



Research Article

Copyright© Ignacio Díez López

A New Tool of Detection Cases Malnutrition: BIG DATA Techniques Over Children Population

Ignacio Díez López^{1*}, Sandra Maeso Mendez¹, Gaspar Sánchez Merino², Iñaki Zorrilla Martínez³ and Ana Gonzalez-Pinto³

¹Department of Pediatrics, BIOARABA Health Research Institute, OSI Araba, University Hospital, UPV/EHU, Vitoria, Spain.

²Coordinator of the Innovation Platform - IIS BIOARABA, Vitoria, Spain

³Department of Psychiatry, BIOARABA Health Research Institute, OSI Araba, University Hospital, CIBERSAM, UPV/EHU, Vitoria, Spain

*Corresponding author: Ignacio Díez López¹, Department of Pediatrics, BIOARABA Health Research Institute, OSI Araba, University Hospital, UPV/EHU, Vitoria, Spain.

To Cite This Article: Ignacio Díez López*, Sandra Maeso Mendez, Gaspar Sánchez Merino, Iñaki Zorrilla Martínez and Ana Gonzalez-Pinto. A New Tool of Detection Cases Malnutrition: BIG DATA Techniques Over Children Population. *Am J Biomed Sci & Res.* 2025 25(5) AJBSR. MS.ID.003365, DOI: 10.34297/AJBSR.2025.25.003365

Received: 📅 January 30, 2025; **Published:** 📅 February 10, 2025

Summary

Big data tools are currently a major tool for assessing population changes. There is a possible causal relationship between a family's economic capacity and social dystocia's and malnutrition. It is also known that a possible cause of malnutrition is eating disorders, which are more frequent in populations with different levels of education.

Main Objective: To assess the possible relationship between family income and prevalence of malnutrition in a child population.

Material and Methods: Using the Cole-Green LMS algorithm with penalized likelihood, implemented in the Ref Curv 0.4.2 software (2020), which allows managing large amounts of data. The hyperparameters have been selected using the BIC (Bayesian Information Criterion). To calculate population deviations from the reference, the reference was taken as being below 1.5 standard deviations from the average according to age.

Results: The data and comparative graphs between districts of the population studied are represented with respect to the variables analyzed.

Conclusions: Big data technology allows for more efficient population studies, selecting populations most in need of health intervention, optimizing scarce health resources.

Introduction

Body Mass Index (BMI) is a commonly used parameter to assess nutritional status [1-3]. In our country, the tables by Carrascosa, *et al.* [4] are widely adopted as a reference for determining BMI status as an indicator of an individual's nutritional condition in relation to a population considered normal. The assessment and detection of changes in BMI are crucial for monitoring and controlling the child population [3]. While most studies tend to focus on the so-called childhood obesity pandemic [2], another important aspect of nu

tritional reality should not be overlooked: low weight for height, represented as a low BMI [3]. Globally, the main cause of a low BMI is malnutrition associated with insufficient intake due to resource scarcity [2-3] or the disease itself. However, in developed countries, another situation should be considered: malnutrition associated with mental health disorders. Therefore, when a case of low weight for height is detected in a child in our context, the main differential diagnoses include an underlying organic process, constitutional



or functional thinness, socio-familial nutritional issues, or mental disorders that lead to weight loss, such as anorexia, bulimia with a restrictive component, etc. [2,3].

Anorexia nervosa is one of the psychiatric conditions with the highest mortality rates [5]. It is defined by an exaggerated assessment of body volume and shape, leading to a relentless pursuit of thinness. It is characterized by excessive voluntary weight loss through a restrictive diet [6,7]. Anorexia nervosa typically manifests in girls in early to mid-adolescence, with a higher prevalence in white populations and those from above-average socioeconomic backgrounds [7,8]. Following the COVID-19 pandemic, these figures appear to have increased significantly, with disorders such as anxiety and depression estimated to have risen by 25-27% [9]. Thinness, in addition to being a manifestation of an underlying disease or physiological condition, can also reflect situations of risk for developing an eating disorder. Eating disorders have been associated with a high level of education, a family history of eating disorders, vigorexia, family conflicts, or even a rejection of puberty itself [8-10]. However, we should not forget that thinness can also indicate a risk of social or economic exclusion within the family.

UNICEF establishes that child poverty is a stark reality [11], not only in developing countries but even in Europe, specifically in Spain and the Basque Country. In Spain, according to this report [11], the prevalence of risk of social exclusion and child poverty could reach 28%. The electronic medical records of current health systems collect numerous variables in clinical practice, including anthropometric and sociodemographic data. Different statistical techniques, such as machine learning, allow for the exploitation of these data from a large number of cases in an almost semi-automated manner, providing valuable statistical insights. Although there are studies on this topic in various countries and even international series [10], there are no studies, at least in our region or nearby populations, assessing the situation of malnutrition in the child and youth population. Moreover, the use of new BIG DATA techniques for these studies has yet to be described.

Material and Methods

Study Design

This is a population-based cross-sectional study.

Study Population

All minors under 18 years of age who are being followed up in the Basque health system, OSAKIDETZA, and who have weight and height data recorded in the electronic medical record system (OSABIDE GLOBAL) in the Álava area.

Inclusion Criteria

- a) Both sexes
- b) Age between 0 and 18 years
- c) Registered or presenting a registered address

Exclusion Criteria

No data registered in OSABIDE GLOBAL

Epidemiological Data

For this study, reliable and official sources are used for variables such as average income per inhabitant, unemployment rate, and immigration rate by district or neighborhood. These data are available at: https://www.eustat.eus/bankupx/pxweb/es/DB/-/PX_010154_cepv1_ep06b.px/table/tableViewLayout1/.

(Accessed on 08/29/2022).

By including the entire registered pediatric population, it is considered unnecessary to calculate a sample size. Data were collected between 01/01/2022 and 30/03/2022.

Variables:

Main Variables

- a) Weight (kg)
- b) Height (cm)
- c) Gender (Male, Female, Binary)
- d) Age (expressed in years and months)
- e) Date of registration
- f) Place of residence – district/neighborhood code
- g) Unemployment rate, per capita income by district

Data Management Plan

A data protection impact assessment has been prepared. The data lifecycle will involve the IT service of OSI Araba, the principal investigator of the project, and collaborating researchers, including professionals from the Basque Center for Applied Mathematics (BCAM), who are part of the research team. There is a collaboration agreement between BCAM and the Bioaraba Health Research Institute.

Statistical Analysis

The method based on Dirichlet Processes (Dirichlet Process, DP) is followed. In this project, we will adopt this approach to build Gaussian Mixture Models (GM). Furthermore, Dirichlet Process Gaussian Mixture Models (DPGMM) will be used. We will also analyze a set of populations using Gaussian mixture models based on hierarchical Dirichlet processes (Hierarchical Dirichlet Process Gaussian Mixture Model, HDPGMM) [12]. Clustering will be performed to inform us about the somatometric similarities and differences of the population based on somatometric variables and the district in which they reside [13]. This will incorporate recent methodological innovations in databases similar to ours, as already described in previous studies [14-16]. BMI will be calculated as weight/height^2 (kg/m^2). These results will be compared with the means and Standard Deviation Scores (SDS) from studies pub-

lished to date, which serve as references for our population [4]. Low weight will be defined as less than 1.5 SDS relative to the reference normality for age and sex [4].

Results

Data were obtained from a total of 67,270 minors. The total number of variables studied (some presented in this work and others reserved) amounts to 1,749,020 variables. The results, categorized by sex, age, BMI, and other variables, are presented in various tables. According to data from the National Institute of Statistics and EUSTAT, the population of our region, Álava province, in the Basque Country, Spain, is 338,765 people. The political territory is divided into several districts. The average disposable income for the resident population in 2021, calculated as total income minus income tax and social security contributions paid by workers, is €19,366. There are significant differences across age groups, gender, and districts. The income of minors depends on the average family income, which in the Basque Country was €47,005 in 2021. Total family income is calculated by aggregating the personal income of all family members, including minors. On average, family income in the Basque Country is about twice the average personal

income. There are notable differences between districts (Source: EUSTAT), with towns in Álava having the lowest average income across the entire region. The unemployment rate in the Basque Country stands at 7.5%, which is well below the national average in Spain. While 6 out of 10 households have all members employed, more than 1 out of 10 households experience complete unemployment. Significant differences exist between districts (Source: EUSTAT), with some towns in Álava and Vizcaya reporting the highest unemployment rates. The distribution of the immigrant population also varies within our territory. The average immigration rate in the Basque Country is 13%, but Vitoria, the capital, has one of the highest rates of immigrants relative to the general population, at 15%. Some districts in the capital exceed 18%, and certain localities even surpass 20%.

After analyzing the variables BMI, per capita income, and the immigration rate in each district separately, an assessment was made of the districts with a higher prevalence of low BMI cases in relation to these other variables. The numerical results are shown in Figures 1&2. Using a cluster study statistical method, districts with fewer than 20 cases with available data during the study period have been excluded.

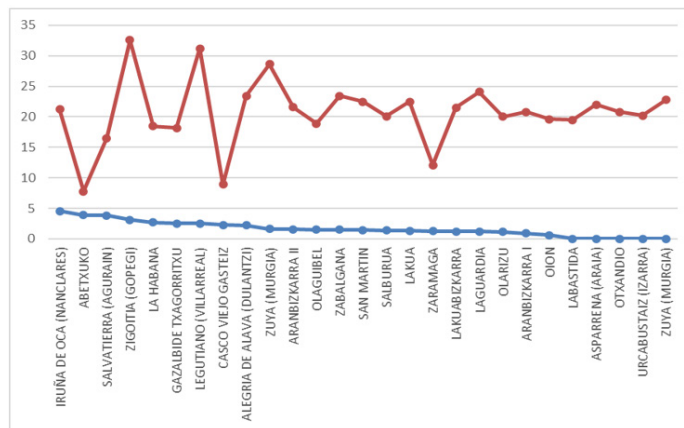


Figure 1: Men < 18 years. Representation of the variable BMI (% of the population < 1.5 SDS) (blue) vs per capita income (thousands € per inhabitant/year) (dark brown line). The orange line represents the average per capita income of the population, which is €19,366.

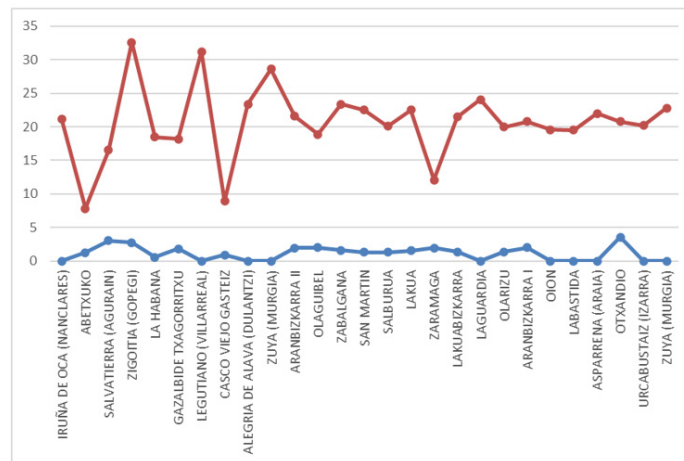


Figure 2: Women < 18 years. Representation of the variable BMI (% of the population < 1.5 SDS) (pink) vs per capita income (thousands € per inhabitant/year) (dark brown line). The orange line represents the average per capita income of the population, which is €19,366.

The MEN graph shows that there is a population with a low BMI in both higher-income districts and in less-privileged ones, and in apparently equal proportion. However, in most districts with a stable average income, the rate of cases with a BMI less than 1.5 SDS is zero or practically zero. In the case of WOMEN, we found a heterogeneous distribution of districts with significant rates of cases with a BMI below 1.5 SDS, but there is apparently a greater number of cases in districts with higher income.

Regarding the assessment of the number of migrants living in a specific district of those studied, in relation to the study of the number of cases with a BMI below 1.5 SDS, it is evident that, in the case of men, a large number of cases appear (reaching more than 3% of the total population) both in districts with a low or almost non-existent immigration rate and, on the contrary, in those with a higher rate (Figure 3).

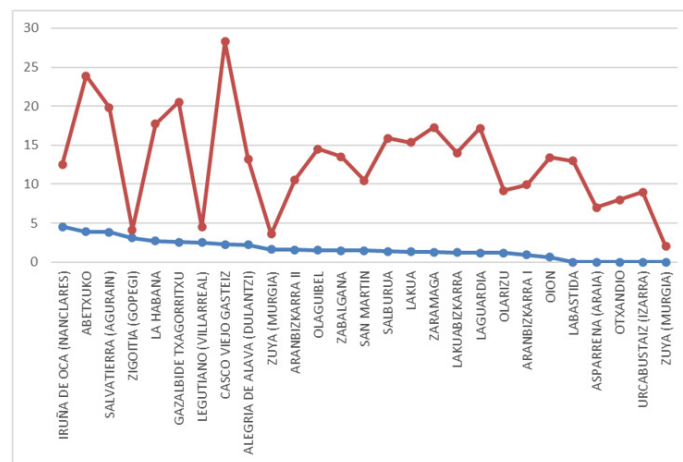


Figure 3: Men < 18 years. Representation of the BMI variable (% of the population > 1.5 SDS) (blue) vs. immigrant population rate living in the district (% of the total) (red line). The orange line represents the average immigrant rate in the territory, 13%.

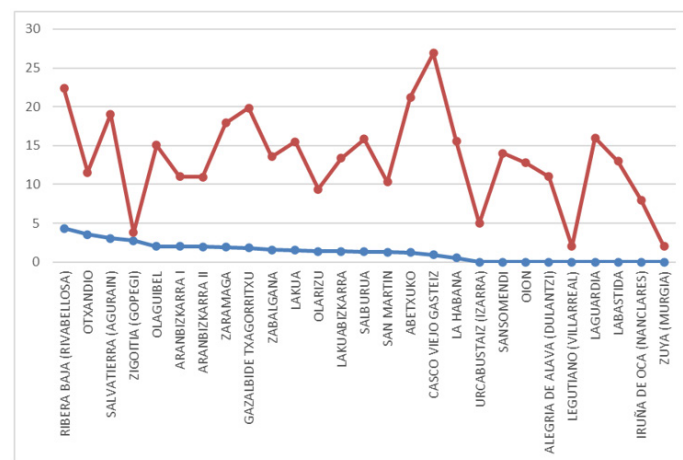


Figure 4: Women < 18 years. Representation of the variable BMI (% of the population > 1.5 SDS) (blue columns) vs. rate of immigrant population living in the district (% of the total) (red line). Orange line represents the average rate of immigrants in the territory, 13%.

In the case of women, a higher number of cases with low BMI is observed in populations with rates of migrants higher than the average for the territory (13%) and in areas with hardly any immigrant population, although there seems to be a predominance of territories with little immigration and a low prevalence of cases with low BMI, except in one of them. On the contrary, high migration rates and the appearance of cases of low BMI are more frequent (Figure 4).

Discussion

The possibilities of conducting epidemiological research using big data are multiplying, offering new strategies for health and healthcare interventions. Machine learning, which has proven effective in other fields [14-17] for interpreting large amounts of real-world data, is emerging as one of the key tools for this purpose. Somatometry in children, as an indicator of health status, and the

issue of eating disorders are clear examples of areas where these techniques can be applied for research. The secular trend of accelerated weight gain [18,19], the rise in eating disorders among youth, and a potential added effect from the COVID-19 pandemic have been discussed by various authors [20-22]. Other proposed causes include the relationship between eating disorders and the socioeconomic status of the family, per capita income, and even the stress caused by unemployment within the family [8,9,11]. According to UNICEF, a child is considered at risk of poverty when their household's disposable income falls below a threshold based on 60% of the median income of all households in the country, also considering the household composition [11].

In our study, we considered these factors and selected variables related to low BMI that were accessible to our research team, such as the average income level of the district and the immigration rate. The latter is seen as an element connected to larger family sizes, lower family income, and higher unemployment rates. Health resources are scarce, making it challenging to decide which population subsectors to intervene with, whether for campaigns or active searches for children at risk of malnutrition. BIG DATA offers a quick and cost-effective way to get an accurate picture of the population's situation, enabling us to determine where, how, and why to allocate these scarce resources [15-17].

Our study reveals that there are indeed risks of malnutrition in our population, with certain towns, neighborhoods, or districts showing an impact rate of nearly 4% of the entire child population in some areas. This finding calls for reflection on both the methodology used and the social nature of our population. However, our work also suggests that factors related to the environment in which a child lives may influence their nutritional status, particularly regarding their recorded weight [8,9,19-20]. The relationship between income level, food quality and quantity, the ability to participate in extracurricular, educational, and sports activities [22], and the general environment where a child grows up seems to play a role in determining whether or not they suffer from malnutrition.

A low BMI can be the result of malnutrition due to a real lack of resources or may be secondary to an underlying organic condition or an eating disorder. On the other hand, we also observed higher rates of low BMI in districts with lower income and higher immigration rates. This may reflect the vulnerability of these families, who likely have fewer food resources. Knowing how and when to intervene in cases of malnutrition risk [23] helps not only to reduce associated mortality but also to minimize resource use, such as hospitalizations and emergency visits. The use of big data is and will continue to be an essential tool in public health decision-making [24] in the coming years. It also presents an opportunity to improve routine clinical practice in the most practical ways.

Ethical Aspects

The study has been prepared in compliance with the principles established in the Declaration of Helsinki (1964) latest version Fortaleza, Brazil 2013, in the Council of Europe Convention on Human Rights and Biomedicine (1997), and in the regulations on

biomedical research, protection of personal data. Law 14/2007 on Biomedical Research. Study approved by the CEIC on 03/24/2023 with CODE File 2022-058

Economic Report

The study will be conducted without funding. The tasks described in the project are undertaken by the principal investigator and his collaborators.

Acknowledgements

This original study has been supported thanks to the work of the Collaborative Group from Basque Center of Applied Mathematics (BCAM). Bilbao, Bizkaia Basque Country, Spain.

- 1) Jose A. Lozano Basque Center for Applied Mathematics BCAM
- 2) Ioar Casado Tellechea Basque Center for Applied Mathematics BCAM
- 3) Aritz Pérez Postdoctoral Fellow BCAM - Basque Center for Applied Mathematics.

References

1. WHO referenced data (2024) Growth reference data from 5-19 years. On <https://www.who.int/tools/growth-reference-data-for-5to19-years>. Date Dec 2024.
2. Mendon P, Witsch M, Becker M, Adamski A, Vaillant M, et al. (2024) Facilitating comprehensive child health monitoring within REDCap - an open-source code for real-time Z-score assessments. *BMC Med Res Methodol* 24(1): 298.
3. Heude B, Pauline Scherdel, Andreas Werner, Morgane Le Guern, Nathalie Gelbert, et al. (2019) A big-data approach to producing descriptive anthropometric references: a feasibility and validation study of pediatric growth charts. *Lancet Digital Health* 1(8): e413-e423.
4. Carrascosa Lezcano A, J M Fernández García, C Fernández Ramos, A Ferrández Longás, J P López-Siguero, et al. (2008) Spanish cross-sectional growth study 2008. Part II: height, weight and body mass index values from birth to adult height. *An Pediatr (Barc)* 68(6): 552- 69.
5. Morales Lopez Maria Jose (2019) Anorexia nervosa in the pediatric population. *Med leg Costa Rica [Internet]* 36 (2): 46-55.
6. van Eeden AE, van Hoeken D, Hoek HW (2021) Incidence, prevalence and mortality of anorexia nervosa and bulimia nervosa. *Curr Opin Psychiatry* 34(6): 515-524.
7. Silén Y, Keski-Rahkonen A (2022) Worldwide prevalence of DSM-5 eating disorders among young people. *Curr Opin Psychiatry* 35(6): 362-371.
8. Schlissel AC, Richmond TK, Eliasziw M, Leonberg K, Skeer MR, et al. (2023) Anorexia nervosa and the COVID-19 pandemic among young people: a scoping review. *J Eat Disord* 11(1): 122.
9. Walsh O, McNicholas F (2020) Assessment and management of anorexia nervosa during COVID-19. *Ir J Psychol Med.* 37(3): 187-191.
10. Silliman Cohen RI, Bosk EA (2020) Vulnerable Youth and the COVID-19 Pandemic. *Pediatrics* 146(1): e20201306.
11. Child poverty report in Spain (2024) UNICEF Report 2023.
12. Rasmussen C (1999) The infinite Gaussian mixture model. *Advances in neural information processing systems* 12.
13. Teh Y W, Jordan M I (2010) Hierarchical Bayesian nonparametric models

- with applications. *Bayesian nonparametrics* 1: 158-207.
14. Van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Machine Learning Res* 9(11).
 15. Kruskal J B (1964) Non metric multidimensional scaling: a numerical method. *Psychometrika* 29(2): 115-129.
 16. Gilholm P, Mengersen K, Thompson H (2020) Identifying latent subgroups of children with developmental delay using Bayesian sequential updating and Dirichlet process mixture modeling. *PLoS one* 15(6): e0233542.
 17. Diana A, Matechou E, Griffin J, Johnston A (2020) A hierarchical dependent Dirichlet process prior for modeling bird migration patterns in the UK. *Annals Applied Statistics* 14(1): 473-493.
 18. Ahrens W, Moreno L A, Pigeot I (2011) Childhood obesity: Prevalence worldwide. In: Moreno LA, editor. *Epidemiology of Obesity in Children and Adolescents*. New York: Springer: 219-235.
 19. Umekar S, Joshi A (2024) Obesity and Preventive Intervention Among Children: A Narrative Review. *Cureus* 16(2): e54520.
 20. Boltri M, Brusa F, Apicella E, Mendolicchio L (2024) Short- and long-term effects of Covid-19 pandemic on health care system for individuals with eating disorders. *Front Psychiatry* 15: 1360529.
 21. Dalle Grave R, Chimini M, Cattaneo G, Dalle Grave A, Ferretti L, Parolini S, et al. (2024) Intensive Cognitive Behavioral Therapy for Adolescents with Anorexia Nervosa Outcomes before, during and after the COVID-19 Crisis. *Nutrients* 16(10): 1411.
 22. Winston A P, Taylor M J, Himmerich H, Ibrahim M AA, Okereke U, Wilson R (2023) Medical morbidity and risk of general hospital admission associated with concurrent anorexia nervosa and COVID-19: An observational study. *Int J Eat Disord* 56(1): 282-287.
 23. Schlissel AC, Richmond TK, Eliasziw M, Leonberg K, Skeer MR (2023) Anorexia nervosa and the COVID-19 pandemic among young people: a scoping review. *J Eat Disord* 11(1): 122.
 24. Wesson P, Hswen Y, Valdes G, Stojanovski K, Handley MA (2022) Risks and Opportunities to Ensure Equity in the Application of Big Data Research in Public Health. *Annu Rev Public Health* 43: 59-78.