TITLE

Robust estimation of infant feeding indicators by longitudinal Electronic Health Records from birth up to two years of life.

AUTHORS

- García-de-León-Chocano R. Department of Information Technology. Hospital Virgen del Castillo. Yecla. Gerencia Área de Salud V - Altiplano. Servicio Murciano de Salud. Spain.
- Sáez C. Biomedical Data Science Lab, Instituto Universitario de Tecnologías de la Información y Comunicaciones (ITACA), Universitat Politècnica de València (UPV). Spain.
- Muñoz-Soler V. Doctoral student Nursing Department, Faculty of Health Sciences, University of Alicante (Spain). Spain.
- Oliver-Roig A. Department of Nursing, Faculty of Health Sciences, University of Alicante (Spain). Spain.
- García-de-León-González R. Department of Paediatrics, Hospital Virgen del Castillo, Yecla, Gerencia Área de Salud V-Altiplano, Servicio Murciano de Salud, Spain.
- García-Gómez Juan M. Biomedical Data Science Lab, Instituto Universitario de Tecnologías de la Información y Comunicaciones (ITACA), Universitat Politècnica de València (UPV). Spain.

ABSTRACT

Background and objective:

The Baby-Friendly Hospital Initiative (BFHI) is an international strategy aimed at improving breastfeeding practices in health care services. Regular monitoring of indicators is key for BFHI implementation and maintenance. Currently, routine data collected in electronic health records (EHR) is an excellent source for infant feeding monitoring, but data quality (DQ) assessment should be undertaken. The aim of this research is to perform a DQ assessment of EHR data to enable robust estimations of infant feeding indicators.

Methods:

We use the longitudinal series of healthcare contacts belonging to 6 427 children born from 2009 to 2018 in the Health Area V of Murcia (Spain). Longitudinal data came from EHR at hospital discharge and community infant health reviews up to 18 months. The data of each healthcare contact contained a 24-hour recall of infant feeding. We perform a DQ process in three phases: 1) an assessment of each-single-contact and the definition of their infant feeding status; 2) a longitudinal DQ assessment of completeness and consistency of the series of contacts to obtain meta-information that guides the duration calculus, for each case, of the different types of breastfeeding: exclusive breastfeeding (EBF), full breastfeeding (FBF) and any breastfeeding (ABF); and finally 3) a robust estimation of indicators and description of DQ of each indicator.

Results:

We found deficiencies of DQ in 34.46% of single contacts for EBF, 26.28% for FBF and 22.50% for ABF that were used to establish the infant feeding status. However, after longitudinal DQ assessment, we had obtained valid and reliable data rates for most indicators such as “median duration of breastfeeding” nearly 90%, both for FBF and ABF, and 64.79% for EBF.

Conclusions:

Despite the DQ deficiencies found in the raw data, the DQ assurance approach by indicators proposed in this work, allowed us to obtain a robust estimation of indicators with a significant percentage of subjects with valid information for monitoring and research. The estimations were consistent with results previously published. The methodology provided with this study allows a

continuous and reliable population monitoring of infant feeding indicators of BFHI from EHR data.

KEYWORDS

Electronic health record. Data quality. Indicator. Monitoring. Baby-Friendly Hospital Initiative. Infant feeding.

ABBREVIATIONS

World Health Organization (WHO)

United Nations Children’s Foundation (UNICEF)

Baby-Friendly Hospital Initiative (BFHI)

Indicators for assessing infant and young child feeding (IYCF)

Electronic health records (EHR)

Data quality (DQ)

Exclusive breastfeeding (EBF)

Full breastfeeding (FBF)

Any breastfeeding (ABF)

Infant feeding status (IFS)

Integrated data repository (IDR)

HIGHLIGHTS

* Infant feeding repositories from EHR routinely collected data are a valuable data source for monitoring and research.

* Routinely clinical data must be analyzed to assure its validity and reliability before reuse.

* A DQ assurance approach by indicators allows to increase the tolerance level to DQ errors and with it, decrease the probability of selection bias for its use in outcomes monitoring.

* The availability of DQ infant feeding repositories facilitates the implementation of the BFHI.

1. BACKGROUND

The first 1000 days of life - from conception to the one's second birthday - is a unique period of opportunity when the foundations of optimum health, growth, and neurodevelopment across the lifespan are established [1]. One of the key aspects of this period is adequate nutrition. Breastfeeding is the biological norm for infant feeding during early life and therefore provide numerous health benefits to the mother and the baby [2,3]. The World Health Organization (WHO) recommends exclusive breastfeeding during the first 6 months of life, and with adequate complementary foods until at least 2 years of age [4]. However, fewer than half of infants are breastfed according to these recommendations. Improving this reality *is critical for achieving global goals on nutrition, health and survival, economic growth and environmental sustainability* [2,5,6].

The Baby-Friendly Hospital Initiative (BFHI) is an international strategy from WHO/UNICEF aimed at improving breastfeeding practices in health care services [5] and is one of the most effective [7] and also cost-effective [8] interventions for the overall improvement in breastfeeding rates. Nevertheless, at present, only 10% of births in the world take place in a BFHI accredited hospital [9].

A major factor for the successful implementation and maintenance of the BFHI was defined by the regular monitoring of process and outcomes indicators [10-12]. The outcomes' indicators recommended for BFHI monitoring are the "indicators for assessing infant and young child feeding (IYCF)" defined by WHO/UNICEF [13]. In particular, *breastfeeding initiation, duration and exclusivity are major outcomes for breastfeeding studies* [14]. Also, to achieve BFHI accreditation, health facilities must demonstrate a rate of at least 75% exclusive breastfeeding at discharge.

Despite this, nowadays, regular monitoring of infant feeding is carried out by a minority of countries [15,16]. One of the main barriers to establishing a regular monitoring system is obtaining relievable population data, that brings to the problem to get a relevant set of variables to measure and assess the quality of large datasets. [10,17-20].

Currently, routine collected electronic health records (EHR) data from infant community care health reviews is an excellent data source candidate for infant feeding monitoring, research and surveillance due to its population coverage and its continuity in time [17,18,21-25]. Concerning this, there are previous research studies in infant feeding based on routine collected EHR [23,24,26,27].

Nevertheless, given that EHR data are routinely collected for clinical purposes rather than for research and monitoring, data quality (DQ) assessment should be undertaken to ensure the validity and reliability of data reuse in research, healthcare quality, or indicator monitoring [22,28-31]. In particular, for monitoring reuse, DQ has a real and direct impact on the validity and reliability of indicators [32]. Despite this, recent studies in infant feeding based in EHR do not specify DQ analysis [23,24,26,27], and there isn't used standard criteria for allocating breastfeeding status [24,26].

In previous works, the authors proposed a DQ assessment and assurance process for EHR data [33,34], including an application to perinatal data [35] based in 8 DQ dimensions: predictive value, correctness, duplication, consistency, completeness, contextualization, temporal-stability, and spatial-stability. Later, they designed a pilot project in Spain [36] that established the basis for reliable monitoring of infant feeding indicators, focusing on data standardization and data quality analysis, mainly of perinatal variables, without addressing the analysis of the longitudinal feeding register.

In this study, we pay attention to the DQ of the infant feeding follow-up registry. We use the longitudinal registry formed by the data at hospital discharge and community healthcare data from infant health reviews. This longitudinal registry contains the required variables to estimate all the IYCF indicators.

The longitudinal nature of these registries introduces several additional quality problems which need to be addressed. First, it is necessary for generating indicators derived from clinical variables that were not developed for this reuse a thoughtful construction of valid variables [30]. Also, due to DQ should be analyzed over time [31,37-40], it is needed to assess longitudinal completeness and consistency problems. Therefore, a simple mapping is not enough for monitoring interventions from follow-up registries [30].

Given this complexity to generate valid and reliable information for monitoring from longitudinal registries, the question is: ¿Is it possible to assure a robust estimation of infant feeding indicators from routinely collected EHR data?

The aim of this research is to perform a DQ assessment, and an assurance process of EHR routine collected data to enable robust estimations of infant feeding indicators. Here, we report the process of derived variables construction and DQ meta-information generation needed for valid and reliable estimation of infant feeding indicators.

2. MATERIAL

We use population infant feeding data from the three public healthcare centres of Health Area V from Murcia (Spain), corresponding to 6 427 children born from 1 January 2009 to 31 December 2018. The 84.59% (5 436) of these children were born in the public hospital (Virgen del Castillo Hospital) of the Health Area V. The infant feeding data series come from the EHR forms filled from the birth up to two years of age during consecutive health care contacts. All children born in the hospital had a first contact filled at hospital discharge. The rest of contacts corresponded to seven community care health reviews of the Healthy Child Programme at 8 specific time points: before the first month, at 2 months, at 4 months, at 6 months, between 9 and 12 months, at 15 months and 18 months. All reviews, except the “before the first month” and “between 9 and 12 months” reviews, coincide with vaccination points.

EHR feeding forms have two types of information: 1) current data formed by a 24-hour recall of infant feeding ("current data" is the usual term in infant feeding research articles to refer to the current state of breastfeeding in a timely survey or health review), and 2) retrospective data formed by infant feeding milestones (the age for the introduction of other types of food different from breast milk and the age of breastfeeding cessation). Different forms have some specific items according to the age of the child in the contact.

Thus, the initial data set was formed by 6 427 subjects with a total of 58 368 healthcare contacts routinely collected from EHR. Table 1 introduces the structure of the raw data of contacts.

Contextual data
Subject ID (common for hospital and primary care information systems)
Date of birth
Healthcare centre where the child belongs
Basic health unit where the child belongs
Date of the health review
Age of the subject of care during the contact
The user that register the registered
Date and time at which the observation is recorded
Somatometry
Weight the day of the health review
Height the day of the health review
Cephalic perimeter the day of the health review
Structured infant feeding data
24-hour recall of infant feeding. ***[5 current data items]:***
Type of breastfeeding? (Exclusive breastfeeding, Predominant breastfeeding, Partial breastfeeding, Bottle-feeding)
Has the baby received breastmilk in the last 24 hours? (Yes, No)
The baby has received water, infusions, oral rehydration solution and/or fruit juices in the last 24 hours? (Yes, No)
Has the baby received complementary feedings in the form of solid or semi-solid food in the last 24 hours? (Yes, No)
Has the baby received complementary feedings in the form of infant formula in the last 24 hours? (Yes, No)
Infant-feeding milestones (age of introduction of different types of food and age of breastfeeding cessation). ***[14 retrospective data items]:***
Liquids, Gluten-free cereals, Cereals containing gluten, Fruits, Vegetables, Formula milk, Chicken, Fish, Legumes, Egg yolk, Whole egg, Yogurt, Cow milk, Breastfeeding cessation

Table 1. Structure of the raw data routinely collected from EHR composed of contextual data of the subject of care and the contact, somatometry data and structured infant feeding data. The structured infant feeding data included both current data and retrospective data.

This structure was used to construct variables and estimate infant feeding indicators, according to most accepted definitions of different types of breastfeeding [13,41,42]. Table 2 introduces the indicators extracted from IYCF indicators defined by WHO/UNICEF that we used for monitoring of breastfeeding [13]: Exclusive breastfeeding (EBF), full breastfeeding (FBF) and any breastfeeding (ABF).

Indicator	Description
EBF at hospital discharge	The proportion of infants at hospital discharge who are fed exclusively with breastmilk (EBF) during the previous day
EBF under 15 days	The proportion of infants 0–15 days of age who are fed exclusively with breastmilk (EBF) during the previous day
EBF under 6 months	The proportion of infants 0–5 months of age who are fed exclusively with

	breastmilk (EBF) during the previous day
FBF under 6 months	The proportion of infants 0–5 months of age who are fed with breastmilk and may also have received non-nutritive liquids (FBF) during the previous day
Continued breastfeeding at 1 year	The proportion of children 12-15 months of age who are fed breastmilk (ABF)
Median duration of breastfeeding (ABF)	The median duration of breastfeeding (ABF) among children less than 24 months of age.
Median duration of FBF	The median duration of FBF among children less than 24 months of age.
Median duration of EBF	The median duration of EBF among children less than 24 months of age.

Table 2. Indicators for infant feeding monitoring.

3. METHODS

An expert group set up by 4 leading clinicians (RGLG, AOR, VMS and MAQ) of the health area in this domain and 3 information technology experts (RGLC, CSS, JMGG) defined the DQ assurance process to achieve a robust estimation of infant feeding indicators.

The inclusion criteria were all children with infant feeding follow-up in any of the three public healthcare centres of Health Area V. We defined a general exclusion criteria for subjects that do not meet a minimum follow-up: less than 3 Healthy Child Programme reviews, first review after 120 days of age or last review before 180 days. All data was anonymized previous to the analyses.

Each subject included (case) in the study had a serial of healthcare contacts. Thus, for the robust estimation of infant feeding indicators, we performed a DQ assurance process over current and retrospective data, both in every single contact and in the whole series of contacts that conform each case. The DQ assurance process consisted in three phases: 1) Assessment of current information on the feeding of each-single-contact and definition of the infant feeding status at every single contact (IFS); 2) Longitudinal DQ assessment of the series of contacts through the analysis of the completeness and consistency of the longitudinal feeding information in the series of contacts (combining current and retrospective information) to calculate the duration of the different types of breastfeeding for each case; and 3) Robust estimation of Indicators for infant feeding monitoring and the description of DQ of each indicator. Figure 1 summarizes this DQ process. The final output is an integrated data repository (IDR) of maternal-child care up to two years of age that incorporates a DQ assured estimation of infant feeding indicators.

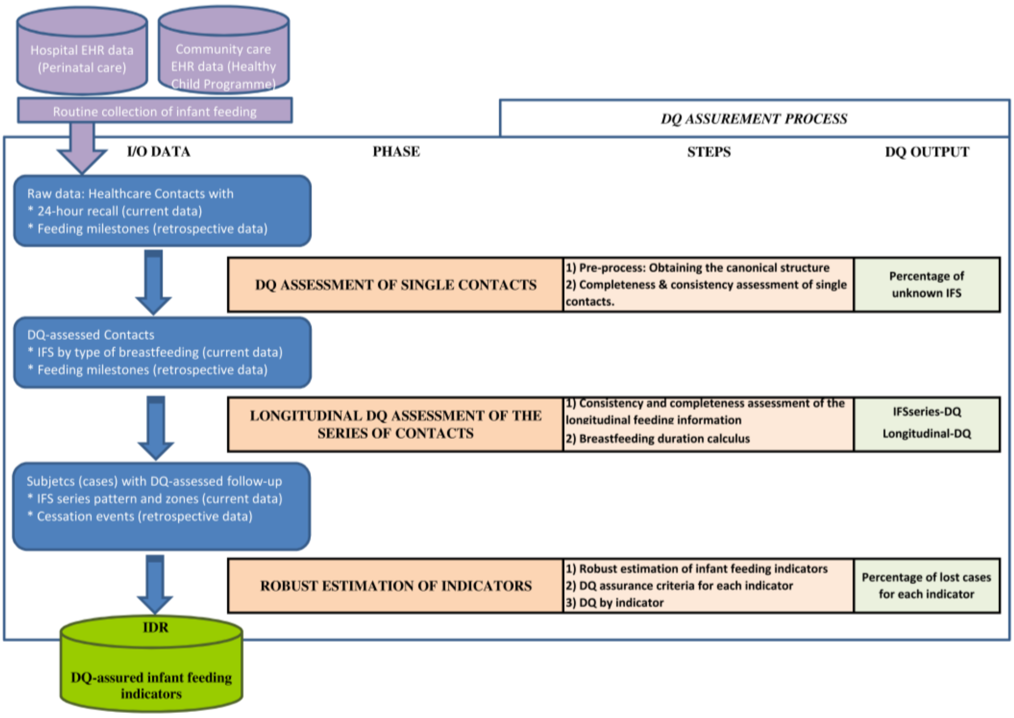

Figure1: DQ assurance process for a robust estimation of infant feeding indicators. On the top, in magenta colour: data sources of the routine collection. For the DQ assurance process are shown: the input/output data involved in each phase, the steps and DQ outputs.

3.1. Phase 1: DQ assessment of single contacts

The DQ assessment of single contacts was performed in two steps. We first performed a pre-process using DQ methods and metrics from previous works [34,35] to assure the structural integrity of the data and solved de-duplication of data, obtaining a canonical structure where the DQ dimensions of contextualization, temporal stability, spatial stability and duplicity were assessed.

Second, we assessed the completeness and consistency of every single contact of each case, specifically the current data. The information about feeding milestones did not require analysis in each contact given its retrospective nature. We assessed the DQ of the 24-hour recall variables (see current data items Table 1) to obtain the infant feeding status (IFS) in 3 derived binary variables, one for the type of breastfeeding (EBF, FBF, ABF). The IFS represents if the infant feeds according to the specific type of breastfeeding in every single contact. If DQ of 24-

hour variables prevents to determine the IFS for a type of breastfeeding, we tagged the contact as “unknown IFS” for the type of breastfeeding.

3.2. Phase 2: Longitudinal DQ assessment of the series of contacts

This phase was performed in two steps: 1) Consistency and completeness assessment of the longitudinal feeding information in the series of contacts (combining current data and retrospective data) using a base of rules that gathers the expert knowledge to assess this information; and 2) Duration calculus of the different types of breastfeeding for each case guided by the DQ meta-information obtained in the previous step.

3.2.1. Consistency and completeness assessment of the longitudinal feeding information in the series of contacts

We analyzed the series of contacts for each case morphologically, both current data (series of values of IFS variables from the previous phase), retrospective data (series of values of infant feeding milestones) and combination information of both.

For the morphological analysis of these series, we defined key elements that allowed establishing a set of DQ attributes for the DQ assessment, detailed in Table 3.

Elements defined for current data series	
IFS series key moments	**Description**
T0	Age (in days) at the hospital discharge.
T1	Age (in days) at the first community care contact where the IFS was positive (the baby was identified as being fed according to the type of breastfeeding).
T2	Age (in days) of the last contact where IFS was positive (always before a negative IFS appears).
T3	Age (in days) of the first contact where the IFS status was negative (the baby was identified as not being fed according to the type of breastfeeding).
IFS series patterns	**Description**
Duration-Uncensored	The breastfeeding begins and ends within the cohort. The series have T1, T2 and T3 moments. T2 and T3 delimit the breastfeeding cessation interval.
Duration-Censored	Continue breastfeeding before the end of the cohort (18 months). The series only have T1 and T2. The breastfeeding cessation interval is delimited by T2 and end of cohort.
No breastfeeding	Breastfeeding is not started (T1 and T2 do not appear, only exist T3 in the series). The breastfeeding cessation interval is delimited by birth and T3.
Null pattern	No valid information about breastfeeding.
IFS series zones	**Description**
Start	Between the birth and T1.
Core	Between T1 and T2.
Breastfeeding cessation interval	See definition in each pattern.
End	From T3 until the end of series.
DQ of current data series	
DQ attributes and score of IFS series	**Description**
Null Start Contacts	The number of null contacts of the start zone.
Valid Core Contacts	The number of contacts set to 1-Yes in the core zone.
Null Core Contacts	The number of null contacts of the core zone.
Completeness Ratio of Core Contacts	The proportion of valid contacts versus null contacts inside the core.
Null breastfeeding cessation interval Contacts	The number of null contacts in the breastfeeding cessation interval.
Null End Contacts	The number of null contacts of the end zone.
Inconsistent End Contacts	The number of contacts set to 1-Yes from T3 until the end of the series.
Valid End Contacts	The number of contacts set to 2-No from T3 until the end of the series.
IFSseries-DQ	The score for DQ summary of IFS series (from 0 to 100)
Elements defined for retrospective data series	
Cessation events	**Description**
Age of liquids introduction	The minimum value of milestones related to non-nutritive liquids.
Age of foods introduction	The minimum value of milestones related to foods (solids, semi-solids or formula).
Breastfeeding cessation	The minimum value of breastfeeding cessation.
Global DQ score (current data & retrospective data combination)	
Longitudinal-DQ	Likert value according to the position of the cessation events in relation to the breastfeeding cessation interval

Table 3. Elements defined for the analysis and DQ assessment of the series of contacts, both for current data of IFS and retrospective data, and the combination of both. The elements, DQ attributes and scores of the table were defined and calculated for each type of breastfeeding.

Specifically, for current data, the series for each type of breastfeeding was composed for the sequence of values of the correspondent IFS. From 4 key moments (T0, T1, T2, T3) described in Table 3, we could identify a set of IFS series patterns: Duration-Uncensored, Duration-

Censored, No breastfeeding and Null pattern (not valid information due to severe problems of completeness in the series due to unknown contacts). Besides, according to the key moments, the series of contacts were divided into four time zones (Start, Core, Breastfeeding cessation interval and End) that delimit the status transitions within the type of breastfeeding. From these elements, a set of DQ attributes were established to assess the completeness and consistency of the IFS series. This DQ assessment was summarized in a DQ score that we call “IFS series-DQ” weighing the attributes and normalizing the result on a scale of 0 (for IFS series with DQ problems that prevent determining the pattern) to 100 (for IFS series without DQ problems).

On the other hand, for retrospective data (infant-feeding milestones), we determined three cessation events expressed in days for the age of liquids introduction, age of foods introduction and age breastfeeding cessation using the minimum value of the correspondent feeding milestones for each case.

Finally, we contrasted the information extracted from current data and retrospective data by comparing the elements of the IFS series with the cessation events. We use a base of rules in the form of a table of values that includes 109 entries according to values of the type of breastfeeding, the pattern and zones of the IFS series and the relative position of the cessation events respect the breastfeeding cessation interval. The result of the evaluation was the global score of DQ, that we called “Longitudinal-DQ”, expressed in the following Likert scale:1- Without longitudinal DQ problems (all events related to the specific type of breastfeeding within the breastfeeding cessation interval of a valid IFS series pattern); 2- Completeness problems in cessation events (cessation events related to the specific type of breastfeeding were not registered); 3- Inconsistency in end zone (the most relevant events for the type of breastfeeding were after breastfeeding cessation interval); 4- Inconsistency in core zone (the most relevant events for the type of breastfeeding were before breastfeeding cessation interval); and 5- IFS series with a null pattern.

3.2.2. Breastfeeding duration calculus for each case

We calculated the breastfeeding duration for each case from current and retrospective data and guided by the DQ meta-information of the longitudinal DQ assessment. As is showed in Figure 2, we established a punctual duration of breastfeeding in two cases: a) when retrospective data is complete and consistent with current data, the punctual duration of each type of breastfeeding was established to the particular cessation event, we mean, from retrospective data; or b) when we had not cessation events, or they were inconsistent with current data, the expert group decided to establish the punctual duration to the midpoint of the breastfeeding cessation interval for Duration-Uncensored patterns and the minimum of the interval for Duration-Censored patterns.

On the other hand, we defined the cases with missing information for a type of breastfeeding when a) data inconsistencies in the core zone and the DQ of current data is poor (IFS series-DQ <75) where the expert group decided not to establish a valid duration value; or b) when we couldn't get the breastfeeding cessation interval due to a null pattern of IFS series a duration value could not be determined.

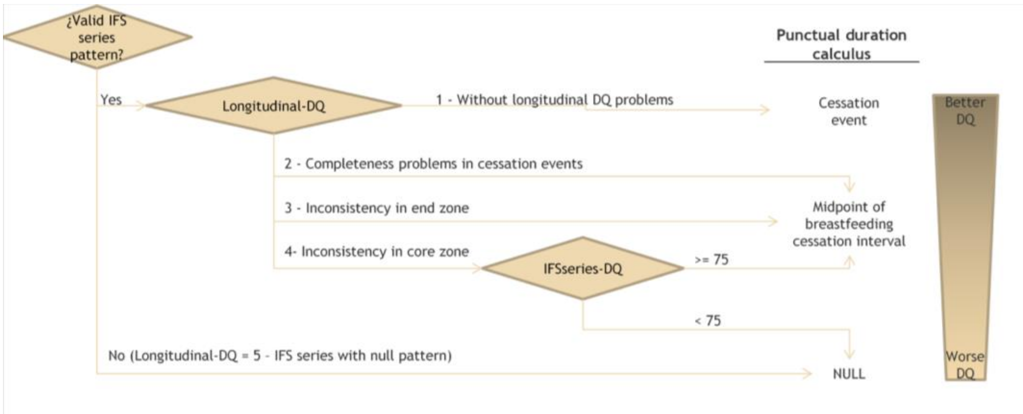

Figure 2: Breastfeeding duration calculus guided by DQ. Algorithm to calculate the duration according to the DQ of the series of contacts (IFS series pattern, Longitudinal-DQ score and IFS series-DQ). The calculation types are ordered from top to bottom from better to worse DQ.

3.3. Phase 3: Robust estimation of indicators

From the duration calculated for each case and type of breastfeeding in the previous phase, we estimated the value of indicators for infant feeding monitoring using the actuarial method and guided by the DQ meta-information (Figure 2). By the actuarial method, we grouped the cessation of breastfeeding events, from birth to 720 days, in intervals of thirty days, assigning to each case, in each interval, a value of 1 when breastfeeding continued, a value of 2 when it was ceased, and a missing value if the data was not available.

The expert group defined for each infant feeding indicator the DQ assurance criteria to obtain three DQ levels: standard DQ, tolerable DQ problems and non-tolerable DQ problems, the latter leading to lost cases.

DQ assurance criteria were established as described below. Non-tolerable DQ problems (lost cases): null breastfeeding duration for the type of breastfeeding of the indicator; for all other cases (valid cases), the standard DQ cases were those where the duration was determined by cessation event, the rest of valid cases (where duration was established to the midpoint of the cessation interval) were cases with tolerable DQ problems.

4. RESULTS

From the 6 427 children, the percentage of excluded subjects for not meeting the minimum follow-up was 10.33% (664 subjects). The final included population was 5 763 subjects with a total of 60 885 healthcare contacts. These contacts are homogeneously distributed between the sequence of contacts: 87.51% of subjects had a contact at hospital discharge, 74.51% had the review before the first month, 97.78% at 2 months, 97.88% at 4 months, 97.83% at 6 months, 96.41% between 9 and 12 months, 83.71% at 15 months and 69.18% at 18 months.

4.1. DQ assessment of single contacts results

All the variables and values of the initial dataset collected from EHR were harmonized and de-duplicated obtaining a homogenous and stable structure of contacts along the period of study.

From the DQ analysis of completeness and consistency of current data of single contacts: the number of contacts with unknown IFS, where it was not possible to determine the IFS (IFS-DQ = 0), according to the type of breastfeeding was: 20 979 contacts (34.46%) for EBF, 16 003 (26.28%) for FBF and 13 698 (22.50%) for ABF. In particular, Figure 3 shows the distribution of unknown IFS throughout the sequence of contacts (at hospital discharge and reviews of the Healthy Child Programme).

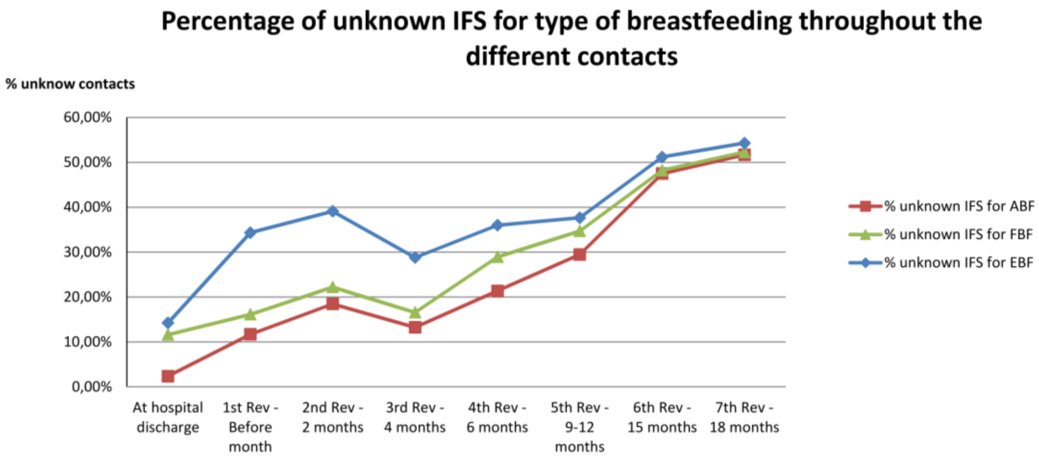

Figure 3: Percentage of unknown IFS by type of breastfeeding throughout the different contacts (at hospital discharge and reviews of the Healthy Child Programme). On the horizontal axis, we show the sequence of contacts. With blue line, we present the percentage of unknown IFS for EBF, the green line for FBF and red line for ABF.

4.2. Longitudinal DQ assessment results

Figure 4 shows the results of the assessment of the series of contacts summarized by the Longitudinal DQ Score (step 1 of phase 2). We show the categorization of the Longitudinal-DQ score for each type of breastfeeding of the 5 763 cases.

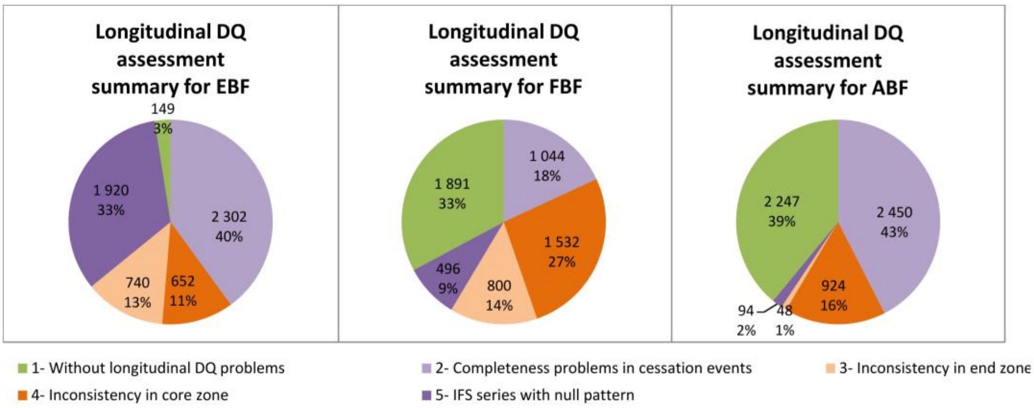

Figure 4. Categorization of longitudinal DQ problems for the 5 763 included subjects. We show a pie chart by type of breastfeeding, from left to right: Longitudinal DQ assessment for EBF, FBF and ABF.

4.3. Results of DQ assurance for robust estimation of indicators

Table 4 introduces the robust estimation indicators results with the corresponding percentage of lost cases (subjects with non-tolerable DQ problems respect the assurance criteria established by the expert group).

Indicator	% of lost cases	Estimation
EBF at hospital discharge	14.24%	77.48%
EBF under 15 days	40.15%	64.75%
EBF under 6 months	38.12%	44.98%
FBF under 6 months	13.31%	55.63%
Continued ABF at 1 year	2.20%	23.86%
Median duration of ABF	2.20%	7 months
Median duration of FBF	10.79%	3 months
Median duration of EBF	35.21%	1 month

Table 4. Robust estimation of indicators for different types of breastfeeding: EBF, FBF and ABF. We show for each indicator the percentage of lost cases and the robust estimation value (for the valid subjects).

Finally, we introduce Figure 5 to show the effect of DQ on the estimations for the example of the median duration of breastfeeding key indicators. We show the percentage of subjects attending to their DQ level and the estimation of the indicator for sub-population of subjects with standard DQ and subjects with tolerable DQ problems.

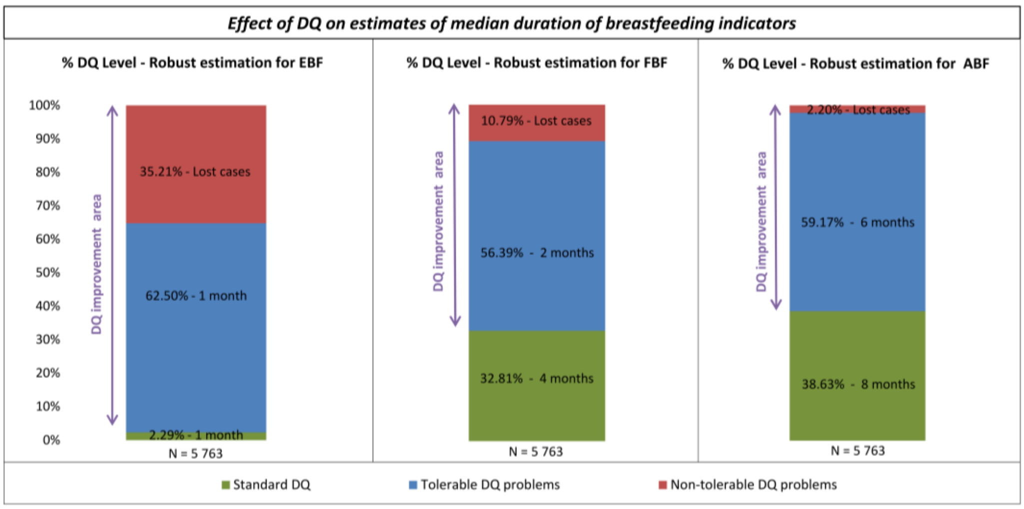

Figure 5. Effect of DQ on estimates of median duration of breastfeeding indicator. We show a column chart for the median duration indicator by type of breastfeeding, from left to right, the

median duration of EBF, FBF and ABF. The columns represent the percentage of subjects attending to their DQ level and the robust estimation for the subpopulations of valid cases. Subjects with standard DQ are showed in green and subjects with tolerable DQ problems in blue. Lost cases are shown in red. We also point out the DQ improvement area.

5. DISCUSSION

The calculation of international infant feeding indicators, as BFHI outcomes, is usually based on costly ad-hoc surveys, which are carried out over long spaced periods. This work introduces a robust estimation of indicators through a DQ assurance process of EHR.

We had obtained valid and reliable data rates for all indicators of FBF and ABF. In particular, for the key indicator “median duration of breastfeeding”, the percentage of valid cases was close to 90%, both for FBF and ABF. Also, a rate of 64.79% of valid cases has been obtained for EBF.

Despite the relevance and advantages of using EHR data for infant feeding monitoring and research [17,18,21-25], to our understanding, this is the first study that deals with the assessment of DQ to obtain a robust estimation of IYCF indicators over a population-based IDR of routine collected EHR. There are two examples of infant feeding population-based studies that use EHR registries exclusively [23,27], but none of them mentions the use of a DQ assessment procedures and one of them needs a complementary manual review of EHRs [27]. We have not found references of other population-based and DQ-assured infant feeding repositories up to 18 months based on EHR for reliable infant feeding outcomes monitoring.

We found DQ problems in single contacts to establish the value of IFS (unknown IFS) in 34.46% for EBF, 26.28% for FBF and 22.50% for ABF. These values are not distributed homogeneously throughout the sequence of contacts but increase according to the age of the review of the Healthy Child Programme as we can see in Figure 3. For the first contact at hospital discharge, the percentage of unknown IFS is low due to the mandatory nature of the EHR hospital variables related to the IFS.

However, after longitudinal analysis, as we can see in Figure 4, the percentage of subjects with a null pattern for the IFS series (Longitudinal-DQ = 5) was only 9% for FBF and 2% for ABF. For EBF, the percentage was 33% due to the higher requirements for this type of breastfeeding. Moreover, we obtained a significant percentage of subjects (33% for FBF and 39% for ABF, not so for EBF) where the retrospective data was complete and consistent with current data (Longitudinal-DQ = 1) for which we could calculate a precise, punctual duration of breastfeeding based on the cessation event. For the rest of the subjects (with Longitudinal-DQ = 2, 3 or 4), only the current information could be used for the duration calculus.

Despite DQ problems founded, the DQ meta-information obtained allowed us to define assurance criteria for each indicator, obtaining a valid population for infant feeding monitoring as we can see in Table 4.

For the particular case of EBF indicators, the percentage of lost cases is higher than the other types of breastfeeding. The problem to measure EBF is that, unlike the formula or solids, the introduction of liquids can be discontinuous without necessarily having to condition a change of type of breastfeeding, which can generate problems of inconsistency. The high percentage of lost cases does not apply to “EBF at hospital discharge” indicator due to the mandatory nature of infant feeding variables in the hospital EHR, as previously mentioned.

Moreover, we can see the effect of DQ on estimates of indicator in Figure 5 for the particular case of the median duration of breastfeeding indicators. Better breastfeeding rates were obtained for subjects with Standard DQ. This may be due both to the more accurate estimate that was made for cases with Standard DQ.

Thus, we had obtained an IDR for infant feeding monitoring and research from birth up to 18 months of age, with routinely collected EHR data both from hospital care and primary care from 2009 to 2018. Despite the DQ problems found in the raw data, the DQ assurance approach by indicators proposed in this work, allowed us to obtain a robust estimation of indicators with a significant percentage of subjects with valid information for monitoring and research.

To control the possible selection biases by excluding subjects with non-tolerable problems (lost cases), in addition to the feeding data, the repository incorporates sufficient perinatal information [35] to compare the distribution of the population of valid subjects and lost subjects according to the factors related to the abandonment of breastfeeding described in the literature [43,44] (maternal age, early skin-to-skin contact, assessment of breastfeeding initiation, type of delivery, gestational age, …).

Finally, the estimation of the indicators (Table 4) was consistent with the results previously published [45]. Moreover, we used data from a previous Patient-Reported Outcomes multicentric study (PI09/90899) performed in our health area for a sample of 241 mothers who gave birth in the period 2009-2011 to compare its results with the results of our indicators, also in the 2009-2011 period, and we obtained similar values for the indicators.

5.1. Limitations

A precise duration could only be obtained for subjects without DQ problems, however, given the low proportion of these subjects it was necessary to incorporate subjects with tolerable DQ where the duration was obtained as the midpoint of the breastfeeding cessation interval.

Regarding the extension in other environments, to generate the base of rules, the follow-up feeding forms of EHR must be composed of: 1) a 24-hour recall with the necessary items to be able to calculate the feeding status according to the definitions of the WHO; and 2) sufficient feeding milestones to determine cessation events. Moreover, in case of not having data at hospital discharge, specific treatment should be considered, since the breastfeeding initiation data are not available (truncated data).

5.2. Future work

The detailed and systematic meta-information obtained during the DQ assessment will allow us to implement a Total Data Quality Management process for the monitoring and evaluation of DQ indicators. This process will allow us to address the important DQ improvement area (Figure 5), focusing on the main problems detected and in particular those related to EBF.

From this repository, a series of epidemiological studies: 1) predictive validity of the perinatal variables on the duration of different types of breastfeeding; 2) co-relation of the duration of breastfeeding with quality of life, childhood obesity, childhood asthma and attention deficit hyperactivity disorder. These studies will be the basis for formal validation of the usability of these repositories.

6. CONCLUSION:

The methodology provided with this study, allows a continuous and reliable population monitoring of the infant feeding indicators of BFHI from EHR routinely collected data.

Routinely collected data show, in general, DQ problems. However, our DQ assurance approach by indicators allowed us to increase the tolerance level to DQ errors and with it, decrease the probability of selection bias for its use in outcomes monitoring.

Conflict of interest

All the authors declare that they do not have any conflict of interest.

Acknowledgements

The authors would like to acknowledgement the clinical and management expert Manuel Alcaraz Quiñonero (MAQ) of his suggestions for data analysis.

This research is partially supported by the Spanish Ministry of Health, Social Services and Equality (2015/07PN0010), within the framework of a pilot project envisaged to the development of a National IDR of maternal-child care information.

REFERENCES

[1] United Nations Children’s Fund, World Health Organization, World Bank, Early Childhood Development Action Network, The Partnership for Maternal, Newborn and Child Health. Nurturing care framework for early childhood development: a framework for helping children survive and thrive to transform health and human potential. Geneva, 2018. https://nurturing -care.org/resources/Nurturing_Care_Framework_en.pdf.

[2] C.G. Victora, R. Bahl, A.J. Barros, G.V. França, S. Horton, J. Krasevec, et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet. 2016; 387: 475-90. https://doi.org/10.1016/S0140- 6736(15)01024-7.

[3] A. Cattaneo. Academy of Breastfeeding Medicine Founder's Lecture 2011: Inequalities and Inequities in Breastfeeding: An International Perspective. Breastfeed Med. 2012; 7 (1): 3-9. https://doi.org/10.1089/bfm.2012.9999.

[4] WHO, UNICEF. Global strategy for infant and young child feeding. Geneva: WHO, 2003. Available at: https://apps.who.int/iris/bitstream/handle/10665/42590/9241562218.pdf?sequence=1 (Accessed January, 2019).

[5] WHO, UNICEF. Implementation guidance: protecting, promoting and supporting breastfeeding in facilities providing maternity and newborn services – the revised Baby-friendly Hospital Initiative. Geneva: WHO, 2018. Available at: https://apps.who.int/iris/bitstream/handle/10665/272943/9789241513807-eng.pdf?ua=1 (Accessed January, 2019).

[6] N.C. Rollins, N. Bhandari, N. Hajeebhoy, S. Horton, C.K. Lutter, J.C. Martines, et al. Why invest, and what it will take to improve breastfeeding practices? Lancet. 2016; 387: 491-504. https://doi.org/10.1016/S0140-6736(15)01044-2.

[7] R. Pérez-Escamilla, J.L. Martínez, S. Segura-Pérez. Impact of the Baby-friendly Hospital Initiative on breastfeeding and child health outcomes: a systematic review. Matern Child Nutr. 2016; 12(3): 402-17. https://doi.org/10.1111/mcn.12294.

[8] S. Horton, T. Sanghvi, M. Phillips, J. Fiedler, R. Pérez-Escamilla, C. Lutter, et al. Breastfeeding promotion and priority setting in health. Health Policy Plan. 1996; 11: 156-68. https://doi.org/ 10.1093/heapol/11.2.156.

[9] WHO. National implementation of the Baby-friendly Hospital Initiative. Geneva: WHO, 2017. Available at: http://apps.who.int/iris/bitstream/10665/255197/1/9789241512381-eng.pdf?ua=1 (Accessed January, 2019).

[10] R. Pérez-Escamilla, D.J. Chapman. Breastfeeding protection, promotion, and support in the United States: a time to nudge, a time to measure. J Hum Lact. 2012; 28(2): 118-21. https://doi.org/10.1177/0890334412436721.

[11] R. Pérez-Escamilla. Evidence based breast-feeding promotion: the Baby-Friendly Hospital Initiative. J Nutr. 2007; 137(2): 484-7. https://doi.org/10.1093/jn/137.2.484.

[12] I. Zakarija- Grković, M. Boban, S. Janković, A. Ćuže, T. Burmaz. Compliance With WHO/UNICEF BFHI Standards in Croatia After Implementation of the BFHI. J Hum Lact. 2018; 34(1): 106-115. https://doi.org/10.1177/0890334417703367.

[13] WHO/UNICEF. Indicators for assessing infant and young child feeding practices. Part 1 Definitions. Washington: WHO/UNICEF, 2007. Available at: https://www.unicef.org/nutrition/files/IYCF_updated_indicators_2008_part_1_definitions.pdf (Accessed January, 2019).

[14] N.K. Wood, N.F. Woods. Outcome measures in interventions that enhance breastfeeding initiation, duration, and exclusivity: A systematic review. MCN Am J Matern Child Nurs. 2018; 43(6): 341–7. https://doi.org/10.1097/NMC.0000000000000472.

[15] United Nations Children's Foundation. UNICEF data: monitoring the situation of children and women. Access the data: infant and young child feeding. Available at: http://data.unicef.org/topic/nutrition/infant-and-young-childfeeding. (Accessed January, 2019).

[16] R. Pérez-Escamilla, V. Hall Moran. Scaling Up Breastfeeding Programmes in a Complex Adaptive World. Matern Child Nutr. 2016; 12(3): 375-80. https://doi.org/10.1111/mcn.12335.

[17] N.C. Nickel, L. Warda, L. Kummer, J. Chateau, M. Heaman, C. Green, et al. Protocol for establishing an infant feeding database linkable with population-based administrative data: a prospective cohort study in Manitoba, Canada. BMJ Open. 2017; 7(10): e017981. https://doi.org/10.1136/bmjopen-2017-017981.

[18] J.A. Paul, J. Chateau, C. Green, L. Warda, M. Heaman, A. Katz, et al. Evaluating the Manitoba Infant Feeding Database: a Canadian infant feeding surveillance system. Can J Public Health. 2019; 110: 649. https://doi.org/10.17269/s41997-019-00211-6.

[19] European Centre for Disease Prevention and Control. Data quality monitoring and surveillance system evaluation: a handbook of methods and applications. Stockholm: ECDC, 2014. Available at: https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/Data-quality-monitoring-surveillance-system-evaluation-Sept-2014.pdf (Accessed January, 2019).

[20] D.J. Chapman, R. Pérez-Escamilla. US National Breastfeeding Monitoring and Surveillance: Current Status and Recommendations. J Hum Lact. 2009; 25(2): 139-150. https://doi.org/10.1177/0890334409332437.

[21] Euro-Peristat project with SCPE and Eurocat. European Perinatal health report. The health of pregnant women and babies in Europe in 2010. European Perinatal Health Report 2010. Available at: https://www.europeristat.com/images/doc/EPHR2010_w_disclaimer.pdf (Accessed January, 2019).

[22] C. Weng, P. Appelbaum, G. Hripcsak, I. Kronish, L. Busacca, K.W. Davidson, et al. Using EHRs to integrate research with patient care: promises and challenges. J Am Med Inform Assoc. 2012; 19(5): 684-7. https://doi.org/10.1136/amiajnl-2012-000878.

[23] M.K. Halvorsen, E. Langeland, G. Almenning, S. Haugland, L.M. Irgens, T. Markestad, et al. Breastfeeding surveyed using routine data. Tidsskr Nor Laegeforen. 2015; 135(3): 236-41. https://doi.org/10.4045/tidsskr.14.0133.

[24] O. Ajetunmobi, B. Whyte, J. Chalmers, M. Fleming, D. Stockton, R. Wood. Informing the 'early years' agenda in Scotland: understanding infant feeding patterns using linked datasets. J Epidemiol Community Health. 2014; 68: 83-92. https://doi.org/10.1136/jech-2013-202718.

[25] American Academy of Pediatrics Steering Committee on Quality Improvement and Management; American Academy of Pediatrics Committee on Practice and Ambulatory Medicine, Hodgson ES, Simpson L, Lannon CM. Principles for the development and use of quality measures. Pediatrics. 2008; 121(2): 411-8. https://doi.org/10.1542/peds.2007-3281.

[26] M. Busck-Rasmussen, S.F. Villadsen, F.N. Norsker, L. Mortensen, A.M. Nybo Andersen. Breastfeeding practices in relation to country of origin among women living in Denmark: a population-based study. Matern Child Health J. 2014; 18: 2479-88. https://doi.org/10.1007/s10995-014-1486-z.

[27] E. Bartsch, A.L. Park, J. Young, J.G. Ray, K. Tu. Infant Feeding Practices Within a Large Electronic Medical Record Database. BMC Pregnancy Childbirth. 2018; 18(1). https://doi.org/10.1186/s12884-017-1633-9.

[28] R.A. Verheij, V. Curcin, B.C. Delaney, M.M. McGilchrist. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res. 2018: 20(5): e185. https://doi.org/10.2196/jmir.9134.

[29] N.G. Weiskopf, C. Weng. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013; 20: 144-51. https://doi.org/10.1136/amiajnl-2011-000681.

[30] R. Khare, L. Utidjian, B.J. Ruth, M.G. Kahn, E. Burrows, K. Marsolo, et al. A Longitudinal Analysis of Data Quality in a Large Pediatric Data Research Network. J Am Med Inform Assoc. 2017; 24 (6): 1072-79. https://doi.org/10.1093/jamia/ocx033.

[31] A.P. Reimer, A. Milinovich, E.A. Madigan. Data quality assessment framework to assess electronic medical record data for use in research. Int. J. Med. Inform. 2016; 90: 40-7. https://doi.org.10.1016/j.ijmedinf.2016.03.006.

[32] K. Dentler, R. Cornet, A. ten Teije, P. Tanis, J. Klinkenbijl, K. Tytgat, et al. Influence of data quality on computed Dutch hospital quality indicators: a case study in colorectal cancer surgery. BMC Med Inform DecisMak. 2014; 14: 32. https://doi.org/10.1186/1472-6947-14-32.

[33] C. Sáez, J. Martínez-Miranda, M. Robles, J.M. García-Gómez. Organizing data quality assessment of shifting biomedical data. Stud Health Technol Inform. 2012; 180: 721-25. https://doi.org/10.3233/978-1-61499-101-4-721.

[34] R. García-de-León-Chocano, C. Sáez, V. Muñoz-Soler, R. García-de-León-González, J.M. García-Gómez. Construction of quality-assured infant feeding process of care data repositories: definition and design (Part 1). Comput Biol Med. 2015; 67: 95-103. https://doi.org/10.1016/j.compbiomed.2015.09.024.

[35] R. García-de-León-Chocano, V. Muñoz-Soler, C. Sáez, R. García-de-León-González, J.M. García-Gómez. Construction of quality-assured infant feeding process of care data repositories: Construction of the perinatal repository (Part 2). Comput Biol Med. 2016; 71: 214-22. https://doi.org/10.1016/j.compbiomed.2016.01.007.

[36] C. Sáez, D. Moner, R. García-De-León-Chocano, V. Muñoz-Soler, R. García-De-León-González, J.A. Maldonado, et al. A Standardized and Data Quality Assessed Maternal-Child Care Integrated Data Repository for Research and Monitoring of Best Practices: A Pilot Project in Spain. Stud Health Technol Inform. 2017; 235: 539-43. https://doi.org/10.3233/978-1-61499-753-5-539.

[37] C.W. Whitney, B.K. Lind, P.W. Wahl. Quality Assurance and Quality Control in Longitudinal Studies. Epidemiol Rev. 1998; 20(1): 71-80. https://doi.org/10.1093/oxfordjournals.epirev.a017973.

[38] V.T. Guidry, C.L. Gray, A. Lowman, D. Hall, S. Wing. Data quality from a longitudinal study of adolescent health at schools near industrial livestock facilities. Ann Epidemiol. 2015; 25(7): 532–8. https://doi.org/10.1016/j.annepidem.2015.03.005.

[39] R. Li, L. Abela, J. Moore, L.M. Woods, U. Nur, B. Rachet, et al. Control of Data Quality for Population-Based Cancer Survival Analysis. Cancer Epidemiol. 2014; 38(3): 314-20. https://doi.org/10.1016/j.canep.2014.02.013.

[40] M.G. Kahn, T.J. Callahan, J. Barnard, A.E. Bauck, J. Brown, B.N. Davidson, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC). 2016; 4(1): 1244. https://doi.org/10.13063/2327-9214.1244.

[41] M. Labbok, K. Krasovec. Toward Consistency in Breastfeeding Definitions. Stud Fam Plann. 1990; 21(4): 226-30.

[42] M.H. Labbok, A. Starling. Definitions of Breastfeeding: Call for the Development and Use of Consistent Definitions in Research and Peer-Reviewed Literature. Breastfeed Med. 2012; 7(6): 397-402. https://doi.org/10.1089/bfm.2012.9975.

[43] E. Mangrio, K. Persson, A.C. Bramhagen. Sociodemographic, Physical, Mental and Social Factors in the Cessation of Breastfeeding Before 6 Months: A Systematic Review. Scand J Caring Sci. 2018; 32(2): 451-65. https://doi.org/10.1111/scs.12489.

[44] E.R. Moore, N. Bergman, G.C. Anderson, N. Medley. Early Skin-To-Skin Contact for Mothers and Their Healthy Newborn Infants. Cochrane Database Syst Rev. 2016; 11(11). https://doi.org/10.1002/14651858.CD003519.pub4.

[45] R. García-de-León-González, A. Oliver-Roig, M. Hernández-Martínez, B. Mercader-Rodríguez, V. Muñoz-Soler, M.I. Maestre-Martínez, et al. Becoming baby-friendly in Spain: a quality-improvement process. Acta Paediatr. 2011; 100(3): 445-50. https://doi.org/10.1111/j.1651-2227.2010.02061.x.