HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 3 (2024)

Demographic Determinants and Disparities in Food Allergies: Dissecting the Impact on Prevalence

Penelope Kim

Glen Allen High School, USA

ABSTRACT

The rising prevalence of food allergies underscores the critical need for research into genetic and environmental influences on this worrying trend. Utilizing data from a retrospective cohort study of over a million children's health records from the Children's Hospital of Philadelphia, this study investigates the onset of food allergies across different demographics. The study employs multivariate logistic regression to model the probability of allergy presence, revealing significant correlations between allergies and factors such as age, gender, race, ethnicity, and socioeconomic status. Specifically, it finds variations in the prevalence of peanut, soy, and milk allergies among different racial groups and genders, with age being a critical factor in the likelihood of having these allergies. The findings suggest that non-Caucasian populations have a higher incidence of certain food allergies, supporting the hypothesis that racial and socioeconomic factors significantly influence allergy prevalence. The study contributes to the growing body of evidence on health disparities in food allergies, emphasizing the need for targeted interventions and equitable access to treatment. Larger, more diverse sample sizes in future research could help refine the understanding of allergies' demographic determinants.

Introduction

Individuals with food allergies have significant negative impacts to their quality of life including economic costs, social inconveniences, and health risks. Food allergy is important in healthcare and even politics with food labeling laws and public safety concerns (Warren et al. 2020). Along with this, the incidence of allergies is on the rise (Saab et al. 2022).

Allergic reactions to food ingestion can manifest in a variety of presentations throughout the body. Commonly including dermatologic, gastrointestinal and respiratory systems, these symptoms can vary from a minor inconvenience to a fatality, necessitating a prompt increase in the understanding of all aspects of allergic reactions. This food allergy response is caused by an immunological response to the proteins found in foods. These responses can be caused by both immunoglobulin or non-immunoglobulin pathways. Food allergy is differentiated from food intolerances or aversions which are unpleasant but not life threatening (Lopez et al. 2024).

Allergies are a plight that has plagued humanity for centuries, which is why it is ever so important to know more about what factors influence their prevalence. Allergic reactions can be fatal which is further reason to strive to understand them further. Despite the existence of food allergies throughout human history, they are increasing in frequency as of modern times, an alarming development that increases the need to research and understand this trend (Saab et al. 2022). Teasing out the influence of genetics versus environment is an important factor in the further understanding of allergies.

Health inequity permeates all of medicine. This holds true in the field of food allergies as well. Disparities in health are multifactorial and can only be understood when the differences are observed (National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division 2017). This paper hopes to shed light on the demographic determinants of disparities in food allergies to further assist in the development

of policies, procedures, medical advancements and proper focus of attention and resources to where the need is the greatest. The need for awareness on this issue will become even more important as allergy treatments evolve and access to this treatment becomes limited due to financial, socioeconomic and cultural influences.

Literature Review/ Research Questions and Hypothesis

According to the Center for Disease Control and Prevention, National Center for Health Statistics, over twenty five percent of children residing in the United States have a form of allergy, with 5.8 percent of all children having a food allergy. This is an increase from 3.4 percent prior to the year 2000. (Zablotsky 2021). This development of allergies begins early with as high as ten percent of infants having been reported to have food allergies in some studies (Osborne, 2011). Race has also been found to impact risk. Children of East Asian or African descent but who were born in the West have been found to be at a higher risk for food allergies than Caucasians born in the same geographical area (Loh et al. 2018). In the United States, studies have shown that Black and Asian children are at higher odds of having peanut allergy than White children (Gupta et al. 2011).

There are a variety of correlations found with peanut allergies. According to a nationwide study conducted in Canada, demographic information and the occurrence of food allergy showed distinct correlations. Among these, there was a notable increase in risk of food allergy within certain high-risk groups: children, females, Canadian-born participants, adults with post-secondary education, and those residing in smaller households (Clarke et al. 2021).

According to an article in the Journal of the American Medical Association, a large survey was taken to investigate the racial ethnic and socioeconomic differences in the United States with regards to the incidence of food allergies. From this it was noted that Asian, Black and Hispanic reported more food allergies than Whites (Jiang, 2023). Other studies have shown that women and individuals of white ethnicity were more prone to report overcoming cow's milk allergy compared to men and respondents from other racial and ethnic backgrounds (Warren 2022). Of children with soy allergy, approximately half are no longer diagnosed with their allergy by the age of seven years old. Absolute soy IgE measurements were helpful indicators if they will outgrow their allergy (Savage et al., 2010).

Using the retrospective cohort study dataset by Hill et al. (2016), this paper aims to test the following hypothesis informed by current literature: Allergies to food groups will be higher in non-caucasian populations. Based on the literature review, it is hypothesized that the incidence of some food allergies will be lower among individuals in the white racial group. Meanwhile it is also anticipated that certain food allergies are prone to decrease in severity as a child matures. This expectation is founded in the observed phenomenon of children outgrowing allergies, which is well documented in the literature (Savage et al., 2010).

Methods

The anonymised data collected by Hill et al.(2016) is publicly available on Kaggle.com. From that data the age of onset of food allergy was collected and compared against race (black, other, unknown, white, non-hispanic), gender, and Medicaid versus non-Medicaid. Data is from a retrospective cohort study of healthcare provider-diagnosed eczema, asthma, allergic rhinitis, and food allergy among children. Their data was obtained from electronic medical records at the Children's Hospital of Philadelphia from over a million children who interacted with their health care system, from this the data two retrospective cohorts were defined including 29,662 children in a closed birth cohort and 333,200 children in a cross-sectional cohort. The diagnoses were identified from the International Classification of Diseases Ninth Revision (ICD-9) diagnosis codes in the charts (Hill et al. 2016). Children with specific food allergies were identified using the allergy module of the electronic medical record. To provide query validation by chart review, the medical charts of 240 subjects were manually

reviewed to gauge how accurately ICD-9 codes reflected the presence of specific food allergies. (Hill et al. 2016)

The summary statistics of each variable from the sample is shown in Table 1 below:

Table 1. Summary Statistics for Numerical Variables

	count	mean	std	min	25%	50%	75%	max
peanut_binary	333190.0	0.025961	0.15902	0.000000	0.000000	0.000000	0.000000	1.000000
milk_binary	333190.0	0.025961	0.15902	0.000000	0.000000	0.000000	0.000000	1.000000
soy_binary	333190.0	0.025961	0.15902	0.000000	0.000000	0.000000	0.000000	1.000000
AGE_START_YEARS	333190.0	3.942203	4.64621	-4.312115	0.021903	1.765914	7.208761	17.984942

Modeling Strategy

Given the binary dependent variable of whether a child has an allergy, the probability of detecting an allergy was modeled through multivariate logistic regression. Formally,

$$P(Allergy = 1|Age, Gender, Race, Ethnicity, PayerFactor)$$
$$= F(\beta_0 + \beta_1 Age + \beta_2 Gender + \beta_3 Race + \beta_4 Ethnicity + \beta_5 PayerFactor + \varepsilon$$

where $F(x) = \frac{1}{1+e^{-x}}$, the logit model.

Gender, race, ethnicity and payer factor are modeled as categorical variables whereas age is modeled as a continuous variable.

Results

Peanut Allergy

Logit Regression Results

Dep. Variable:	peanut_binary	No. Observations:	333190
Model:	Logit	Df Residuals:	333181
Method:	MLE	Df Model:	8
Date:	Mon, 18 Mar 2024	Pseudo R-squ.:	0.02503
Time:	18:20:20	Log-Likelihood:	-39115.
converged:	True	LL-Null:	-40119.
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-2.9571	0.080	-37.039	0.000	-3.114	-2.801
C(GENDER_FACTOR)[T.Male]	0.3781	0.022	16.966	0.000	0.334	0.422
C(RACE_FACTOR)[T.Black]	-0.4487	0.051	-8.775	0.000	-0.549	-0.349
C(RACE_FACTOR)[T.Other]	-0.6183	0.140	-4.401	0.000	-0.894	-0.343
C(RACE_FACTOR)[T.Unknown]	-0.9170	0.059	-15.476	0.000	-1.033	-0.801
C(RACE_FACTOR)[T.White]	-0.7801	0.049	-16.018	0.000	-0.876	-0.685
C(ETHNICITY_FACTOR)[T.Non-Hispanic]	0.2113	0.063	3.328	0.001	0.087	0.336
C(PAYER_FACTOR)[T.P1 - Medicaid]	-0.5865	0.030	-19.336	0.000	-0.646	-0.527
AGE_START_YEARS	-0.0895	0.003	-30.322	0.000	-0.095	-0.084

Figure 1. Relationship Between Demographic Variables and Peanut Allergy Test Readings

The gender factor male is statistically significant and has a coefficient of 0.3781 this means if someone is male, they are likely to score 0.3781 higher on the log-odds of having a peanut allergy than if someone was a female. The race factor Black is statistically significant with coefficient estimates -0.4487; in other words, Black Americans are likely to score -0.4487 lower on the log-odds of having a peanut allergy compared to Asian Americans (the hold-out group). Similarly, with a statistically significant coefficient estimate of -0.6183, Other Americans are subject to a log-odds of -0.6183 lower than Asian Americans for peanut allergy. The race factor Unknown is statistically significant with coefficient estimates -0.9170 ; in other words, Unknown Americans are likely to score -0.9170 lower on the log-odds of having a peanut allergy compared to Asian Americans. The race factor White shows statistical significance, expressed with coefficient estimates of -0.7801. This suggests that compared to Asian Americans, White Americans are expected to score -0.7801 lower on the log-odds of having a peanut allergy. The payer factor of medicaid children shows statistical significance, expressed with coefficient estimates of -0.5865; this means that medicaid children are likely to score -0.5865 lower on the log odds compared to non-medicaid children. The ethnicity factor of non hispanic children shows statistical significance, expressed with coefficient estimates of 0.2113; this means that non hispanic children are likely to score 0.2113 higher on the log odds compared to hispanic children. The age factor is statistically significant with a coefficient estimate of -0.0895 this means that for every year of age, it is likely that people will score -0.0895 lower on the log-odds of having a peanut allergy. Overall, according to this model estimates, Unknown is the most impactful factor that impacts the results of a log-odds of having a peanut allergy. The age factor, meanwhile, has the smallest statistically significant impact on log-odds of having a peanut allergy. The race factor Black, has a fairly low statistically significant impact and the race factor White has a higher statistically significant impact.

Soy Allergy

Logit Regression Results

Dep. Variable:	soy_binary	No. Observations:	333190
Model:	Logit	Df Residuals:	333181
Method:	MLE	Df Model:	8
Date:	Mon, 25 Mar 2024	Pseudo R-squ.:	0.03316
Time:	17:25:09	Log-Likelihood:	-13845.
converged:	True	LL-Null:	-14320.
Covariance Type:	nonrobust	LLR p-value:	1.152e-199

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.0256	0.139	-28.896	0.000	-4.299	-3.753
C(GENDER_FACTOR)[T.Male]	0.4342	0.042	10.329	0.000	0.352	0.517
C(RACE_FACTOR)[T.Black]	-0.7459	0.097	-7.710	0.000	-0.936	-0.556
C(RACE_FACTOR)[T.Other]	-0.1578	0.209	-0.755	0.451	-0.568	0.252
C(RACE_FACTOR)[T.Unknown]	-0.9473	0.111	-8.557	0.000	-1.164	-0.730
C(RACE_FACTOR)[T.White]	-0.6048	0.089	-6.784	0.000	-0.780	-0.430
C(ETHNICITY_FACTOR)[T.Non-Hispanic]	0.0053	0.107	0.049	0.961	-0.203	0.214
C(PAYER_FACTOR)[T.P1 - Medicaid]	-0.4052	0.057	-7.099	0.000	-0.517	-0.293
AGE_START_YEARS	-0.1478	0.007	-21.975	0.000	-0.161	-0.135

Figure 2. Relationship Between Demographic Variables and Soy Allergy Test Readings

The gender factor male is statistically significant and has a coefficient of 0.4342 this means if someone is male, they are likely to score 0.4342 higher on the log-odds of having a soy allergy than if someone was a female. The race factor Black is statistically significant with coefficient estimates -0.7459 ; in other words, Black Americans are likely to score -0.7459 lower on the log-odds of having a soy allergy compared to Asian Americans. The race factor Other is not statistically significant. The race factor Unknown is statistically significant with coefficient estimates -0.9473 ; in other words, Unknown Americans are likely to score -0.9473 lower on the log-odds of having a soy allergy compared to Asian Americans. The race factor White shows statistical significance, expressed with coefficient estimates of -0.6048. This suggests that compared to Asian Americans, White Americans are expected to score -0.6048 lower on the log-odds of having a soy allergy. The payer factor of medicaid children shows statistical significance, expressed with coefficient estimates of -0.4052; this means that medicaid children are likely to score -0.4052 lower on the log odds compared to non-medicaid children. The ethnicity factor of non hispanic children does not show statistical significance.. The age factor is statistically significant with a coefficient estimate of -0.1478 this means that for every year of age, it is likely that people will score -0.1478 lower on the log-odds of having a soy allergy. Overall, according to this model estimates, Unknown is the most impactful factor that impacts the results of a log-odds of having a soy allergy. The age factor, meanwhile, has the smallest statistically significant impact on log-odds of having a soy allergy. The race factor Black, has a fairly low statistically significant impact and the race factor White has a higher statistically significant impact.

Milk Allergy

Logit Regression Results

Dep. Variable:	milk_binary	No. Observations:	333190
Model:	Logit	Df Residuals:	333181
Method:	MLE	Df Model:	8
Date:	Mon, 25 Mar 2024	Pseudo R-squ.:	0.04079
Time:	17:19:40	Log-Likelihood:	-33625.
converged:	True	LL-Null:	-35055.
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-3.0808	0.084	-36.679	0.000	-3.245	-2.916
C(GENDER_FACTOR)[T.Male]	0.3035	0.024	12.583	0.000	0.256	0.351
C(RACE_FACTOR)[T.Black]	-0.4873	0.062	-7.881	0.000	-0.608	-0.366
C(RACE_FACTOR)[T.Other]	-0.1130	0.134	-0.845	0.398	-0.375	0.149
C(RACE_FACTOR)[T.Unknown]	-0.5537	0.068	-8.165	0.000	-0.687	-0.421
C(RACE_FACTOR)[T.White]	-0.3749	0.058	-6.445	0.000	-0.489	-0.261
C(ETHNICITY_FACTOR)[T.Non-Hispanic]	0.0406	0.060	0.672	0.502	-0.078	0.159
C(PAYER_FACTOR)[T.P1 - Medicaid]	-0.2919	0.032	-9.104	0.000	-0.355	-0.229
AGE_START_YEARS	-0.1716	0.004	-41.161	0.000	-0.180	-0.163

Figure 3. Relationship Between Demographic Variables and Milk Allergy Test Readings

The gender factor male is statistically significant and has a coefficient of 0.3035 this means if someone is male, they are likely to score 0.3035 higher on the log-odds of having a milk allergy than if someone was a female. The race factor Black is statistically significant with coefficient estimates -0.4873 ; in other words, Black Americans are likely to score -0.4873 lower on the log-odds of having a milk allergy compared to Asian Americans. The race factor Other is not statistically significant. The race factor Unknown is statistically significant with coefficient estimates -0.5537 ; in other words, Unknown Americans are likely to score -0.5537 lower on the log-odds of having a milk allergy compared to Asian Americans. The race factor White shows statistical significance, expressed with coefficient estimates of -0.3749. This suggests that compared to Asian Americans, White Americans are expected to score -0.3749 lower on the log-odds of having a milk allergy. The payer factor of medicaid children shows statistical significance, expressed with coefficient estimates of -0.2919; this means that medicaid children are likely to score -0.2919 lower on the log odds compared to non-medicaid children. The ethnicity factor of non hispanic children does not show statistical significance. The age factor is statistically significant with a coefficient estimate of -0.1716 this means that for every year of age, it is likely that people will score -0.1716 lower on the log-odds of having a milk allergy. Overall, according to this model estimates, Unknown is the most impactful factor that impacts the results of a log-odds of having a milk allergy. The age factor, meanwhile, has the smallest statistically significant impact on log-odds of having a milk allergy.

Now that more treatments are becoming available, further inquiry is needed into if there are racial disparities among those receiving the treatments. From a healthcare equity perspective it is important that the populations that suffer from food allergies the most have access to treatments that are becoming available.

Conclusion

As a result of testing the relationship between things and things, there is a correlation that cannot be dismissed between, age and milk allergy severity, age and peanut allergy severity, and age and soy allergy severity. Because of this it can be interpreted that many different factors have a significant impact on milk, soy, and peanut allergy severity. Additionally, the results underscore the importance of further research to explore the underlying mechanisms driving these demographic differences, paving the way for more effective prevention and treatment approaches tailored to the specific needs of diverse communities.

Limitations

The sample size of the dataset from Hill et al. (2016) consists of 300,000 children, with only 7288 having a milk allergy, 2418 having a soy allergy, and 8652 having a peanut allergy. This limitation of sample size may confound the estimates of any causal relationship between variables based on the data; perhaps with a larger sample, more accurate estimates could be established. There could also have been some errors when collecting the data which is another limitation as it reduces the accuracy and reliability of the dataset. More relevant data that could augment the support of the incidence of allergies among different demographic measures in future cohort studies could better disentangle the potential causal relationship between persistence and incidence of allergies among children. Furthermore, there is a concern for omitted variable bias arising from variables that were not controlled for. For instance, the timing of a subject's exposure to allergenic food is known to influence the prevalence of pediatric allergies. However, since there is no data on this variable, the coefficient estimates are subject to the risk of being biased by the omission of this control variable that also influences the outcome variable in this study.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Clarke, A. E., Elliott, S. J., St Pierre, Y., Soller, L., La Vieille, S., & Ben-Shoshan, M. (2021). Demographic characteristics associated with food allergy in a Nationwide Canadian Study. Allergy, asthma, and clinical immunology : official journal of the Canadian Society of Allergy and Clinical Immunology, 17(1), 72. https://doi.org/10.1186/s13223-021-00572-z

Hill, D. A., Grundmeier, R. W., Ram, G., & Spergel, J. M. (2016). The epidemiologic characteristics of healthcare provider-diagnosed eczema, asthma, allergic rhinitis, and food allergy in children: a retrospective cohort study. BMC pediatrics, 16, 133. https://doi.org/10.1186/s12887-016-0673-z

Jackson, K. D., Howie, L. D., & Akinbami, L. J. (2013). Trends in allergic conditions among children: United States, 1997-2011. NCHS data brief, (121), 1–8.

Jiang J, Warren CM, Brewer A, Soffer G, Gupta RS. Racial, Ethnic, and Socioeconomic Differences in Food Allergies in the US. JAMA Netw Open. 2023;6(6):e2318162. doi:10.1001/jamanetworkopen.2023.18162

Loh, W., & Tang, M. L. K. (2018). The Epidemiology of Food Allergy in the Global Context. *International journal of environmental research and public health*, *15*(9), 2043. https://doi.org/10.3390/ijerph15092043

Lopez CM, Yarrarapu SNS, Mendez MD. Food Allergies. [Updated 2023 Jul 24]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK482187/

National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Committee on Community-Based Solutions to Promote Health Equity in the United States, Baciu, A., Negussie, Y., Geller, A., & Weinstein, J. N. (Eds.). (2017). Communities in Action: Pathways to Health Equity. National Academies Press (US).

Osborne N.J., Koplin J.J., Martin P.E., Gurrin L.C., Lowe A.J., Matheson M.C., Ponsonby A.L., Wake M., Tang M.L., Dharmage S.C., et al. Prevalence of challenge-proven IgE-mediated food allergy using population-based sampling and predetermined challenge criteria in infants. *J. Allergy Clin. Immunol.* 2011;**127**:668–676. doi: 10.1016/j.jaci.2011.01.039. [PubMed] [CrossRef] [Google Scholar] [Ref list]

Ruchi S. Gupta, Elizabeth E. Springston, Manoj R. Warrier, Bridget Smith, Rajesh Kumar, Jacqueline Pongracic, Jane L. Holl; The Prevalence, Severity, and Distribution of Childhood Food Allergy in the United States. Pediatrics July 2011; 128 (1): e9–e17. 10.1542/peds.2011-0204

Saab, I. N., & Jones, W. (2022). Trends in Food Allergy Research, Regulations and Patient Care. *Nutrition today*, *57*(2), 64–69. https://doi.org/10.1097/NT.0000000000000530

Savage, J. H., Kaeding, A. J., Matsui, E. C., & Wood, R. A. (2010). The natural history of soy allergy. The Journal of allergy and clinical immunology, 125(3), 683–686. https://doi.org/10.1016/j.jaci.2009.12.994

Warren, C. M., Agrawal, A., Gandhi, D., & Gupta, R. S. (2022). The US population-level burden of cow's milk allergy. The World Allergy Organization journal, 15(4), 100644. https://doi.org/10.1016/j.waojou.2022.100644

Warren, C.M., Jiang, J. & Gupta, R.S. Epidemiology and Burden of Food Allergy. Curr Allergy Asthma Rep 20, 6 (2020). https://doi.org/10.1007/s11882-020-0898-7

Zablotsky B, Black LI, Akinbami LJ. Diagnosed allergic conditions in children aged 0–17 years: United States, 2021. NCHS Data Brief, no 459. Hyattsville, MD: National Center for Health Statistics. 2023. DOI: https://dx.doi.org/10.15620/cdc:123250