MedCrave
Step into the World of Research
Endocrinology & Metabolism International Journal

Research Article
Open Access
CrossMark

Was the COVID-19 pandemic and home confinement responsible for a childhood obesity pandemic? responses from big data

Abstract

It is suspected that the confinement due to the COVID-19 virus of the general population, and especially children, throughout 2020; and the increase in the use of subsidiary technologies led to an increase in cases of childhood obesity. Big data tools are nowadays postulated as a tool of the first magnitude to assess observed population changes.

Main objective: To assess the effect on the somatometric parameters of confinement in a child population throughout the year 2020 and 2021.

Material and methods: Data collected from episodes of computerized medical records, studying the variables sex, age, weight, height, of the same pediatric population, buying two periods PRE (year 2020) and POST pandemic (year 2022). To calculate the curves and percentile tables we have used the Cole-Green LMS algorithm with penalized likelihood, implemented in the RefCurv 0.4.2 (2020) software, which allows managing large amounts of data. The hyper parameters have been selected using the BIC (Bayesian information criterion). To calculate population deviations from the reference, being above 1.5 standard deviations from the mean according to age has been taken as a reference.

Results: 66,975 computerized episodes of children under 16 years of age and a total of 1,205,000 variables studied. Comparative data and graphs are represented pre and post pandemic with respect to the standards. Regional somatometrics, There are significant differences, with an increase in overweight in the entire sample of men and women in our population than the usual standards. Big data technology surpasses classic population studies in power and is an innovative tool compared to auxological studies (limited in N) carried out to date. COVID-19 and its confinement led to an increase in obesity in the child population.

Note: CEIC OSI ARABA Approval Expte 2022-058

Volume 12 Issue 3 - 2024

Ignacio Díez López,[1] Sandra Maeso Mendez,[2] Gaspar Sánchez Merino[3]

[1]Basque Country University, Pediatric Department, Collaborative Group from Basque Center of Applied Mathematics (BCAM), Spain
[2]Child and adolescent endocrinology Unit, Pediatric Department, Collaborative Group from Basque Center of Applied Mathematics (BCAM), Spain
[3]Coordinator of the Innovation Platform - IIS BIOARABA, Spain

Correspondence: Ignacio Díez López, Basque Country University, Pediatric Department, collaborative group from Basque center of applied mathematics (BCAM), Spain, Email ignacio.diezlopez@osakidetza.eus

Received: September 04, 2024 | **Published:** September 20, 2024

Introduction

Knowing the weight and height situation is considered a relevant indicator of health status in childhood. It is therefore common clinical practice to weigh and measure 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 increased in the last decade in all regions of the world.[4,5]

Our country, Spain, is no stranger to this problem and local authors have already demonstrated an increase in the weight of our population in 2010 compared to the most classic standards (Carrascosa et al. – Spanish studies 2010 vs those prepared by the Orbegozo Foundation years 1988, 2004 and 2011).[6,7] These longitudinal studies are by nature long, expensive and have a limited number of subjects.

Current electronic medical records include, within normal clinical practice, the collection of multiple variables, including the variables date of collection, weight and height. Different statistical techniques, such as machine learning, allow this data to be exploited from a large number of cases in an almost semi-automated way, providing data of great statistical value.

The confinement due to COVID-19 throughout the year 2020 throughout the world and throughout 2021 in some countries, plus the effects of the pandemic, seems to have had a harmful effect on various aspects of health, including the probable increase in the prevalence of overweight. Although there are studies on this matter (Pediatrics 2020, Children 2021),[8,9] there are no studies on this matter, at least in our environment and nearby population ; and those referred to have a studied population limited in number of cases. The causes of this increase have been postulated to be diverse, from a decrease in physical activity, overeating or the use of new technologies[8,9]

Our national and local authorities established a general confinement of the child population from April to the end of June 2020; as well as different restrictions on mobility, group and sports activities from that date until the end of 2021. This work is presented as an opportunity to use the data already existing in the computerized network to know the situation of the entire pediatric population in our region regarding its obesity situation before and after confinement.

Goals

Main objective: Describe the situation of the prevalence of overweight in the pediatric population of our environment, Alava, Country Vasco, Spain, before and after the COVID-19 pandemic by extracting variables from the electronic medical record and their subsequent analysis using a new big data approach in two different historical moments separated by the period of global confinement and limitations and mobility restrictions (from April 2020 to December 2021). Compare whether there are differences in these two periods in the BMI variable (kg/m2) with respect to the reference standards used (Orbegozo 2004 and Españoles 2010) through a comparison of paired means.

Material and methods

Design: It is a cross-sectional population study.

Study population: All children under 18 years of age under follow-up in the Basque health system, OSAKIDETZA, who present weight and height records in the OSABIDE GLOBAL tool in the Alava area.

Inclusion criteria:

a) Both sexes

b) Ages between 0 and 18 years

c) Be registered or present a registration address (according to GLOBAL data for the Alava /Araba area) belonging to the OSI Araba (which includes the entire rural area of the province except the LLODIO tax area) belonging to one of the centers of health of the OSI itself, collecting what it is in each case.

d) Have this data collected in the OSABIDE GLOBAL

Exclusion criteria: Not having data registered in GLOBAL

Sample size calculation: The study will include all people between 0 and 18 years old who reside in the historical territory of Araba (except as described). According to data from the Basque Institute of Statistics (Eusstat), in 2021 there were 47,853 people between 0-19 years of age in Vitoria (Basque Institute of Statistics (Eusstat). Population of the AC of Euskadi by territorial area, large age groups, sex and period. Available at: https://www.eustat.eus/bankupx/pxweb/es/DB/-/PX_010154_cepv1_ep06b.px/table/tableViewLayout1/ (Accessed 08/29/2022). Araba de Vitoria pediatric population, it is considered that it is not necessary to calculate the sample size. However, it is possible that after the data purification process there will be a loss in the number of participating subjects, in those cases where. Those that have not recorded the data necessary to carry out the study or are not well recorded. The data recorded from 01/01/2020 to 03/30/2020 is chosen for the PRE-pandemic (PRECOV) group. Data from the POST-pandemic (POSTCOV) group are recorded as those registered between 01/01/2022 and 03/30/2022.

Variables

Main variables: 1) Weight (kg), 2) Size (cm), 3) Sex (Male, Female, Binary), 4) Age (expressed in years and months), 5) Registration date

Data management plan: A data protection impact assessment has been prepared. The OSI Araba IT service, the main researcher of the project and the collaborating researchers will participate in the data life cycle, including professionals from the Basque Center for Applied Mathematics (BCAM) who are part of the research team. There is a collaboration agreement between the BCAM and the Bio araba Health Research Institute.

The principal investigator requests the extraction of data (date of birth in month and year format, sex, weight, height, date of registration and health center) from the electronic medical record (EHR) to the OSI Araba IT service. Osakidetza's electronic medical records for the time necessary to carry out the following actions:

a) Data purging

b) Statistical analysis of data

Once the investigation is completed, said database will be completely destroyed by all people involved in this study. Specific security measures were adopted to prevent re-identification and access by unauthorized third parties. The database obtained from the IT Service comes with a patient ID that is neither the CIC, nor medical history number, nor any other data that can be used for patient re-identification. Only the IT Service will be aware of that ID.

Statistical analysis: Hierarchical Dirichlet process mixture model Dirichlet processes (Dirichlet process, DP)[10] are a family of stochastic processes whose realizations (the values it takes) are probability distributions. DPs are used in Bayesian inference to describe prior knowledge about the distribution of random variables, that is, the probability that random variables are distributed according to a particular distribution. The Dirichlet process is specified by a base distribution and a positive real number, called the concentration parameter. The base distribution is the expected value of the process. A DP establishes distributions around the base distribution. Although the base distribution is continuous, the distributions extracted from a DP are discrete. The concentration parameter controls the number of PD values: the realizations are discrete distributions with decreasing concentration as the concentration parameter increases.

Dirichlet processes is as a priori probability distribution in infinite mixture models.[11] In this project we will adopt this approach, and the DPs will allow us to build Gaussian - averaged models (GM).[11] These models are known as Gaussian-averaged models based on Dirichlet processes (Dirichlet process Gaussian mixture models, DPGMM).

DPGMMs are especially of interest for modeling populations in which the number of groups (Clusters) is unknown, because they are capable of establishing the number of components and their parameters (means and covariance matrices) of the Gaussian averaging model. Automatically. The number of components will be determined by the data and by the value of the concentration parameter. The DPGMM will allow two problems to be solved simultaneously: performing a probabilistic segmentation (clustering) of the study population and at the same time modeling its underlying distribution (density estimation) in terms of GMs.

Application to the study of populations: Gaussian averaging models based on hierarchical Dirichlet processes (Hierarchical Dirichlet process Gaussian mixture model, HDPGMM).[12] This will allow us to address the following problems: 1) learn a GM model for the set of populations, 2) establish the different clusters of individuals with differentiated behavior within the total population, 3) establish clusters of populations with similar behavior, 4) determine the strength of the components of the GM in each population, 5) classify a new individual in one of the components of the GM and infer the value of some of its variables from the value of the rest, and 6) compare the results obtained at two time points and analyze the evolution of the populations.

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 depending on the date of data collection, age, sex, or health center.[13] The study will also be an opportunity to study and incorporate recent methodological innovations on databases similar to ours.[14–16]

Throughout the process, open source Python libraries are used: Pandas for data management and preprocessing, Sklearn for the basic algorithms (DSNE, MDS) and Matplotlib for visualization of the results. Regarding the HDPGMM model, we propose our own open source implementation based on public domain libraries such as https://github.com/blei-lab/online-hdp. We proceed from each studied variable to carry out MEDIAS and SDS studies.

The BMI calculation is performed 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 (Orbegozo 2004 and Española's 2011). Overweight is defined from +1. 5 SDS with respect to the referenced normality for age and sex.

Results

Data has been obtained from a total of 67,270 minors. The sum of all the variables studied (some exposed in this work and others reserved) add up to 1,749,020 variables. Although data is available for the age range of 16-18 years, the number of available data being scarcer and with the dispersion presented, it was advised by the collaborative study team, to avoid bias, to be eliminated from this presentation.

Next, we proceed to study using the so-called Hierarchical model or method. Dirichlet process Gaussian mixture model, applied to our population vs reference graphs most used to date in the region (Orbegozo 2011) and the most extensive study in number of cases, most recent (Spanish study 2010) and used in the country. It will be assessed if there are differences at a significance of $p<0.05$.We present in various tables the results obtained by sex, age and the BMI variable (Table 1–4).

Table 1 Numerical representation of data by age of the BMI variable (Kgrs). Means and SDS in men compared to the two most common reference studies of our environment, Carrascosa and Orbegoso growth studies

Spanish Carrascosa Study 2010			
BMI Man (kg/m2)			
Age (and)	**No.**	**mean**	**OF**
0	2974	13.17	1.18
0.25	233	16.69	1.4
0.5	214	17.71	1.84
0.75	213	17.68	1.9
1	169	17.99	1.49
1.25	166	17.64	1.71
1.5	149	17.67	1.65
1.75	153	17.15	1.37
2	182	16.55	1.37
2.5	263	16.57	1.42
3	545	16.24	1.56
3.5	588	16.03	1.82
4	598	16.03	1.74
4.5	564	16.04	1.81
5	497	15.88	1.97
5.5	501	16.12	2.11
6	454	16.16	2.08
6.5	446	16.36	2.29
7	448	16.54	2.33
7.5	445	16.75	2.5
8	418	16.91	2.48
8.5	477	17.73	3.14
9	466	18.01	3.11
9.5	491	18.39	3.07
10	488	18.41	3.14
10.5	519	18.72	3.51
11	493	19.3	3.57
11.5	456	19.44	3.47
12	455	19.72	3.5
12.5	394	20.26	3.53
13	410	20.09	3.54
13.5	404	20.64	3.39
14	359	21.24	3.72
14.5	349	21.13	3.71
15	391	21.41	3.52

Orbegozo Study 2011			
BMI Man (kg/m2)			
Age (and)	**No.**	**mean**	**OF**
0	162	13	0.91
0.25	70	16.67	1.42
0.5	63	17.52	1.62
0.75	84	17.92	1.47
1	68	18.12	1.71
1.25	76	17.66	1.49
1.5	60	17.85	1.57
1.75	77	17.31	1.28
2	104	16.71	1.02
2.5	64	16.59	1.2
3	95	16.12	1.19
3.5	69	16.46	1.83
4	79	16.22	1.46
4.5	72	16.16	1.29
5	71	16,19	1.86
5.5	77	16.26	1.98
6	81	16.23	1.75
6.5	94	16.37	1.81
7	74	16.68	1.61
7.5	71	16.97	2.19
8	78	16.96	2.31
8.5	98	17.69	2.54
9	68	18.54	2.68
9.5	85	18.77	2.63
10	96	18.34	2.8
10.5	93	19.08	3.49
11	85	18.93	2.77
11.5	84	19.04	2.75
12	91	19.64	3.19
12.5	84	19.71	2.56
13	75	19.98	2.84
13.5	82	20.32	3.09
14	82	20.67	2.89
14.5	61	21.12	3.38
15	81	20.89	2.89

Citation: Díez-López, Maeso-Mendez, Sánchez-Merino G. Was the COVID-19 pandemic and home confinement responsible for a childhood obesity pandemic? responses from big data. *Endocrinol Metab Int J.* 2024;12(3):83–90. DOI: 10.15406/emij.2024.12.00353

Table 2 Numerical representation and p-value variation of the BMI variable cohort 2022 vs 2020 for men according to age (Kgrs), with respect to the two most common reference studies in our environment, Carrascosa and Orbegoso growth studies. Green increase of the variable or significant value

Age (and)	Comparative (2022-2020) with ref. CARRASCOSA	P-value (t-test)	Age (and)	Comparative (2022-2020) with ref. ORBEGOZO	P-value (t-test)
0	1.28	0	0	1.45	0
0.25	0.03	0.76313	0.25	0.05	0.77396
0.5	-0.72	0	0.5	-0.53	0.01126
0.75	-0.61	0.0000 1	0.75	-0.85	0
1	-1.14	0	1	-1.27	0
1.25	-1.01	0	1.25	-1.03	0
1.5	-1.3	0	1.5	-1.48	0
1.75	-1.04	0	1.75	-1.2	0
2	-0.58	0	2	-0.74	0
2.5	-0.86	0	2.5	-0.88	0.00004
3	-0.53	0	3	-0.41	0.00321
3.5	-0.35	0.00235	3.5	-0.78	0.00106
4	-0.24	0.00693	4	-0.43	0.01244
4.5	-0.27	0.07716	4.5	-0.39	0.05355
5	-0.1	0.61262	5	-0.41	0.14824
5.5	0.16	0.51493	5.5	0.02	0.95377
6	0	0.98708	6	-0.07	0.74826
6.5	0.62	0.00807	6.5	0.61	0.0292
7	0.62	0.02816	7	0.48	0.13537
7.5	0.93	0.00117	7.5	0.71	0.05398
8	0.25	0.19834	8	0.2	0.51251
8.5	0.08	0.74873	8.5	0.12	0.71357
9	0.41	0.19969	9	-0.12	0.78776
9.5	0.41	0.23432	9.5	0.03	0.95211
10	0.01	0.97115	10	0.08	0.80608
10.5	0.23	0.3637	10.5	-0.13	0.74943
11	0.23	0.43885	11	0.6	0.12486
11.5	0.43	0.20437	11.5	0.83	0.05019
12	0.29	0.35373	12	0.37	0.38517
12.5	0.19	0.63545	12.5	0.74	0.10287
13	0.53	0.08703	13	0.64	0.12409
13.5	-0.53	0.02745	13.5	-0.21	0.57935
14	-1.09	0.00074	14	-0.52	0.20594
14.5	1.2	0.10169	14.5	1.21	0.14621
15	3.97	0.00197	15	4.49	0.00079

Table 3 Numerical representation of data by age of the BMI variable (Kgrs). Means and SDS in women compared to the two most common reference studies of our environment, Carrascosa and Orbegoso growth studies

Spanish Carrascosa Study 2010			
BMI Woman (kg/m2)			
Age (and)	No.	mean	OF
0	146	12.97	1.17
0.25	65	16.08	1.38
0.5	67	17.15	1.46
0.75	62	17.58	1.5
1	70	17.61	1.59
1.25	65	17.11	1.31
1.5	55	16.96	1.45
1.75	51	16.77	1.47
2	81	16.58	1.35
2.5	67	16.37	1.26

Orbegozo Study 2011			
BMI Woman (kg/m2)			
Age (and)	No.	mean	OF
0	146	12.84	1.15
0.25	65	16.2	1.34
0.5	67	17.03	1.37
0.75	62	17.66	1.44
1	70	17.62	1.81
1.25	65	17.24	1.2
1.5	55	17.13	1.43
1.75	51	16.9	1.63
2	81	16.79	1.16
2.5	67	16.3	1.27

Table 3 Continued..

Spanish Carrascosa Study 2010				Orbegozo Study 2011			
BMI Woman (kg/m2)				BMI Woman (kg/m2)			
Age (and)	**No.**	**mean**	**OF**	**Age (and)**	**No.**	**mean**	**OF**
4	63	15.63	1.64	4	63	15.85	1.39
4.5	55	15.85	1.63	4.5	55	15.94	1.68
5	70	15.73	1.76	5	70	16.12	1.6
5.5	70	15.84	2.16	5.5	70	15.9	1.72
6	86	15.97	2.14	6	86	15.75	1.85
6.5	88	16.36	2.27	6.5	88	16.55	1.98
7	84	16.36	2.41	7	84	16.51	1.98
7.5	76	16.94	2.53	7.5	76	17.07	2.5
8	98	17.14	2.84	8	98	17.48	2.45
8.5	87	17.55	2.93	8.5	87	17.32	2.14
9	73	17.74	3.12	9	73	17.91	2.18
9.5	70	18.05	3	9.5	70	17.51	2.66
10	69	18.36	3.3	10	69	18.67	2.36
10.5	78	18.65	3.39	10.5	78	18.57	2.87
11	87	19.42	3.69	11	87	19.23	2.94
11.5	69	19.33	3.27	11.5	69	18.71	2.47
12	70	19.51	3.36	12	70	18.91	2.45
12.5	64	20.04	4.06	12.5	64	19.02	2.87
13	51	20.58	4.03	13	51	19.96	2.79
13.5	58	20.82	3.84	13.5	58	20.55	2.7
14	54	20.77	3.55	14	54	20.15	2.69
14.5	71	21.17	3.63	14.5	71	20.52	2.96
15	64	21.19	3.7	15	64	21.29	3.16

Table 4 Numerical representation and p-value variation of the BMI variable cohort 2022 vs 2020 for women according to age (Kgrs), with respect to the two most common reference studies in our environment, Carrascosa and Orbegoso growth studies. Green increase of the variable or significant value

Age (and)	Comparative (2022-2020) with ref. CARRASCOSA	P-value (t-test)	Age (and)	Comparative (2022-2020) with ref. ORBEGOZO	P-value (t-test)
0	1.14	0	0	1.27	0
0.25	-0.06	0.54226	0.25	-0.18	0.29184
0.5	-0.63	0	0.5	-0.51	0.00342
0.75	-0.9	0	0.75	-0.98	0
1	-1.11	0	1	-1.12	0
1.25	-0.76	0	1.25	-0.89	0
1.5	-0.93	0	1.5	-1.1	0
1.75	-0.87	0	1.75	-1	0.00007
2	-0.81	0	2	-1.02	0
2.5	-0.73	0.00002	2.5	-0.66	0.00212
3	-0.02	0.84275	3	-0.45	0.00241
3.5	-0.2	0.14542	3.5	-0.18	0.40462
4	0.19	0.05259	4	-0.03	0.87309
4.5	0.06	0.7268	4.5	-0.03	0.9129
5	0.09	0.63838	5	-0.3	0.24921
5.5	0.25	0.35877	5.5	0.19	0.56132
6	0.07	0.58414	6	0.29	0.17757
6.5	0.43	0.04158	6.5	0.24	0.38719
7	0.91	0.00015	7	0.76	0.01273
7.5	0.41	0.1266	7.5	0.28	0.45602
8	0.16	0.39348	8	-0.18	0.52709
8.5	0.35	0.17196	8.5	0.58	0.06438
9	0.25	0.34552	9	0.08	0.81303
9.5	-0.18	0.46746	9.5	0.36	0.33913

Table 4 Continued..

Age (and)	Comparative (2022-2020) with ref. CARRASCOSA	P-value (t-test)	Age (and)	Comparative (2022-2020) with ref. ORBEGOZO	P-value (t-test)
10	-0.1	0.614	10	-0.41	0.1839
10.5	0.32	0.23353	10.5	0.4	0.30638
11	-0.44	0.14295	11	-0.25	0.53399
11.5	0.53	0.12266	11.5	1.15	0.00705
12	0.51	0.12133	12	1.11	0.00702
12.5	0.61	0.18166	12.5	1.63	0.00281
13	-0.39	0.21551	13	0.23	0.61219
13.5	-0.17	0.54339	13.5	0.1	0.80203
14	0.13	0.7113	14	0.75	0.10594
14.5	1.24	0.09093	14.5	1.89	0.01746
15	2.1	0.05514	15	2	0.08332

This variable is represented using graphs in percentile format (Chart 1&2).

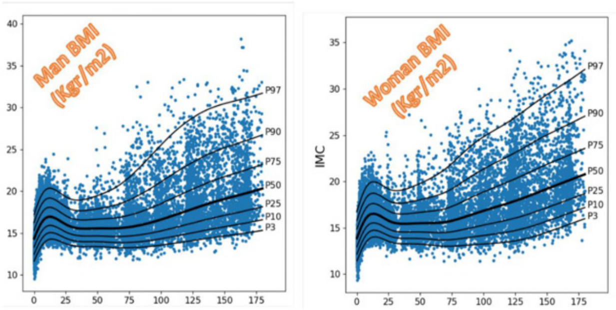

Chart 1 Percentile representation by age of the BMI variable (Kgs /m2). Abscissa age in months. PRECOVID 2020 Cohort

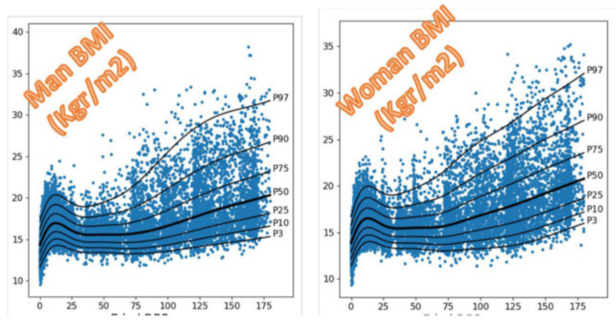

Chart 2 Percentile representation by age of the BMI variable (Kgs /m2). Abscissa age in months. POSTCOVID Cohort – 2022

Overall, it is evident that between the PRECOVID cohort and the POSTCOVID cohort, there is an increase in the BMI variable in both sexes in our population. In the case of men, this variable increases between both cohorts, especially from 6 years of age and extends throughout puberty. The significance is highest in the pre- pubertal age group and the middle ages of puberty.

In the case of women, the BMI increases between both cohorts from 4 years of age and extends throughout childhood until the supposed onset of puberty (around age 9), with a subsequent rebound. Puberty and again stabilization of the increase in BMI around the average age of menarche (13 years). The significance is more, however, in middle childhood and post –menarche (Chart 3&4).

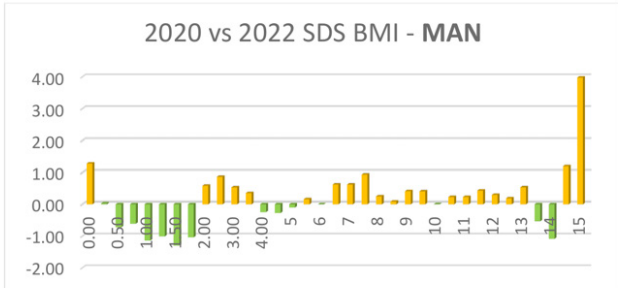

Chart 3 Representation in SDS by age of the BMI variable (Kgs /m2). Men. PRE and POST COVID Cohort 2020-2022

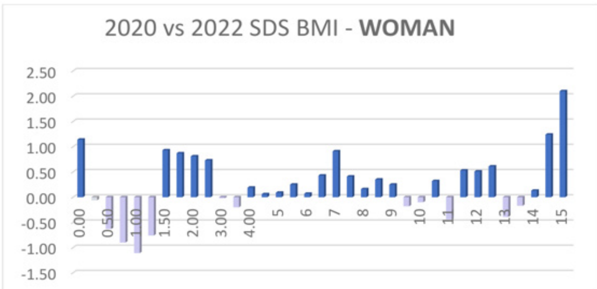

Chart 4 Representation in SDS by age of the BMI variable (Kgs /m2).Women. PRE and POST COVID Cohort 2020-2022

Discussion

Current electronic medical records include the collection of multiple data and clinical constants as part of normal clinical practice. Among them, aspects of children's somatometry. Different statistical techniques, such as machine learning They have been shown in other fields[14–16] to be effective in interpreting a large amount of data generated in real life and being able to make decisions about it.

With works like ours, we postulate the possibility of real use of this technology to obtain population studies and temporal demographic trends. In this work we present this possibility made reality as a methodological approach.

Although data is available for the age range of 16-18 years, the smaller number of cases in relation to the other ages and their dispersion, requires that they be eliminated through the statistical procedure of the study and to avoid biases of this study. This is because adolescents go to the doctor less and therefore the number of registrations is lower.

The secular acceleration of weight in relation to height[4,5] is seen in our population with a possible added effect due to the pandemic and its confinement. This is revealed in the study and may be due to various causes, such as the childhood obesity pandemic that we are experiencing, the effect of confinement/COVID-19[8,9] in 2022 on children's health, changes in food or even the typology of the area's population (immigration rate, socioeconomic level).[2–4]

This study method with BIG DATA is postulated as a faster and more economical way to have updated regional graphs than classic studies. This point should be verified with other studies. We point out that our study reveals the increase in obesity in key stages of development, such as pre-pubertal age or the end of puberty, which can contribute to maintaining and making the problem of overweight chronic.

This work is the basis for developing community intervention strategies pending corroboration by our own team and seeing population dynamics.

Bias and limitations of the study: The main limitation of the study has to do with the fact that the data to be used comes from the electronic medical record and therefore has not been generated for research purposes. This is why, as described in the literature, errors can occur in the measurement and transcription of the data (Heude B et al. A big -data approach to producing descriptive anthropometric references: a feasibility and validation study of pediatric growth charts. Lancet Digital Health. 2019 Dec; 1(8): e 413-e423). To minimize this limitation, data extracted from electronic medical records will be cleaned before proceeding with the statistical analysis of the data.

Project impact: The expected impact of the project results, in terms of the capacity for modification in healthcare processes, to improve the health and quality of life of patients is of great importance. It is estimated that the current cost of preparing an updated, regional, longitudinal growth study, with the consequent limit of cases (<1,000) is more than 8-10 years per project and with an economic cost greater than 60,000 euros in said period, taking into account takes into account published studies (Orbegozo) in its methodology. This project has been developed in a much shorter time and at a lower cost. Furthermore, the data obtained is not limited to a limited population (although it is supposed to be representative) but rather quasi-real as it encompasses most of the data available on the computer servers in the area. The nature of this study allows it to be repeated periodically, detecting areas of improvement in different subpopulations.

On the other hand, by having variables associated with the growth of another type such as a health center, it will allow detecting areas of socio-health risk, where to implement other types of studies or intervention measures...

Conclusion

Our work shows that the COVID-19 pandemic had an impact on the health status of our child population; at least in the nutritional situation. Confinement, increased use of technology, lack of sport or increased intake could be behind this increase. BIG DATA technology is shown to be a useful tool for population epidemiological studies. The use of this technology allows us to respond in a shorter time to population questions with health determinants. This will allow us to be more efficient and detect higher-risk populations for more effective interventions, given the scarce health resources.

Ethical aspects

The study has been prepared respecting 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 Expte 2022-058

Economic report

The study will be carried out without funding. The tasks described in the project are assumed by the main researcher and his collaborators.

Acknowledgments

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

a) Jose A. Lozano Basque Center for Applied Mathematics BCAM

b) Ioar Married Tellechea Basque Center for Applied Mathematics BCAM

c) Aritz Pérez Postdoctoral Fellow BCAM - Basque Center for Applied Mathematics

d) Mónica Machón Sobrado, Methodology and Statistics unit OSI Araba, Collaborative Group from Basque Center of Applied Mathematics (BCAM), Spain

References

1. Zamlout A, Kamal A, Ali K et al. Syrian national growth references for children and adolescents aged 2-20 years. *BMC Pediatrics*. 2022;22(1):282.
2. Tarupi W, Yvan L, Maria L et al. Growth references for weight, height, and body mass index for ecuadorian children and adolescents aged 5-19 years. *Arch Argent Pediatr*. 2020;118(2):117–124.
3. Heude B, Pauline S, Andreas W, et al. A big-data approach to producing descriptive anthropometric references: a feasibility and validation study of pediatric growth charts. *Lancet Digital Health*. 2019;1(8):e413–e423.
4. WHO Multicenter Growth Reference Study Group. WHO child growth standards based on length/height, weight and age. *Acta Paediatr Suppl*. 2006;450:76–85.
5. Onis M, Adelheid WO, Elaine B et al. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ*. 2007;85(9):660–667.
6. Sánchez GE, Carrascosa L, Fernandez Garcia JM et al. [Spanish growth studies: current situation, usefulness and recommendations for their use]. *An Pediatr (Barc).* 2011;74(3):193.e1–16.
7. Carrascosa LA, Fernandez Garcia JM, Fernandez R, et al. [Spanish cross-sectional growth study 2008. Part II: height, weight and body mass index values from birth to adulthood]. *An Pediatr (Barc).* 2008;68(6):552–569.
8. Loucia A, Will C, Christine J. The indirect impact of covid-19 on child health. *Paediatr and Child Health(Oxford)*. 2020;30(12):430–437.
9. Stavridou A, Kapsali E, Panagouli E, et al. Obesity in children and adolescents during covid-19 pandemic. *Children*. 2021;8(2):135.
10. Ferguson TS. A Bayesian analysis of some nonparametric problems. *Ann. statist.* 1973;1(2):209–230.
11. Rasmussen C. The infinite gaussian mixture model. *MIT Press.* 2000;554–560.

12. The YW, Jordan MI. Hierarchical bayesian nonparametric models with applications. *Bayesian nonparametrics*. 2009;1:158-207.
13. Van der ML, Hinton G. Visualizing data using t-SNE. *Journal of machine learning research*. 2008;9(11):2579–2605.
14. Kruskal JB. Non metric multidimensional scaling: a numerical method. *Psychometrika*. 1964;29(2):115–129.
15. Gilholm P, Mengersen K, Thompson H. Identifying latent subgroups of children with developmental delay using bayesian sequential updating and dirichlet process mixture modeling. *PloS one*. 2020:15(6):e0233542.
16. Diana A, Matechou E, Griffin J, et al. A hierarchical dependent dirichlet process prior for modeling bird migration patterns in the uk. *The Annals of Applied Statistics*. 2020;(1):473–493.