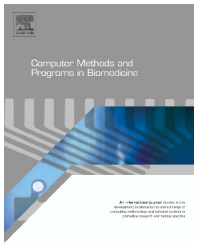
journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Risk groups in children under six months of age using self-organizing maps

A.O.C. Schilithz[a,*], P.L. Kale[b], S.G.N. Gama[c], F.F. Nobre[a]

[a] PEB/UFRJ, Rio de Janeiro, Brazil
[b] IESC/UFRJ, Rio de Janeiro, Brazil
[c] ENSP/FIOCRUZ, Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Received 22 July 2013
Received in revised form
22 January 2014
Accepted 20 February 2014

Keywords:
Infant growth
Artificial neural network
Self-organizing maps
Cesarean delivery
Preterm birth
Exclusive breastfeeding

ABSTRACT

Fetal and infant growth tends to follow irregular patterns and, particularly in developing countries, these patterns are greatly influenced by unfavorable living conditions and interactions with complications during pregnancy. The aim of this study was to identify groups of children with different risk profiles for growth development. The study sample comprised 496 girls and 508 boys under six months of age from 27 pediatric primary health care units in the city of Rio de Janeiro, Brazil. Data were obtained through interviews with the mothers and by reviewing each child's health card. An unsupervised learning, know as a self-organizing map (SOM) and a *K*-means algorithm were used for cluster analysis to identify groups of children. Four groups of infants were identified. The first (139) consisted of infants born exclusively by cesarean delivery, and their mothers were exclusively multiparous; the highest prevalences of prematurity and low birthweight, a high prevalence of exclusive breastfeeding and a low proportion of hospitalization were observed for this group. The second (247 infants) and the third (298 infants) groups had the best and worst perinatal and infant health indicators, respectively. The infants of the fourth group (318) were born heavier, had a low prevalence of exclusive breastfeeding, and had a higher rate of hospitalization. Using a SOM, it was possible to identify children with common features, although no differences between groups were found with respect to the adequacy of postnatal weight. Pregnant women and children with characteristics similar to those of group 3 require early intervention and more attention in public policy.

1. Introduction

Infant growth during the early months of life is affected by several factors, including ethnic differences, intrauterine development [1,2], genetic patterns, the consumption of foods and cultural, environmental, and social factors [1,3–5]. Unfavorable living conditions and their interactions with complications during pregnancy are known to be associated with poor fetal growth and child development in developing countries [6–8]. Additionally, there is evidence that rapid growth during the first six months of life, associated with several socioeconomic factors and maternal characteristics, is an important predictor of childhood obesity [9,10]. The assessment

* *Corresponding author*. Tel.: +55 2127521706.
E-mail address: arthur@peb.ufrj.br (A.O.C. Schilithz).

http://dx.doi.org/10.1016/j.cmpb.2014.02.011
0169-2607/

of growth is an important component for monitoring child development. The detection of changes in the rate of growth and weight gain is important markers of health in early childhood [11,12]. Anthropometric measurements, such as length and weight, are commonly used to evaluate the nutritional status of children under two years of age [13]. Atypical growth patterns may characterize children that are at nutritional risk, and the identification of those children may help prevent inadequate growth.

Several growth charts have been developed for monitoring child growth [14,15]. Recently, the World Health Organization [16] developed a growth chart using information from five continents through a multicenter study with children aged zero to five years, regardless of ethnic origin, socioeconomic status, or type of food. This multicenter study included data from Brazil, representing children in Latin America. Growth curves are primarily used for monitoring and assessing the health and development of children and infants, and they are known to be a cost-effective approach for identifying nutritional deficiencies. The associations of socioeconomic indicators, maternal characteristics and prenatal care with delays in development, including physical growth, are most often mediated by poor nutrition [17,18]. Thus, growth, which is a proxy for nutritional status, is correlated with socioeconomic status, maternal indicators, and prenatal care.

Most studies that investigate infant health and development assess growth failure and related factors by simply characterizing infant growth changes and by comparing the infant's growth with standard growth charts [19,20]. A recent study that evaluated the potential of geographically targeted nutrition programs to reduce the number of infants at nutritional risk concluded that the childhood nutritional status is determined by individual and household factors [21]. The study suggested that new interventions focusing on individual needs are required. Nutritional risk factors tend to cluster in individuals, and predicting which children will be at risk using factors such as socioeconomic indicators, maternal indicators, and prenatal care can help health services develop more efficient preventive measures. One approach for identifying groups of infants with similar risk factors is to use clustering techniques.

Clustering is based on the idea that infants that have similar risk factors should exhibit similar growth patterns. This relationship allows for the possibility that more than one pattern of infant growth will be established and associated with different profiles of socioeconomic and maternal indicators.

Among the available clustering techniques, the Kohonen network [22], which is an unsupervised artificial neural network, also known as a self-organizing map (SOM), is recognized as effective due to its ability to visualize multidimensional data in a low-dimensional space and to extract the essential features of a complex dataset by generating prototype vectors. This technique has been widely used in pattern recognition problems in engineering, and more recently, it has received attention in the field of epidemiology, with applications involving the clustering of patients with infectious diseases, such as dengue fever [23], and patients with chronic diseases [24–26].

The purpose of this study was to identify groups of children with different maternal characteristics, prenatal care factors, and socioeconomic indicators that might be of importance in characterizing growth development. The aim of this report also is to illustrate how a machine learning technique can provide a relatively simple framework with which to visualize and interpret multidimensional data relevant to child growth and nutritional problems.

2. Methods

2.1. *Sample design*

The data were obtained from a cross-sectional study that evaluated 1082 children under six months of age who required pediatric care in 27 primary health care units of the Brazilian Unified Health System (SUS) in the city of Rio de Janeiro from June to September 2007. This study was conducted by researchers from the National School of Public Health, Oswaldo Cruz Foundation.

The sampling plan included two stages. In the first stage, 27 health care units were selected based on both the Euclidean distance between the unit and the administrative center of the municipality of Rio de Janeiro and the cumulative frequency of the monthly average of pediatric consultations for children under six months of age during the first half of 2005. In the second stage, we used a systematic sampling strategy, selecting from the list of children enrolled for routine pediatric evaluation. From each health unit 40 children under six months of age were selected during the period of January to June 2007, resulting in a total of 1080 interviews.

2.2. *Collected variables*

The data were initially obtained from each child's health booklet and supplemented through interviews with the mother. The use of child health booklets was implemented by the Brazilian Ministry of Health in 2005, and these booklets contains information on events related to child health, including obstetric history and neonatal indicators of growth and development; information on breastfeeding; and clinical complications. The booklet is intended to be used for all those born in Brazil and is an appropriate instrument to monitor the health of the child. Once we had all the information recorded in a database, the database was linked to SINASC (Information system of live births in Brazil) to retrieve information on birth weight and gestational age. The SINASC is a national information system of all live births and includes additional information about maternal and infant characteristics, pregnancy, labor, and delivery (www.datasus.gov.br/sinasc). Seven variables related to maternal, pregnancy, and delivery history were collected. These measures include three continuous factors: maternal age, years of maternal schooling, and a socioeconomic attribute based on the presence of household assets using the HAI (household assets index). The HAI evaluates the frequency of the presence of an item in the household, giving more weight for rarer items [27]. Additionally, five categorical variables were considered: marital status (with or without live-in partner), satisfaction with prenatal care (excellent, good, fair/poor/very poor), parity (primiparous or multiparous), type of delivery (vaginal or cesarean), and the