

The Rise of Allergies: An Investigation on Why It Is Occurring and How to Stop It

Jin B. L. Schofield

St. Robert Catholic High School, 8101 Leslie St., Thornhill, Ontario L3T 7P4

KEYWORDS. Allergies, Allergy, Anaphylaxis, Rhinitis.

BRIEF. This paper analyzes correlations between the presence of allergies and various childhood and adult lifestyle factors of 70 participants.

ABSTRACT. The rapid increase in the development of allergies in the world population is an issue that is unaddressed and treated in a reactionary manner, in which children are given EpiPens but not raised in a way to possibly avoid the development of allergies. This paper aims to better understand what factors contribute to the development of allergies in order to allow scientists and society to take a more preventative approach towards treating allergy development by preventing development in the first place. This was done by analyzing survey responses from 70 participants from Richmond Hill, Ontario regarding what types of allergies they possess, and lifestyle factors from their childhood and present life, such as how often they played outside, where they lived, and whether their family members possessed allergies. The study made use of descriptive statistics to improve data, and inferential statistics including linear regression and logistic regression to reach conclusions. Major findings include that factors resulting in more childhood exposure to allergens decrease the chances of developing allergies and that having a family possessing allergies increases the chances of developing allergies. Next steps include recreating the experiment with a vastly larger sample size of greater geographical diversity to reach more specific prescriptive conclusions.

INTRODUCTION.

Since 1997, the amount of food allergy occurrences among children in Canada has increased by 50% [1]. This paper aims to investigate the causes of the development of allergies and generate recommendations for how one can raise children to avoid the development of allergies. It accomplishes this by using statistical analysis to test the most relevant factors in allergy development, as well as test the conventional wisdom and theories on this topic. The ultimate goal of this research is to spread awareness of the causes of allergies in order to ultimately minimize it within our population. According to current research, I predict that if I collect responses from a sample of the Richmond Hill population and analyze the data, then I will find that increased exposure of one's pregnant mother or oneself as an infant to allergens leads to a decreased likelihood of developing allergies, because people will be desensitized to the allergen and therefore not develop an allergy to it. I make use of statistical models to test this statement, among others, in this study.

In the last sixty years, various allergies have become more common in developed, as well as rapidly developing, countries among both children and adults, although children are far more likely to develop them. These allergic diseases include asthma, rhinitis, anaphylaxis, drug allergies, food allergies, insect allergies, eczema, hives, and angioedema [2]. For example, between 1970 and 1990, the number of doctor consultations in the U.K. regarding asthma rose by 400%, accompanied by a rise in hay fever occurrence as well. Additionally, in the twelve years after 1990, the occurrence of food allergies among children had risen by 50%. In several large cities in China, asthma rates rose by 500% in young children between 1990 and 2011 [3]. One theory known as the "old friends" hypothesis states that because humans in developed or rapidly developing countries spend less time

outdoors, they are exposed to less of the organisms that train the human immune system to differentiate between threats and harmless substances. This ultimately causes the body to overreact and become sensitized to otherwise harmless chemicals, such as those found in peanuts [3]. An additional well-known theory, the "hygiene hypothesis", states that the characteristics of one's family, ownership of a pet, exposure to infection as a child, and method of delivery - or the "cleanliness" of one's childhood impacts their development of allergies [4]. Environmental allergens, including those related to helminth infections and other parasitic infections, are known to cause immune responses that can contribute to allergy development [5].

Additional miscellaneous research has found that mice who were fed diets containing higher amounts of fat were more susceptible to food allergies, and that the heaviness of air pollution can contribute to the hindrance of immune development, impacting allergy development [6] [7]. Other research has also linked allergy development to the diet of one's mother, whether one was breastfed, and food processing, especially in the development of food allergies [8]. Another 2018 study showed that the prevalence of allergies in children from the ages of 7 to 9 is negatively correlated with the number of childhood pets (cats and dogs, specifically) they have at the age of 1, implying that the exposure to the allergens pets carry also reduces the likelihood of children developing allergies [9]. Additional ways research has shown help prevent the development of allergies is being born within a hospital, infant weaning, as well as feeding children slightly increasing amounts of an allergen to immunize the child [10].

Overall, the conventional wisdom about allergy causes seems to highlight the following factors: exposure to previous infections, method of birth, air pollution, ownership of pets, location of one's home, whether one was breastfed, and the diet of oneself and one's mother. These factors all pertain to the immediate lifestyle factors of one's mother and oneself during infancy, and not genetic/hereditary factors.

MATERIALS AND METHODS.

Before beginning to conduct my own research, I used the Concordia University online library, the US National Library of Medicine, as well as other miscellaneous sources to research what information had already been identified in academia in regards to the causes of allergies. A summary of what I found was written above in the introduction,

In order to gather my research, I conducted a survey involving 151 participants of all ages in the Richmond Hill area. The survey asked them about their allergies, their family's allergies, and multiple factors from their childhood that research and my own speculation proposed could contribute to allergy development (e.g. diet, exposure to allergens, country of origin, etc...). I was only able to enter 70 of my participants' responses for this model. However, my models are fairly significant, as signified by the collected t-significances. Therefore once I enter all of my participants' answers, I expect to find the same models accurate, but with possibly stronger significance, generalizability, and implications. After conducting a linear regression using the SPSS IBM program and identifying the most relevant variables, I was able to distill a response to the conventional wisdom's

model for the causation of allergy development, discover my own findings about the causation of allergies, and create a linear regression equation (using logit regression methods) that allowed me to discover the specific causes for select allergies. The use of this relevant regression equation allows the calculation of one's either increased or decreased risk of developing allergies based on responses to certain questions. All necessary diagnostics (skewness, kurtosis, Pearson's R, multicollinearity, etc...) were run before proceeding to linear regressions and logit regressions.

The adjusted R-square is a number that indicates the amount of the variance in the dependent variable that is explained by the model. This number can range from 0 (indicating 0%) to 1 (100%). The t-significance indicates the probability that, for a given independent variable, the correlation identified with the dependent variable is caused by chance. This number can range from 0 (indicating 0%) to 1 (100%), with 0.05 (5%) being the maximum acceptable range for significance. The f-significance indicates the probability that an entire model's correlation with a dependent variable is a result of chance. This number can range from 0 (indicating 0%) to 1 (100%), with 0.05 (5%) being the maximum acceptable range for significance. Finally, the B-coefficient is a number associated with each of the independent variables indicating the type of correlation it has with the dependent variable. It has no particular maximums or minimums. For example, a B-coefficient of -3 would indicate that there is a negative correlation.

RESULTS.

For the purpose of testing the conventional wisdom, my dependent variable would be "Number of Allergies", which is simply a summation of every allergy that one case/participant has.

Pre-Model Analysis.

My independent variables for this model would include the variation of their diet as an infant, which is a summation of the different common allergens that they ate (Variation1Diet), whether they were breastfed (Breast_1), the variation of their parents' diet (Variation2Diet), whether they contracted a helminth infection as a child (Helmin_1), the percentage of their food as a baby was solely organic (BabyFo_1), how many pets they had as a child (Pets_1), whether they were born via Cesarean Section (CSecti_1), how often they got sick as a child (Sick_1), how severe the air pollution in their childhood home was (AirPol_1), and how urban their childhood neighbourhood was (Urban_1).

The chance of my results of this first regression being a false positive was 70.1%, due to my F-significance being 0.701, while my adjusted R-Squared, which was -0.042, indicated that my model did not explain the variance of the dependent variable.

Table 1. This shows the significances and coefficients of different models in this study.

Variables	Model I	Model II	Model Iii	Model IV	Model V
Constant	5.581	0.956*	0.804*	0.155	-9.055
Religious	-0.065				
NoNews	1.667				
Views	-0.289				
BabyFood		-0.006	-0.005		-0.026
Pets		-0.007			
HowManyParents			0.602*		
ArOlderRelatives				0.581*	
Travel				0.065*	
Sick				-0.068*	
Peanuts					-2.441
WhenBorn					1.005
C-Section					0.756
Adj R^2	0.061	-0.003	0.100	0.242	
F-Sig	0.064	0.406	0.011	0.000	
Percent Correct					95.7

*t-sig > 0.05

Model I: All Variables Considered, Improved Data.

I removed the three most insignificant variables (those with the highest t-significance (chance of a false positive)) each time I ran a new regression until I was able to maximize the adjusted R square, minimize the F-significance, and minimize the t-significance. This resulted in Model I, which summarizes the results of my raw data. The adjusted R-square is 0.061, denoting that the variance of the independent variables does not significantly explain the variance in the dependent variable of allergy development. The F-significance of 0.064 denotes that there is a 6.4% chance of a false positive, which is above the desired threshold of 5%.

Model I: All Variables Considered, Improved Data.

I removed the three most insignificant variables (those with the highest t-significance (chance of a false positive)) each time I ran a new regression until I was able to maximize the adjusted R square, minimize the F-significance, and minimize the t-significance. This resulted in Model I, which summarizes the results of my raw data. The adjusted R-square is 0.061, denoting that the variance of the independent variables does not significantly explain the variance in the dependent variable of allergy development. The F-significance of 0.064 denotes that there is a 6.4% chance of a false positive, which is above the desired threshold of 5%.

Model I: All Variables Considered, Improved Data.

I removed the three most insignificant variables (those with the highest t-significance (chance of a false positive)) each time I ran a new regression until I was able to maximize the adjusted R square, minimize the f-significance, and minimize the t-significance. This resulted in Model I, which summarizes the results of my raw data. The adjusted R-square is 0.061, denoting that the variance of the independent variables does not significantly explain the variance in the dependent variable of allergy development. The F-significance of 0.064 denotes that there is a 6.4% chance of a false positive, which is above the desired threshold of 5%.

Model II: Testing the Conventional Wisdom.

The purpose of this model is to test the conventional wisdom. The independent variables included the most significant variables that current theory would predict explain allergy development, such as exposure to pets and how organic one's baby food is.

According to the F-significance (chance of false-positive when generalizing, 0.406), the correlations I have found between these two variables and the number of allergies is very insignificant. Therefore, my findings do not support the conventional wisdom. The adjusted R-square of -0.003 shows that the variance of the dependent variable is not explained by the independent variables denoted in the conventional wisdom.

Model III: Countermodel.

After processing all of my data, I found a model that, although questionably significant, explained more about the causes of allergies overall than the conventional wisdom.

Beginning with my countermodel for the number of overall allergies, my highest-performing variables included how many allergies one's parents had (HowManyParents) and what percentage of one's diet as a baby consisted of solely organic food (BabyFo_1). There is a 1.1% chance of a false positive - well below the threshold)

The effect that my independent variables have upon my dependent variable is weak, but stronger than the conventional wisdom, according to the adjusted R-square (0.100).

The table for Model III shows through the t-significances of the variables that there is a low chance of each variable being caused by

chance. As will be further discussed in my conclusion, the variable regarding parent allergies appears to be correlated with the number of allergies one has (can increase the likelihood of obtaining allergies). This is not seen in the conventional wisdom.

For this model, my regression equation would be:

$$\text{NumberofAllergies} = 0.602(\text{HowMany[allergies belong to your]Parents}) - 0.005(\text{BabyFo}_1 [\text{what percentage of your baby food was purely organic, not processed, not packaged, not canned, etc...}]) + 0.804$$

I found that two separate models about specific types of allergies stood out as the most significant of all my models.

Model IV - Allergies Causing Allergic Rhinitis (Pollen, dust, etc....) not Including Pet Allergies.

With my dependent variable as the number of allergies one has that cause allergic rhinitis, but are not triggered by pets, I included the following as independent variables: how many older relatives also have allergies that cause allergic rhinitis (AROlderRelatives), how often one became sick as a child (Sick_1), and how often one travelled abroad every five years (Travel_1).

The F-significance was 0.000, meaning there was no chance of the relationship being a false positive. Though still in the weak-moderate range, the highest adjusted r-square value yet was achieved with 0.242.

My model ultimately suggests that having older relatives with the same type of allergy can increase your own chance of having them, and that becoming sick as a child decreased the chance of developing these allergies, and that travelling more, which possibly denoted the socioeconomic status of participants, increased the likelihood of developing allergies.

The regression equation for this model is:

$$\text{Allergiesrhinitis [Number of Allergies Causing Allergic Rhinitis]} = 0.581(\text{AROlderRelatives}) + 0.065(\text{Travel}_1) - 0.068(\text{Sick}_1) + 0.155$$

Model V - Predicting Peanut Allergies.

For the dependent variable 'Peanut' (whether or not someone has a peanut allergy), my independent variables included whether they ate peanuts as a child (Peanuts1_1), the length of time they were in their mother's womb before being born (WhenBo_1), whether they were given birth to via a Cesarean section (CSection_1), and the percentage of their baby food that was purely organic (no packaging, etc...) (BabyFo_1). I used logistic regression to calculate a regression equation that would allow me to predict whether someone would have peanut allergies based on these variables. According to the significance, there is a 9.8% chance that the results of the linear regression are caused by chance. This is definitely not optimal, but with more cases, this number can improve.

Using the B co-efficients, my linear regression equation would be:

$$\text{Peanut (whether or not someone has a peanut allergy)} = 1/(1+e^{-(2.441(\text{Peanuts1}_1) + 1.005(\text{WhenBo}_1) + 0.756(\text{CSecti}_1) - 0.026(\text{BabyFo}_1) - 9.055)})$$

The model correctly predicted that none of 66 allergy-free people had allergies, and that 3 of 4 allergy-positive people had allergies. This results in a predictive accuracy of 95.7%.

DISCUSSION.

My findings include the following:

The conventional wisdom is flawed in that it does not properly recognize how much allergies form in clusters, consisting of families. Whether this grouping of allergies within families is caused by similar

environmental factors or actual heredity, my research cannot yet tell. However, this means that, in order to slow the growth of allergies within our population, we should begin generations in advance. Overall, I can conclude that my hypothesis was correct in that more exposure to allergens decreases the chances of developing allergies and in that having family possessing allergies increases the chance of developing allergies. However, my hypothesis was incorrect in estimating the importance of family members having allergies, as I had thought the effect would be minimal compared to the lifestyle/environmental factors.

My model suggests that whether one develops an allergy that causes allergic rhinitis is heavily dependent upon whether their parents have the allergy, how often they travel (possibly an indicator of socioeconomic status), and how often they became ill as a child.

My model suggests that peanut allergies can be predicted by whether someone has eaten peanuts as a child, the length they were in their mother's womb, how they exited their mother's womb, and how much of the baby food they consumed was purely organic.

Limitations.

I would consider my largest limitation to be my sample size. Although collecting 151 participants took me a considerable amount of time and energy, being only able to enter 70 of these responses hindered the strength of my findings. Having only 70 participants resulted in t-significances and F-significances in my data that were lower than they could have been given I had a larger sample size.

Additionally, as I collected the vast majority of my responses from Richmond Hill Centre, many participants were only able to respond to two-thirds of my responses, as they had to leave to catch a bus. I had to compensate for much of this lost information with SPSS functions that allowed me to fill in these blanks with the averages of other cases. Therefore, the richness of my data was not as high as it could have been.

I believe my greatest limitation was due to how most participants had to approximate details about their childhoods. I believe that much of the information was estimated and that this led to inaccuracy, and therefore less significance in my findings.

Lastly, none of my models identify causality and simply showcase correlations between variables, highlighting suggested areas for further research.

Further Research

I believe that additional research should be conducted in sample collection in order to strengthen my current findings, as well as to possibly find more connections between the various different indexes of allergies, and the different lifestyle and hereditary factors that cause them.

In addition, I believe that sophisticated research on the topic of how allergies are transferred genetically from parents to offspring, as well as research about at what ages sensitization to allergens usually occurs and how previous exposure to allergens impacts this, would be incredibly beneficial to creating a process within the medical system to stop the rise in occurrence of allergies. Additionally, the clustering of allergies I discovered could be investigated to pinpoint whether the cause is because of genetics, or because of the similar environmental factors and lifestyles people from the same family live in.

ACKNOWLEDGMENTS.

I owe this project to all of my participants, who donated their time so I could collect my research and make my project a success.

REFERENCES

1. "The Allergy Crisis" [Online]. Available: <https://time.com/the-allergy-crisis/> [Accessed: 29-January-2021]
2. R. Pawankar, et al., World Allergy Organization (WAO) White Book On Allergy (WAO, United Kingdom, 2011).
3. P. Sarchet, The Allergy Epidemic. *New Scientist*. **239**, 28–33 (2018).
4. G. M. Alexandre-Silva et al., The hygiene hypothesis at a glance: Early exposures, immune mechanism and novel therapies. *Acta Tropica*. **188**, 16–26 (2018).
5. M. A. Afifi, A. A. Jiman-Fatani, S. E. Saadany, M. A. Fouad, Parasites–allergy paradox: Disease mediators or therapeutic modulators. *Journal of Microscopy and Ultrastructure*. **3**, 53-61 (2015).
6. M. Hussain et al., High dietary fat intake induces a microbiota signature that promotes food allergy. *Journal of Allergy and Clinical Immunology*. **144**, 157-170 (2019).
7. Y. Vallès, M. P. Francino, Air Pollution, Early Life Microbiome, and Development. *Current Environmental Health Reports*. **5**, 512–521 (2018).
8. R. V. Neerven, H. Savelkoul, Nutrition and Allergic Diseases. *Nutrients*. **9**, 762 (2017).
9. B. Hesselmar et al., Pet-keeping in early life reduces the risk of allergy in a dose-dependent fashion. *Plos One*. **13** (2018).
10. "What's Behind The Rise in Food Allergies?" [Online] Available: <https://www.ft.com/content/03d258a9-b666-4460-98c1-fc9fd4987521> [Accessed: 29-January-2021]



Jin Schofield is a student at St. Robert Catholic High School in Thornhill, Ontario.