD-10 (International Classification of Diseases -10) O00 to O99 codes of Chapter XV (Pregnancy, Childbirth and Puerperium), and zero otherwise. Likewise, we assigned the value 1 to a procedure flag field if the principal procedure field was filled out with any obstetric procedure codes (Supplementary Appendix 1) and zero otherwise [19].

The completeness of all fields used by the deterministic matching routine or data analysis was 100%, and all of them, including date fields, presented a consistent format throughout the entire database.

Data Processing

We carried out various sequential processes to create two national study population databases: (1) the AIH Obstetric Isolated Hospitalisations database, formed by isolated AIH records with an admission date in 2018 or 2019, and (2) the AIH Obstetric Hospital Episodes database, including episodes with at least one AIH record with an admission date in 2018 or 2019. The majority of the processes were executed using the tidyverse R package [20] running the R version 4.3.1 for Linux [16], with the R Studio interface (version 2023.06.1 for Linux) [17]. Whenever that was the case, we cited the other computational tools used.

First step

We merged all the federative units' databases to create a national type 1 AIH (acute care) database. As we aimed to analyse the AIH records in 2018 and 2019, we selected records with admission years from 2017 to 2020. We included the years 2017 and 2020 because hospital episodes could have begun in 2017 and ended in 2018. Likewise, episodes could have begun in 2019 and ended in 2020. The selection process resulted in a type 1 AIH database with 16,087,490 records representing AIH records of women of reproductive age (10 to 49 years old) for treating acute conditions.

We created a national type 5 AIH (long-term care) database using the same procedures used to create the type 1 AIH (acute care) database, but selecting type 5 AIH records. As previously mentioned, as a patient can have multiple type 5 AIH records with the same AIH code, we selected just one record, resulting in a type 5 AIH database with 45,258 records. Next, we linked the national type 1 AIH database to the type 5 AIH database, using the AIH code as the linkage key, and updated the field type of discharge, indicating a transfer to long-term chronic care when it occurred. We also grouped the original administrative discharge code with the hospital stay code, as both indicate that the patient needs to stay in the hospital to complete the treatment (Figure 1). Hence, in this analysis, we assigned the following attributes to the field reason for the hospital discharge: discharge alive, hospital stay, inter-hospital transfer, transfer to long-term care, and death in hospital.

Second step

Using the type 1 AIH database generated in the first step, we ran a deterministic matching routine to identify the hospital

episodes of care. Since the operation requires processing large amounts of data, sorting and iterating through it, we opted to work with a SQL database (SQLite, widely available free and open source database) [21] as the backend for the routine developed in C++ [22] for faster execution.

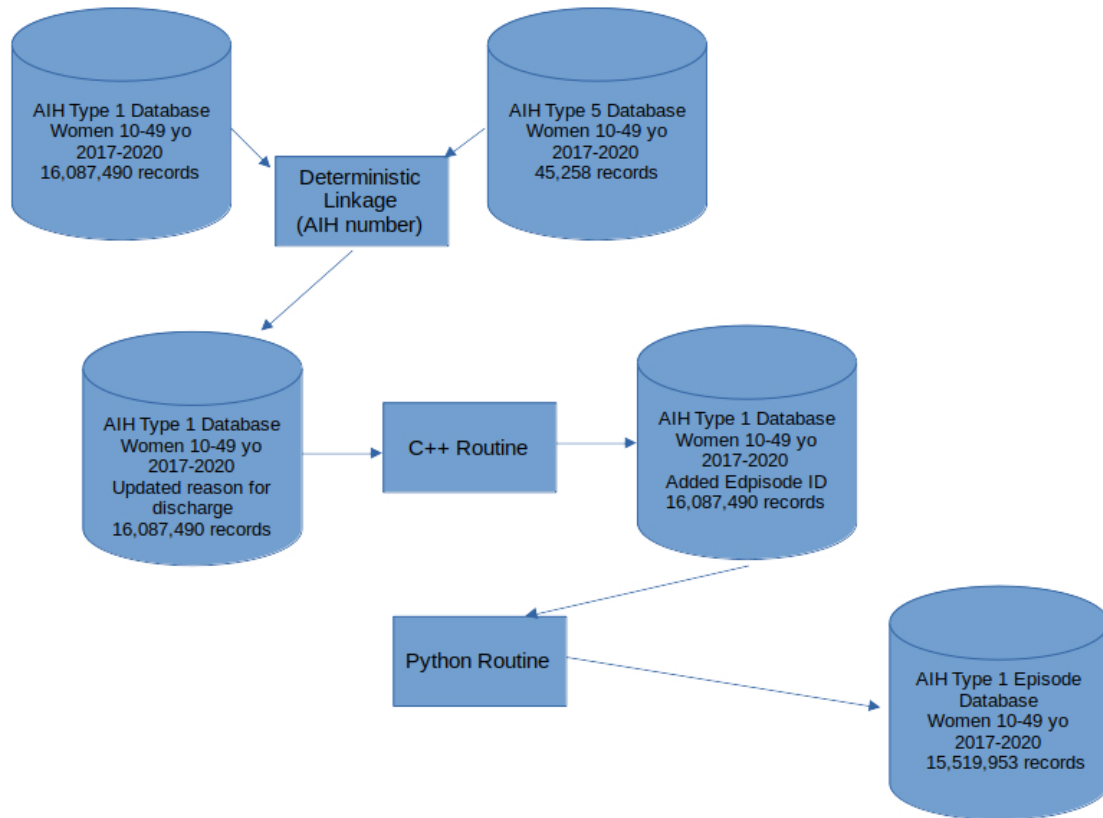
Details of the routine description and the link for the source code are presented in Supplementary Appendix 2. In short, the routine first identifies all AIH records pertaining to the same episode of the same patient in the same hospital using a matching key formed by the hospital code, the woman's municipality of residence, and the date of birth. All the AIH records that exactly agree on the matching key and have the admission date equal to or at most one day after the discharge date of the previous AIH record get the same hospital episode code. Second, whenever an inter-hospital transfer was recorded as the reason for discharge, the routine additionally searches for AIH records that exactly agreed on the woman's postal code, municipality of residence and date of birth. All the AIH records that exactly agree on this second matching key and have the admission date equal to or at most one day after the discharge date of the previous AIH record get the same hospital episode code. The episode codes assigned in each of the two steps are different. Hence, whenever a hospital episode comprises multiple AIH records in the same hospital and one or more AIH records in a different hospital, the routine regroups them, assigning the same episode code.

The routine output CSV file has the same number of records of the type 1 AIH database ($N = 16,087,490$ records), including three new fields created by the routine: an episode code assigned to all AIH records of the same episode; a flag marking whether the AIH record initiates or is part of a sequence of AIH records at the same hospital, or an inter-hospital transfer; and a number that indicates the order of the record in the sequence of the AIH records that compose the episode (Figure 1).

We only assigned the same episode code to the AIH records when the difference between the admission date of the subsequent AIH and the discharge date of the previous AIH is equal to or at most one day after. Hence, the matching routine aims to identify AIH records that comprised one acute episode of care from admission to discharge, not allowing the identification of readmissions or other AIH records unrelated to the acute episode of care. As the matching key to identify inter-hospital transfers did not include the hospital code, it allows the identification of hospital transfers irrespective of the municipality where the destination hospital is located.

The final process in this step was to run a Python [23] routine with the database generated by the C++ routine mentioned above to produce a table in which each record represents a patient hospital episode with the attributes: the episode code, which identifies each hospital episode; the patient's age (calculated by the formula [admission date - date of birth]/ 365.25, rounded to an integer); the admission date of the first AIH record; the date of the discharge of the last AIH record; the length of stay in days (calculated by the formula discharge date - admission date, assign the value zero for hospitalisation stays less than one day); the reason for discharge in the last AIH record; the total days in ICU; the principal diagnosis at the first hospitalisation; the principal procedure at the first hospitalisation, the episode total amount

Figure 1: Type 1 AIH Episode Database processing flowchart



reimbursed; the total number of AIH records in the episode; and the number of AIH records in the episode that met the obstetric criteria (according to diagnosis or procedure). This process resulted in an episode database in CSV format with 15,519,953 records (Figure 1).

The routine description and the link for the source code are presented in Supplementary Appendix 2.

Third step

To create the AIH Obstetric Isolated Hospitalisation database, we selected records from the type 1 AIH database ($N = 16,087,490$ AIH records) that met the criteria for obstetric admission (diagnosis or procedure) and had an admission date in 2018 or 2019, resulting in a database with 5,018,350 AIH records. Whereas to create the AIH Obstetric Hospital Episodes database, we selected records from the hospital episodes database created with the Python routine ($N = 15,519,953$ episode records) whose date of discharge in the last AIH record in the sequence was not 2017 and the date of admission in the first AIH record in the sequence was not 2020 ($N = 8,173,725$). We applied these criteria because we aimed to include only hospital episodes that had at least one AIH record with admission date in 2018 or 2019.

Among the 8,173,725 episode records we selected those with at least one AIH record in the episode that met the obstetric criteria, resulting in a database with 4,926,877 episode records (Figure 2).

Each record in the AIH Obstetric Isolated Hospitalisation database refers to one isolated AIH record. In contrast, in the AIH Obstetric Hospital Episodes database, each record refers

to a hospital episode, aggregating data for all AIH records that comprise the episode.

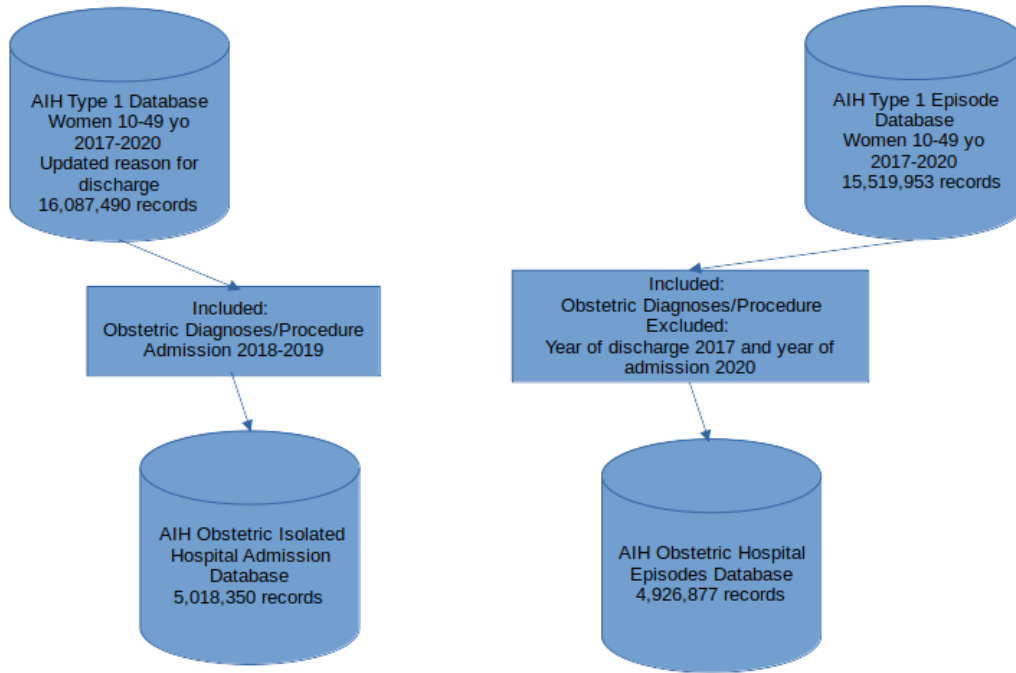
Data Analysis

To analyse the data, we used the following R (version 4.3.1 for Linux) [16] packages with the R Studio interface (version 2023.06.1 for Linux) [17]: tidyverse [20], labelled [24], gtsummary [25], effectsize [26], and DescTools [27]. We used tidyverse to process and transform data, labelled to manipulate metadata, gtsummary to create tables, effectsize and DescTools to calculate effect size measures.

First, we presented the characteristics of the hospital episodes, including the woman's age; the number of AIH records per hospital episode; LOS; the amount reimbursed (in US\$); the reason for the hospital episode discharge, informed in the last AIH record; and the five more frequent principal diagnoses and principal procedures, informed in the first episode AIH record. Next, we compared selected characteristics of the hospital episodes according to their type (multiple vs single AIH records per episode). We evaluated the differences between the two groups, estimating the effect size using the rank-biserial correlation for continuous variables and the Cramer V for categorical variables.

We also evaluated four markers of case severity, as follows: use of ICU, prolonged LOS (above three days, which is the 75th percentile of the hospital episode distribution), high reimbursement (above US\$186.19, which is the 75th percentile of the hospital episode distribution), and hospital death. As we evaluated obstetric hospitalisations with various reasons for admission (principal diagnosis and procedure), we could

Figure 2: AIH Obstetric Isolated Hospitalisations and AIH Obstetric Hospital Episodes Databases processing flowchart



not use specific benchmarking to define the prolonged length of stay and high reimbursement. Hence, we used the 75th percentile cutoff point to determine these outcomes. We estimated the odds ratio (OR) and 95% confidence interval (CI) to compare episodes according to their type (multiple vs single AIH records per episode), using the single-AIH record episode as the reference.

Next, we compared the interpretation of the case severity marks (use of ICU, prolonged LOS, high reimbursement, and hospital death) when using the hospital episode as the unit of analysis (the AIH Obstetric Hospital Episodes database) to that of isolated hospitalisation (the AIH Obstetric Isolated Hospitalisation database). For this analysis, we estimated the indicators in each database separately. To assess the relative difference of indicators calculated in the two databases, we used the formula: Relative difference = (the indicator in the AIH Obstetric Hospital Episodes database - the indicator in the AIH Obstetric Isolated Hospital Admissions database) / (the indicator in the AIH Obstetric Isolated Hospital Admissions database) * 100.

To additionally assess the information gain by using the deterministic matching computer routine, we evaluated: (1) the number of AIH records that initially did not meet the criteria of obstetrics but were identified as part of an obstetric hospital episode; (2) the computer routine ability to decrease the number of AIH records with a reason for discharge that did not indicate the final outcome (hospital stay or inter-hospital transfer).

The lack of a gold standard precluded us from evaluating measures of linkage quality for both false negative and false positive errors, so we resorted to using two different strategies to evaluate potential inconsistencies within the multiple AIH records episodes (117,253 records). First, we examined the occurrence of AIH records subsequent in time to a record indicating a death, considering that these should be the last record in the series. Second, we evaluated the differences in

the postal codes among the multiple AIH records episodes. We excluded hospital episodes that only comprised inter-hospital transfers ($N = 291$), as the postal code was included in the matching key used to identify inter-hospital transfers. This might indicate that the computer routine grouped AIH records that do not belong to the same woman. We use the postal code because, except for inter-hospital transfer, it was not part of the matching key. The episodes with inconsistencies were manually examined.

Results

Table 1 depicts the characteristics of the 4,926,877 obstetric hospital episodes, with at least one AIH record presenting an admission date between 2018 and 2019.

The median age of hospitalised women was 25 years. Most hospital episodes had only one AIH record (97.6%), whereas for the 117,253 remaining, 93% had two. The median LOS was two days, the median amount reimbursed was US\$ 154.36, and around 99% of the women were discharged alive. Single spontaneous delivery was the principal reason for the admission.

Compared to single-AIH record episodes, multiple-AIH record episodes had longer LOS, higher amount reimbursed, and lower proportion of discharge alive (Table 2). In addition, multiple-AIH record episodes were more likely associated with the use of ICU, prolonged LOS, high amount reimbursed, and hospital death (Table 3). These results explain the relative difference observed when comparing the markers of case severity calculated using the AIH Obstetric Hospital Episodes database to the one obtained using the AIH Obstetric Isolated Hospitalisation database. An increase was observed for all indicators, especially for hospital deaths, which presented an increment of 13.15%. The number of hospitalisations was overestimated by 1.82% (Table 4).

Table 1: Characteristics of obstetric hospital episodes, Brazil 2018–2019

Characteristic	N = 4,926,877 ¹
Woman age Median (IQR)	25 (21, 31)
Number of AIH records	
<i>Multiple</i>	117,253 (2.4%)
<i>Single</i>	4,809,624 (97.6%)
Length of stay (Days) Median (IQR)	2 (2, 3)
Amount reimbursed (US\$) Median (IQR)	154.36 (125.49, 186.19)
Reason for episode discharge	
<i>Discharge alive</i>	4,888,771 (99.23%)
<i>Hospital stay</i>	11,791 (0.24%)
<i>Inter-hospital transfer</i>	24,155 (0.49%)
<i>Transfer to long-term care</i>	12 (0.00%)
<i>Death in hospital</i>	2,148 (0.04%)
Five most frequent principal diagnoses²	
<i>Single spontaneous delivery</i>	2,127,634 (43.2%)
<i>Single delivery by caesarean section</i>	608,822 (12.0%)
<i>Spontaneous abortion</i>	180,272 (3.7%)
<i>Premature rupture of membranes</i>	130,105 (2.6%)
<i>Labour and delivery complicated by fetal stress</i>	111,183 (2.3%)
Five most frequent principal procedures³	
<i>Vaginal delivery</i>	1,969,312 (40.0%)
<i>C-section</i>	1,367,407 (27.7%)
<i>Treatment of complications of pregnancy</i>	433,467 (8.8%)
<i>Post-abortion uterine curettage</i>	327,007 (6.6%)
<i>C-section in high-risk pregnancy</i>	314,050 (6.4%)

¹n (%) or Median (Interquartile Range - IQR) when indicated.

²Categories of the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10).

³The SUS Procedures, Medications, Orthoses, Prostheses and Special Materials Management System Table [12].

Table 2: Characteristics of obstetric hospital episodes according to the episode type (multiple vs single AIH records per episode), Brazil 2018–2019

Characteristic	Hospital episode type		Effect size ² (CI 95%) ³
	Multiple N = 117,253 ¹	Single N = 4,809,624 ¹	
Age	26 (21, 32)	25 (21, 31)	0.062 (0.059, 0.065)
Length of stay (Days)	6(4, 10)	2(2, 3)	0.760 (0.759, 0.761)
Amount reimbursed (US\$)	251.01 (195.37, 347.35)	153.20 (124.89, 184.67)	0.673 (0.671, 0.674)
Reason for episode discharge	n (%)	n (%)	0.042 (0.036, 0.067)
<i>Discharge alive</i>	114,887 (97.98%)	4,773,884 (99.26%)	
<i>Hospital Stay</i>	1,145 (0.98%)	10,646 (0.22%)	
<i>Inter-hospital transfer</i>	629 (0.54%)	23,526 (0.49%)	
<i>Transfer to long-term chronic care</i>	11(0.009%)	1 (0.000%)	
<i>Death in Hospital</i>	581 (0.496%)	1,567 (0.033%)	

¹Median (Interquartile range - IQR) or n (%) when indicated.

²Continuous variables: rank-biserial correlation; categorical variables: Cramer V.

³CI 95% - Confidence Interval 95%.

The deterministic matching computer routine identified additional 21,752 AIH records that did not meet the obstetrics criteria initially. The principal diagnoses in those records were urinary tract infection, site not specified (1,403/ 21,752; 6.4%), sterilisation (1,400/ 21,752; 6.4%), and essential (primary) hypertension (899/21,752; 4.1%).

The computer routine decreased AIH records with a reason for discharge that did not indicate the outcome (hospital stay or inter-hospital transfer) from 2.29% (115,038/5,018,350) in the AIH Obstetric Isolated Hospital Admissions database to 0.73% (35,946/4,926,877) in the AIH Obstetric Hospital Episodes database, a relative decrease of around 68%.

Table 3: Markers of cases severity of obstetrics hospital episode, according to the episode type (multiple vs single AIH records per episode), Brazil 2018–2019

Characteristic	Hospital episode type		Odds ratio (CI 95%) ³
	Multiple ¹ N = 117,253	Single ¹ N = 4,809,624 [REF] ²	
Use of Intensive Care Unit	7,677 (6.55%)	19,838 (0.41%)	16.91 (16.46, 17.38)
Prolonged length of stay ⁴	91,434 (77.98%)	662,408 (13.77%)	22.17 (21.86, 22.48)
High reimbursement ⁵	92,904 (79.23%)	1,138,661 (23.67%)	12.30 (12.13, 12.48)
Death in Hospital	581 (0.496%)	1567 (0.033%)	15.27 (13.9, 16.81)

¹n (%).²Reference category.³Confidence Interval 95%.⁴ > 75th percentile of the distribution of the length of stay.⁵ > 75th percentile of the distribution of the value reimbursed (US\$).

Table 4: Markers of cases severity according to the type of analysis, Brazil 2018–2019

Characteristic	Type of analysis ¹		Relative difference ³ (%) –1.82%
	Hospital episode ² N = 4,926,877	Isolated hospitalisation ² N = 5,018,350	
Use of intensive care unit	27,515 (0.56%)	27,333 (0.54%)	+3.70%
Prolonged length of stay ⁴	753,842 (15.30%)	728,721(14.52%)	+5.37%
High reimbursement ⁴	1,231,565 (25.00%)	1,193,840(23.79%)	+5.09%
Death in hospital	2,148 (0.044%)	1,892 (0.038%)	+13.15%

¹Analysis without (Isolated) or with (Hospital Episode) aggregation of records related to the same hospital episode.²n (%).³Relative Difference: (%Episode of Care-%Isolated)/%Isolated*100.⁴ >75th percentile of the distribution of episodes of care.

Among the multiple AIH records episodes (117,253 records) we found 30 (0.03%) with recorded deaths dated before the last AIH record. The manual examination of those episodes showed that 15 the subsequent records to the registered death record had an organ donor principal diagnose, indicating that this was not an error, leaving 15 episodes (0.01%) with unexplained inconsistencies.

Of the 116,962 hospital episodes evaluated for postal code inconsistencies, we identified differences in 6,300 (5.4%). Among them, 4,705 (74.7%) had all AIH records with a principal diagnosis of the ICD-10 - Chapter XV (Pregnancy, Childbirth and Puerperium), suggesting they are related. Among the 110,953 hospital episodes with multiple AIH records without differences in postal code, 87,691 (79.0%) had all AIH records with a principal diagnosis of the ICD-10 - Chapter XV (Pregnancy, Childbirth and Puerperium).

Discussion

This study's findings showed that the deterministic matching computer routine allowed for the identification of obstetric hospital episodes in a de-identified administrative database. This not only improved the assessment of severe cases but also decreased AIH records without information on the outcome. These results have a direct impact on data quality,

potentially enhancing the monitoring of maternal potentially life-threatening conditions and the fitting of predictive models for severe maternal morbidity and mortality.

The analysis based on hospital episodes has been used to study selected diagnoses [28], hospital records combined with ICU records [29], hip fracture surgery [30], and all causes of hospitalisation [31]. Using this approach, we observed that the number of AIH records was overestimated by 1.82% when failing to identify the AIH records that comprise a single hospital episode. This figure is lower than the ones observed by other studies that assessed adult hospitalisations [28, 29]. These studies showed that the patient number overestimation varies according to the record linkage technique used [28], the principal diagnosis [28], and the method used to define the episode of care [29]. In addition to these explanations, our study looked at obstetric admissions, which are more restricted by pregnancy and relevant time frames, which may also explain the lower overestimation observed.

In our study, only 2.4% of hospital episodes involved multiple AIH records. Although linkage errors could not be ruled out [32], a small proportion of hospital episodes with multiple AIH records was expected as we studied obstetric hospitalisations of young women (median 25 years old, percentile 75% 31 years old). However, these episodes were more severe cases, presenting higher odds of ICU use, prolonged LOS, and higher reimbursement amounts.

Compared to the isolated hospitalisations analysis, the evaluation of hospital episodes showed around a 5% increase in prolonged LOS and a higher amount reimbursed, as well as a 3.7% increase in ICU use. These increments are relevant, as SMM is a rare condition [8].

A recent validation study [19] compared data from the SIH-SUS with data from medical records, showing a high identification of obstetric admissions in the SIH-SUS when using the criteria proposed by the study but underreporting the main morbidities that are part of the WHO criteria for potentially life-threatening conditions [2], mainly related to hypertensive and haemorrhagic complications. Our study showed that essential hypertension was recorded in 4.5% of hospitalisations not initially identified as obstetric, which is an example of poor recording in the AIH form of the reason for hospital admission when an ICD-10 hypertension code that is not specific to pregnancy was used, resulting in the hospitalisation not being captured as an obstetrics event. Therefore, we chose to use severity markers that did not depend on the recording of morbidity (ICU admission, length of stay, mortality) and the assessment of the reimbursement amounts due to their relevance for the management of services. Nevertheless, using SIH-SUS for SMM surveillance will benefit from improvements in data quality.

A comparison of our results with those of other national studies is limited [33–36] as, in general, these studies evaluated maternal complications rather than the need for additional care. Not every woman with severe maternal morbidity will have a hospital episode with multiple AIH records, and therefore, these numbers are not directly comparable. Furthermore, these studies differ in study design (population or hospital base, prospective or sectional), the definition of maternal morbidity cases, and data source (interview with the woman or medical record data). In any case, the frequency of severe maternal morbidity in Brazilian studies ranges from 1.5% [35] to 21.2% [33], while our frequency of hospital episodes with multiple AIH records was estimated at 2.4%, therefore being within the range of values estimated in these other studies.

Maternal death is a rare event, with studies demonstrating under-reporting due to incomplete data or classification errors [37]. Therefore, the 13.15% increase in death as the reason for discharge when adopting an episode-of-care approach is very relevant. It is known that the registry of deaths in the SIH-SUS has limitations, with an estimated registration coverage of 71.8% of deaths of women of childbearing age when compared to the Brazilian Mortality Information System. This coverage varies according to hospital characteristics and location, with higher values in hospitals located in the more developed South region, in hospitals with greater complexity, managed by the federal government, with a higher proportion of adult beds available for hospitalisations publicly funded, and in teaching hospitals. On the contrary, the coverage is lower in hospitals with emergency wards [38]. Despite these limitations, a study in Brazil that compared the national maternal mortality ratio estimated using the Brazilian Mortality database and the estimate based on the SIH-SUS found no significant differences between the two data sources, concluding that the SIH-SUS could be an important complementary source for the study of maternal mortality in the country [39]. One advantage of using SIH-SUS is the more timely death identification, as the

data is made available monthly, emphasising the importance of applying strategies, such as our matching routine, that result in greater identification of deaths in SIH-SUS.

This study has potential limitations. First, we did not evaluate the routine information gain for identifying potentially life-threatening conditions (PLTC), which would demand processing a second SIH-SUS table that records all services provided to the patient [12, 13]. However, we analysed important case severity marks related to the PLTC elected by our team to conduct the SMM surveillance using SIH-SUS [7]. Second, the lack of a gold standard precluded us from evaluating measures of linkage quality [32], for both false negative and false positive errors. Nevertheless, we evaluated data inconsistencies in the multiple AIH records episodes, showing that only a small proportion presented unexplained discrepancies. Third, this analysis only includes hospitalisations publicly funded, and its results do not apply to women assisted in the supplementary private medical system. Finally, although the routines we developed are rather specific to the databases that underlie this study, the overall logic can be adapted to processing other databases that similarly have tables with a set of fields that can be used as a compound key for a deterministic linkage process. Under circumstances where more personal identifiers are available, such as the full name and social security numbers, and similar characteristics are available to the researcher, the overall results may be even more robust.

Conclusion

In conclusion, the deterministic matching computer routine proved to be a valuable tool for identifying AIH records that comprise the same hospital episode. This not only improved the assessment of severe cases but also has the potential to enhance the overall quality of maternal healthcare data. These findings can be directly applied to improve the monitoring and prediction of severe morbidity and mortality in maternal healthcare.

Acknowledgments

This work was supported by the Bill & Melinda Gates Foundation [INV-027961] and the Brazilian Ministry of Health/DECIT/CNPq [445116/2020-0]. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission Cláudia Medina Coeli (CMC) and Kenneth Rochel de Camargo Jr (KRCJ), were partially supported by research fellowship grants from the National Council for Scientific and Technological Development- CNPq (CMC -306668/2023-8; KRCJ 306228/2021-1).

Statement on conflicts of interest

The authors have no conflicts of interest to declare.

Ethics statement

The study was conducted using databases that were de-identified and made publicly available (open access). According to The Brazilian National Health Council (CNS) ethics Resolution n° 510/2016 (April 7, 2016), the research ethics committee approval is waived.

Data availability statement

The SIH-SUS databases can be directly accessed in <https://datasus.saude.gov.br/transferencia-de-arquivos/>. The files analysed in the manuscript can be accessed in <https://doi.org/10.7303/syn64313527>.

References

- Kuklina EV, Goodman DA. Severe Maternal or Near Miss Morbidity: Implications for Public Health Surveillance and Clinical Audit. *Clinical Obstetrics & Gynecology*. 2018; 61(2):307–18. <https://doi.org/10.1097/GRF.0000000000000375>
- World Health Organization. Evaluating the quality of care for severe pregnancy complications: the WHO near-miss approach for maternal health. Geneva: World Health Organization; 2011. [Accessed 2024 Apr 30]. Available from: https://apps.who.int/iris/bitstream/handle/10665/44692/9789241502221_eng.pdf.
- Castro MC, Massuda A, Almeida G, Menezes-Filho NA, Andrade MV, De Souza Noronha KVM, Rocha R, Macinko J, Hone T, Tasca R, Giovanella L, Malik AM, Werneck H, Fachini LA, Atun R. Brazil's unified health system: the first 30 years and prospects for the future. *The Lancet*. 2019;394(10195):345–56. [https://doi.org/10.1016/S0140-6736\(19\)31243-7](https://doi.org/10.1016/S0140-6736(19)31243-7)
- Instituto Brasileiro de Geografia e Estatística (IBGE). Pesquisa nacional de saúde: 2019: informações sobre domicílios, acesso e utilização dos serviços de saúde: Brasil, grandes regiões e unidades da federação. Coordenação de Trabalho e Rendimento. Rio de Janeiro: IBGE. 2020.[Accessed 2024 Apr 30]. Available from: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101748>.
- Travassos Veras CM. Equity in the use of private hospitals contracted by a compulsory insurance scheme in the city of Rio de Janeiro, Brazil, in 1986. PhD thesis, London School of Economics and Political Science. [cited 2024 Apr 30]. Available from: <http://etheses.lse.ac.uk/id/eprint/2431>.
- Instituto Brasileiro de Geografia e Estatística (IBGE). Sistema de informações hospitalares do SUS (SIH/SUS). [Accessed 2024 Apr 30]. Available from: <https://ces.ibge.gov.br/base-dados/metadados/ministerio-da-saude/sistema-de-informacoes-hospitalares-do-sus-sih-sus>.
- Domingues RMSM, Dias MAB, Saraceni V, Pinheiro RS, Paiva NS, Coeli CM. Vigilância da morbidade materna no Brasil: contribuições para o debate. *Cad Saúde Pública*. 2023;39(11):e00151123. <https://doi.org/10.1590/0102-311XPT151123>
- Kuklina EV, Ewing AC, Satten GA, Callaghan WM, Goodman DA, Ferre CD, et al. Ranked severe maternal morbidity index for population-level surveillance at delivery hospitalization based on hospital discharge data. Garzon S, editor. *PLoS ONE*. 2023;18(11):e0294140. <https://doi.org/10.1371/journal.pone.0294140>
- Osman M, Quail J, Hudema N, Hu N. Using SAS® to Create Episodes-of-Hospitalization for Health Services Research. SAS. 2015. Paper 3281-2015.[Accessed 2024 Apr 30]. Available from: <https://support.sas.com/resources/papers/proceedings15/3281-2015.pdf>.
- Manitoba Centre for Health Policy. The Manitoba Population Research Data Repository. Hospital Episode / Hospital Episodes. Glossary Definition. [Accessed 2024 Apr 30]. Available from: <http://mchp-appserv.cpe.umanitoba.ca/viewDefinition.php?printer=Y&definitionID=103925>.
- Cerqueira DRC, Alves PP, Coelho DSC, Reis MVM, Lima AS. Uma análise da base de dados do sistema de informação hospitalar entre 2001 e 2018: dicionário dinâmico, disponibilidade dos dados e aspectos metodológicos para a produção de indicadores sobre violência. Ipea; 2019. Accessed [2024 Apr 30]. Available from: https://repositorio.ipea.gov.br/bitstream/11058/9409/1/Uma_analise_da_base_de_dados_do_sistema_de_informacao_hospitalar.pdf.
- Brasil. Ministério da Saúde. SIGTAP - Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS. Accessed [2024 Apr 30]. Available from: <http://sigtap.datasus.gov.br/tabela-unificada/app/sec/inicio.jsp>.
- Brasil. Ministério da Saúde/ Secretaria de Atenção à Saúde/ Departamento de Regulação, Avaliação e Controle/Coordenação-Geral de Sistemas de Informação – 2010. Manual técnico operacional do Sistema de Informação hospitalar – orientações técnicas. Versão 01.2012.2012. Accessed [2024 Apr 30]. Available from: https://bvsmis.saude.gov.br/bvs/publicacoes/manual_tecnico_sistema_informacao_hospitalar_sus.pdf.
- Domingues RMSM, Rodrigues AS, Dias MAB, Saraceni V, Francisco RPV, Pinheiro RS, et al. Maternal health surveillance panel: a tool for expanding epidemiological surveillance of women's health and its determinants. *Rev bras epidemiol*. 2024;27:e240009. <https://doi.org/10.1590/1980-549720240009>
- Saldanha RDF, Bastos RR, Barcellos C. Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS

- (DATASUS). *Cad Saúde Pública*. 2019;35(9):e00032419. <https://doi.org/10.1590/0102-311X00032419>
16. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2023. <https://www.R-project.org/>.
 17. Posit team. RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. 2023. Accessed [2024 Apr 30]. Available from: <http://www.posit.co/>.
 18. Brasil, Ministério da Saúde. Nota Técnica nº 2/2024-CACRIAD/CGACI/DGCI/SAPS/MS. Accessed [2024 Dec 02] Available from: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/notas-tecnicas/2024/nota-tecnica-no-2-2024.pdf/view>.
 19. Domingues RMSM, Meijinhos LS, Guillen LCT, Dias MAB, Saraceni V, Pinheiro RS, Paiva NS, Coeli CM. Estudo de validação das internações obstétricas no Sistema de Informações Hospitalares do Sistema Único de Saúde para a vigilância da morbidade materna: Brasil, 2021-2022. *Revista Epidemiologia e Serviços de Saúde*; 2024. In press.
 20. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. *J Open Source Softw*. 2019;4(43):1686. <https://doi.org/10.21105/joss.01686>
 21. SQLite: SQLite Home Page. Accessed [2024 Apr 30]. Available from: <https://www.sqlite.org/>
 22. C++: GCC online documentation. Accessed [2024 Apr 30]. Available from: <https://gcc.gnu.org/onlinedocs/>.
 23. Python: The Python Language reference. Accessed [2024 Apr 30]. Available from: <https://docs.python.org/3/reference/index.html>
 24. Larmarange J. labelled: Manipulating Labelled Data. R package. Accessed [2024 Apr 30]. Available from: <https://CRAN.R-project.org/package=labelled>.
 25. Sjoberg DD, Whiting K, Curry M, Lavery JA, Larmarange J. Reproducible summary tables with the gtsummary package. *The R Journal*. 2021;13:570–80. <https://doi.org/10.32614/RJ-2021-053>
 26. Ben-Shachar M, Lüdtke D, Makowski D. effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*. 2020; 5(56), 2815. <https://doi.org/10.21105/joss.02815>
 27. Signorell A. DescTools: Tools for Descriptive Statistics. R package version 0.99.54. 2024. Accessed [2024 Apr 30]. Available from: <https://CRAN.R-project.org/package=DescTools>.
 28. Tyndall RM, Clarke JA, Shimmings J. An automated procedure for determining patient numbers from episode of care records. *Medical Informatics*. 1987;12(2):137–46. <https://doi.org/10.3109/14639238709003563>
 29. Fransoo R, Yogendran M, Olafson K, Ramsey C, McGowan KL, Garland A. Constructing episodes of inpatient care: data infrastructure for population-based research. *BMC Med Res Methodol*. 2012 Dec;12(1):133. <https://doi.org/10.1186/1471-2288-12-133>
 30. Sheehan KJ, Sobolev B, Guy P, Bohm E, Hellsten E, Sutherland JM, et al. Constructing an episode of care from acute hospitalization records for studying effects of timing of hip fracture surgery. *Journal Orthopaedic Research*. 2016 Feb;34(2):197–204. <https://doi.org/10.1002/jor.22997>
 31. Peng M, Li B, Southern DA, Eastwood CA, Quan H. Constructing Episodes of Inpatient Care: How to Define Hospital Transfer in Hospital Administrative Health Data? *Medical Care*. 2017 Jan;55(1):74–8. <https://doi.org/10.1097/MLR.0000000000000624>
 32. Doidge JC, Harron KL. Reflections on modern methods: linkage error bias. *International Journal of Epidemiology*. 2019 Oct 21;dyz203. <https://doi.org/10.1093/ije/dyz203>
 33. Rosendo TS, Roncalli AG, Azevedo GD. Prevalence of Maternal Morbidity and Its Association with Socioeconomic Factors: A Population-based Survey of a City in Northeastern Brazil. *Rev Bras Ginecol Obstet*. 2017 Nov;39(11):587-595. <https://doi.org/10.1055/s-0037-1606246>
 34. Moreira DDS, Gubert MB. Healthcare and sociodemographic conditions related to severe maternal morbidity in a state representative population, Federal District, Brazil: A cross-sectional study. *PLoS One*. 2017 Aug 3;12:e0180849. <https://doi.org/10.1371/journal.pone.0180849>
 35. Moraes AP, Barreto SM, Passos VM, Golino PS, Costa JA, Vasconcelos MX. Incidence and main causes of severe maternal morbidity in São Luís, Maranhão, Brazil: a longitudinal study. *Sao Paulo Med J*. 2011 May;129(3):146-52. <https://doi.org/10.1590/s1516-31802011000300005>
 36. Cecatti JG, Costa ML, Haddad SM, Parpinelli MA, Souza JP, Sousa MH, Surita FG, Pinto E Silva JL, Pacagnella RC, Passini R Jr; Brazilian Network for Surveillance of Severe Maternal Morbidity study Group. Network for Surveillance of Severe Maternal Morbidity: a powerful national collaboration generating data on maternal health outcomes and care. *BJOG*. 2016 May;123(6):946-53. <https://doi.org/10.1111/1471-0528.13614>
 37. Ahmed SMA, Cresswell JA, Say L. Incompleteness and misclassification of maternal death recording: a systematic review and meta-analysis. *BMC Pregnancy Childbirth*. 2023;23(1):794. <https://doi.org/10.1186/s12884-023-06077-4>
 38. Marques JA, Domingues RMSM, Dias MAB, Coeli CM, Pinheiro RS, Saraceni V. Predictive factors for the registration of deaths of women of childbearing age in the Hospital Admission System (SIH/SUS), Brazil,

2012-2020. Rev Bras Epidemiol. 2024; 27::e240051.
<https://doi.org/10.1590/1980-549720240051>

39. Ranzani OT, Marinho MDF, Bierrenbach AL. Usefulness of the Hospital Information System for maternal mortality surveillance in Brazil. Rev bras epidemiol. 2023;26:e230007. <https://doi.org/10.1590/1980-549720230007>

Abbreviations

AIH: from Portuguese, *Autorização de Internação Hospitalar* (Hospitalisation Authorization)
CI: Confidence Interval

ICD-10: International Classification of Diseases -10
ICU: Intensive Care Unit
IQR: Interquartile Range
LOS: Length of Stay
MCHP: Manitoba Centre for Health Po
PLTC: Potentially Life-Threatening Conditions
SIH-SUS: from Portuguese *Sistema de Informação Hospitalar do SUS* (SUS Hospital Information System)
SMM: Severe Maternal Morbidity
SUS: from Portuguese *Sistema Único de Saúde* (The Brazilian Unified Healthcare System).
WHO: World Health Organisation



Supplementary Appendix 1

List of obstetric procedures according to the Management System of Procedures, Medications, Orthoses, Prosthesis and Special Materials of the Unified Health System Table. <http://sigtap.datasus.gov.br/tabela-unificada/app/sec/inicio.jsp>

#0201010011 Amniocentesis
 #0211040010 Amnioscopy
 #0211040061 Antepartum tococardiography
 #0310010012 Birth assistance without dystocia
 # 0310010020 Care for newborns at birth
 #0310010039 Normal delivery
 # 0310010047 Normal delivery in high-risk pregnancies
 # 0310010055 Normal delivery at normal delivery centre
 # 0303100010 Treatment of intercurrents predominantly related to the postpartum period
 #0303100028 Eclampsia treatment
 # 0303100036 Treatment of edema, proteinuria and hypertensive disorders in pregnancy, childbirth and the postpartum period.
 # 0303100044 Treatment of clinical intercurrents during pregnancy
 #0303100052 Hydatidiform mole treatment
 #0409060011 Cervical cerclage
 #0409060054 Uterine curettage in hydatidiform mole
 # 0409060070 Post-abortion uterus emptying by manual intrauterine aspiration
 #0411010018 Manual placental abruption
 # 0411010026 Caesarean delivery in high-risk pregnancies
 #0411010034 Caesarean delivery
 #0411010042 Caesarean delivery with tubal ligation
 #0411010050 Manual reduction of acute postpartum uterine inversion
 #0411010069 Postpartum episiotomy resuture
 # 0411010077 Suture of pelvic tract lacerations (in childbirth before admission)
 #0411010085 Surgical treatment of acute postpartum uterine inversion
 #0411020013 Post-abortion/puerperal curettage
 #0411020021 Embryotomy
 #0411020030 Puerperal hysterectomy
 #0411020048 Surgical treatment of ectopic pregnancy
 #0411020056 Treatment of other maternal disorders predominantly related to pregnancy
 #0417010028 Obstetric analgesia for normal delivery
 #0417010010 Obstetric anaesthesia for caesarean section
 # 0417010036 Obstetric anaesthesia for caesarean section in high-risk pregnancy

Supplementary Appendix 2

C++ deterministic matching routine to identify hospitalizations (AIH records) belonging to the same hospital episode in the AIH database from the SIH-SUS and Python routine to create a hospital episodes database.

Since the operation requires processing large amounts of data, sorting and iterating through it, we opted to work with a SQL database (SQLite, widely available free and open source database) as backend for the routine, which was developed as a command line (character mode) program in C++ for faster execution. The gcc version 11.4.0 toolchain was used. The source code and respective makefile are available at github: https://github.com/kencamargo/cuidados_multiplos/tree/main/processaih.

The processing was done under Linux, but since both the C++ and Python codes only used standard features of both languages, the programs can be run in any OS that supports a standard C++ toolchain and a Python 3 interpreter.

First, the CSV file is imported into a table, at the same time the necessary indexes are created, in order to speed up processing later. Only the fields that are going to be used are imported. Two additional fields, which will hold the episode code as reference

for the cluster (AIHREF) and a flag marking whether the event initiates or is part of a sequence or transference are added. The fields of the table are all variable length text fields, named as follows:

Field name	Description
CNES	Identifier of the health care facility
CEP	Postal code
MUNIC_RES	Municipality where the patient resides
NASC	Date of birth
DT_INTER	Date of admission
DT_SAIDA	Date of hospital discharge (AIH closure)
COBRANCA	Reason for hospital discharge
N_AIH	Individual identifier of the AIH form
AIHREF	Common identifier for all rows for a full hospital episode (added when imported)
FLAG	Flag marking whether row initiates or is part of a sequence or transfer
DIAG_PRINC	Main diagnosis on admission
PROC_REA	Principal Procedure on admission
US_TOT	Reimbursement value in US dollars
UTI_MES_TO	Days in an intensive care unit
alta	Reason for hospital episode discharge (consolidated)
criterio_primario	Flag marking whether episode constitutes an obstetric event

Second, the table is ordered by CNES, MUNIC_RES, NASC and DT_INTER. The table is traversed from beginning to end, and all the rows that have the same value for the first three fields, and have DT_INTER equal to or at most one day after DT_SAIDA of the previous rows get the AIHREF. In this step, we reconstitute the sequence of events happening in a single healthcare facility. The resulting table is exported as a CSV file for further processing if necessary.

In the next steps the program searches for all the rows that have COBRANCA showing a transferto another hospital, in order to add rows subsequent to the transfer to the episode group. The following steps are repeated until all the transfers have been processed:

The table is reordered by CEP, MUNIC_RES, NASC and DT_INTER. A second table is created with the following fields (all variable size text):

Field name	Description
ORIGAIHREF	Original common identifier
AIHREF	New common identifier
DT_INTER	Date of admission

Every row with a transfer discharge is compared with the next row in the sequence, and if the latter has the same value for the first three fields, has DT_INTER equal to or at most one day after DT_SAIDA of the previous row, and a different AIHREF from the row with the transfer discharge generates a row in the second table. For easier comprehension of the following steps, the row that originated the transfer will be referred to as the origin row and the following row in the sequence will be referred to as the destination row.

If the destination row was not previously flagged as being a sequence to a transfer (see below), the second table will have ORIGAIHREF set to the destination row AIHREF, and AIHREF set to the origin row AIHREF. If, however, the destination row is already part of a transference sequence, ORIGAIHREF in the second table will be set to the origin row AIHREF, and AIHREF in the second table will be set to AIHREF of the origin row. This is necessary because some odd transference sequences break already ordered permanence sequences. In both cases DT_INTER will be taken from the destination row.

In the next step, the first (main) table is ordered by AIHREF and DT_INTER, and the second table is ordered by DT_INTER. The second table is then traversed, and for every row ORIGAIHREF and AIHREF are retrieved. All the rows in the first table that have the same AIHREF as the ORIGAIHREF of the second table have that changed to the AIHREF of the second table, thus

grouping all the rows connected by the transfer. Those rows are also flagged as being part of a transfer sequence, changing the corresponding FLAG field.

After successive passes over the last two steps exhaust all the unconnected transfer rows, this part of the algorithm ends. At the end of processing, the resulting table is exported to a comma-separated values (CSV) file, with an added field with the sequential number of the row in its respective cluster.

Further processing takes place with a Python version 3.11 scripts accessing the SQLite database and generating an additional CSV file.

First, the processed table resulting from the processing with the C++ routine is ordered by AIHREF and DT_INTER. The first row initialises several variables in memory that will contain, respectively:

- AIHREF
- initial date (from DT_INTER)
- last date (from DT_SAIDA)
- COBRANCA
- total reimbursement (from US_TOT)
- total days in ICU (from UTI_MES_TO)
- number of episodes
- total number of rows which fit the obstetric admission criterion (from criterio_primario)

The table is traversed, and for each row:

If AIHREF is equal to the one stored in memory:

- if DT_INTER is smaller than the initial date in memory, the latter is replaced.
- if DT_SAIDA is greater than the initial date in memory, the latter is replaced.
- COBRANCA is stored in memory.
- US_TOT is added to total reimbursement.
- UTI_MES_TO is added to total days in ICU.
- number of episodes is incremented.
- criterio_primario is added to total rows fitting obstetric criterion.

On the other hand, if AIHREF is different, a record is added to the CSV file containing all the variables above, plus the calculated difference in days from initial date to last date, and all the variables are reinitialized.

