Gangemi et al. World Allergy Organization Journal (2023) 16:100764
http://doi.org/10.1016/j.waojou.2023.100764

WORLD ALLERGY ORGANIZATION JOURNAL

Open Access

Multimodal, open-source big data analysis in asthma: A novel approach to inform public health programming

Sebastiano Gangemi, MD, PhD[a], Alessandro Tonacci, PhD[b], Giulia Costanzo, MD[c], Davide Firinu, MD[c*,1] and Stefano Del Giacco, MD, PhD[c]

ABSTRACT

Asthma is a chronic respiratory disease affecting over 358 million people worldwide; for this reason analysing big data on asthma from different countries could give a more detailed picture of current disease burden.

We aim to investigate the correlations between asthma and key socio-demographic parameters from different world databases. We found a direct correlation with the gross domestic product (GDP) per capita and its nominal counterpart, with wealthiest countries seen to have the highest prevalence of asthma, as also confirmed by a similar correlation with the human development index (HDI). A positive correlation was also seen between asthma prevalence and a number of socio-cultural data being representative of a good life quality index and prevalent in more developed and wealthier countries. Concerning medical data, an inverse relationship was seen between asthma prevalence and helminthiasis.

Those data indicate a higher prevalence for asthma in more developed countries, where socio-economic status is higher and also the access to medical care is more ubiquitous. The approach used in our study highlighted the role of medical literacy and access to healthcare facilities in the correct diagnosis of asthma and vice versa. Our data appear to be suitable in terms of a health programming approach because of the high burden of disease worldwide.

Keywords: Asthma, Big data, Socio economic status, Health programming, Epidemiology, Real data

INTRODUCTION

Asthma is one of the most common chronic diseases worldwide, affecting over 358 million individuals. These individuals are heterogeneously distributed throughout the world, and such distribution appears to be irrespective of the degree of development.[1] As reported by the GINA guidelines, asthma is a heterogeneous disease in terms of symptoms, age of onset, and endotyping and phenotyping, with phenotypes to be defined as "observable characteristics that

[a]School and Operative Unit of Allergy and Clinical Immunology, Department of Clinical and Experimental Medicine, University of Messina, Via Consolare Valeria 1, 98125 Messina, Italy
*Corresponding author. Department of Medical Sciences and Public Health, University of Cagliari and Unit of Internal Medicine, Policlinico Universitario – AOU di Cagliari, Italy. E-mail: davide.firinu@unica.it
[1] Permanent address: Azienda Ospedaliero Universitaria, SS 554-Bivio Sestu, 09042 Monserrato (CA), Italy. Telephone: +39 070 51096128 - Fax +39 070 51096227.
Full list of author information is available at the end of the article

http://doi.org/10.1016/j.waojou.2023.100764
Received 29 August 2022; Received in revised from 8 March 2023; Accepted 15 March 2023
Online publication date xxx

http://doi.org/10.1016/j.waojou.2023.100764

result from a combination of hereditary and environmental influences".[2]

During the last years, several studies have been published on asthma worldwide prevalence and risk factors in large cohorts from different countries.[3,4] Despite this, there are still scarce data on correlations of asthma with key socio-economic indicators.

Big data about large populations may generate models more robust and generalizable. Indeed, as the population studied or sample size approaches the majority of the whole population, then the observations used begin to be always closer to a trustworthy description of the population.[5]

A recent paper by Fishe and colleagues evaluated the prodromal correlates of asthma using big data for a multi-domain analysis of demographic, clinical, and socio-ecological predictors of asthma in Florida, United States.[6] We here aim to investigate the correlations between asthma and key socio-demographic parameters from different world databases.

MATERIALS AND METHODS

Data sources

Several databases and scientific articles were used for searching the different parameters used in the present work. To the benefit of the reader, we tried to summarize the databases used for this analysis in Table 1.

In more detail, concerning data for asthma prevalence around the world, the paper by To and colleagues was employed,[4] as actually being the

Variable	Countries	Reference(s)
Asthma	70 worldwide countries	4
Population, Population Density, Mean age, Gross Domestic Product (GDP), Nominal GDP, Gini, HDI, yearly average temperature	Global	Wikipedia (see previously[7])
Education Index	Global	United Nations Development Programme, 2016
Total Fertility Rate	Global	World Bank Data, https://data.worldbank.org/indicator/SP.DYN.TFRT.IN
Social development indices	Global	International Institute of Social Studies, https://www.indsocdev.org/
Pollution index	Global	WHO, https://www.numbeo.com/pollution/rankings_by_country.jsp?title=2020
Pets ownership	22 worldwide countries	GFK, https://cdn2.hubspot.net/hubfs/2405078/cms-pdfs/fileadmin/user_upload/country_one_pager/nl/documents/global-gfk-survey_pet-ownership_2016.pdf
Green areas	European countries	European Environment Agency, https://www.eea.europa.eu/data-and-maps/figures/percentage-of-green-and-blue
Vaccination	Global	WHO, https://www.who.int/immunization/monitoring_surveillance/data/en
Calories consumption	Global	https://ourworldindata.org/diet-compositions

Table 1. Databases used to extract the relevant variables for the present investigation

most recent and complete collection of quantitative evidences in the specific field.

As for economic indicators, as already done previously,[7] the majority of them were retrieved using the related Wikipedia pages, including those for countries' Population, Population Density, mean age, Gross Domestic Product (GDP) and Nominal GDP (both Per Capita), the Gini Coefficient of wealth distribution (an index of overall income inequality, ranging from 0, where everyone has the same income, to 1, where inequality is maximal), the Human Development Index (HDI, a composite index of life expectancy, education, and per capita income indicators ranging from 0 to 1). The same free, web-based source was also used for yearly average temperature, taking into account the period between 1961 and 1990, based on gridded climatologies from the Climatic Research Unit elaborated in 2011, and for the Education Index, dated back to 2015 and published in the Human Development Report 2016 (United Nations Development Programme, 2016).

On the other hand, information about the fertility rate (Total Fertility Rate, TFR, defining the total number of children born or likely to be born to a woman in her life time) was retrieved per each country according to the World Bank data (https://data.worldbank.org/indicator/SP.DYN.TFRT.IN). The different social development indices were taken by the data from International Institute of Social Studies (https://www.indsocdev.org/).

Data about pollution were retrieved from the publicly available website Numbeo (https://www.numbeo.com/pollution/rankings_by_country.jsp?title=2020) referring to 2020 data provided by the World Health Organization (WHO). As for pet ownership by country, a survey conducted in 2015 by the GFK in 22 countries was taken as a reference (https://cdn2.hubspot.net/hubfs/2405078/cms-pdfs/fileadmin/user_upload/country_one_pager/nl/documents/global-gfk-survey_pet-ownership_2016.pdf).

Green areas are often calculated throughout the world in a different manner; therefore, it is quite hard to find uniform criteria for their quantification. Since the aim of the present study is to retrieve possibly existing correlations between clinical conditions and environmental, social, and economic factors, we decided to pick up data from just 1 database per indicator, avoiding the risk of dis-omogeneity of data sources. In light of this, green areas were considered for European countries only, and retrieved, updated in 2012, via the official European Environment Agency website (https://www.eea.europa.eu/data-and-maps/figures/percentage-of-green-and-blue).

Data about vaccines were retrieved by the WHO databases (https://www.who.int/immunization/monitoring_surveillance/data/en). Finally, data concerning food consumption (total calories, proteins and fats) were collected by the work.

Statistical analysis

In order to correlate clinical data with the other, possibly related features, we applied a bivariate correlation using Spearman's Correlation Test, with a Bonferroni post-hoc analysis correcting for multiple comparison. In all cases, statistical significance was assumed at $p < 0.05$. Correlation level was classified as: small (when $0.1 \leq r < 0.3$), medium ($0.3 \leq r < 0.5$), large ($r \geq 0.5$).

RESULTS

At first, asthma prevalence was correlated with demographic data. In particular, in this specific category, data included are: i) overall population, ii) population density, iii) GDP, iv) GDP per capita, v) GDP nominal, vi) GDP nominal per capita, vii) GINI, viii) HDI, ix) fertility rate, and x) mean population age.

Table 2 reports correlation data in terms of "r" and "p" values.

The second correlation analysis was performed with socio-cultural indicators, including: i) civic activism, ii) intergroup cohesion, iii) clubs and associations, iv) interpersonal safety trust, v) gender quality, vi) social inclusion, vii) education index, viii) owning of pets, ix) owning of dogs, x) owning of cats, xi) owning of fishes, xii) owning of birds, xiii) and owning of other pets.

The related results are reported in Table 2.

Environmental data were taken into account in the correlation with asthma. Notably, such features included: i) mean yearly country temperature, ii) pollution index, iii) green areas, iv) food

	Population	Density	GDP	GDP Per Capita	GDP Nominal	GDP Nominal Per Capita	GINI	HDI	Fertility	Mean Age
R	−.225	.098	.199	.577	.314	.590	−.320	.549	−.304	.349
P	.065	.424	.103	<.001**	.009	<.001**	.008	<.001**	.012	.004*
	Civic Activism	Intergroup Cohesion	Clubs and Associations	Interpersonal Safety Trust	Gender Equality	Inclusion	Education Index	Pets	Dogs	Cats
R	.615	.474	.169	.402	.406	.570	.517	0	.145	.376
P	<.001**	<.001**	.209	.001*	.001*	<.001**	<.001**	1	.593	.151

Table 2. Correlation values between asthma and demographic data (*: p < .05, **: p < .01 after Bonferroni correction) and between asthma and socio-cultural data (*: p < .05, **: p < .01 after Bonferroni correction)

	Temperature	Pollution Index	Green Areas	Food	Protein	Fat
R	−.149	−.679	.165	.235	.436	.460
P	.226	<.001**	.500	.055	<.001**	<.001**
	Vaccination Index	Cysticercosis	Helminthiasis			
R	0.107	−0.172	−0.346			
P	0.384	0.162	0.004*			

Table 3. Correlation values between asthma and environmental data (**: p < .01 after Bonferroni correction) and between asthma and medical data (*: p < .05 after Bonferroni correction)

consumption, v) protein consumption, and vi) fat consumption. Related results are displayed in Table 3.

Finally, medical data were correlated with asthma, with respect to: i) vaccination index, ii) cysticercosis, iii) helminthiasis, with results displayed in Table 3.

DISCUSSION

All the considerations drawn from the current data are intended to be based on a global level analysis of the results obtained. According to the data retrieved here, the country-based prevalence of asthma is correlated with some socio-economic indicators. In particular, a direct correlation with the GDP per capita and its nominal counterpart was observed, with wealthiest countries in the entire globe seen to have the highest prevalence of asthma, as also confirmed by a similar correlation with the HDI. A weak correlation with the mean age of the country population was also seen, possibly as a consequence of the relationship between a better economic status and higher life expectancy worldwide.

Reasonably concerned with that, a positive correlation for asthma prevalence was also seen with a number of socio-cultural data, including civic activism, intergroup cohesion, gender equality, social inclusion, and education index, all of them being representative of a good life quality index and, naturally, prevalent in more developed and wealthier countries throughout the world.

When it comes to environmental data, it is important to stress the presence of a strong, negative correlation between asthma prevalence and pollution index, with least polluted countries showing higher amounts of asthma cases. Concerning eating habits, countries with the highest consumption of proteins and fats per capita were the ones where asthma was deemed more prevalent.

Finally, concerning medical data, a moderate, inverse relationship was seen between asthma prevalence and helminthiasis, possibly due to the high prevalence of this medical issue in underdeveloped countries.

Taken together, those data probably indicate a higher prevalence for asthma in more developed countries, where socio-economic status is higher and also the access to medical care is more ubiquitous, making asthma diagnosis easier than in other parts of the globe. Albeit apparently counterintuitive, this possible explanation is made reasonable checking the actual literature. In fact, most of the published works have found higher prevalence of asthma and respiratory diseases in those subjects with a lower socio-economic status.[8-11] However, those works are mainly conducted on cohorts that are living in a somewhat uniform geographical area, and therefore with similar access levels to healthcare facilities. These data are also reported in a study conducted to investigate the socioeconomic and environmental predictors of Asthma-Related mortality worldwide.[12]

On the other hand, our research made use of worldwide data concerning country-based prevalence and variables taken into account, therefore highlighting the possible role of medical literacy and access to healthcare facilities in the correct diagnosis of asthma and vice versa.[13] Under this light, taking into account the various limitations of our approach including the statistical methods employed as well as the large heterogeneity of the data sources, our data appear to be suitable in terms of a health programming approach because of the high burden of disease worldwide and possibly suggest a need for a more in-depth investigation of asthma and similar conditions in underdeveloped countries, where the disease seems to be less prevalent, possibly due to underdiagnosis occurrence.

Abbreviations

GINA, global initiative for asthma; GDP, Gross Domestic Product; HDI, Human Development Index; TFR, Total Fertility Rate.

Funding

The authors did not received funding for this study.

Availability of data and materials

The dataset may be available upon reasonable request contacting the corresponding author.

Author contributions

Study conception and design contributed to SG, AT. Extracted the data contributed to SG, AT; Analysis and interpretation of data contributed to SG, AT; Drafting of manuscript contributed to SG, AT, GC,DF. Critical revision contributed to SDG, GC, DF, SG, AT. All authors read and worked on the manuscript.

Ethics approval

The study protocol conforms to the ethical guidelines of the Declaration of Helsinki.

Authors' consent for publication

All authors agree and give their consent for publication.

Declaration of competing interest

The authors state no relevant conflict of interest.

Author details

[a]School and Operative Unit of Allergy and Clinical Immunology, Department of Clinical and Experimental Medicine, University of Messina, Via Consolare Valeria 1, 98125 Messina, Italy [b]Institute of Clinical Physiology, National Research Council of Italy (IFC-CNR), Via G. Moruzzi 1, 56124 Pisa, Italy [c]Department of Medical Sciences and Public Health, University of Cagliari, Italy.

REFERENCES

1. Collaborators GBDCRD. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med*. 2017;5: 691–706. https://doi.org/10.1016/S2213-2600(17)30293-X.

2. Pembrey L, Barreto ML, Douwes J, et al. Understanding asthma phenotypes: the World Asthma Phenotypes (WASP) international collaboration. *ERJ Open Res*. 2018:4. https://doi.org/10.1183/23120541.00013-2018.

3. Almonacid Sanchez C, Melero Moreno C, Quirce Gancedo S, et al. PAGE study: summary of a study protocol to estimate the prevalence of severe asthma in Spain using big data methods. *J Investig Allergol Clin Immunol*. 2021;31:308–315. https://doi.org/10.18176/jiaci.0483.

4. To T, Stanojevic S, Moores G, et al. Global asthma prevalence in adults: findings from the cross-sectional world health survey. *BMC Publ Health*. 2012;12:204. https://doi.org/10.1186/1471-2458-12-204.

5. Diver S, Brightling CE. Big asthma data: getting bigger and more beautiful? *Thorax*. 2018;73:311–312. https://doi.org/10.1136/thoraxjnl-2017-211148.

6. Fishe JN, Bian J, Chen Z, et al. Prodromal clinical, demographic, and socio-ecological correlates of asthma in adults: a 10-year statewide big data multi-domain analysis. *J Asthma Off J Assoc Care Asthma*. 2020;57:1155–1167. https://doi.org/10.1080/02770903.2019.1642352.

7. Gangemi S, Billeci L, Tonacci A. Rich at risk: socio-economic drivers of COVID-19 pandemic spread. *Clin Mol Allergy CMA*. 2020;18:12. https://doi.org/10.1186/s12948-020-00127-4.

8. Azeez IA, Ladipo MM, Ige OM. Assessment of socioeconomic status and control of asthma in adults. *Ann Ib Postgrad Med*. 2016;14:85–91.

9. Ekerljung L, Sundblad BM, Ronmark E, Larsson K, Lundback B. Incidence and prevalence of adult asthma is associated with low socio-economic status. *Clin Respir J*. 2010;4:147–156. https://doi.org/10.1111/j.1752-699X.2009.00164.x.

10. Largent J, Nickerson B, Cooper D, Delfino RJ. Paediatric asthma hospital utilization varies by demographic factors and area socio-economic status. *Publ Health*. 2012;126:928–936. https://doi.org/10.1016/j.puhe.2012.04.011.

11. Schyllert C, Lindberg A, Hedman L, et al. Low socioeconomic status relates to asthma and wheeze, especially in women. *ERJ Open Res*. 2020:6. https://doi.org/10.1183/23120541.00258-2019.

12. Sinharoy A, Mitra S, Mondal P. Socioeconomic and environmental predictors of asthma-related mortality. *J Environ Public Health*. 2018;2018, 9389570. https://doi.org/10.1155/2018/9389570.

13. Zarmouh A. Does clinic-based education have a sustainable impact on asthma patient awareness? *Libyan J Med*. 2019;14, 1635843. https://doi.org/10.1080/19932820.2019.1635843.