

Big Data Applications in Population Epidemiology: Social and Economic Variables and their Influence on the Weight of Children's Population

Ignacio Diez Lopez^{1-2*}, Sandra Maeso Mendez² and Gaspar Sánchez Merino³

¹Basque Country University UPV-EHU, Pediatric Department, Spain

²Child and adolescent endocrinology Unit, Pediatric Department, OSI Araba. Osakidetza. Vitoria Alava Spain

³Coordinator of the Innovation Platform - IIS BIOARABA, Spain

*Corresponding author: Ignacio Diez Lopez, Basque Country University, Pediatric Department, Vitoria, Spain

ARTICLE INFO

Received: 📅 October 21, 2024

Published: 📅 October 29, 2024

Citation: Ignacio Diez Lopez, Sandra Maeso Mendez and Gaspar Sánchez Merino. Big Data Applications in Population Epidemiology: Social and Economic Variables and their Influence on the Weight of Children's Population. Biomed J Sci & Tech Res 59(2)-2024. BJSTR. MS.ID.009272.

SUMMARY

Big data tools are currently a major tool for assessing population changes. Could be a causal relationship between low economic levels and a higher prevalence of conditions associated with obesity? Big data could give us answers about.

Main Objective: Study effect of the unemployment rate, average income and immigration rate as a possible effect of increasing the prevalence of malnutrition associated with childhood obesity.

Material and Methods: Data collected from computerized clinical history episodes, studying the variables of sex, age, weight, height, of a pediatric population (compare 2020 vs 2022), comparing it with the average income of their residential district, unemployment rate and immigration rate. Use of big data methods for the study of variables. Using the Cole-Green LMS algorithm with penalized likelihood, implemented in the RefCurv 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 above 1.5 standard deviations from the average according to age.

Results: 66,975 computerised episodes of children under 16 years of age and a total of 1,205,000 variables studied. The data and comparative graphs between districts of the population studied are represented with respect to the variables analysed. There are significant differences, with an increase in the rate of overweight in those areas with lower economic income and higher unemployment and immigration rates. Big data technology allows for more efficient population studies, selecting populations most in need of health intervention, optimizing scarce health resources.

Note: CEIC OSI ARABA Approval Expte 2022-058.

Abbreviations: BCAM: Basque Center for Applied Mathematics; DP: Dirichlet Process; BMIBIC: Bayesian Information Criterion; SES: Socioeconomic Status

Introduction

Health programs to carry out checks on children throughout childhood [1,2] to assess their growth and development status [3]. The body mass index (BMI) is a common parameter to calculate and assess the degree of overweight [3], whether or not it is a criterion of health. Childhood overweight has been seen to increase in the last de-

cade in all regions of the world [4,5]. In Spain, the tables of Carrascosa et al [6] are most applied about. This group of authors (Diez Lopez, et al. [7]) published how the use of these methods would allow population studies to be carried out with greater statistical power than classic longitudinal studies. There are various studies that correlate obesity, especially childhood obesity, with the most disadvantaged sectors of society [8]

The causes of the higher prevalence of obesity in the most disadvantaged strata have been postulated as diverse, from decreased physical activity, overeating, less education, use of lower-cost and higher-calorie foods [8,9]. Although after COVID-19 pandemic a high prevalence of obesity could be appeared [7]. For another hand, is know than COVID-19 could have an import role for increasing prevalence of obesity among children [10,11]. Home confinement, lack of physical activity, increased screen time were key factors [11], but how much did the family's economic situation condition it? what role have social and economic status of family in all this story? We try ask about in this original.

Goals

Main Objective

To describe the situation of the prevalence of overweight in the pediatric population of our area, Álava, Basque Country, Spain, using a new big data approach in relation to the place of residence and the unemployment rate, average income per person and rate of immigrant population compare 2020 vs 2022 To compare whether there are differences in the BMI variable (kg/m²) by comparing paired means between districts and neighborhoods.

Material and Methods

Design

This is a population-based cross-sectional study.

Study Population: All minors under 18 years of age being followed up in the Basque health system (OSAKIDETZA) who present weight and height records in the electronic clinical history tool of OSABIDE GLOBAL in the Alava area. Inclusion criteria: Ages between 0 and 18 years.

Exclusion Criteria: Not datas registered.

Epidemiological Data

Source is used on the variables average income per inhabitant, unemployment rate and immigration rate by district/neighborhood.

Available at EUSTAT

https://www.eustat.eus/bankupx/pxweb/es/DB/-/PX_010154_cepv1_ep06b.px/table/tableViewLayout1/.

(Accessed 08/29/2022).

Data is recorded between 01/01/2020 and 30/03/2020, and the same period 2 years later, between both COVID-19 pandemic was occurred.

Variables

- Weight (Kgrs)
- Size (cm)

- Gender (Male, Female, Binary)
- Age (expressed in years and months)
- Date of registration
- Place of residence – district/neighborhood code
- Unemployment rate, per capita income and immigration rate by district

Data Management Plan

A data protection impact assessment has been prepared. The data life cycle will involve the IT service of OSI Araba, the project's principal investigator 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 Institut. The method already described by Diez- Lopez et al [7] is followed, based on the Dirichlet processes (Dirichlet process, DP) [10]. In this project we will adopt this approach that allows to build Gaussian mixture models (GM) [7-13]. In addition, Gaussian averaging models based on Dirichlet processes (Dirichlet) are used. process Gaussian mixture models, DPGMM). Gaussian averaging models based on hierarchical Dirichlet processes (Hierarchical Dirichlet). Dirichlet process Gaussian mixture model, HDPGMM) [14]. Specifically, by grouping the data according to the different variables, clusters will be obtained that will inform us about the somatometric similarities and differences of the population based on the somatometric variables and the district in which they live [15], incorporating recent methodological innovations on databases similar to ours already described [7,16-18]. The BMI is calculated as weight/height² (kg /m²). These data are compared with the means and SDS of the studies published to date and reference of our population [6]. Overweight is defined as +1.5 SDS with respect to the normal reference for age and sex.

Results

Data has been obtained from a total of 67,270 cases. The sum of all variables studied (some presented in this work and others reserved) amounts to 1,749,020 variables. We present in various tables the results obtained by sex, age and BMI and other variables. Data from the National Institute of Statistics and EUSTAT indicate that 338,765 people live in our territory. Of these inhabitants, 166,437 are men and 172,328 are women. In addition, 52,241 of these people were born abroad. In the last year, our territory has gained 3,199 foreign-born inhabitants: a number higher than the total population growth. Although the rate of immigrants in the Basque Country is on average 13% of the total population, there are significant differences between districts (Source EUSTAT), with the towns of Álava having the highest average percentage of immigrants of the entire population of the region. Not significal changes from 2020 vs 2022. About average incomes at the Basque Country was calculated as the total income minus income tax and social security contributions paid by

workers, of the resident population in 2021 is 19,366 euros. There are significant differences between age groups, sex and districts. The income of minors depends on the average family income. The average family income of the Basque Country is 47,005 euros in 2021. Total family income is obtained by aggregating the personal income of family members, including minors.

The average income for all families in the Basque Country is around double the average personal income. There are significant differences between districts (Source EUSTAT), with the towns of Álava having the lowest average income of the entire population of the region. The unemployment rate in the Basque Country is 7.5%, well below the average for the country, Spain. Although 6 out of 10 house-

holds have all their people employed, in more than 1 out of 10 they are all unemployed. There are significant differences between districts (Source EUSTAT), with some towns in Álava and Vizcaya having the highest unemployment rate. Based on the above, it is clear that the territory under study is the one with the lowest per capita income, the highest prevalence of migrants and some of the areas most affected by structural unemployment. At the same time, within the territory we have studied, Álava, there are significant differences between different districts. If we focus on the BMI data, establishing as a significant cut-off point those that exceed more than 10% of the entire population affected by obesity, differences are evident between districts within the territory of Alava (Table 1) (Charts 1-6).

Table 1: Numerical representation of data by districts for the variable BMI (Kgrs) according to sex. Reference normal population (P50 Carrasco study). The percentages indicate the amount of population whose BMI is 1.5 standard deviations above the average for their age. Age 2020 and 2022. Differences between both years. In red is represented the district with more than 10% of the population > 1.5 SDS.

Increase Number of Subjects - Women - % of Population with BMI>1.5 SDS by Neighborhood or District - Different Period				
District	2020 (%)	2022 (%)	Difference x100	No. of samples
Saint Martin	7.24	7.58	0.34	541
OLAGUIBEL	7.10	7.38	0.24	542
ZABALGAÑA	7.98	8.22	0.24	1386
OLARIZU	6.00	6.32	0.32	364
SALBURUA	6.27	6.37	0.10	832
ARABIZKARRA I	6.53	6.53	0	199
OLD TOWN GASTEIZ	9.80	10.20	0.4	330
LOWER RIVER BAY (RIVABELLOSA)	4.40	4.35	-0.05	23
SANSOMENDI	4.98	5.00	0.02	36
ABETXUKO	16.30	17.28	0.98	81
GAZALBIDE TXAGORRITXU	11.00	11.25	0.25	329
Urcabustaiz (Izarra)	2.04	2.22	0.18	45
Zuya (Murgia)	6.32	6.25	0.07	48
LAKUABIZKARRA	7.99	8.06	0.07	1240
ARABIZKARRA II	9.58	9.72	0.24	257
JOY OF ALAVA (DULANTZI)	12.01	12.50	0.49	80
HABANA	6.58	6.49	-0.09	185
IRUÑA DE OCA (NANCLARES)	4.02	3.95	-0.07	76
LAKUA	5.77	5.81	0.04	396
ZIGOITIA (GOPEGI)	0.00	0.00	0	36
SALVATIERRA (AGURAIN)	7.22	7.36	0.14	163
OTXANDIO	10.02	10.71	0.69	28
ZARAMAGA	14.42	15.47	1.05	362
LABASTIDA	10.85	11.47	0.62	61
OION	6.68	6.55	-0.19	168
LEGUTIANO (VILLARREAL)	12.20	12.50	0.30	32
LAGUARDIA	2.70	2.97	0.27	101
Saint Martin	5.02	5.13	0.11	546
OLAGUIBEL	9.05	9.21	0.16	532

ZABALGAÑA	5.80	5.77	-0.03	1335
OLARIZU	9.85	9.95	0.10	432
SALBURUA	7.00	7.03	0.03	939
ARABIZKARRA I	6.54	6.88	0.34	218
OLD TOWN GASTEIZ	9.52	10.09	0.57	352
LOWER RIVER BAY (RIVABELLOSA)	17.98	18,18	0.20	77
SANSOMENDI	13.50	13.97	0.47	315
ABETXUKO	6.5	7.50	1	40
GAZALBIDE TXAGORRITXU	4.64	4.84	0.2	62
Urcabustaiz (Izarra)	8.57	8.68	0.11	1221
Zuya (Murgia)	8.07	8.20	0.13	256
LAKUABIZKARRA	6.54	6.67	0.13	90
ARABIZKARRA II	9.54	9.90	0.36	222
JOY OF ALAVA (DULANTZI)	10.27	10.61	0.35	66
HABANA	6.51	6.67	0.14	375
IRUÑA DE OCA (NANCLARES)	0.05	0.00	0	32
LAKUA	12.32	12.64	0.32	182
ZIGOITIA (GOPEGI)	7.28	7.41	0.13	27
SALVATIERRA (AGURAIN)	16.25	17.63	1.38	397
OTXANDIO	8.75	9.62	0.67	52
ZARAMAGA	9.54	9.94	0.40	161
LABASTIDA	2.25	2.50	0.25	40
OION	5.68	5.95	0.27	84
LEGUTIANO (VILLARREAL)	5.02	5.13	0.11	546
LAGUARDIA	9.12	9.21	0.09	532

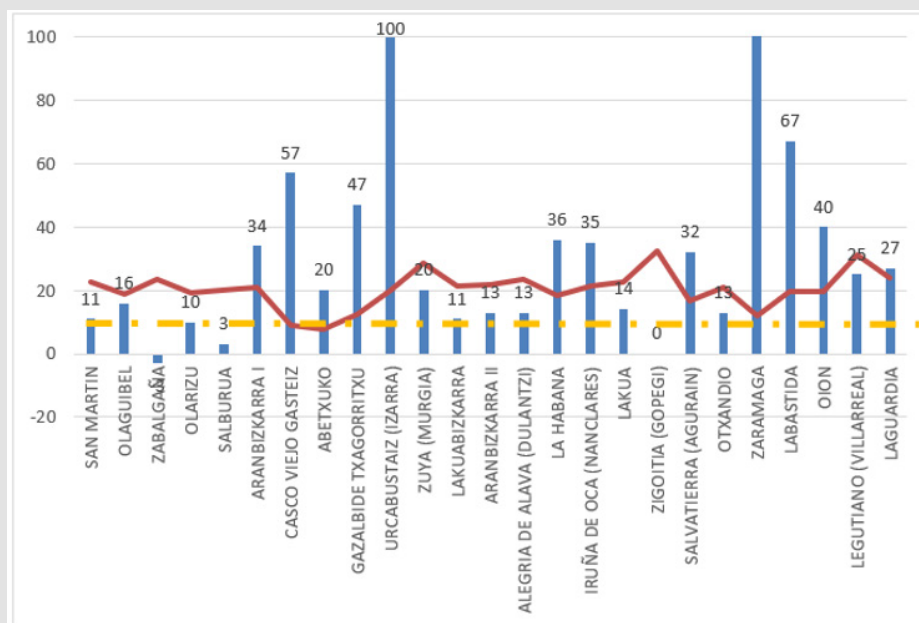


Chart 1: Representation of the increase of MEN (x100) with a BMI % > 1.5 SDS) (blue columns) 2020 /2022 vs per capita income (thousands € per inhabitant/year) (red line). The orange line represents the average per capita income of the population, which is €19,366.

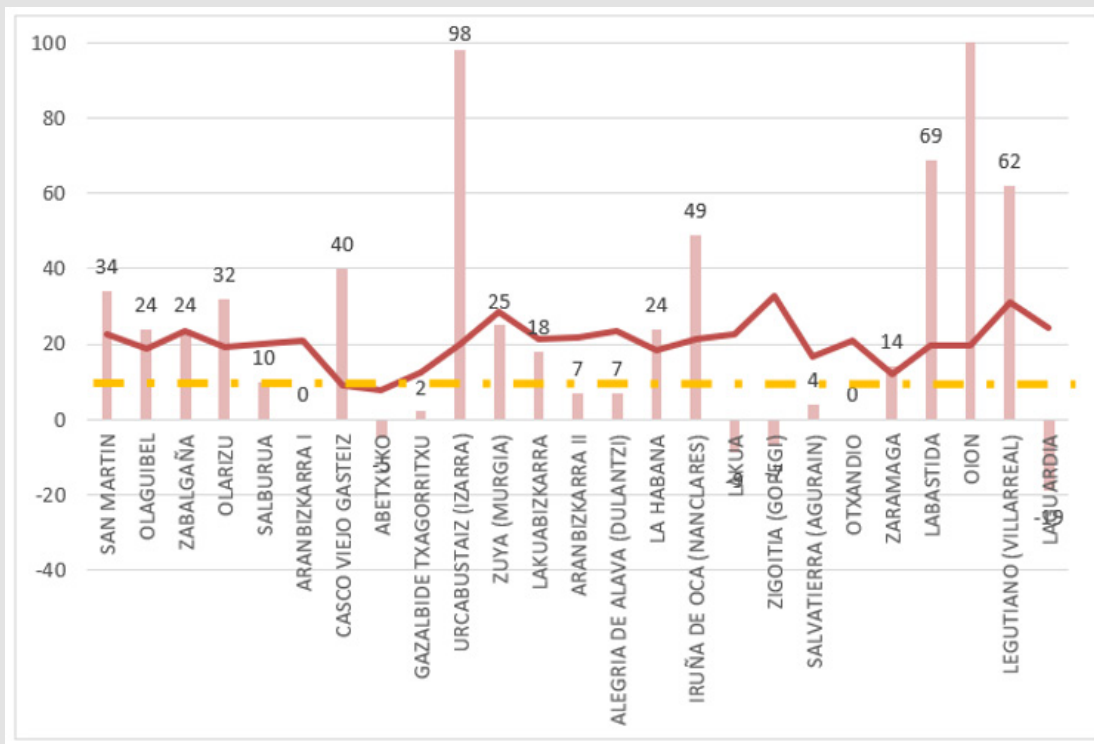


Chart 2: Representation of the increase of WOMAN (x100) with a BMI % > 1.5 SDS (pig columns) 2020 /2022 vs per capita income (thousands € per inhabitant/year) (red line). The orange line represents the average per capita income of the population, which is €19,366.

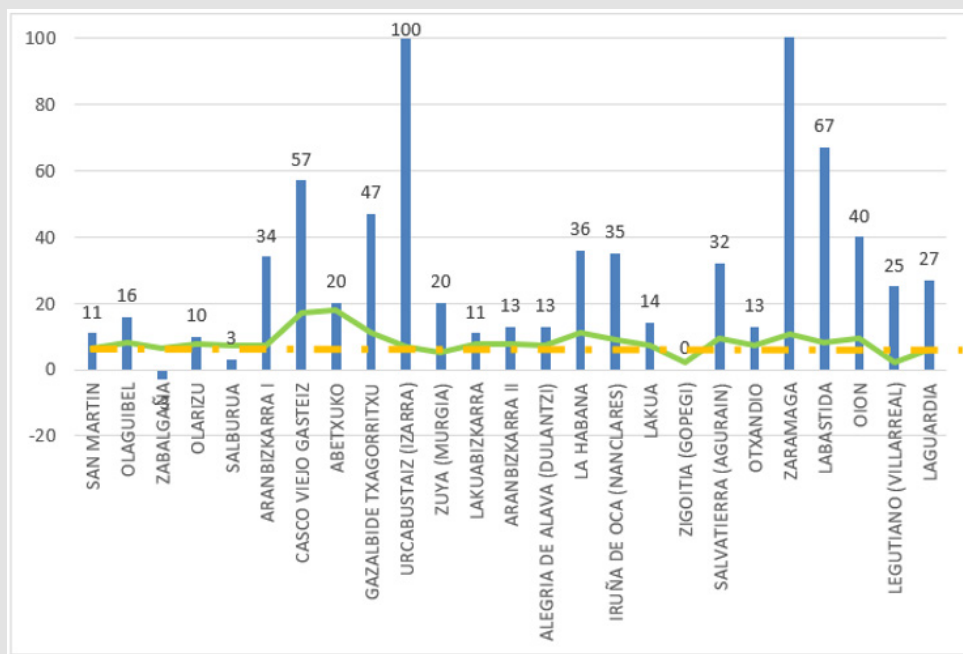


Chart 3: Representation of the increase of MEN (x100) with a BMI % > 1.5 SDS (blue columns) 2020 /2022 vs unemployment rate (in % of active population) (green line). The orange line represents the average unemployment rate in the population, which is 7.5%.

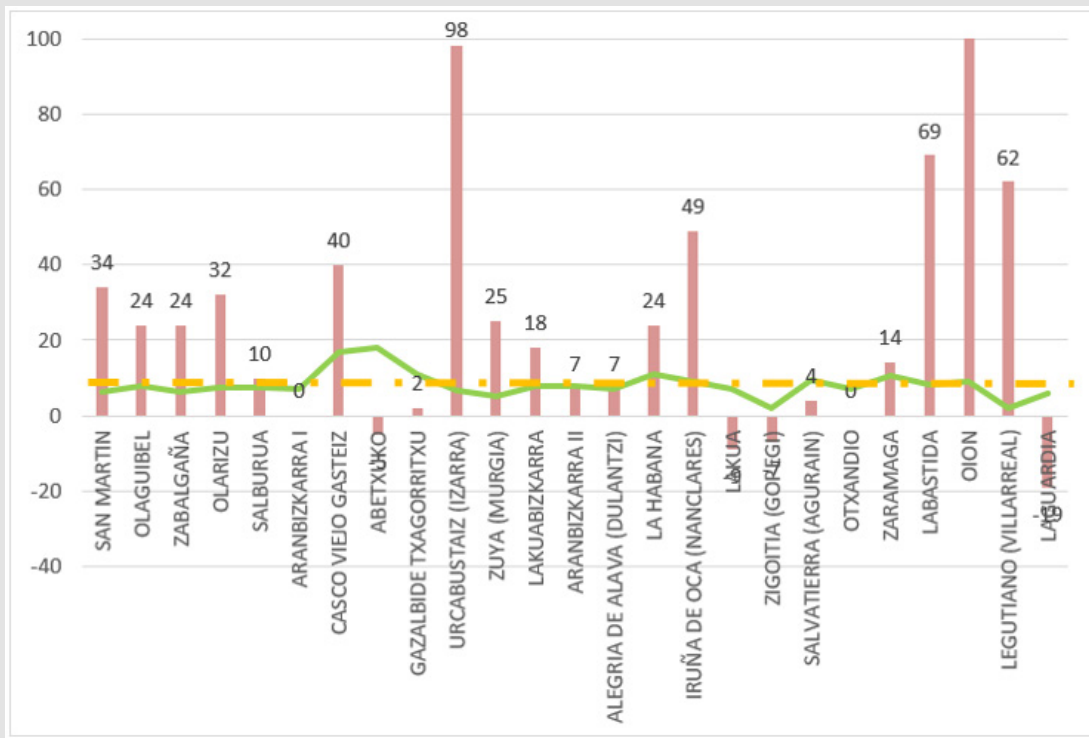


Chart 4: Representation of the increase of WOMAN (x100) with a BMI % > 1.5 SDS (pig columns) 2020 /2022 vs unemployment rate (in % of active population) (green line). The orange line represents the average unemployment rate in the population, which is 7.5%.

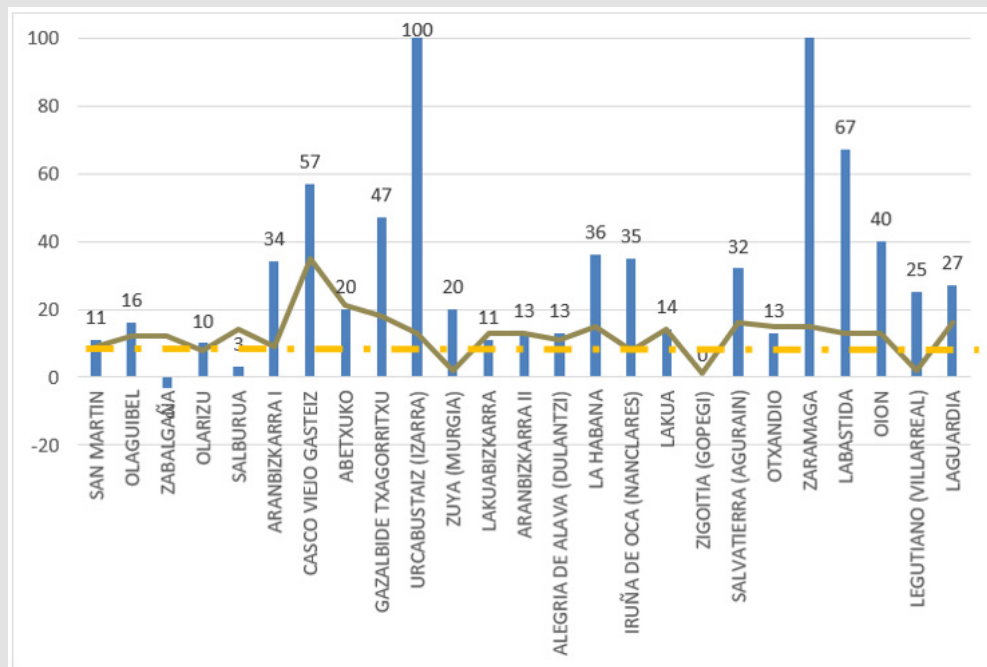


Chart 5: Representation of the increase of MEN (x100) with a BMI % > 1.5 SDS (blue columns) 2020 /2022 vs % of migrated people in relation to the total (brown line). The orange line represents the average rate of the immigrant population, which is 13%.

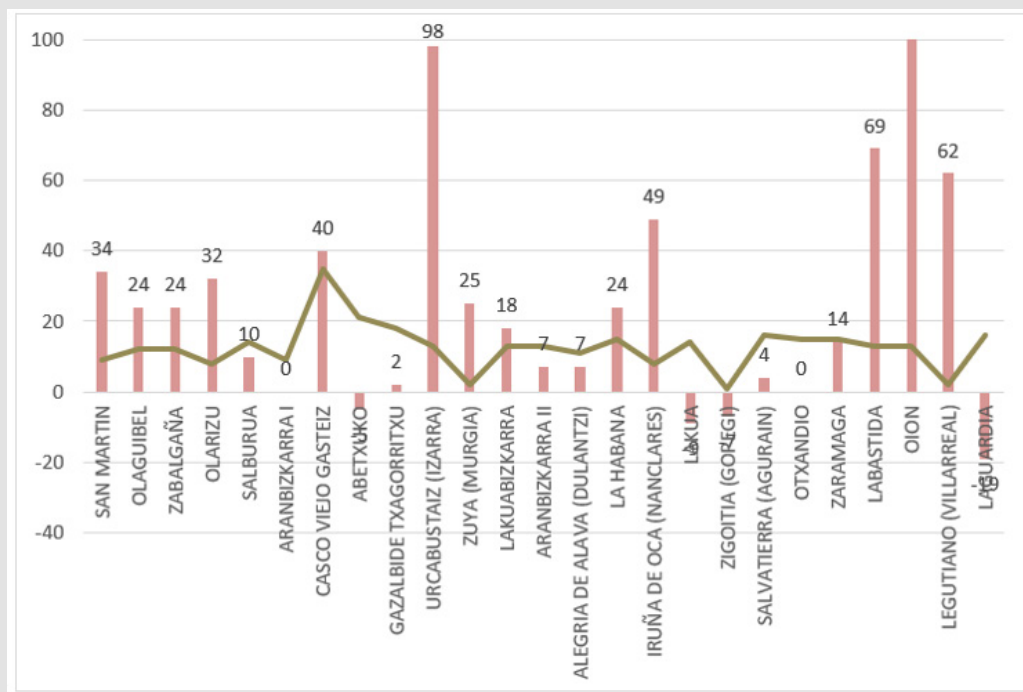


Chart 6: Representation of the increase of WOMAN (x100) with a BMI % > 1.5 SDS (blue columns) 2020 /2022 vs % of migrated people in relation to the total (brown line). The orange line represents the average rate of the immigrant population, which is 13%.

After studying the variables BMI, unemployment rate, per capita income and immigrant population separately, an assessment is made of those districts in the territory that have a higher prevalence of childhood obesity (>10%) in relation to these other variables. We also study where was highest improving ratio It is observed that there is an indirect relationship between a higher prevalence of childhood obesity expressed as >1.5 SDS according to district and the per capita income recorded in said district expressed in thousands of euros. Most of cases increase their rate of overweight from 2020 to 2022. A direct relation sheep is observant about per capita income. At woman is possible to show a decrease of overweight on most rich district. Regarding the unemployment rate by district, the opposite is evident, that with per capita income there is a direct relationship between the obesity rate and the unemployment rate. This suggests that both variables are therefore in a possible cause-effect relationship in their effect on the state of child nutrition. So, for children, we find a relationship between incomes family and unemployment rate. Both of them they are relationated too with changes of obesity in the las period, after ant before COVID19. As charts show us there are a direct relationship between rate of migrated people in the different part of our region, economic level and increase of overweight in the last years.

Discussion

Modern clinical records, combined with new epidemiological research techniques and the use of big data, allow us to propose new

health strategies. Different statistical techniques, such as machine learning, have been shown in other fields [7,14-16] to be effective in interpreting a large amount of data generated in real life and making decisions about it. Somatometry in children in general and the problem of obesity in particular are proposed as one of the fields of research. The secular acceleration of weight relative to height [4,5] is observed, and a possible additional effect to that observed is that generated by the pandemic and its confinement. Other postulated causes are the relationship with the socioeconomic level of the family [8,9]. Health resources and BIG DATA are proposed as a faster and cheaper way to obtain a real picture of the population situation, and therefore to determine where, how and why to invest these scarce resources [15-17]. We point out that our study shows that childhood obesity is present in our population, with towns, neighborhoods or districts with an affectation rate of more than 10% of the entire child population. Obesity occurs at key stages of development, such as pre-pubertal or late pubertal age, which can contribute to maintaining the problem of overweight [4,7,18]. Children with immigrant backgrounds generally have higher rates of overweight and obesity compared to native-born children:

- In Spain, children of immigrant origin had significantly higher prevalence of overweight/obesity than native children (40.5% vs 29.5% for boys, 44.8% vs 30.3% for girls) economic Factors. Surprisingly, family economic status does not fully explain the higher obesity rates among immigrant children: Children of im-

migrants tend to have higher obesity rates across all socioeconomic status (SES) groups, including high SES families. Among children from the most economically successful families, those of newly arrived immigrants (1.0 generation) are significantly more likely to be overweight or obese. Other Contributing Factors as acculturated immigrants tend to have the highest obesity rates. Some studies show mixed results regarding obesogenic behaviors:

- 1) In New Zealand, children of foreign-born mothers had lower odds of consuming fast food and soft drinks, but higher odds of inadequate sleep duration
- 2) In Spain, immigrant children were at higher risk of consuming sugary soft drinks, exercising less, and using screens more

These findings suggest that addressing childhood obesity among immigrant populations requires a nuanced approach that goes beyond economic factors. Public health interventions should consider:

- Language barriers and cultural factors
- Acculturation processes
- Specific risk factors and behaviors within immigrant communities

By targeting these areas, policymakers and health professionals can work towards reducing health inequalities and improving outcomes for children in immigrant families. But in addition, our work shows how there are situations not foreign to the environment where a child lives that seem to condition his or her situation regarding the recorded weight [8,9,19,20]. The relationship between income level, quality (and quantity) in food purchases, the possibility of attending extracurricular, educational, and sports activities [20,21] and in general the environment where a child grows up also seems to mark the possibility of suffering or not from obesity. Knowing first-hand where, how and in what way to act within the global epidemic of childhood obesity [22] will allow us to optimize the scarce resources we have and to carry out health intervention policies that are as effective as possible [23,24].

Biases and Limitations of the Study

The main limitation of the study is related to the fact that the data used come from the electronic medical record and therefore have not been generated for research purposes. This is why, as described in the literature, errors may occur in the measurement and transcription of the data (Heude B, et al. [3]). The nature of this study allows it to be repeated periodically, detecting areas of improvement in different subpopulations.

Ethical Aspects

The study has been prepared in compliance with the principles established in the Declaration of Helsinki (1964) latest version For-

taleza, 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

- Jose A. Lozano Basque Center for Applied Mathematics BCAM
- Ioar Married Tellechea Basque Center for Applied Mathematics BCAM
- Aritz Pérez Postdoctoral Fellow BCAM - Basque Center for Applied Mathematics.

References

1. Zamlout A, Kamal Alwannous, Ali Kahila, Majd Yaseen, Raneem Albadish, et al. (2022) Syrian national growth references for children and adolescents aged 2-20 years. *BMC Pediatr* 22(1): 282.
2. Tarupi W, Yvan Lepage, María L Felix, Claude Monnier, Roland Hauspie, et al. (2020) Growth references for weight, height, and body mass index for Ecuadorian children and adolescents aged 5-19 years. *Arch Argent Pediatr* 118(2): 117-124.
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. (2006) WHO Child Growth Standards based on length/height, weight and age. WHO Multicentre Growth Reference Study Group. *Minutes Paediatr Suppl* 450: 76-85.
5. de Onis M, Adelheid W Onyango, Elaine Borghi, Amani Siyam, Chizuru Nishida, et al. (2007) Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ* 85(9): 660-667.
6. Carrascosa Lezcano A, J M Fernández García, C Fernández Ramos, A Fernández 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. *Pediatr (Barc)* 68(6): 552-569.
7. Díez López I, Maeso Mendez S, Machón Sobrado M (2024) A new paradigm in the construction of growth charts in pediatrics. Why not use big data? *Endocrinol Metab Int J* 12(3): 92-99.
8. Wang Y, Lim H (2012) The global childhood obesity epidemic and the association between socio-economic status and childhood obesity. *Int Rev Psychiatry* 24(3): 176-188.
9. Wang Y (2001) Cross-national comparison of childhood obesity: the epidemic and the relationship between obesity and socioeconomic status. *Int J Epidemiol* 30(5): 1129-1136.

10. Suárez Reyes M, Fernández Verdejo R, Quintiliano D, Pinheiro AC, Pizarro T (2024) Effects of school closure on lifestyle behaviours and health outcomes in children during the COVID-19 pandemic in Chile: A time-matched analysis. *Pediatr Obes* 8: e13182.
11. Arayess L, Knockaert N, Winkens B, Lubrecht JW, Verweij M, et al. (2022) The Side-Effects of the COVID-19 Pandemic: Increased BMI z-Score in Children with Overweight and Obesity in a Personalised Lifestyle Intervention One Year after the Start of the Pandemic in The Netherlands. *Nutrients* 14(9): 1942.
12. Stavridou A, Kapsali E, Panagouli E, Thirios A, Polychronis K, et al. (2021) Obesity in Children and Adolescents during COVID-19 Pandemic. *Children* 8(2): 135.
13. Ferguson T S (1973) A Bayesian analysis of some nonparametric problems. *The annals of statistics* 1(2): 209-230.
14. Rasmussen C (1999) The infinite Gaussian mixture model. *Advances in neural information processing systems* 12: 554-560.
15. Teh Y W, Jordan MI (2010) Hierarchical Bayesian nonparametric models with applications. *Bayesian nonparametrics* 1: 158-207.
16. Van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. *Journal of machine learning research* 9(11): 2579-2605.
17. Kruskal J B (1964) Non metric multidimensional scaling: a numerical method. *Psychometrika* 29(2): 115-129.
18. 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.
19. Diana A, Matechou E, Griffin J, Johnston A (2020) A hierarchical dependent Dirichlet process prior for modeling bird migration patterns in the UK. *The Annals of Applied Statistics* 14(1): 473-493.
20. Ahrens W, Moreno LA, Pigeot I (2011) Childhood obesity: Prevalence worldwide. In: Moreno LA, editor: *Epidemiology of Obesity in Children and Adolescents*. New York: Springer, pp. 219-235.
21. Wang Y, Monteiro C, Popkin BM (2002) Trends of obesity and underweight in older children and adolescents in the United States, Brazil, China, and Russia. *Am J Clin Nutr* 75(6): 971-977.
22. Umekar S, Joshi A (2024) Obesity and Preventive Intervention Among Children: A Narrative Review. *Cureus* 16(2): e54520.
23. Goel A, Reddy S, Goel P (2024) Causes, Consequences, and Preventive Strategies for Childhood Obesity: A Narrative Review *Cureus* 16(7): e64985.
24. Antwi F, Fazylova N, Garcon MC, Lopez L, Rubiano R, et al. (2012) The effectiveness of web-based programs on the reduction of childhood obesity in school-aged children: A systematic review. *JBIR Book Syst Rev* 10(42 Suppl): 1-14.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2024.59.009272

Ignacio Diez Lopez. Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>