JOURNAL OF
EMERGING INVESTIGATORS

VOLUME 3, ISSUE 9 | SEPTEMBER 2020
emerginginvestigators.org

R-S-V-Bee
How a new program improved the geographic distribution of Scripps contestants

Electromagnetism and plants
Do our devices impact plant growth?

Asthma genetics
Evaluating a gene potentially linked to respiratory disorders

Green chemistry
Using more environmentally-friendly solvents for industrial processes

JOURNAL OF EMERGING INVESTIGATORS

The Journal of Emerging Investigators is an open-access journal that publishes original research in the biological and physical sciences that is written by middle and high school students. JEI provides students, under the guidance of a teacher or advisor, the opportunity to submit and gain feedback on original research and to publish their findings in a peer-reviewed scientific journal. Because grade-school students often lack access to formal research institutions, we expect that the work submitted by students may come from classroom-based projects, science fair projects, or other forms of mentor-supervised research.

JEI is a non-profit group run and operated by graduate students, postdoctoral fellows, and professors across the United States.

Contents

VOLUME 3, ISSUE 9 | SEPTEMBER 2020

Jingze Li, Hien Tran, and Jie Li
The Shattuck St. Mary's School, Faribault, Minnesota

JOURNAL OF
EMERGING INVESTIGATORS

Article

FCRL3 Gene Association with Asthma and Allergic Rhinitis

Shanzeh Sheikh, Emily Adams
The Walker School, Marietta, Georgia

SUMMARY

This study sought to determine if there is an association between the single nucleotide polymorphism rs7528684 of the Fc receptor-like-3 (*FCRL3*) gene and asthma or allergic rhinitis (AR). Based on previous studies in an Asian population, we hypothesized that participants with an AA genotype of *FCRL3* would be more likely to have asthma and/or allergic rhinitis. To test the hypothesis, surveys were administered to participants, and genotyping was performed on spit samples via PCR, restriction digest, and gel electrophoresis. Our results identified a statistically significant association between the GG genotype of *FCRL3* and an increased risk of asthma with comorbid AR. Additionally, we found a statistically significant association between the presence of other allergies with asthma and comorbid AR. These results suggest that having the GG genotype of *FCRL3* or having a history of other allergies predisposes one to having asthma with comorbid AR. The results of the study are important in more clearly identifying asthma and AR-related genes such as *FCRL3* and the variation in such genes across ethnicities. These findings could lead to earlier diagnosis and advancements in targeted therapies for the highly prevalent diseases of asthma and allergic rhinitis.

INTRODUCTION

Asthma is a chronic lung disease that causes inflammation of the airways, which can result in coughing, wheezing, and difficulty breathing (1). Thirteen percent of children under eighteen years of age and 13.4% of adults eighteen or older have asthma in the United States (2). These numbers translate to over 25 million people, 6 million of them children (3). In addition, 40 to 60 million Americans have allergic rhinitis (AR), a sensitivity to inhaled particles such as pet dander, pollen, or dust that can occur seasonally or throughout the year, causing sinus congestion, coughing, sneezing, nasal and ocular itching, and rhinorrhea (4). About two-thirds of asthma cases are allergy related (1). Asthma and AR can also occur simultaneously, worsening symptoms for both (5). In most cases, the exact cause of asthma and a possible cure are unknown (3). Patients must instead rely on various treatments that can provide relief from their symptoms (1). However, around 50% of children with asthma in the United States have uncontrolled asthma, meaning their symptoms can occur frequently and in intense episodes, potentially resulting in death (6). Further study into asthma and allergic rhinitis and their genetic determinants is essential for improving diagnosis, increasing awareness of symptom control options, and implementing effective, targeted treatment strategies.

Cookson and Moffatt noted that the etiology of asthma is 50% genetic and 50% environmental (7). Numerous studies have researched the genetic component of asthma, particularly genes connected to the immune system, because asthma and AR are considered to result from an overly-sensitive immune response (1). The Fc receptor-like-3 (or *FCRL3*) gene in particular has been associated with autoimmune diseases like rheumatoid arthritis, autoimmune thyroid disease, and systemic lupus erythematosus (8). This gene is believed to play an important role in the regulation of the immune system by promoting the proliferation, activation, and survival of B-cells while suppressing the differentiation of plasma cells and inhibiting the production of antibodies (9). *FCRL3* is expressed in B-cells, particularly memory B-cells, as well as T-cells and cells of the lymphoid organs and spleen (10).

Due to the relation of *FCRL3* with various autoimmune diseases, Gu et al. (2015, 2019) have researched single nucleotide polymorphisms (SNPs) in *FCRL3* in patients with asthma and/or AR, along with healthy controls in the Chinese Han population. They found a significant association between the SNP rs7528684 (-169 A/G) and AR, as well as asthma with comorbid AR, and concluded that SNPs of *FCRL3* may be related to regulating B- and T-cell proliferation and function, although the actual pathways are unclear, and it is unknown if these SNPs increase or decrease function (11, 12). However, based on a study by Swainson et al. (2010), it is possible that these SNPs may result in increased *FCRL3* expression, leading to T regulatory cell dysfunction, thus causing the development of an unregulated immune response (8).

No research on the association between asthma and/or AR and the SNP rs7528684 has been reported to date in populations outside of Asia, which is important because the relationship between SNPs in various immune system-related genes and asthma can vary greatly between ethnic groups (11). The purpose of this study was to determine if there is an association between the SNP rs7528684 of *FCRL3* and asthma and/or allergic rhinitis in a North American population. The results indicate an association between the GG genotype of *FCRL3* and an increased risk of asthma with comorbid AR, as well as an association between participants who had other

allergies and an increased risk of asthma with comorbid AR. The results of the study will aid in identifying asthma and AR-related genes such as *FCRL3* and determining the variation in such genes across ethnicities, which could lead to earlier diagnosis and advancements in targeted therapies (7).

RESULTS

There were 38 participants in the study, including 17 healthy controls who had neither asthma nor AR. Summaries of the demographic results obtained from the surveys completed by the study population as well as the demographic characteristics in relation to asthma alone, allergic rhinitis (AR) alone, and asthma with comorbid AR were compiled **(Table 1, Table 2)**. The patients were divided into three age ranges: young adult (16-26), adult (27-55), and senior adult (56+). The type of location (suburban, city, multi, or rural) that participants lived in for 7 or more years was also recorded **(Table 1)**. We collected this information in order to determine a possible correlation between where one has lived and asthma and/or AR. None of the participants were related or lived together. Two participants did not provide information on their AR diagnosis or other allergies (**Table 2,** designated as (-)). Only information about a diagnosis of asthma was known. Both of these participants did not have asthma. Their samples were not included in the genotype or risk ratio calculations.

Characteristic		Percentage Breakdown % (#)
Gender:	Male	28.95% (11)
	Female	71.05% (27)
Age:	16-26 years	60.05% (23)
	27-55 years	28.95% (11)
	56+ years	10.53% (4)
Race:	White	73.68% (28)
	Asian	15.79% (6)
	Multi-racial	7.89% (3)
	Black	2.63% (1)
Location:	Suburb	81.57% (31)
	City	7.89% (3)
	Multi	7.89% (3)
	Rural	2.63% (1)
Other Allergies:	No	73.68% (28)
	Yes	21.05% (8)
	No Answer	5.26% (2)

Other Allergies means allergies other than asthma or AR reported; *No Answer* indicates no answer given for other allergies.

Table 1: General demographic characteristics of the study population.

There was a significant association between males and asthma with comorbid AR (RR 5.2, 95% CI 1.12-24.08). Males were 5.2 times as likely to have asthma with comorbid AR compared to females. Additionally, there was a significant association between other allergies and asthma with comorbid AR (RR 7, 95% CI 1.56-31.51). Participants with other allergies were 7 times as likely to have asthma with comorbid AR compared to participants without other allergies. No other demographic such as age, race, or location was significantly associated with asthma, AR, or asthma with comorbid AR.

In the study population, we found 16 (42.11%) individuals with an AA genotype, 17 (44.74%) with an AG genotype, and 5 (13.16%) with a GG genotype. Frequency distributions of each genotype with regards to asthma, AR, asthma with comorbid AR, or neither asthma nor AR, were calculated, along with the corresponding risk ratios and confidence intervals (**Table 3, Table 4)**.

Characteristic		Asthma (#)	AR (#)	Asthma-AR (#)
Gender (male/female)		0/2	0/11	4/2
Age:	16-26 years	2	5	5
	27-55 years	0	5	1
	56+ years	0	1	0
Race:	White	0	8	5
	Asian	0	2	1
	Multi-racial	1	1	0
	Black	1	0	0
Location:	Suburb	2	10	3
	City	0	0	2
	Multi	0	1	0
	Rural	0	0	1
Other Allergies:	No	1	8	2
	Yes	1	3	4
	No Answer	0	-	0

AR, allergic rhinitis; *Other Allergies*, indicated experience of allergies other than asthma and/or AR; *No Answer* indicates no answer given for other allergies; - indicates unknown.

Table 2: Frequencies of demographic characteristics of study population in relation to asthma alone, AR alone, and asthma with comorbid AR (Asthma-AR).

Genotype	Asthma (%)	AR (%)	Asthma-AR (%)	None (%)
AA	0	4 (26.67%)	2 (13.33%)	9 (60%)
AG	2 (11.76%)	7 (41.18%)	2 (11.76%)	6 (35.29%)
GG	0	0	2 (50%)	2 (50%)

None, neither asthma or AR.

Table 3: Frequencies and percentages of genotypes of SNP rs7528684 in relation to asthma alone, AR alone, and asthma with comorbid AR (Asthma-AR).

Genotype	Asthma (RR (95%CI); p-value)	AR (RR (95%CI); p-value)	Asthma-AR (RR (95%CI); p-value)
AA	0.07 (0.00-38.48; 0.13)	0.8 (0.28-2.25; 0.33)	0.7 (0.15-3.34; 0.33)
AG	22.47 (0.04-12450; 0.07)	1.96 (0.69-5.53; 0.10)	0.56 (0.12-2.68; 0.23)
GG	0.39 (0.00-205.6; 0.38)	0.07 (0.00-32.93; 0.09)	4* (1.05-15.3; 0.03)

RR, risk ratio; *CI*, confidence interval; *, statistically significant (95% Confidence Interval (Taylor series) and p-value (chi-square test)).

Table 4: Risk Ratios for genotypes in relation to asthma alone, AR alone, and asthma with comorbid AR (Asthma-AR).

		Asthma-AR		
		Present	Not Present	Total
GG Genotype	Present	2	2	4
	Not Present	4	28	32
	Total	6	30	36

Table 5: Risk Ratio of FCRL3 GG Genotype and Asthma with comorbid AR (Asthma-AR).

Risk ratios containing counts of zero were substituted with 0.1 to allow completion of the calculation, conducted using OpenEpi, online software for epidemiological calculations (13). The GG genotype was significantly associated with an increased risk for asthma with comorbid AR (RR 4, 95% CI 1.05-15.3) **(Table 4).** The risk ratio calculation and analysis was completed using a risk ratio table **(Table 5).**

DISCUSSION

Asthma and allergic rhinitis cause significant morbidity worldwide, with etiologies linked to both genetics and the environment. The hypothesis for this study was developed based on previous research studying the association between SNPs of *FCRL3* and asthma as well as AR in a Chinese Han population. Unlike previous research, this study was based in a predominantly Caucasian population. It was hypothesized that the AA genotype would be associated with an increased risk of asthma and/or AR. However, in this North American study, the AA genotype seemed to be associated with a decreased risk for asthma, AR, and asthma with comorbid AR, although this was not statistically significant. Previous research had indicated a possible protection against asthma and/or AR with the GG genotype in the Chinese Han population, but in this United States study, the GG genotype was significantly associated with an increased risk for asthma with comorbid AR (11). People with the GG genotype were 4 times as likely to have asthma with comorbid AR than people without the GG genotype (RR 4, 95% CI 1.05-15.3). Previous studies have mentioned the possible differences across ethnicities in the expression of genes related to the immune system (7). Results from this study suggest that those differences are present in *FCRL3*. The variation in results between ethnicities may be due to the multi-factorial nature of asthma and AR. Asthma and AR involve multiple genes and proteins interacting with each other and with the environment, which can result in different consequences for various populations (7).

Another significant result was that participants with allergies other than asthma or AR were 7 times as likely to have asthma with comorbid AR as compared to participants without other allergies (RR 7, 95% CI 1.56-31.51). This result corroborates previous research that indicates AR is a complex allergic disease associated with other atopic diseases such as asthma, eczema, and food allergies (14). Additionally, males were 5.2 times as likely to have asthma with comorbid AR as compared to females (RR 5.2, 95% CI 1.12-24.08). Although this significant result may have been due to a small sample size rather than an actual correlation, a higher incidence of asthma with comorbid AR in males has been described in an Italian population of a similar age range to this study (15).

One limitation of the study is the small study sample size of 38 drawn from a primarily Caucasian population located in the southeastern United States. Research by Gu et al. (2015, 2019) was performed in a predominantly Chinese Han population with a sample size of 1140 and 506, respectively (11, 12). Based on power analysis for a one sample, dichotomous outcome, future studies with sample sizes of around 383 participants in a larger, more racially and geographically diverse population would enhance the significance of the results. Additionally, this study did not take into account different types of asthma or AR, primarily because it sought only to identify a general association between a *FCRL3* SNP and asthma and/or allergic rhinitis in a population where no such study had been reported to date. Therefore, the proportion of participants with asthma who had atopic asthma is unknown. Furthermore, the physician diagnosis of asthma and AR was self-reported by the participants, and could be subject to reporting error. Finally, the sample was restricted to people sixteen years of age and above and thus did not include a younger pediatric population.

The results of this study could be important in identifying genes related to asthma and allowing earlier diagnosis and targeted therapy. For instance, if one can identity a risk of asthma early on, physicians can prescribe immunotherapy for allergic rhinitis to prevent the development of asthma later in life (16). To establish significance of the findings, we recommend further study of *FCRL3* and asthma as well as allergic rhinitis in a larger, more diverse population, with a wider age range and physician-documented diagnoses.

MATERIALS AND METHODS

Volunteers were recruited from a suburban Atlanta school's junior and senior classes as well as teachers during school assembly. Volunteers included both those without a history or diagnosis of allergies and those with allergies, particularly asthma and/or allergic rhinitis. Participants gave their informed consent through consent forms. A 1% NaCl solution was used to collect saliva samples. During the study, the researcher was blinded to sample identification.

DNA extraction was completed by centrifuging the spit samples, mixing together 50 µL of DNA and 200 µL of 10% Chelex, and placing mixed samples in a dry bath at 95ºC for 10 minutes. *FCRL3* SNP rs7528684 (-169A/G) was analyzed using PCR-RFLP (polymerase chain reaction-restriction fragment length polymorphism). The PCR tubes were each labeled with the participant identification number. The PCR-RFLP methodology was modeled after a study by Jin et al (2015) (17). The tubes were placed in a rack over ice, with 12.5 µL GoTaq Green Master Mix (Promega), 2.5 µL of the forward primer, 2.5 µL of the reverse primer, and 7.5 µL of DNA pipetted into each tube (Forward: 5'-CCCTTCACTACCTTGTCTTCACAC-3'; Reverse: 5'-GGGTGGAACCTCTTTGATTGC-3'). The tubes were placed into the thermal cycler with the following conditions: pre-denaturation at 95ºC for 5 minutes, 95ºC for 30 seconds, annealing at 58ºC for 20 seconds, extension at 72ºC for 60 seconds (40 cycles), and a final extension at 72ºC for 10 minutes.

Restriction enzyme digestion was performed with a mix of 7.5 µL of PCR product, 36 µL nuclease-free water, 5 µL 10X Buffer Tango, and 1 µL FaqI (BsmFI) restriction enzyme (New England Biolabs). The reaction tube was then incubated for 60 minutes at 65ºC and another 20 minutes at 80ºC to deactivate the enzyme. The products were detected using 2% agarose gel electrophoresis. The finished gels were stained overnight on a rocker in 1X FastBlast stain (Bio Rad).

Two samples from each subject were genotyped to confirm results. Samples were genotyped by the number of

lines in each lane of the gel **(Figure 1)**. There were three possible genotypes: AA, GG, and AG. The AA genotype corresponded to one fragment of 296 bp; the GG genotype to two fragments of 184 bp and 112 bp; and the AG genotype to three fragments of 296 bp, 184 bp, and 112 bp. Results were used to determine the frequency distribution of genotypes and alleles for each group (asthma, AR, asthma with comorbid AR, none), which was recorded in a Google spreadsheet. Results were found by calculating risk ratios and determining confidence intervals using OpenEpi.

Received: Janurary 28, 2020
Accepted: June 6, 2020
Published: August 5, 2020

REFERENCES

1. Scherzer, Rebecca, and Mitchell H. Grayson. "Heterogeneity and the Origins of Asthma." *Annals of Allergy, Asthma & Immunology,* vol. 121, no. 4, 2018, pp. 400-405., doi:10.1016/j.anai.2018.06.009
2. "Lifetime Asthma Prevalence Percents by Age, NHIS, 2017." *Centers for Disease Control and Prevention,* www.cdc.gov/asthma/nhis/2017/table2-1.htm.
3. "Asthma Surveillance Data." *Centers for Disease Control and Prevention*, www.cdc.gov/asthma/asthmadata.htm.
4. Meltzer, Eli O. "Allergic Rhinitis: Burden of Illness, Quality of Life, Comorbidities, and Control." *Immunology and Allergy Clinics of North America,* vol. 36, no. 2, 2016, pp. 235-248., doi:10.1016/j.iac.2015.12.002
5. Ciprandi, Giorgio, and Ignazio Cirillo. "The Lower Airway Pathology of Rhinitis." *The Journal of Allergy and Clinical Immunology*, vol. 188, no. 5, 2006, pp. 1105-1109., doi:10.1016/j.jaci.2006.05.010
6. "AsthmaStats: Uncontrolled Asthma among Children, 2012–2014." *Centers for Disease Control and Prevention,* www.cdc.gov/asthma/asthma_stats/uncontrolled-asthma-children.htm.
7. Cookson, William, and Miriam Moffat. "Making Sense of Asthma Genes." *The New England Journal of Medicine*, vol. 351, no. 17, 2004, pp. 1794–1796.
8. Swainson, Louise A et al. "Expression of the Autoimmune Susceptibility Gene *FcRL3* on Human Regulatory T Cells is Associated with Dysfunction and High Levels of Programmed Cell Death-1." *Journal of Immunology* vol. 184, no. 7, 2010, pp. 3639-47., doi:10.4049/jimmunol.0903943
9. Li, Fu Jun et al. "*FCRL3* Promotes TLR9-induced B-cell Activation and Suppresses Plasma Cell Differentiation." *European Journal of Immunology* vol. 43, no. 11, 2013, pp. 2980-92., doi:10.1002/eji.201243068
10. Davis, R S et al. "Identification of a Family of Fc Receptor Homologs with Preferential B Cell Expression." *Proceedings of the National Academy of Sciences of the United States of America,* vol. 98, no. 17, 2001, pp. 9772-7., doi:10.1073/pnas.171308498
11. Gu, Zheng et al. "*FCRL3* Gene Polymorphisms Confer Autoimmunity Risk for Allergic Rhinitis in a Chinese Han Population." *PloS one* vol. 10, no. 1, 2015., doi:10.1371/journal.pone.0116419
12. Gu, Zheng et al. "Genetic Risk of FCRL3 and *FCRL5* Polymorphisms in Children with Asthma and Allergic Rhinitis in a Chines Han Population." *International Journal of Pediatric Otorhinolaryngology* vol. 120, 2019, pp. 58-63., doi:10.1016/j.ijporl.2019.02.015
13. OpenEpi. www.openepi.com/Menu/OE_Menu.htm.
14. Cingi, C et al. "Multi-Morbidities of Allergic Rhinitis in Adults: European Academy of Allergy and Clinical Immunology Task Force Report." *Clinical and Translational Allergy* vol. 7, no. 17, 2017., doi:10.1186/s13601-017-0153-z
15. Cazzola, M et al. "Asthma and Comorbid Medical Illness." *European Respiratory* Journal vol. 38, no. 1, 2011, pp. 42-49., doi:10.1183/09031936.00140310
16. Van Bever, H.P., Samuel, S.T. and Lee, B.W. "Halting the Allergic March." *World Allergy Organization* Journal vol. 1, no. 57, 2008., doi: 10.1097/WOX.0b013e31816ddbc1
17. Jin, Guo-Xi et al. "Correlation Between Single Nucleotide Polymorphism of FCRL-3 Gene and Graves' Disease in Han Population of Northern Anhui Province, China." *International Journal of Clinical and Experimental Medicine* vol. 8, no. 8, 2015, pp. 12624-30.

ACKNOWLEDGEMENTS

This research was conducted between August 2019 and December 2019 at The Walker School. Funding for this research was provided by The Walker School.

JOURNAL OF
EMERGING INVESTIGATORS

Article

Kinetic monitoring and Fourier-Transform Infrared (FTIR) spectroscopy of the green oxidation of (-)-menthol to (-)-menthone

Anvi Surapaneni[1], Atri Surapaneni[1], Jeslyn Wu[2], Ayush Bajaj[1], Katrina Reyes[3], Rohan Adwankar[4], Ananya Vittaladevuni[5], and Edward Njoo[6]

[1]The Quarry Lane School, Dublin, California
[2]Mission San Jose High School, Fremont, California
[3]American High School, Fremont, California
[4]Irvington High School, Fremont, California
[5]Dougherty Valley High School, San Ramon, California
[6]Department of Chemistry, Biochemistry & Physics, Aspiring Scholars Directed Research Program, Fremont, California

SUMMARY

Green chemistry involves reducing the production of hazardous substances in chemical processes to be more sustainable, thereby decreasing pollution and other environmental damage. Solvent is a major contributor of waste in organic synthesis, so the greenness of a chemical reaction can be improved by changing the solvent. The oxidation of (-)-menthol to (-)-menthone in different solvent systems was monitored using Fourier-transform infrared (FTIR) spectroscopy. Our study conducted this reaction in solvent systems of acetic acid and other solvents as replacements for acetonitrile: acetone, ethyl acetate, and dichloromethane. Solvent choice was determined based on the principles of green chemistry, and kinetics and yield of the reactions in different solvents were investigated. Through FTIR spectroscopy, the products of all the reactions were characterized as (-)-menthone. Since a limiting factor in hypochlorite-mediated oxidations of alcohols to carbonyls is the solubility of the hypochlorite salt, we hypothesize that the most polar solvent systems will give the fastest reaction rate and highest yields. Our hypothesis was incorrect, as the reaction time was shortest and product yield was greatest for the oxidation performed in ethyl acetate and acetic acid, which are less polar than acetonitrile. The green oxidation of (-)-menthol to (-)-menthone was optimized in this solvent system.

INTRODUCTION

Green chemistry is defined as the design of chemical operations that utilize the most economical usage of reagents by maximizing yields, exploring safer methodology to conduct chemical reactions, and reducing unnecessary byproducts and hazardous waste (1). The concept of green chemistry is important because it reduces pollution on a molecular scale and reduces the harmful effects of hazardous byproducts and reagents on people and the environment by increasing the safety and efficiency of chemical processes. Here, we investigated the oxidation of (-)-menthol to (-)-menthone in a variety of solvent systems, which were selected based on the distribution of cost, availability, and sustainability. In 2014, reactions involving the oxidation of alcohols were used in 10.4% of medicinal chemistry papers and appeared in 46% of manuscripts regarding the chemical synthesis of natural products (2). It is clear that oxidation of alcohols is an important step in the diversification and synthesis of natural products. Therefore, an efficient and green oxidation would be beneficial in the large-scale synthesis of complex chemical compounds. The initial reaction procedure developed by Nwaukwa and Keehn used a solvent system of acetonitrile and acetic acid (3). The principles of green chemistry were important to further optimize these initial conditions to conduct a greener reaction overall.

(-)-Menthol is a natural product that can be extracted from peppermint (*Mentha piperita*). The compound is a chiral crystalline white solid with a strong peppermint odor. (-)-Menthol is a biologically active compound and also has industrial uses in cosmetics and perfumes because of its minty fragrance (4). (-)-Menthol is a useful test substrate for other reactions in industry given its structural simplicity. (-)-Menthone, another biologically active compound, shares similar characteristics with (-)-menthol, most notably its slight peppermint odor (5).This molecule also has similar uses in the cosmetic and perfume industries. Menthone's natural form is a colorless, viscous liquid, and it can be obtained by oxidation of the secondary alcohol in (-)-menthol to a ketone (**Figure 1**) (6). (-)-Menthone, a reported natural product, is a useful synthetic intermediate, which can be used in a variety of other chemical reactions. It is a versatile chiral building block for the

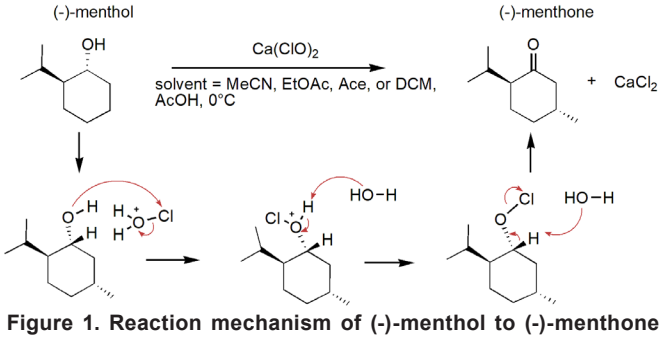

Figure 1. Reaction mechanism of (-)-menthol to (-)-menthone oxidation tested in four different solvent systems. The arrow pushing mechanism of the calcium hypochlorite oxidation is depicted.

Oxidation Reaction	Reagents	Atom Economy	Molecular Weight of discarded reagents	Hazardous byproducts
Dess-Martin (7)	Dess-Martin Periodinane	26.58%	424.14 g/mol	Iodinane
Swern (8)	Oxalyl chloride, DMSO, triethylamine	33.35%	306.25 g/mol	Carbon monoxide, dimethyl sulfide
Jones (9)	Chromium trioxide, Sulfuric acid	43.53%	198.07 g/mol	Chromium (IV) oxide
Corey-Kim (10)	N-chlorosuccinimide, Dimethyl sulfide, triethylamine	34.04%	296.85 g/mol	dimethyl sulfide
Calcium Hypochlorite (3)	Calcium hypochlorite	51.55%	142.98 g/mol	None

Table 1. Comparison of atom economy of oxidation reactions with different mild oxidizing agents.

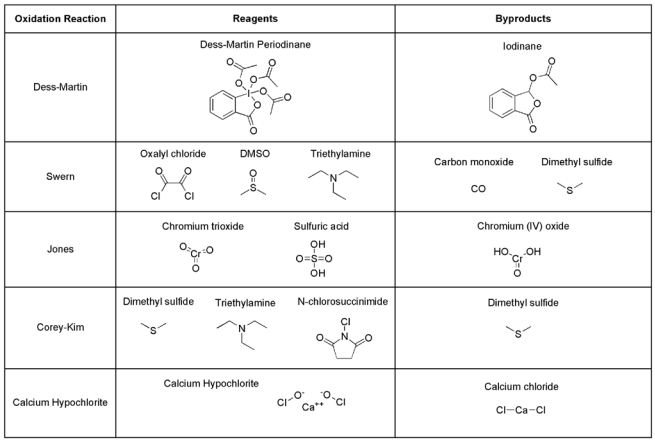

Figure 2. Chemical structures of reagents and byproducts of common oxidation reactions. The byproducts of four other common reactions to oxidize secondary alcohols to ketones were compared to a calcium hypochlorite oxidation.

synthesis of other compounds.

There are a number of reactions that can efficiently oxidize secondary alcohols into ketones; we focused on optimizing green conditions for this transformation. Calcium hypochlorite was chosen as a mild oxidizing agent because it is relatively inexpensive, and involves minimal production of hazardous byproducts. Hypochlorite salts can perform oxidations with high atom economy, as it produces only calcium chloride as a byproduct, which can be removed through an extraction rather than via column chromatography, thereby minimizing the use of solvent in the purification process (**Table 1**). Most suffer from the coproduction of other substances during an oxidation reaction, many of which are themselves highly reactive, toxic, or challenging to remove (**Figure 2**).

A limitation in hypochlorite-mediated oxidations of alcohols to carbonyls is the solubility of the hypochlorite salt, so more polar solvent systems for these reactions will result in faster reaction rates and higher yields. The oxidation of (-)-menthol to (-)-menthone was tested in a variety of solvents, chosen based on their polarity and relative greenness. All the reactions produced the desired product, (-)-menthone, confirmed by Fourier-transform infrared (FTIR) spectroscopy. The reaction performed in ethyl acetate, the least polar solvent, completed first with the greatest yield as opposed to the reaction with acetonitrile, so our hypothesis was incorrect.

RESULTS

Solvent optimization

The solvents screened for the oxidation reaction included acetonitrile, ethyl acetate, acetone, or dichloromethane, which are of different polarities (**Table 2**). We hypothesized that the most polar solvent systems will result in the fastest reaction rate and highest yield because the solubility of the hypochlorite salt, a limiting factor, will be greater. All reactions were conducted at room temperature and in an ice bath. The starting material, (-)-menthol, was first characterized through Fourier-transform infrared (FTIR) spectroscopy, with a broad O-H stretch at 3250 cm-1, representing an alcohol, and 3 medium C-H stretches from 2850 to 2950 cm^{-1} (**Figure 3**). Reasonable yields of (-)-menthone were produced from every reaction with an average yield of 75%. During synthesis, reaction progress was tracked by thin-layer chromatography (TLC) because the polarity of the alcohol in (-)-menthol is greater than the polarity of the ketone in (-)-menthone,. The silica gel plate used in TLC shows the less polar spot for the (-)-menthone traveled farther along the plate than the spot for (-)-menthol, verifying the alcohol to ketone transformation. The products were characterized by FTIR spectroscopy, which showed the lack of a broad alcohol stretch at 3250 cm^{-1} and the appearance of a ketone stretch at 1706cm^{-1} in the spectrum of the product (**Figure 4**). The other IR spectra includes all products from (-)-menthol oxidations; all products have 3 matching C-H stretches near or at 2955, 2928, and 2870cm-1 in their spectra and strong C=O peaks at 1706cm^{-1} (**Figure 3**). The 3 C-H and C=O stretches in IR spectra of all products near 3000cm^{-1} and 1700cm^{-1} match the strong C-H stretches at 2953cm^{-1}, 2926cm^{-1}, and 2869cm^{-1}, and the strong C=O stretch at 1706cm^{-1} in the IR spectrum of isolated (-)-menthone, confirming that (-)-menthone was produced (**Figure 5**). The spectra of (-)-menthone from the oxidation conducted in dichloromethane contains the same

Solvent used with acetic acid	Relative Cost (lowest - highest)	'Greenness' Score (GSK solvent guide)	Relative Polarity (least - most polar)	Production of (-)-Menthone
Acetonitrile (MeCN)	4	Some known issues	0.46	Yes
Ethyl acetate (EtOAc)	2	Few known issues	0.308	Yes
`Acetone (Ace)	3	Some known issues	0.355	Yes
Dichloromethane (DCM)	1	Major known issues	0.309	Yes (with an impurity)

Table 2. Comparison of each solvent tested in oxidation reaction of (-)-Menthol to (-)-Menthone.

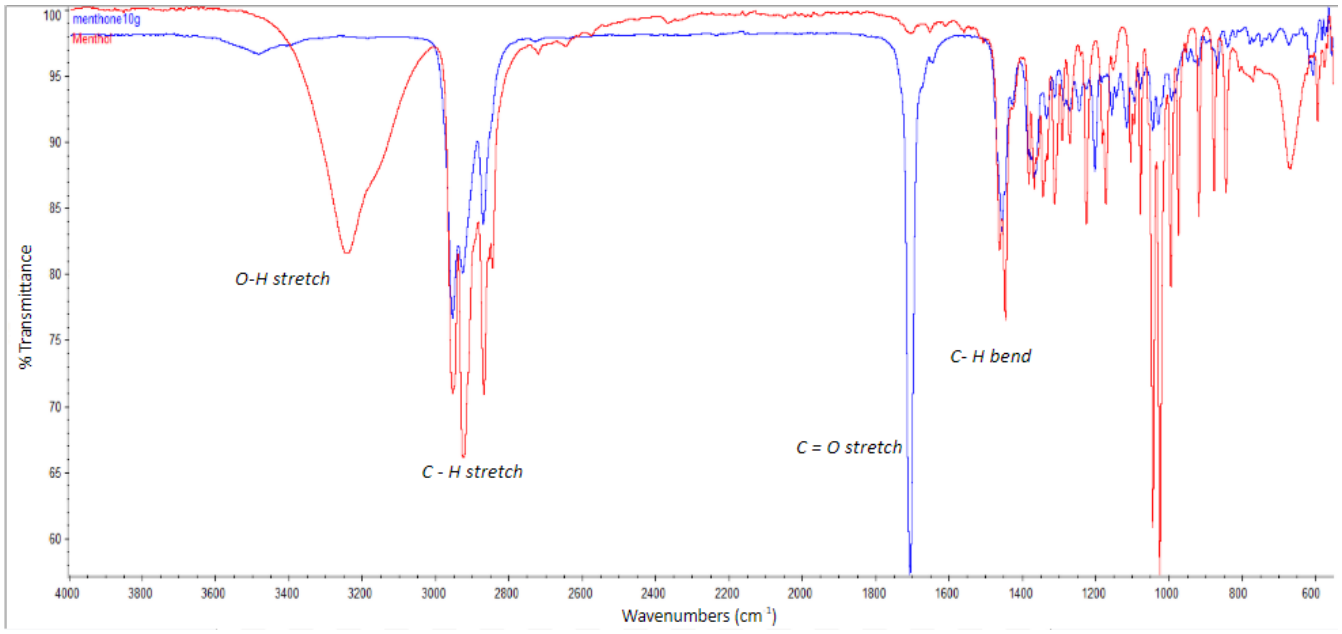

Figure 3. IR Spectra of (-)-menthol. (-)-menthol is seen in red with broad O-H stretch at 3250 cm^{-1} and 3 medium C-H stretches from 2850 to 2950 cm^{-1} and (-)-menthone (blue) with strong C=O peak at 1706cm^{-1}.

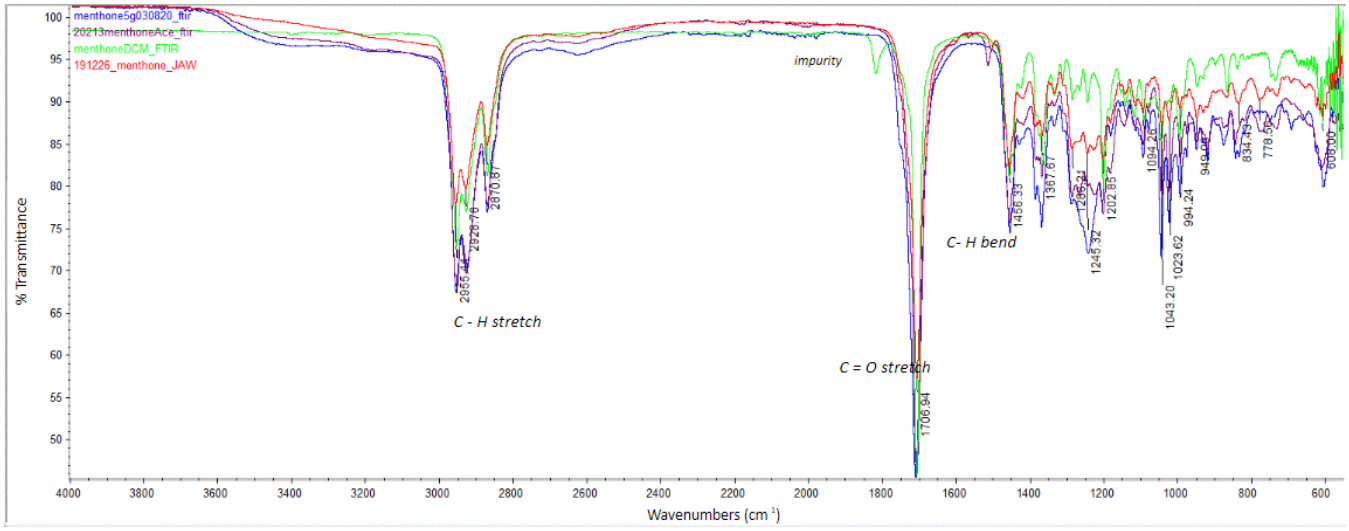

Figure 4. IR Spectra of product from all (-)-menthol oxidation. The oxidation product from reaction performed in DCM (green), (-)-menthone from oxidation in acetone (purple), (-)-menthone from oxidation in acetonitrile (red), (-)-menthone from oxidation in ethyl acetate (blue) with C-H stretches near/at 2955, 2928, and 2870cm^{-1} and strong C=O peaks at 1706cm^{-1}.

major stretches as the other products, but includes a weak peak at 1800cm^{-1} , suggesting an impurity in the product (**Figure 4**). The yields of the reactions are also reported (**Figure 6**). Because there was an impurity in the reaction in dichloromethane (DCM), yield of that reaction was not taken. All four reactions had at least two trials, once at room temperature and once in an ice bath, so each solvent was tested for percent yield twice and the kinetics experiments were done twice total. The solvents acetonitrile and ethyl acetate were tested two more times beyond the initial two trials. Overall, the greatest yield of (-)-menthone was produced in the reaction with ethyl acetate, the least polar solvent, which does not support our hypothesis.

Kinetics Study

Two kinetics studies were conducted for this reaction: at room temperature and in an ice bath. The completion of each reaction was monitored using thin-layer chromatography. At room temperature, there was no significant difference in kinetics of the reactions in different solvent systems, with all the reactions completing within minutes of each other, so this data was not recorded. The reactions in an ice bath were performed on a smaller scale than the reactions at room temperature due to a decreasing supply of the starting

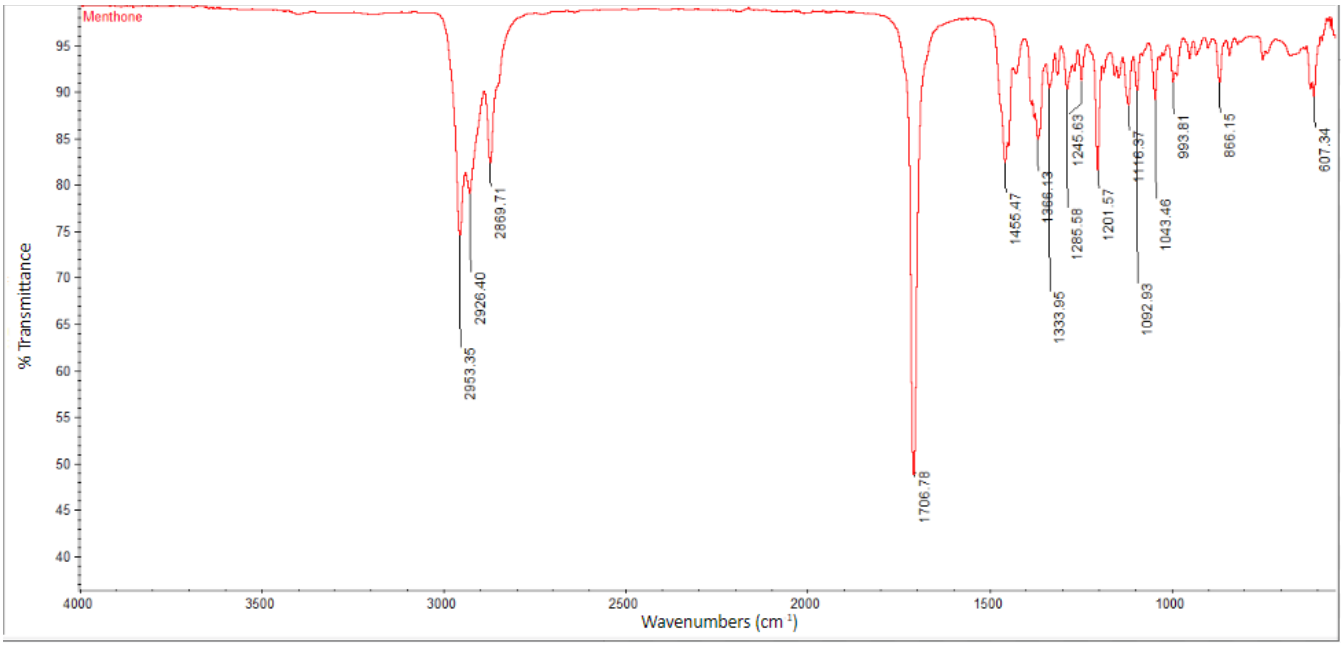

Figure 5. IR Spectrum of isolated (-)-menthone. Strong C-H stretches at 2953cm^{-1}, 2926cm^{-1}, and 2869cm^{-1} and strong C=O stretch at 1706cm^{-1} are observed.

material, so comparisons between the reaction times cannot be made. Of the reactions performed in an ice bath, the oxidation in the solvent system of ethyl acetate and acetic acid completed first in 18 minutes, followed by the reactions in dichloromethane and acetic acid, acetonitrile and acetic acid, and acetone and acetic acid, respectively (**Figure 6**). This also contradicts our initial hypothesis of the reaction completing fastest in the most polar solvent because the reaction completed fastest in ethyl acetate, the least polar of all the solvent systems. Still, from these results, it appears that this reaction works to produce (-)-menthone with a relatively green solvent system and a green oxidant.

DISCUSSION

Optimization of solvent scope & kinetics study

The oxidation of (-)-menthol to (-)-menthone converts a secondary alcohol into a ketone (**Figure 1**). The initial solvent system for this reaction reported by Nwaukwa and coworkers is acetonitrile and acetic acid (3). Solvents that were considered in place of acetonitrile included dichloromethane, ethyl acetate, and acetone (**Figure 1**). In a paper by Byrne et al., GlaxoSmithKline's, Pfizer's, and Sanofi's solvent selection guides were compared, providing a comprehensive assessment of certain solvents (12). Green solvents are evaluated based on their environmental, health, and safety (EHS) effects and energy demand. Based on the updated GSK's solvent sustainability guide and the solvent selection guides from Pfizer and Sanofi, ethyl acetate and acetone were considered more green replacements for acetonitrile (13). Furthermore, ethyl acetate and acetone cost less than acetonitrile. Although dichloromethane was not recommended as a green solvent in comparison, it was chosen in our study due to its widespread availability. While changing the solvent system, we also compared the kinetics of the reaction. Each reaction condition was then assessed on the basis of yield and reaction time. In all conditions screened, acetic acid was added as a green acid source to catalyze the reaction.

We report the synthesis of (-)-menthone from (-)-menthol in four different solvent systems, acetonitrile and acetic acid, ethyl acetate and acetic acid, acetone and acetic acid, and dichloromethane and acetic acid. This was confirmed by Fourier-transform infrared (FTIR) spectroscopy. We initially hypothesized that the more polar solvents would give higher yields due to better solubility of the hypochlorite salt. Contrary to initial expectations, the oxidation reaction conducted in ethyl acetate gave the greatest yield of (-)-menthone and completed the fastest, so this solvent system was confirmed as the most efficient. We believe the reaction progressed fastest in ethyl acetate because it can harbor a greater concentration of the active oxidizing agent, hypochlorous acid. Therefore, the oxidation of (-)-menthol to (-)-menthone using calcium hypochlorite is optimal in a solvent system of acetic acid and ethyl acetate. Although the reaction in dichloromethane did take considerably less time compared to the reactions in acetone and acetonitrile, attempts to oxidize (-)-menthol consistently contained an impurity that we were not able to characterize, but was detected by FTIR.

We did face certain experimental limitations in that only two trials were run for each experiment reaction, and the kinetics study results for the reactions at room temperature could not be fully quantified. Due to a decreasing supply of (-)-menthol, the reactions for the experiment in an ice bath were run at a smaller scale, which made it difficult to compare to the reactions at room temperature. Also, the product was

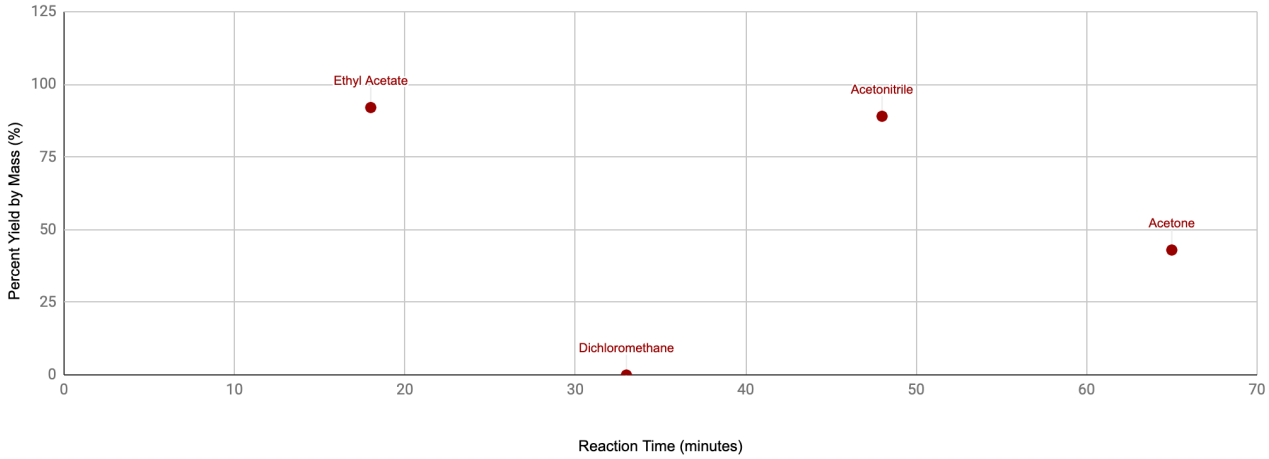

Figure 6. Percent yield by mass of each oxidation reaction product vs. time for completion. Each data point is labeled according to the solvent the reaction was completed in. (The percent yield for the reaction in DCM was not taken due to an impurity, so its data point shows as 0% yield).

initially extracted in DCM, which is not considered a "green" solvent. Instead, the solvent-based extraction could be performed with cyclohexane or better, ethyl acetate, both of which are organic solvents that are not miscible with water. This change would contribute to an even more green reaction.

Despite these setbacks, data was still collected although it did not support our hypothesis. The oxidation reaction of (-)-menthol to (-)-menthone conducted in a solvent system of ethyl acetate and acetic acid resulted in the greatest yield, with the least amount of time for the reaction to complete. The product of our reaction was characterized through FTIR to monitor the conversion of the alcohol on (-)-menthol to a ketone. The transformation of the alcohol to ketone was confirmed through IR spectra of the products, which either contained or did not contain the peak for the functional group. Considering that the natural product is the stereoisomer (-)-menthol, the product of its oxidation was (-)-menthone. To further characterize the product, we can use proton nuclear magnetic resonance spectroscopy (1H NMR) and a polarimeter to measure optical rotation, which will provide more information about the structure of the products. Further, the impurity produced in DCM may be a higher frequency carbonyl, and this could be investigated by tracking the reaction progress of the reaction mixture with FTIR and determining which compounds were formed during the reaction.

Ultimately, our research optimizes an oxidation reaction by utilizing a green solvent system, green oxidant, and column-free purification. Further, the reaction can be scaled-up easily while using greener solvents for extractions. As a more sustainable chemical reaction, this work contributes to the promotion and development of green chemistry principles in the future.

METHODS

Oxidation of (-)-menthol to (-)-menthone

In a beaker, (-)-menthol (1g, 0.0064 mol, 1 equivalence) was dissolved in a solvent system of concentrated acetic acid (5mL, 0.084 mol) and then 1 molar equivalent of acetonitrile, ethyl acetate, acetone, or dichloromethane. The dissolved solution was added dropwise to a cooled solution of calcium hypochlorite (0.09g, 0.0064 mol, 1 equivalence) and water in a round bottom flask equipped with a Teflon stir bar. The reaction was stirred in an ice bath, and reaction progress was monitored by thin-layer chromatography (TLC) in a solvent system of hexane to ethyl acetate (90:10). Since the polarity of the alcohol in (-)-menthol, the starting material, is greater than the polarity of the ketone in (-)-menthone, the assumed product, TLC can be used to track this reaction. The silica gel plate visually shows how the less polar spot for the product would have traveled farther along the plate than the spot for starting material, thus confirming the alcohol to ketone transformation. The TLCs were visualized using a potassium permanganate stain. Upon completion, water was added, and the product was extracted in four portions of DCM (4x30mL). The organic layers were washed with a saturated sodium carbonate aqueous solution, combined, dried over anhydrous magnesium sulfate and concentrated *in vacuo* to afford (-)-menthone as light-yellow colored oil with a slight peppermint odor. The product was then characterized with FTIR. IR spectra were collected on a Thermo Nicolet iS5 Fourier-transform infrared (FTIR) spectrometer equipped with an iD5 attenuated total reflectance (ATR) sample assembly. IR spectra were processed on OMICS software from Thermo Scientific. Then, after taking the mass of the final product, percent yield of the reaction was determined by dividing the mass of the product by the mass of the starting material used and then multiplying by 100. The percent yield for each reaction was calculated based on the assumption that the final product following the extractions is pure. This

assumption was made because no solvent peaks were found in the IR spectra of the products. The percent yield reported in Figure 6 was the greater yield of each reaction between the initial two trials.

Kinetics Study

The oxidation reaction described above was conducted for all solvent systems simultaneously at a 1-gram scale at room temperature. Four stir plates were arranged side by side and solutions of 1g of (-)-menthol (1g, 0.0064 mol, 1 equivalence) were dissolved into solvent systems of acetic acid and the solvents in Table 2 in the specified quantities. Each solution was added dropwise to a round bottom flask containing 1g of calcium hypochlorite (0.09g, 0.0064 mol, 1 equivalence) and 13.33 mL of water over a period of 10 minutes at room temperature. Reaction progress was monitored by TLC immediately, 15 minutes, 30 minutes, 1 hour, and 2 hours after stirring began.

The (-)-menthol oxidation reaction was also performed on a 100mg scale with (-)-menthol (0.01g, 0.00064 mol, 1 equivalence) and calcium hypochlorite (0.009g, 0.00064 mol, 1 equivalence) in an ice bath for all solvent systems. These reactions were performed in vials in an ice bath. Reaction progress was monitored by TLC immediately after the reactions were stirred, until up to 1.5 hours when the last reaction completed.

Received: April 20, 2020
Accepted: June 16, 2020
Published: August 6, 2020

REFERENCES:

1. "12 Principles of Green Chemistry." *American Chemical Society*, www.acs.org/content/acs/en/greenchemistry/principles/12-principles-of-green-chemistry.html.
2. Brown, Dean G., and Jonas Boström. "Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone?" *Journal of Medicinal Chemistry*, 26 May 2016, pp. 4443–4458., doi:10.1021/acs.jmedchem.5b01409.s001.
3. Nwaukwa, Stephen O., and Philip M. Keehn. "The Oxidation of Alcohols and Ethers Using Calcium Hypochlorite [Ca(OCl)2]." *Tetrahedron Letters*, Pergamon, 15 May 2001, www.sciencedirect.com/science/article/pii/S0040403900975257.
4. Kamatou, Guy P.p., *et al*. "Menthol: A Simple Monoterpene with Remarkable Biological Properties." *Phytochemistry*, vol. 96, 2013, pp. 15–25., doi:10.1016/j.phytochem.2013.08.005.
5. Oumzil, H, *et al*. "Antibacterial and Antifungal Activity of Essential Oils of Mentha Suaveolens." *Phytotherapy Research : PTR*, U.S. National Library of Medicine, Dec. 2002, www.ncbi.nlm.nih.gov/pubmed/12458474.
6. "Menthol." National Center for Biotechnology Information. PubChem Compound Database, U.S. National Library of Medicine, pubchem.ncbi.nlm.nih.gov/compound/Menthol.
7. Li, Jie Jack. "Dess–Martin Periodinane Oxidation." *SpringerLink*, Springer, Cham, 1 Jan. 1970, link.springer.com/chapter/10.1007/978-3-319-03979-4_86#citeas.
8. Li, Jie Jack. "Swern Oxidation." *SpringerLink*, Springer, Cham, 1 Jan. 1970, link.springer.com/chapter/10.1007/978-3-319-03979-4_270.
9. Li, Jie Jack. "Jones Oxidation." *SpringerLink*, Springer, Cham, 1 Jan. 1970, link.springer.com/chapter/10.1007/978-3-319-03979-4_143.
10. Li, Jie Jack. "Corey–Kim Oxidation." *SpringerLink*, Springer, Cham, 1 Jan. 1970, link.springer.com/chapter/10.1007/978-3-319-03979-4_72.
11. Frontier, Alison J. "Not Voodoo X.4." Reagents & Solvents, chem.chem.rochester.edu/~nvd/pages/reagents.php?page=solvent_polarity. http://chem.chem.rochester.edu/~nvd/pages/reagents.php?page=solvent_polarity
12. Byrne, Fergal P., *et al*. "Tools and Techniques for Solvent Selection: Green Solvent Selection Guides." *Sustainable Chemical Processes*, vol. 4, no. 1, 2016, doi:10.1186/s40508-016-0051-z.
13. Alder, Catherine M., *et al*. "Updating and Further Expanding GSKs Solvent Sustainability Guide." *Green Chemistry*, vol. 18, no. 13, 2016, pp. 3879–3890., doi:10.1039/c6gc00611f.

JOURNAL OF
EMERGING INVESTIGATORS

Article

Geographic Distribution of Scripps National Spelling Bee Spellers Resembles Geographic Distribution of Child Population in US States upon Implementation of the RSVBee "Wildcard" Program

Rosy E. Kannankeril[1] and Prince J. Kannankeril[2]
[1]Franklin Classical School, Franklin, TN, [2]Vanderbilt University, Nashville, TN

SUMMARY

The Scripps National Spelling Bee (SNSB) is an iconic academic competition for United States (US) schoolchildren, held annually since 1925. Historically, children qualified for the SNSB by winning a sponsored regional spelling bee. However, the sizes and geographic distributions of sponsored regions are uneven. One state may send more than twice as many spellers as another state, despite similar numbers in child population. In 2018, the SNSB introduced a wildcard program known as RSVBee, which allowed students to apply to compete as a national finalist, even if they did not win their regional spelling bee. This allowed more students to compete, but it was unknown how this would affect the geographic distribution of spellers. The purpose of our experiment was to test our hypothesis that the geographic distribution of SNSB national finalists more closely matched the child population of the US after RSVBee was implemented. We compared the number of sponsored (non-RSVBee) national finalists to the US child population from each of the 50 states, the District of Columbia, and Puerto Rico, for the years 2012–2019 and found a stable, strong correlation in each year, with correlation coefficients ranging from 0.741–0.774. With RSVBee in 2018 and 2019, we found the correlations increased significantly, with correlation coefficients ranging from 0.916–0.917 for all spellers (*p*-value = 0.032 by paired *t*-test). We conclude that the RSVBee program significantly improved the geographic distribution of the SNSB by matching the geographic distribution of SNSB finalists to the child population of the US.

INTRODUCTION

The Scripps National Spelling Bee (SNSB) is a high profile, annual academic competition primarily for United States (US) schoolchildren, run by the E.W. Scripps Company (1). To qualify for the SNSB, a speller must attend a school enrolled in the SNSB program and win their final regional spelling bee sponsored by a newspaper, a university, or another community organization. The sponsor typically pays for the speller and one chaperone to attend the week-long SNSB, which is held at the Gaylord National Resort and Convention Center in National Harbor, Maryland (2). There are, however, some geographic inequalities in the regional sponsor system, and Scripps has made recent attempts to address them. For instance, even though there are more than 200 sponsors every year, some regions of the US may not have any sponsors. Starting in 2015, spellers living in a region with no local sponsor could compete in the SNSB through the self-sponsorship program, in which the families of spellers who won their school spelling bee could pay their own way to attend the SNSB (3). Self-sponsorship did provide more spellers an opportunity to compete in the SNSB, however, from 2015—2017, only 30 spellers self-sponsored in all 3 years combined. The low number of self-sponsorships could possibly be attributed to the high cost of self-sponsorship, which, in 2016 cost $3,450 plus travel expenses. Furthermore, self-sponsoring was only available to those living in a region with no local sponsor. Another disparity with the regional sponsor system lies in the unequal size and geographic distribution of sponsored regions across the US. For instance, in 2018, the state of Massachusetts had 4 regional sponsors while the state of Maryland, which has a slightly smaller child population, had 11. Furthermore, the size of individual regions is highly variable—the median number of schools competing in a regional bee is 40, while the largest region has more than 1,000—leading to variability in the competitiveness of regional spelling bees (4). Thus, geographical factors can affect advancement to the SNSB as opposed to merit alone. In 2018, the SNSB introduced a wildcard program known as RSVBee with the intent of giving more spellers the experience of competing at the SNSB and providing more opportunities for students to improve their spelling and vocabulary. RSVBee allowed students who were either previous national finalists or winners of their school spelling bee to apply for a chance to compete as a national finalist, even if they did not win their regional spelling bee. In 2018, over 230 spellers competed through RSVBee, and the champion was an RSVBee speller (5). In 2019, the program continued with a slight increase in financial contribution from the spellers (6). Despite the increase, the number of RSVBee spellers increased to 285, once again giving more spellers the opportunity to compete at the SNSB. As illustrated, through the RSVBee program, the SNSB aimed to increase the number of spellers who could compete at the SNSB.

In this manuscript, we tested the hypothesis that the RSVBee program increased the strength of the correlation between the geographic distribution of national finalists and the child population of the United States. An increase in this correlation might suggest that RSVBee helped the 52 regions

become more appropriately represented at the SNSB. Overall, we were able to show that the RSVBee program improved the correlation between the geographic distribution of SNSB spellers and the child population of the US.

RESULTS

We obtained child population data and the number of spellers at the SNSB for the years 2012 through 2019 for each of the 50 states, the District of Columbia (DC), and Puerto Rico (PR) from publicly available sources. For each year, we compared the percentage of sponsored spellers from a state/territory with the percentage of children residing in each state/territory as a ratio of sponsored spellers to child population, with 1 being the ideal ratio. For the years during which either self-sponsorship (2015–2017) or RSVBee (2018–2019) was available, we also calculated the ratio of the total spellers to child population. We found the Pearson correlation coefficient for each year between sponsored spellers and child population. For years during which self-sponsorship or RSVBee was available, we also found the correlation between the total spellers and the child population. We used the Fisher *z*-transformation and a paired sample *t*-test to assess the improvement in correlations attributable to self-sponsorship and RSVBee. We found that the RSVBee program improved the correlation between the geographic distribution of SNSB spellers and the child population of the US.

From 2012 through 2014, there was no form of self-sponsorship or RSVBee, so we only calculated a sponsored speller to child population ratio **(Table 1)**. There was a strong positive correlation between the percentage of sponsored spellers and the percentage of child population by state in 2012 **(Figure 1)**. If spellers were distributed proportionally to the population of children in the 52 regions, the ratios would be approximately 1, and the overall correlation would approach 1. We defined underrepresented states as those with a speller to population ratio less than 0.5 and overrepresented states as those with a speller to population ratio greater than 1.5. In 2012, there were 11 underrepresented regions, 16 overrepresented regions and the Pearson correlation coefficient between sponsored speller percentage and child population percentage was 0.772. In 2013, there were 11 underrepresented regions, 15 overrepresented regions and the Pearson correlation coefficient was 0.769. In 2014, there were 11 underrepresented regions, 15 overrepresented regions and the Pearson correlation coefficient was 0.763. Thus, the geographic representation in the SNSB was relatively stable across the 50 US states, DC, and PR from 2012–2014.

From 2015 through 2017, self-sponsorship was available only for spellers with no local sponsor. We calculated a sponsored speller to child population ratio and a total speller (sponsored plus self-sponsored) to child population ratio, with 1 being the ideal ratio. In 2015, when 6 spellers self-sponsored (3 from CT, 1 each from NJ, NY, and MO), 11 regions were underrepresented and 16 regions were overrepresented. The Pearson correlation coefficient was 0.744 for sponsored spellers to child population and increased minimally to 0.746 for total speller to child population. In 2016, when 8 spellers self-sponsored (2 each from MO and NY, 1 each from NJ, CA, OR, and TN), 10 regions were underrepresented and 14 regions were overrepresented. The correlation of sponsored speller to child population was 0.754 and increased slightly to 0.763 for total speller to child population. In 2017, when 16 spellers self-sponsored (7 from FL, 2 each from NJ and NY, 1 each from CA, OR, IL, IN, and MO), 11 regions were underrepresented and 15 regions were overrepresented. The correlation between sponsored speller and child population was 0.741 and increased to 0.766 for total speller and child population **(Figure 2).** We Fisher *z*-transformed the correlations and performed a paired *t*-test comparing the correlations before and after self-sponsorship in order to test for a statistically significant effect of self-sponsorship on the correlation. The sponsored to population and total to population correlations for 2015–2017 were not statistically significantly different (*p*-value = 0.219). Thus, the geographic representation across the 50 US states, DC, and PR was relatively stable from 2015–2017, with no statistically significant increase in the correlation between spellers and child population attributable to self-sponsorship.

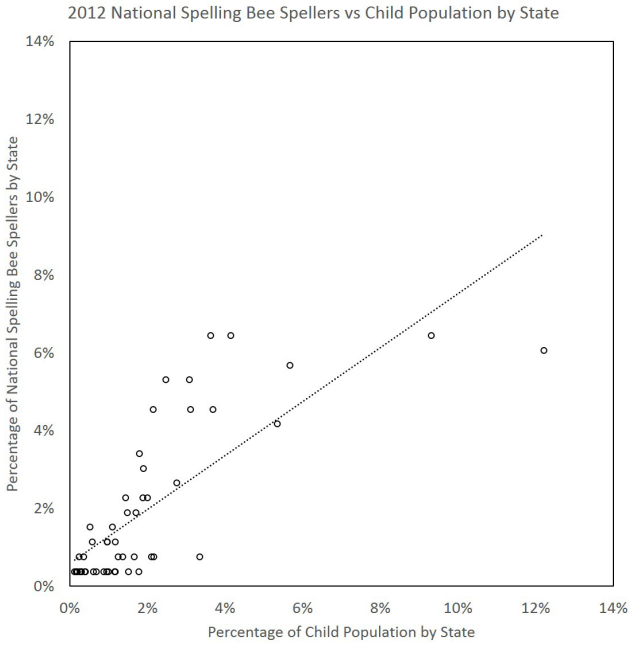

Figure 1: Scatterplot of percentage of SNSB spellers by state vs percentage of child population by state in 2012 shows a strong positive correlation (Pearson $r = 0.772$).

For both 2018 and 2019, we calculated a sponsored speller to child population ratio (before RSVBee) and a total speller to child population ratio (after RSVBee), with 1 being the ideal ratio **(Table 2)**. In 2018, there were 259 sponsored spellers and 242 RSVBee spellers. Before RSVBee, there

were 12 underrepresented regions; after RSVBee, 8 of those 12 had appropriate representation (speller to population ratio between 0.5 and 1.5). Before RSVBee, there were 16 overrepresented regions; after RSVBee, 8 of those 16 had appropriate representation. The correlation coefficient between sponsored spellers and child population was 0.747, and the correlation coefficient of total speller to child population was 0.916 **(Figure 3)**. In 2019, there were 256 sponsored spellers and 285 RSVBee spellers from the 52 regions. Before RSVBee, there were 10 underrepresented regions; after RSVBee, 7 of those 10 were appropriately represented (1 became overrepresented). There were 14 overrepresented regions; after RSVBee, 7 of those 14 regions were appropriately represented. The correlation coefficient between sponsored spellers and child population was 0.774 and increased to 0.917 for total speller to child population. We compared the Fisher *z*-transformed correlations before and after RSVBee with a paired *t*-test and observed a statistically significant difference (*p*-value = 0.032) for 2018 and 2019. Thus, we observed a large and statistically significant increase in the correlation between the geographic distribution of national finalists and the child population of the United States attributable to the RSVBee program **(Figure 4).** With the RSVBee program, more individual states had appropriate representation based on the size of their child population **(Figure 5).** Overall, these results demonstrate that RSVBee improved the correlation between the geographic distribution of national finalists and the US child population.

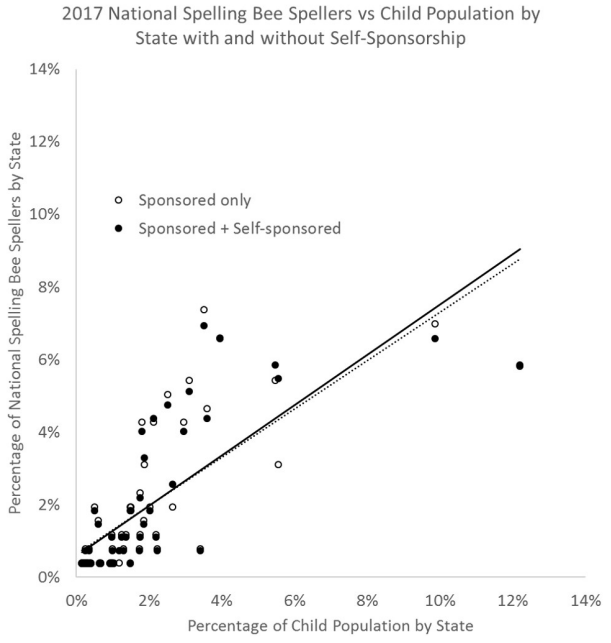

Figure 2: Scatterplot of percentage of sponsored (open circles) and total (filled black circles) SNSB spellers by state vs percentage of child population by state in 2017 shows a positive correlation (r = 0.741) which does not change significantly with self-sponsorship (r = 0.766).

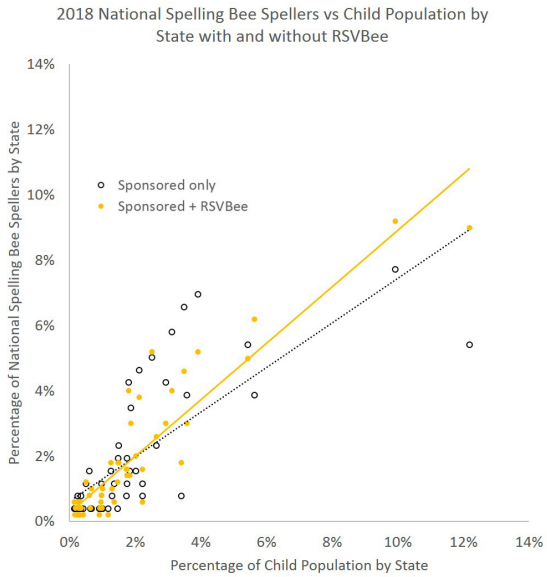

Figure 3: Scatterplot of percentage of sponsored (open circles) and total (filled yellow circles) SNSB spellers by state vs percentage of child population by state in 2018 shows a positive correlation (r = 0.747), which increases significantly with RSVBee (r = 0.916).

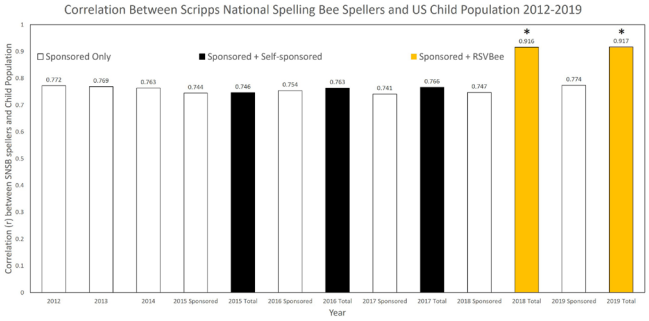

Figure 4: Correlation between geographic distribution of SNSB spellers and US child population from 2012–2019. Sponsored speller to child population correlations are in white, total speller to population correlations are in black (self-sponsor years) or yellow (RSVBee years). The asterisks indicate statistical significance (p-value < 0.05).

DISCUSSION

The main findings of our analysis were that the geographic distribution of sponsored spellers from the 50 US states, DC, and PR at the SNSB was relatively stable from 2012–2019, that self-sponsorship in 2015–2017 had minimal impact on the geographic distribution of SNSB spellers, and that the RSVBee program (2018–2019) resulted in a statistically significant change in the geographic distribution of spellers that more closely resembled the geographic distribution of the US child population. The increased strength of correlation reflected that states and regions that were underrepresented by sponsors relative to their child population generally had increased speller to population ratios after counting RSVBee spellers. Similarly, states and regions that were relatively

overrepresented had more appropriate speller to population ratios after RSVBee. The wildcard program implemented in 2018 and 2019 appears to be a robust mechanism for SNSB national finalists to reflect the geographic distribution of school-aged children in the US more accurately.

Academic contests can take one of several approaches to determine qualifiers for a national competition. Some allow a single state-wide winner to compete at the national competition. The SNSB relies on regional sponsors to support qualification to the national finals. Even though the SNSB makes no explicit claim that national finalists should be proportional to child population, this regional system is already reasonably well-matched to the US child population, as evidenced by the strong positive correlations (0.741– 0.774) we observed between the percentage of sponsored spellers and percentage of child population by state from 2012–2019. However, inequalities are still present. The two states with the largest percentage of child population are relatively underrepresented by sponsors at the SNSB **(Figure 1).** With the implementation of the RSVBee wildcard system, some of these inequities are mitigated as most of the under and overrepresented states had more appropriate representation at the SNSB.

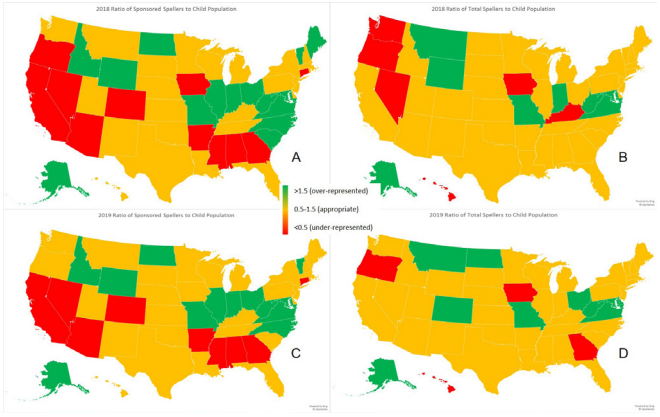

Figure 5: Maps of the United States showing 2018 sponsored (panel A), 2018 total (panel B), 2019 sponsored (Panel C) and 2019 total (Panel D) speller/child population ratios for each state in the SNSB. Ratios < 0.5 (relative underrepresentation) are colored red, ratios between 0.5-1.5 (appropriate) are yellow, and ratios > 1.5 (relative overrepresentation) are colored green.

Certain demographic factors can have large impacts on the results. States with large child populations such as California and Texas would need a very large number of spellers to reach a speller to population ratio of 1. Regions with very small populations, such as Alaska and the District of Columbia, can experience a large increase in total speller to child population ratio from sponsored speller to child population ratio because of just one or two RSVBee spellers. Regions closer to National Harbor, Maryland may be more likely to have RSVBee spellers because in 2018, RSVBee spellers paid for their own travel and accommodations and had no requirement to stay at the Gaylord National Hotel, making the cost of RSVBee much less expensive for those who live near National Harbor (7). In 2019, spellers were responsible for their own travel and accommodations and were required to stay at the Gaylord National or pay a $600 non-participation fee, somewhat offsetting the disparity in costs for spellers who live near National Harbor (6).

Our experiment had several limitations . First of all, certain factors, such as the number of enrolled schools in each region may be more important than overall child population, but such data is not publicly available. Second, we only had access to the child population data for the 50 US states, the District of Columbia, and Puerto Rico, so any SNSB spellers outside of these regions are not included in this analysis.

In conclusion, the RSVBee wildcard program positively impacted the SNSB by allowing more spellers to compete in the SNSB and significantly improving the correlation between geographic location of national finalists and child population of the United States. In 2018, 242 spellers came to the SNSB through RSVBee. In 2019, there was more awareness for RSVBee, and 285 spellers competed through RSVBee. In 2020, plans were underway to have a reduced number (about 140) of RSVBee spellers limited to 7th and 8th graders, with a $7.50 application fee and up to 18 financial aid packages (which would include the $1,500 participation fee, six-night hotel, and $1,000 for travel and meals) (8). The 2020 SNSB was canceled due to the COVID-19 pandemic (9).

MATERIALS AND METHODS

Child Population Data

We collected child population data from the 50 US states, the District of Columbia, and Puerto Rico from the United States Census Bureau, Population Division (datacenter.kidscount.org) (10). We added the data for children ages 5–11 and 12–14 to determine the number of children aged 5–14 years in each state/territory. We used the population data from the previous year for each corresponding SNSB (for example, 2011 population data for the 2012 SNSB) to estimate the number of children ages 6–15 years during the SNSB. This spans the ages of the youngest (6 years) and oldest (15 years) spellers at the SNSB to date. For example, for New Jersey (NJ) from the 2011 population data, the number of children from the age 5–11 cohort was 792,356 and the number of children from the age 12–14 cohort was 350,977. We added these numbers to get 1,143,333 children ages 5–14. We added the number of children ages 5–14 from each of the 52 regions to find the total number of children ages 5–14 in the 52 regions combined. When using the 2011 population data, this number was 41,534,282. We then calculated a percentage for each region by dividing the number of children from each region ages 5–14 by the total number of children ages 5–14. The population percentage from the 2011 population data (used for the 2012 SNSB) for NJ was 1,143,333/41,534,282 = 2.75%.

SNSB Data

We used the SNSB round results from each year to find the number of spellers from each state at the SNSB. The results from 2019 are available online (11). Results from 2012 through 2018 are available using the internet archive: Wayback Machine (12–18). For each year, we counted the number of sponsored spellers and, if present, self-sponsored spellers or RSVBee spellers from each region. The sponsored spellers could be distinguished from self-sponsored and RSVBee spellers because the name of the sponsor was listed for each speller—self-sponsored and RSVBee spellers have their school listed as their sponsor. In cases where it was not clear whether a speller was sponsored, we used open source intelligence to collect information regarding regional spelling bee winners to identify sponsored spellers. In cases where a sponsor covered more than one state or sponsored spellers from a state other than where the sponsor was based, the speller was counted as being from the state in which they attend school.

Analysis

For each year, we calculated the percentage of sponsored spellers from each region by dividing the number of sponsored spellers from each region by the total number of sponsored spellers. For example, in 2012, there were 7 sponsored spellers from NJ and 264 sponsored spellers from the 52 regions. The sponsored speller percentage for NJ was 7/264 = 2.65%. In NJ for the 2012 SNSB, the sponsored speller to population ratio was 2.65%/2.75% = 0.96. For the years 2015–2017, in addition to finding the sponsored speller to population ratio, we also added the number of self-sponsored spellers to the number of sponsored spellers to find the new total number of spellers. From there, we calculated a new percentage by dividing the number of total spellers (with any self-sponsored spellers added) from each region by the new total. We then found the total speller to population ratio. For the years 2018 and 2019, in addition to finding the sponsored speller to population ratio, we added the number of RSVBee spellers from each region to the number of sponsored spellers from each region. We calculated a new percentage of spellers from each region by dividing the new number of spellers from each region by the new total number of spellers. We then found the total speller to population ratio. For 2012–2014, we found the correlation between sponsored spellers and population. For 2015–2017, we found the correlation between sponsored spellers and population and the correlation between the total spellers (sponsored plus self-sponsored) and population. For 2018 and 2019, we found the correlation between sponsored spellers and population and the correlation between the total spellers (sponsored plus RSVBee) and population. We used a Fisher *z*-transformation on the correlation coefficients and a paired sample *t*-test to compare sponsored spellers/population to total spellers/population (2015–2017) to assess the effect of self-sponsorship and sponsored spellers/population to total spellers/population (2018–2019) to assess the effect of RSVBee. A two-tailed *p*-value of < 0.05 was considered statistically significant. Calculations and correlations were performed using Microsoft Excel 2019. Fisher *z*-transformation and paired *t*-test were performed using SPSS for Windows version 26.

Region	2011 population age 5-14 (%)	2012 sponsored spellers (%)	2012 speller/pop ratio	2013 speller/pop ratio	2014 speller/pop ratio
AL	1.51%	0.38%	0.25	0.25	0.25
AK	0.25%	0.76%	3.08	3.02	3.02
AZ	2.17%	0.76%	0.35	0.34	0.34
AR	0.96%	0.38%	0.40	0.39	0.39
CA	12.22%	6.06%	0.50	0.49	0.49
CO	1.66%	0.76%	0.46	0.44	0.44
CT	1.10%	1.52%	1.38	1.03	1.39
DE	0.27%	0.38%	1.39	1.37	1.37
DC	0.12%	0.38%	3.06	2.92	2.81
FL	5.34%	4.17%	0.78	0.76	0.76
GA	3.36%	0.76%	0.23	0.22	0.22
HI	0.40%	0.38%	0.95	0.93	0.92
ID	0.58%	1.14%	1.97	1.28	1.28
IL	4.15%	6.44%	1.55	1.54	1.55
IN	2.15%	4.55%	2.11	2.26	2.43
IA	0.97%	1.14%	1.17	1.15	1.14
KA	0.97%	1.14%	1.17	1.15	1.15
KY	1.37%	0.76%	0.55	0.82	0.55
LA	1.49%	1.89%	1.27	1.25	1.25
ME	0.36%	0.76%	2.08	2.07	2.10
MD	1.79%	3.41%	1.90	2.29	2.28
MA	1.89%	2.27%	1.20	0.99	0.80
MI	3.11%	4.55%	1.46	1.33	1.35
MN	1.71%	1.89%	1.11	1.09	1.08
MS	1.00%	0.38%	0.38	0.37	0.37
MO	1.89%	3.03%	1.60	1.58	1.59
MT	0.30%	0.38%	1.28	1.26	1.24
NE	0.61%	0.38%	0.62	0.60	0.60
NV	0.88%	0.38%	0.43	0.42	0.42
NH	0.38%	0.38%	0.99	0.99	1.01
NJ	2.75%	2.65%	0.96	0.96	0.96
NM	0.69%	0.38%	0.55	0.54	0.54
NY	5.67%	5.68%	1.00	1.06	1.00
NC	3.08%	5.30%	1.72	1.69	1.68
ND	0.20%	0.38%	1.92	1.83	1.76
OH	3.63%	6.44%	1.77	1.86	1.87
OK	1.25%	0.76%	0.61	0.59	0.59
OR	1.16%	0.38%	0.33	0.32	0.32
PA	3.69%	4.55%	1.23	1.22	1.23
PR	1.18%	0.38%	0.32	0.33	0.34
RI	0.30%	0.38%	1.28	1.28	1.29
SC	1.44%	2.27%	1.58	1.54	1.53
SD	0.27%	0.38%	1.42	1.37	1.35
TN	2.01%	2.27%	1.13	1.11	1.11
TX	9.31%	6.44%	0.69	0.71	0.70
UT	1.18%	1.14%	0.97	0.94	0.92
VT	0.17%	0.38%	2.21	2.20	2.24
VA	2.48%	5.30%	2.14	2.09	2.09
WA	2.10%	0.76%	0.36	0.53	0.70
WV	0.52%	1.52%	2.93	2.90	2.91
WI	1.79%	0.38%	0.21	0.21	0.21
WY	0.18%	0.38%	2.12	2.05	2.02

Table 1: Child population aged 5-14 years (2011), 2012 spellers, and speller to child population ratios for 2012–2014 at the Scripps National Spelling Bee (SNSB) by US region. Column 2 shows the percentage of US children aged 5–14 years living in a specific state or territory in 2011. Column 3 shows the percentage of spellers from each state or territory at the 2012 SNSB. Column 4 shows the ratio of 2012 speller percentage to the 2011 child population percentage (column 3 divided by column 2) for each state or territory for 2012. Columns 5 and 6 show the speller to population ratios for 2013 and 2014 respectively.

Region	2018 population age 5-14 (%)	2019 sponsored spellers (%)	2019 total spellers (%)	2019 sponsored speller/pop ratio	2019 total speller/pop ratio
AL	1.47%	0.39%	0.92%	0.27	0.63
AK	0.25%	0.78%	0.55%	3.18	2.26
AZ	2.24%	0.78%	1.66%	0.35	0.74
AR	0.95%	0.39%	0.92%	0.41	0.97
CA	12.16%	5.86%	8.87%	0.48	0.73
CO	1.72%	0.78%	2.77%	0.45	1.61
CT	1.00%	0.39%	0.55%	0.39	0.55
DE	0.27%	0.39%	0.55%	1.42	2.02
DC	0.16%	0.39%	0.18%	2.44	1.16
FL	5.71%	3.52%	4.62%	0.62	0.81
GA	3.41%	0.78%	1.66%	0.23	0.49
HI	0.41%	0.39%	0.18%	0.96	0.45
ID	0.61%	1.56%	0.74%	2.55	1.21
IL	3.86%	6.25%	5.18%	1.62	1.34
IN	2.12%	4.30%	2.96%	2.03	1.39
IA	0.99%	0.78%	0.37%	0.79	0.37
KA	0.96%	1.17%	1.11%	1.22	1.15
KY	1.36%	1.17%	0.74%	0.86	0.54
LA	1.47%	1.95%	1.11%	1.33	0.75
ME	0.34%	0.39%	0.18%	1.15	0.54
MD	1.81%	3.52%	3.51%	1.94	1.94
MA	1.83%	1.17%	1.11%	0.64	0.61
MI	2.91%	4.30%	2.96%	1.48	1.02
MN	1.76%	1.95%	0.92%	1.11	0.52
MS	0.97%	0.39%	0.74%	0.40	0.77
MO	1.86%	3.13%	3.33%	1.68	1.79
MT	0.31%	0.39%	0.74%	1.26	2.38
NE	0.64%	0.39%	0.55%	0.61	0.86
NV	0.94%	0.39%	0.74%	0.42	0.79
NH	0.35%	0.39%	0.18%	1.11	0.53
NJ	2.64%	1.95%	2.77%	0.74	1.05
NM	0.66%	0.39%	0.74%	0.59	1.11
NY	5.40%	6.64%	6.47%	1.23	1.20
NC	3.12%	6.25%	3.33%	2.00	1.07
ND	0.24%	0.39%	0.37%	1.65	1.56
OH	3.49%	6.64%	5.36%	1.90	1.54
OK	1.30%	0.78%	0.74%	0.60	0.57
OR	1.19%	0.78%	0.37%	0.66	0.31
PA	3.57%	3.91%	2.77%	1.09	0.78
PR	0.84%	0.39%	0.37%	0.47	0.44
RI	0.27%	0.39%	0.18%	1.42	0.67
SC	1.51%	1.95%	1.48%	1.29	0.98
SD	0.29%	0.39%	0.37%	1.33	1.26
TN	2.03%	1.56%	1.29%	0.77	0.64
TX	9.99%	8.20%	12.57%	0.82	1.26
UT	1.27%	1.56%	1.11%	1.23	0.87
VT	0.16%	0.39%	0.18%	2.49	1.18
VA	2.52%	5.08%	4.62%	2.02	1.84
WA	2.24%	1.17%	1.66%	0.52	0.74
WV	0.50%	1.17%	0.92%	2.36	1.86
WI	1.74%	1.17%	1.48%	0.68	0.85
WY	0.19%	0.39%	0.18%	2.10	1.00

Table 2: 2018 child population aged 5-14 years, 2019 SNSB sponsored and total spellers, and speller to child population ratios for 2019 at the SNSB by US region. Column 2 shows the percentage of US children aged 5–14 years living in a specific state or territory in 2018. Column 3 shows the percentage of sponsored spellers from each state or territory at the 2019 SNSB. Column 4 shows the percentage of total (sponsored plus RSVBee) spellers from each state or territory at the 2019 SNSB. Column 5 shows the ratio of 2019 sponsored speller percentage to the 2018 child population percentage (column 3 divided by column 2) for each state or territory for 2019. Column 6 shows the ratio of 2019 total speller percentage to the 2018 child population percentage (column 4 divided by column 2) for each state or territory for 2019.

Received: April 2, 2020
Accepted: July 20, 2020
Published: August 17, 2020

REFERENCES

1. Shankar, Shalini. "Coming in First: Sound and Embodiment in Spelling Bees." *Journal of Linguistic Anthropology,* vol. 26, no. 2, 2016, pp. 119–140., doi:10.1111/jola.12117.
2. Remer, Scott B. Words of Wisdom: Keys to Success in the Scripps National Spelling Bee. 3rd ed., *CreateSpace Independent Publishing Platform,* 2014, pp. 5, 13, 288.
3. Hofstra University. Adm_spelling_bee_2015.pdf. 2015, pg. 11, https://www.hofstra.edu/pdf/admission/adm_spelling_bee_2015.pdf.
4. Communication with the E.W. Scripps Company.
5. The E.W. Scripps Company. Scripps National Spelling Bee announces changes, new opportunities through RSVB. *The Denver Channel*, 13 Nov. 2018, https://www.thedenverchannel.com/news/national/scripps-national-spelling-bee-announces-changes-new-opportunities-rsvbee.
6. The E.W. Scripps Company. *RSVBee FAQ Expanded List 2019.pdf.* 2019, https://spellingbee.com/sites/default/files/inline-files/RSVBee%20FAQ%20expanded%20list%202019.pdf.
7. The E.W. Scripps Company. *RSVBee FAQ.pdf.* 2018, https://spellingbee.com/sites/default/files/inline-files/RSVBee%20FAQ.pdf.
8. The E.W. Scripps Company. 2020_RSVBee_Fact_Sheet.pdf. 2020, https://www.spellingbee.com/sites/default/files/inline-files/2020_RSVBee_Fact_Sheet.pdf.
9.The E.W. Scripps Company. *Updates Regarding COVID-19.* 21 Apr. 2020, https://spellingbee.com/updates.
10. The Annie E. Casey Foundation. "Child Population by Age Group in the United States." *Kids Count Data Center,* 2020, https://datacenter.kidscount.org/data/tables/101-child-population-by-age-group?loc=1&loct=1#detailed/2/2-53/false /37,871,870,573,869,36,868,867,133,38/62,63,64,6,4693/419,420.
11. The E.W. Scripps Company. *Summary of Round 2.* 2019, https://secure.spellingbee.com/public/results/2019/round_results/summary/2.
12. The E.W. Scripps Company. *Summary of Round 2.* 2012, https://web.archive.org/web/20150222112938/http://spellingbee.com/public/results/2012/round_results/summary/2.
13. The E.W. Scripps Company. *Summary of Round 2.* 2013, https://web.archive.org/web/20150530022411/http://spellingbee.com/public/results/2013/round_results/summary/2.
14. The E.W. Scripps Company. *Summary of Round 2.* 2014, https://web.archive.org/web/20150221092953/http://spellingbee.com/public/results/2014/round_results/summary/2.
15. The E.W. Scripps Company. *Summary of Round 2.* 2015, https://web.archive.org/web/20191216015348/http://spellingbee.com/public/results/2015/round_results/summary/2.
16. The E.W. Scripps Company. *Summary of Round 2.* 2016, https://web.archive.org/web/20191216021003/http://

spellingbee.com/public/results/2016/round_results/summary/2.
17. The E.W. Scripps Company. *Summary of Round 2.* 2017, https://web.archive.org/web/20191205071227/http://spellingbee.com/public/results/2017/round_results/summary/2.
18. The E.W. Scripps Company. *Summary of Round 2.* 2018,https://web.archive.org/web/20191216015718/http://spellingbee.com/public/results/2018/round_results/summary/2.

ACKNOWLEDGEMENTS

"We would like to thank Ann Vessey Kannankeril and Sarah Kannankeril for their continual support. Above all, we would like to thank our Lord and Savior Jesus Christ."

JOURNAL OF
EMERGING INVESTIGATORS

Article

Bacteria and Antibiotic Resistance in School Bathrooms

Isabella Ciarlet[1], Youssef Gehad[2], Vinamra Jain[3], Emma Liu[3], Eileen Ablondi[4], Michael Mazzola[4], Caroline B. Palavicino-Maggio[4] and Eddie W. Rodriguez[4]

[1]St. Peter's School, 96 Concord Ave, Cambridge, MA 021381
[2]Medford High School, 489 Winthrop Street, Medford, MA 021552
[3]Acton-Boxborough Regional School, 36 Charter Road, Acton, MA 017203
[4]Harvard Medical School, 220 Longwood Ave, Boston, MA 021154

SUMMARY

Antibiotics are a class of compounds that can be used to treat patients with bacterial infections. Although these drugs are largely effective, the widespread use of antibiotics has led to the development of antibiotic-resistant bacteria. Since school bathrooms are widely suspected to be unsanitary, we wanted to compare the total amount of bacteria with the amount of bacteria that had ampicillin or streptomycin resistance across different school bathrooms in the Boston area. We hypothesized that because people interact with the faucet, outdoor handle, and indoor handle of the bathroom, based on whether or not they have washed their hands, there would be differences in the quantity of the bacteria presented on these surfaces. Therefore, we predicted certain surfaces of the bathroom would be less sanitary than others. After plating samples with or without antibiotics, we concluded that the only swabbed surface with bacteria was the faucet. As a whole, our results demonstrate that we need to clean our bathroom better in some places more than others. Future research questions could investigate more surfaces in the bathrooms, more bathrooms across Boston, or different times of the day.

INTRODUCTION

Although humans interact with bacteria every day, some strains are pathogenic and can lead to diseases like tuberculosis or cholera. The rise of antibiotic-resistant bacteria is making it harder for doctors to prescribe effective antibiotics (1). The global emergence and spread of extensively drug-resistant bacteria exemplifies the threat that antibiotic resistance presents. Some studies indicate that antibiotic resistance found in clinical settings is closely related to mechanisms found in environmental bacteria (2). This mechanism of antibiotic resistance involves part of the global microbial population and predates the modern selective pressure of clinical human use and abuse of antibiotics (3). Interestingly, bacteria's antibiotic resistance increases healthcare costs by $20 billion in the United States alone (4).

An estimated 15.8 million students attend public schools every day, and each of these students typically use the school restroom at least once a day (5). Therefore, we would expect to find a lot of bacteria in school bathrooms because humans touch bathroom surfaces. We wanted to see how many antibiotic-resistant bacteria live in our school bathrooms. We hypothesized that surfaces with the most human contact prior to hand-washing would be most likely to have antibiotic-resistant bacteria. Each member of our group swabbed the doorknob going into the bathroom, the doorknob going out of the bathroom, and the faucet of their sink in their own school bathrooms, representing three different schools in the Boston area. We could only detect bacteria from the faucets of two schools. In fact, one of the plates not only showed a bacterial lawn, but also showed significant growth when plated on both ampicillin and streptomycin containing plates, suggesting that the bacteria are antibiotic resistant. We chose to use ampicillin because it is an antibiotic that is used to treat different types of infections, including bacterial infections. Similarly, we chose to use streptomycin because it is an antibiotic used to prevent bacterial infection and contamination. Our data hint that while school bathroom faucets should be cleaned properly to prevent the spread of antibiotic-resistant bacteria, in general, school bathroom handles are surprisingly free of bacteria.

RESULTS

Initial visual assessment of the bacterial plates grown from swabbed samples revealed differences in bacterial growth between samples collected from different schools. The plates containing samples taken from the female and male restrooms in school #1 had no bacterial colonies visible on any plates (Figure 1). Although the sample collected from the bathroom faucet (F) at school 3 generated a single bacterial colony on the no-antibiotic plate, none of the other samples taken from this school generated observable bacterial growth (Figure 1-3). In contrast, the samples collected from the bathroom faucet at school 2 produced bacterial lawns that were green and had an odor not only when grown without antibiotics, but also when grown in the presence of ampicillin or streptomycin (Figure 1). Chemicals released from dying bacteria were probably the source of odorant (6). We confirmed that bacteria known not to be resistant to ampicillin and streptomycin would not grow on plates with either antibiotic (Figure 2) and that bacteria transformed with a plasmid conferring antibiotic

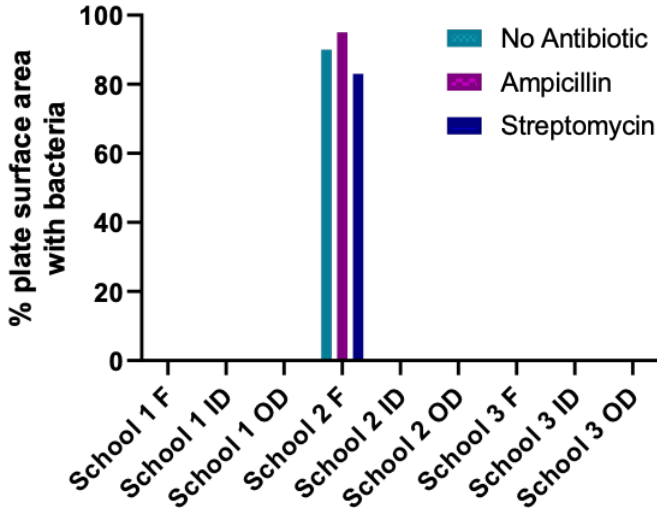

Figure 1: Bacterial growth. Bacterial swabs were collected from three schools around Boston (school 1, 2 & 3), plated without antibiotic, 20 µg/mL ampicillin, or 10 µg/mL streptomycin, and assessed for the percentage of the plate covered in bacterial growth after 24 h incubation.

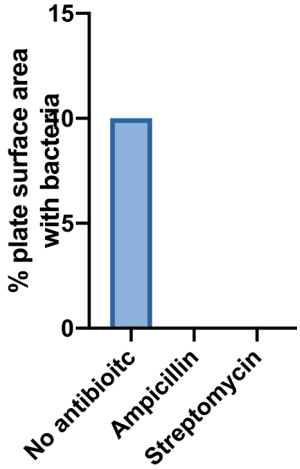

Figure 2: E. coli growth. E. coli with no known resistance to antibiotics were grown on plates without antibiotic, 20 µg/mL ampicillin, or 10 µg/mL streptomycin, and the percentage of the plate covered in bacterial growth was calculated after 24 h incubation.

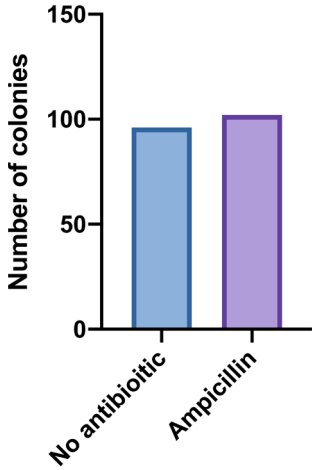

Figure 3: E. coli growth with ampicillin resistance. E. coli transformed with a plasmid known to confer antibiotic resistance to ampicillin were grown on plates without antibiotic or 20 µg/mL ampicillin, and colonies were counted after 24 h incubation.

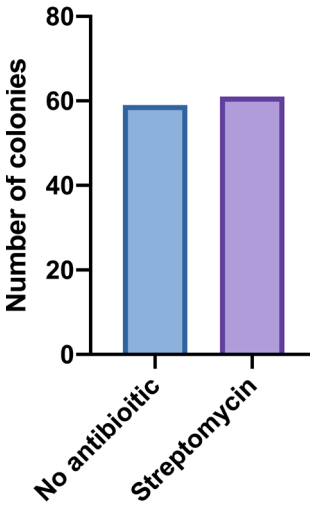

Figure 4: E. coli growth with streptomycin resistance. E. coli transformed with a plasmid known to confer antibiotic resistance to streptomycin were grown on plates without antibiotic or 10 µg/mL streptomycin, and colonies were counted after 24 h incubation.

resistance to ampicillin and streptomycin would grow on both no antibiotic and antibiotic-containing plates.

DISCUSSION

In the experiment, we found that only 1 school had major growth that was in the faucet. Additionally, another school had little growth in the faucet. The remaining school and the rest of the surfaces for the two previously mentioned schools were clean.

Our experiment found that there was no bacterial growth on the door knobs on either side of bathroom doors, regardless of if samples were plated on soft agar with streptomycin, ampicillin, or no antibiotics. For three of the four samples of faucets, there was also little to no growth. However, there were antibiotic resistant bacteria isolated from the faucet of one bathroom.

Essentially, our data suggested that, despite our initial hypothesis that significant differences would be found between F, OD, and ID samples, the most dramatic differences were actually observed between samples collected from different schools on faucet samples alone. These results indicate that faucets may be a place that can harbor lots of bacteria. Moreover, a considerable amount of the bacteria on bathroom faucets may be antibiotic resistant, as suggested by their ability to grow on plates with streptomycin and ampicillin. These data highlight the importance of cleaning bathroom faucets in schools.

Future experiments should control for the time of sample collection. It is possible that bacterial growth is dependent on the time the bathroom was last cleaned. Another improvement that could be made to our experiment would be to incubate the swabs as soon as samples were collected to ensure that no bacteria died while the samples were kept at room temperature. Future research could use the same experimental method to sample other public places. This work could reveal if dangerous bacterial species inhabit public areas. Ultimately, we have identified that bacteria resistant to streptomycin and ampicillin exist in school bathrooms, suggesting the prevalence of these strains in widely used environments.

MATERIALS AND METHODS

Sample Collection

Each experimenter was provided with three swabs and three sample collection tubes to swab each of the three locations. First, we swabbed using a sterile cotton swab three times. Following collection, we placed the swab into a sterile

Eppendorf tube containing 500 uL phosphate-buffered saline (PBS) for storage. After collecting all samples, we refrigerated the tubes at 2-5ºC for storage until platting. Two of our sample sets came from female student bathrooms, and one from male student bathroom.

Plating Samples

Tryptic soy agar plates were prepared and supplemented with 20 µg/mL ampicillin, 10 µg/mL streptomycin, or no antibiotic. Each group member swabbed the inside door knobs of their school restroom (ID), the outside of the doorknob of their school restrooms (OD), and the faucet of their school restroom (F). Samples from these swabs were suspended in PBS as above, spread on these plates, and incubated for 16 h at 37ºC. Throughout our experiment, we followed sterile procedures when handling samples and limiting our samples' exposure to air to avoid contamination. Escherichia coli transformed with a plasmid known to confer resistance to ampicillin and streptomycin were plated on ampicillin and streptomycin plates, respectively, to confirm antibiotic resistant bacteria could grow on the plates (positive control). Additionally, a plate without antibiotic and without sample was grown and shown to not be contaminated (negative control).

Plate Analysis

We did not count the individual colonies because some plates had lawns; therefore, we calculated the ratio of the surface area of the bacterial growth and compared it to the total surface area of the plate. This ratio was calculated by drawing bacterial growth on graph paper and calculating the area of the plate covered by bacteria

REFERENCES

1. WHO. "Antibiotic Resistance." World Health Organization, World Health Organization, 5 Feb. 2018, www.who.int/news-room/fact-sheets/detail/antibiotic-resistance.
2. Vincenti, Sara, et al. "Non-Fermentative Gram-Negative Bacteria in Hospital Tap Water and Water Used for Haemodialysis and Bronchoscope Flushing: Prevalence and Distribution of Antibiotic Resistant Strains." Science of The Total Environment, vol. 499, 2014, pp. 47–54., doi:10.1016/j.scitotenv.2014.08.041
3. Harnisz, Monika, et al. "Tet Genes as Indicators of Changes in the Water Environment: Relationships between Culture-Dependent and Culture-Independent Approaches." Science of The Total Environment, vol. 505, 2014, pp. 704-711, doi: 10.1016/j.scitotenv.2014.10.048
4. Dadgostar, Porooshat. "Antimicrobial Resistance: Implications and Costs: IDR." Infection and Drug Resistance, vol.12, 2019, pp 3903–3910, doi: 10.2147/IDR.S234610
5. NPWF. "Healthy Kids, Healthy Schools: The Case for a National Paid Sick Days Standard." National Partnership for Women & Families. Sept. 2019, https://www.nationalpartnership.org/our-work/resources/economic-justice/paid-sick-days/healthy-kids-healthy-schools-the-case-for-paid-sick-days.pdf.
6. Wolfe, Benjamin E. "Why Does Rotting Food Smell Bad?" Tufts Now, 15 Dec. 2014, now.tufts.edu/articles/why-does-rotting-food-smell-bad.

Article submitted: January 20, 2020
Article accepted: March 23, 2020
Article published: June 28, 2020

JOURNAL OF
EMERGING INVESTIGATORS

Article

Electromagnetic Radiation From Electronics Does Affect Plant Growth

Akshya Mahadevan[1] & Mr. Gregory Young[1]
[1] Roberto Clemente Middle School, Germantown, Maryland

SUMMARY

Global warming is becoming an increasingly prominent issue. It is commonly disputed whether electronics and the radiation they emit are harmful towards the growth of organisms. As this cannot be tested on a human, we chose plants which are a key element of Earth. To be specific, this research was conducted on basil plants. The purpose of this experiment is to find out if electromagnetic (EM) radiation from electronics affects plant growth. After experimentation, it can be concluded that EM radiation did affect basil plant growth. The plants exposed to EM radiation were taller, thinner, and a much lighter shade of green compared to the unexposed plants. With the help of this experiment, the effects of EM radiation on the growth of organisms can be seen and studied. Further study is necessary to determine if the EM radiation affected the plants on a molecular level, as well as understand how they adapt or if they die. This research can be used to determine the potential detrimental effects that electronics can have on living organisms and the environment to eventually help create a solution to reduce or eliminate this harm.

INTRODUCTION

Much of the ecosystem relies on plant growth. Plants are the main producers of oxygen and organic compounds. Ensuring the health of these organisms is vital. Recent technologies could be damaging them, which is what this research intends to find out. Cell phones emit microwave (type of light) radiation. Cell phones transmit about one to two watts of power in the range of 824 - 1780 MHz. Computers emit infrared radiation in the range of 410.0048 - 900.0074 THz due to processes such as the device becoming heated, the Bluetooth antenna, and the display (1). Multiple experiments have been conducted, using lentils to soy beans to chili plants, to be able conclude EM radiation does affect plant growth with varying results, mostly negative. Due to mixed results from previous studies, scientists still debate much on whether EM radiation affects plant growth or not.

An experiment conducted with fenugreek seed and pea plants to see which one grew faster and healthier consisted of exposing multiple plants of one species to different amounts of EM radiation. This study concluded that mobile phone radiation does cause changes in the morphology and biochemistry of the plants since the plant height, germination, seedling vigor, root length, and biomass % decreased (2). Scientists conducted a similar experiment with lentils. There were several discrepancies in the EM radiation-exposed plants in comparison to the unexposed plants. For example, only 10.35% of the seeds germinated, while the control group had a germination rate of 22.85%. They also observed many mitotic abnormalities in the plants exposed to EM radiation (3). In another study, researchers analyzed plants near the remains of a nuclear reactor and found that the soybeans had adapted to the new conditions. To retest this, they planted soybeans at the 5 km point and the 100 km point from the remains of the reactor. There was 163 times less cesium-137 in that area. The results of this experiment deduced the soybeans next to the nuclear site had adapted to their environment. The soybeans weighed half as much and took up water much slower compared to the soybeans exposed to little to no radiation. On the molecular level, they were even stranger. The soybeans exposed to high levels of radiation had three times more cysteine synthase, which protects the plant by binding heavy metals. They also had 32% more betaine aldehyde dehydrogenase, a compound found in human blood when exposed to radiation that reduces chromosomal abnormalities. They also had fluctuating amounts of seed storage proteins, a source of nitrogen for germinating seeds, compared to a regular soy plant (4). An experiment in Kolkata, India, tested 50 chili plants with exposure to EM radiation from the seed stage to fully grown. They exposed the plants to over 71,954 minutes of radiation over the course of 50 days. This experiment had highly mixed results as the plants and leaves grew taller and longer, but were duller in color and wrinkly (5). All the results corroborated that EM radiation negatively affects plant growth. Blocking EM radiation exposure to the best extent possible is necessary for maintaining normal plant growth, as these studies have demonstrated diverse effects that negatively impact the plant life cycle.

A few plants can absorb EM waves such as the cactus, betel leaf, spider plant, sansevieria, and the stone lotus flower. NASA has proved that the cactus can absorb radioactive waves with extensive research. Betel leaf extract has proven to protect DNA from radiation by preventing radiation-induced DNA strand breaks in a concentration-dependent matter (6). The spider plant, also known as Chlorophytum comosum, can absorb a large number of air pollutants including formic acid and aldehyde. This same concept of absorption also applies to EM radiation (7). Sansevieria can harmlessly absorb 100

different types of poison and is an anti-pollutant (6). It is an antidote to certain forms of radiation. Lastly, the stone lotus flower, though small, can absorb radiation effectively (7).

This project aims to determine whether nearby sources of EM radiation affects household plants. We can also use this research to identify if the devices used by agricultural workers, assuming those devices put out EM radiation, affect crop yields, which could potentially waste money, time, and resources. Based on previous findings already mentioned, I hypothesize that if a plant is exposed to the EM radiation of many electronics, the plant will have smaller leaves, slower growth and have a weaker structure because the EM radiation will cause the plant to adapt or die. The results of the experiment supported the hypothesis. The EM-exposed plants were visibly weaker. They had thinner stems and smaller leaves. Not only that, but they were also a much lighter shade of green, though they did grow taller at a much quicker rate.

RESULTS

This study examined the effects of EM radiation on plant growth by placing a cup near electronics and a cup away from all electronics. Two identical cups were set up in each area and plant growth was measured daily for a period of 42 days. This experiment was conducted as a triplicate. The physical characteristics observed in the EM plants were weak stems, brittle leaves, and an extremely light shade of green. Plant height was significantly higher but was maintained for a longer period in the control plants compared to the EM plants for all studies (Fig. 1, Fig. 2). More germination occurred in the EM plants compared to the control plants for all studies (Fig. 3).

There was an increase in height in all of the EM plants compared to the control plants. However, as time went by, the EM plants died while the control plant heights started wilting but later started becoming taller again. Control Plant 3 never grew; none of its eight seeds germinated. Due to the death of plants earlier in the study, the data is skewed, since about 66% of the plants did not live long. Overall, the data suggests that EM radiation affects plant growth factors such as height and length of life of the plant (Fig. 2). The EM plants were taller, germinated first, and grew faster than the control plants, but did not live long. On the other hand, the control plants germinated later, grew slower, and lived much longer. The EM plants have an average maximum height of 0.958 inches while the control plants have an average maximum height of 0.6875 inches. The control plants were approximately 0.2705 cm taller than EM plants overall and taller by 0.1149 inches on a daily average. The results of the one-way ANOVAs conducted for each EM and control set shows that there is a significant statistical difference (EM 1 and control 1 : f-ratio = 6.28096, EM 2 and control 2 : f-ratio = 25.7887, EM 3 and control 3 : f-ratio = 8.89215). All of the results proved to have a p-value less than 0.05, which proved there to be a significant difference. A statistical difference can also be proven as all of the f-ratio is greater than one, proving there is more variance between groups rather than within.

EM radiation also affected plant germination since all of the EM plants had more sprouts overall, though they died much faster (Fig. 3). The average of the maximum number of sprouts for the EM cups was 6.3, while the average of maximum sprouts for the control cups was 4. Though this

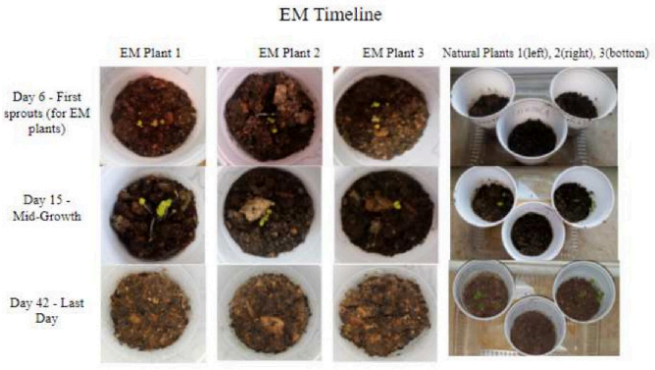

Figure 1: Timeline of EM plants and control plants. Images of EM and control plants are shown for Day 6, 15, and 42. The EM plants sprouted before control plants, the first ones appearing on Day 6. Moderate growth is visible in both groups on Day 15. On Day 42, the last day of the study, all of the EM plants had died, but control plants 1 and 2 appeared healthy. Control plant 3 never sprouted.

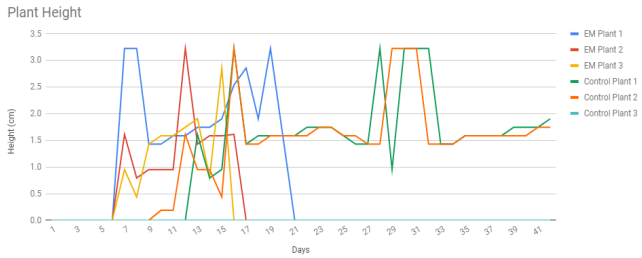

Figure 2: Timeline of EM plants and control plants. Images of EM and control plants are shown for Day 6, 15, and 42. The EM plants sprouted before control plants, the first ones appearing on Day 6. Moderate growth is visible in both groups on Day 15. On Day 42, the last day of the study, all of the EM plants had died, but control plants 1 and 2 appeared healthy. Control plant 3 never sprouted.

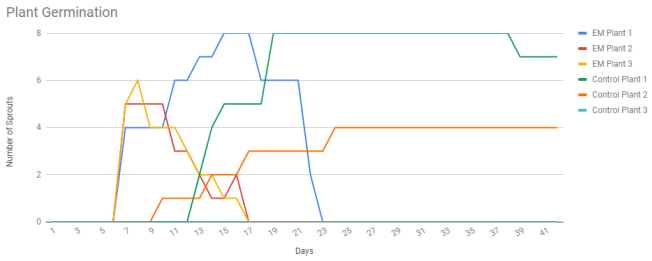

Figure 3: The line graph represents the number of sprouts and germination in the cup every day for 42 days. All trials were carried out indoors in the designated EM exposed and natural, unexposed area. Each cup can grow up to 8 sprouts. Control plant 1 and EM plant 1 were observed to hit 8 sprouts, though Control maintained it much longer.

data provides evidence against the hypothesis, human error could have played a role in the lack of germination in control plant 3. Even with the small sample size, there are significant statistical differences observed between the three sets (EM 1 and control 1 : f-ratio = 16.54675, EM 2 and control 2 : f-ratio = 25.65636, EM 3 and control 3 : f-ratio = 9.64706). All of the results also proved to have a p-value less than 0.05 and a f-ratio greater than one, which proves there to be a significant statistical difference.

DISCUSSION

EM radiation affects plants in many different ways, both positive and negative. The germination levels increased by 57.5%, though human error could have played a role since none of the control plant 3 seeds sprouted. The height of the plants also differs greatly between the groups. This shows that EM radiation tends to make plants grow a lot faster, though they died much quicker. The results both contradict and agree with previous experiments. The basil grew slightly taller when exposed to EM radiation, similar to the experiment conducted in Kolkata. However, the plants also died much quicker, supporting the other experiments. This could have been due to the fact that basil has not been tested yet, as all of the experiments used other plant species. The physical characteristics of the EM plants could have been due to destruction or mutations of chloroplasts in the plant's leaves as chloroplasts allow for photosynthesis and the green color of plants. EM radiation could also be changing the plants' genetic makeup since the stems grew faster but were much weaker. In an experiment conducted on the effects of EM radiation on the photosynthesis of a plant, the results concluded damage in the membrane of photosynthetic cells in the leaf, the halt of the transmission process of electrons in photosynthesis, and the decrease in potential active and photochemical efficiency (8). To confirm these results, it would be necessary to repeat this experiment multiple times, as human error and a small sample size could have played a role in the outcome of this study. Examination of the plants under a microscope would also be valuable in future studies. If a change in the makeup of the cells is present, it could have negative effects on all of the organisms that consume them. As agriculture feeds future generations, we must ensure the safety of their growth and consumption. This can be done with the use of future studies to be able to take the necessary precautions. Along with this, it is important to see if these devices affect other organisms similarly, such as humans, since it could potentially impact many lives.

MATERIALS AND METHODS

This experiment was conducted indoors with the use of two different areas. The first area, identified for the EM plants, had electronics and was in front of a window for adequate sunlight. In this experiment, we used three laptops and two phones that added up to a minimum of 2878 MHz of radiation. The devices were all functional and were active to the accordance of a 9-5 workday to model real-life circumstances. A second area, identified for the control plants, was completely devoid of electronics. This area was also in front of a window to make sure the plants would have sufficient lighting. We used six plastic drinking cups to plant basil seeds within. Each cup had three small holes the size of a pencil tip at the bottom in a triangular shape to ensure proper drainage of water. Then, we filled each cup halfway with potting soil and placed eight basil seeds on top of the soil. On top of the seeds, 1/4 cup more of soil was poured. After these steps were completed, half of the cups (three cups) were placed in the first area and the other half in the second area. The plants were routinely watered ⅛ cup of water once in the morning around 8 a.m. The soil should be damp an inch down into the soil to ensure enough water for the plant. Along with this, the number of sprouts and the height of the tallest sprout was recorded into a spreadsheet and with pictures.

To analyze the results, two line graphs and two ANOVAs were made: one of each to compare seed germination and another to compare sprout heights. The graphs make it possible to observe and note down significant trends, identify outliers, and possible errors in the data. This is how we came to the conclusion that EM radiation does affect plant growth and that human errors could have played a role in the growth of control plant 3. One-way ANOVAs were conducted to ensure there was a significant difference between the EM and control plants. As this experiment was conducted in triplicates, the respective EM and control pair were cross analyzed and compared to each other (EM plant 1 and control plant 1). If the p-value was less than 0.05, the group can be marked statistically significantly different. A statistical difference can also be proven if there is more variance between groups rather than within, meaning the f-ratio must be greater than one. Else, the groups do not have a strong enough statistical difference to be considered so.

Received: July 3, 2020
Accepted: July 31, 2020
Published: August 28, 2020

REFERENCES

1. “Why Doesn't My Laptop Emit Radiation?” Science Questions with Surprising Answers, wtamu.edu/~cbaird/sq/2014/12/04/why-doesnt-my-laptop-emit-radiation/.
2. Parihar, Leena. “Effect of Mobile Phone Radiation on Nodule Formation in the Leguminous Plants.” Current World Environment, vol. 9, no. 1, 2014, pp. 145-155. doi: 10.12944/CWE.9.1.21
3. Akbal, Ayhan & Kiran, Y & Sahin, Ahmet & Turgut-Balik, Dilek & Balik, Hasan. Effects of Electromagnetic Waves Emitted by Mobile Phones on Germination, Root Growth, and Root Tip Cell Mitotic Division of Lens culinaris Medik. Polish Journal of Environmental Studies, vol. 21, no. 1, 2010, pp 23-29.
4. WadmanSep, Meredith, et al. “How Plants Survived

Chernobyl." Science | AAAS, American Association for the Advancement of Science, 10 Dec. 2017, www.sciencemag.org/news/2009/05/how-plants-survived-chernobyl.
5. "Cell Phone Radiation Affects Plants, Vegetables: Research." Https://Www.hindustantimes.com/, Hindustan Times, 15 Feb. 2016, www.hindustantimes.com/india/cell-phone-radiation-affects-plants-vegetables-research/story-tWhZEXYsxutPOJN4KqJqpJ.html.
6. "5 Indoor Plants That Absorb Radiation." EMF Academy, 20 June 2018, emfacademy.com/5-indoor-plants-that-absorb-radiation/.
7. Daniel. "10 Plants That Can Absorb Electromagnetic Radiation." EMF Advice, 25 May 2020, emfadvice.com/plants-absorb-eliminate-radiation/.
8. Li, Y., Xi, G., Song, Q., Cao, Yi. "Difference of Chlorophyll Fluorescence Dynamics Process and the System of Photosynthetic Pigment in Leaf of Spinach and Tobacco under the Action of Low Level Microwave Electromagnetic Field." ResearchGate, 1 July 2005, www.researchgate.net/publication/292107432_Difference_of_chlorphyll_fluorescence_dynamics_process_and_the_system_of_photosynthetic_pigment_in_leaf_of_spinach_and_tobacco_under_the_action_of_low_level_microwave_electromagnetic_field.

ACKNOWLEDGEMENTS

I would like to sincerely thank Dr. Mohanraj Manangeeswaran for giving me some amazing advice. I would also like to thank my parents and sister for motivating me no matter what. Last, but not least, I would like to thank the JEI reviewers and publishers who helped me through this process and allowed for the publishing of this paper.

JOURNAL OF
EMERGING INVESTIGATORS

Article

Population Forecasting by Population Growth Models based on MATLAB Simulation

Jingze Li[1], Hien Tran[2], Jie Li[3]

[1] The Shattuck St. Mary's School, 1000 Shumway Ave, Faribault, MN, USA
[2] The Department of Mathematics, North Carolina State University, North Carolina, USA
[3] The School of Mechatronical Engineering, Beijing Institute of Technology, Beijing, China

SUMMARY

Population forecasting is a necessary effort to understand population growth, which affects various aspects of a country's society and economy, including future demand for food, water, energy, and services. Mathematical models are commonly used to understand the interplay of the migration, birth and death rates on population growth. Mathematical models help population forecasting by capturing statistical trends from historical datasets. However, these need to be carefully compared to understand the implications of different model formulations in predicting future population, which the models have not seen or were trained on. In this work, we have compared two common population growth models, namely Malthusian law and Logistic law, using US population data from 1951 to 2019. By formulating a least-squared curve fitting problem, the birth and death rates can be estimated using MATLAB software. MATLAB simulations showed that the Logistic law of population growth yields smaller sum of squared residuals than the Malthusian law. In this case, a better population model may be beneficial in the social science, such as political science and sociology.

INTRODUCTION

In view of the world population continuously evolving in the last century, population prediction becomes more and more important in policy making, economic planning, education, and so on [1]. Population is affected by numerous factors such as policy, economy, and culture. Therefore, it is hard for demographic researchers to analyze each factor. Among all principles, historical data are the significant foundation in forecasting. By analyzing the inherent tendency inside the historical data, a comprehensive and reasonable forecast can be made without discussing each factor that affects the population. In order to achieve satisfying forecast performance, authentic historical data are needed. The Malthusian model and the Logistic model are two widely used methods for population forecast and have shown good performance. The Logistic model was firstly introduced in 1837 by the Dutch biomathematician Pierre Verhulst [2, 3]. Miranda [4] later developed the Logistic model which has proved to be well suited for population forecasting. As shown in this paper, predictions obtained via the Logistic model are in quite acceptable agreement with current official forecasts from the US Census Bureau.

The matrix laboratory (MATLAB) is primarily intended for mathematical and scientific computation and is a powerful programming software for problem solving. MATLAB allows sophisticated functions, matrix manipulations, algorithmic design, and so on to be implemented in a relatively simple manner [5]. The growth of a population is often described by ordinary differential equations (ODEs), which can be effectively solved by MATLAB. Given the importance of forecasting population and the competitive ability of MATLAB on dealing with mathematical problems encountered in real life, the goal of our research was to figure out an accurate population growth model by utilizing MATLAB. We hypothesized that the Logistic law, which considers the death rate to be a function of population size, would be more accurate and realistic than the Malthusian law in which the death rate is a constant for a long term prediction. To test this hypothesis, we obtained the US population data from 1951 to 2019 from the US Census Bureau. Then, we adopted the data for numerical verification and to compare the accuracy of the mathematical models by the Malthusian law and the Logistic law for population predictions. Furthermore, to determine whether the Logistic law gives a statistically significant improvement in the fit as compared to the Malthusian law, we conducted an F-test. In order to verify and bolster the conclusion, we applied the F-test to 20 different countries to test if the Logistic model can be widely used. The population data of the countries were obtained originally from United Nations - World Population Prospects [10].

RESULTS

We utilized the mathematical models, described by ODEs using the Malthusian law and the Logistic law, to describe the US population growth and to predict future population. The Logistic law considers the death rate as a function of population size and the Malthusian law assumes a constant death rate. We adopted the ODE45 solver in MATLAB to solve these two models, utilizing for data the US population from 1951 to 2019 [6] to determine the birth and death rates and to show how well the model fit the data.

Compared to the actual populations sizes, the Logistic law is more accurate and realistic than the Malthusian law for predicting the US population from 1951 to 2019 as sum of squared residual for the Logistic law 8.78×10^{14} is smaller than the sum of squared residual for the Malthusian law

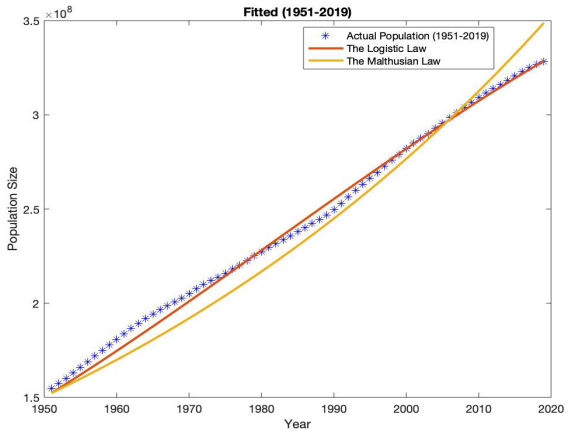

Figure 1: Fitted solutions of the mathematical models using the Malthusian law and the Logistic law using the US resident population data from 1951 to 2019. The parameters values found from *fminsearch* for the Malthusian law (r = 0.0122) and the Logistic law (b = 0.0235, n = 5.0471×10^{-11}).

6.12 x 10^{15} (**Figure 1**). In addition, near the year 2019, the predicted population by the Malthusian model continues to grow exponentially, while the predicted population by the Logistic model starts to exhibit diminishing growth. To further demonstrate that the Logistic growth model gives a statistically significant improvement in the fit, we performed an F-test. Using the degrees of freedom of 1 and 67 in the numerator and denominator, respectively, we calculated an F-statistic value of 400.22. The F-value is much larger than the F-critical value of 3.984 at 5% significance level; hence, we can reject the null hypothesis at 95% level of confidence. In other words, the Logistic model fits the actual data better than the Malthusian model does according to the F-test.

Using the birth and death rates obtained from data fitting, we obtained the predicted US population for the year 2020 from both models, which yielded 328.60 million and 348.80 million from the Logistic and the Malthusian models, respectively. Compared to the prediction value of 329.73 million from the US Census Bureau [6], the Logistic model gives a better prediction than the Malthusian model.

To further analyze the fidelity of the Logistic model, we repeated the above analysis for ten countries with high Human Development Index (HDI) and ten countries with low to medium HDI (see Method Section, "Analysis with different countries"). **Table 1** summarizes the sum of squared residuals corresponding to the model using the Malthusian law (SSR_M) and the Logistic law (SSR_L) and F-values of 20 countries. **Table 1** shows that only one country, Norway, has an F-value (2.62) lower than the critical value of 3.984. In this case, Norway's population is still in the period of accelerating growth and, hence, the Malthusian model is a good model for its population prediction (see the discussions in Section

Table 1: F-test results

	Country	SSRM	SSRL	F value	Result
Very high HDI countries	Norway	6.67*10^{11}	6.42*10^{11}	2.62	fail to reject
	Switzerland	5.62*10^{12}	3.36*10^{12}	45.15	reject
	Argentina	1.28*10^{14}	2.65*10^{12}	3169.23	reject
	Canada	2.73*10^{14}	2.93*10^{13}	557.27	reject
	United States	6.13*10^{15}	8.78*10^{14}	400.22	reject
	Australia	5.23*10^{13}	1.34*10^{13}	194.57	reject
	Iceland	1.21*10^{10}	2.17*10^{9}	306.63	reject
	Sweden	1.83*10^{12}	1.27*10^{12}	29.54	reject
	Singapore	4.06*10^{12}	1.88*10^{12}	77.69	reject
	South Korea	1.69*10^{15}	2.94*10^{12}	38446.61	reject
Low and medium HDI countries	Kenya	1.12*10^{14}	9.18*10^{12}	750.43	reject
	Myanmar	7.86*10^{14}	2.68*10^{13}	1898	reject
	Philippines	2.10*10^{15}	1.48*10^{13}	9439.76	reject
	India	9.21*10^{16}	1.51*10^{16}	341.66	reject
	Guatemala	1.50*10^{13}	9.80*10^{11}	958.51	reject
	Morocco	3.89*10^{14}	7.02*10^{12}	3645.68	reject
	Vietnam	3.14*10^{15}	3.38*10^{13}	6157.26	reject
	Guyana	1.16*10^{12}	1.69*10^{10}	4531.81	reject
	Bhutan	1.53*10^{11}	1.01*10^{10}	947.95	reject
	Bangladesh	3.74*10^{15}	2.34*10^{14}	1003.85	reject

The sum of squared residuals corresponding to the model using Malthusian law (SSR_M) and Logistic law (SSR_L) and F values of 20 countries in terms of the F-test (F-critical value at the confidence level of 95% equals to 3.984).

"Methods"). For all other nineteen countries sampled, F-values were much larger than the critical value. This result further confirms our hypothesis that, in general, the Logistic model provides a better population prediction than the Malthusian model.

DISCUSSION

Comparing the actual US population and the predicted US populations from the Malthusian law and the Logistic law of population growth, we can see that the logistic model is more in accord with the actual population than the Malthusian model is (**Figure 1**). In addition, the Logistic model also predicted a carrying capacity value of 465.61 million. Furthermore, we can visualize the simulation results that when the population size is smaller than half the carrying capacity, the population's growth trend is in an accelerated growth period. After crossing this cut-off point, the population's growth rate decreases and eventually reaches zero near the carrying capacity, which is the period of diminishing growth. This characteristic of population growth using the Logistic law, which is depicted by an S-shaped curve, seems to make sense practically. The major difference between the Malthusian law and the Logistic law is that the Logistic law takes the dynamic death rate into consideration. This death rate can be influenced by many factors. As the population size gets very large, individual members are now competing for limited resources such as food, living space, economic conditions, and medical care. Hence, population does not grow as fast as it does at the beginning, when the population size is small. One way to slow down the growth of the population for large population size is to make the death rate a function of population size as was done with the Logistic law. This means that as population grows larger, the death rate will also become larger which, consequently, will slow down population growth, so the population cannot grow at a constant or an increasing growth rate. In contrast, the model of the Malthusian law does not show the diminishing growth. We can see the estimation based on Malthusian law exceeds the actual population around 2005 and keeps growing at an exponential rate from **Figure 1**. This means that the model of the Malthusian law can only predict a very short period of population pattern.

We assumed that the actual population of US resident population from the US Census Bureau was reliable to test our predictions based on the Malthusian Law and the Logistic law. According to the US Census Bureau, the net growth rate in US population size changes yearly, from -0.06% in 1918 to 2.12% in 1910 [6]. The mean US population growth rate is 1.24%, and the median is 1.16%. Data fitting yielded a net growth rate of 1.22% for the Malthusian law model which is close to the US Census Bureau's estimation. However, the exponential growth characteristic of the Malthusian law yields gross exaggeration of population value for long term prediction into the future.

MATERIALS AND METHODS

In the literature, there are two common approaches for modeling the population size: the Malthusian law of population growth and the Logistic law of population growth. They are both theoretically sound and can be applied to practical censuses. We start with the US population in 1950 (about 152.27 million) as an initial condition and fit the two mathematical models to the US population data from 1951 to 2019.

The Malthusian Law for Population Growth

Let $P(t)$ denotes the population size at any given time t. To simplify this sophisticated problem, we assume two major impacts that would change the population, *i.e.*, the birth rate and the death rate. In the Malthusian law for population growth, the birth rate and the death rate are assumed to be proportional to the population size with the proportionality constants denoted by b and d, respectively. The change in population at any time would be described by:

$$\frac{dP(t)}{dt} = (b-d)P(t) = rP(t) \qquad (1)$$

In Equation (1), the birth rate term, $bP(t)$, is positive and will increase the population size. On the other hand, the death rate, $dP(t)$, is negative, and will decrease the size of the population. The net effect, $r = b - d$, will increase the population size if b is larger than d (more birth than death) and will decrease the population size if b is smaller than d. When $b = d$, the rate of change of the population is zero, and the population will remain constant in size.

For the initial condition, we input the US resident population in 1950. That is, $t = 1950$ is the initial time and $P(0) = 152.27\times10^6$.

The Logistic Law for Population Growth

The Logistic model was first developed by Pierre Verhulst [2, 3]. As discussed in this paper, for a large population it is more realistic to consider the death rate coefficient d to be a function of population size since more people will die due to limited resources, diseases, epidemics, and so on. A simple model to account for this observation is to assume the death rate coefficient to be proportional to the population size; that is, $d = nP(t)$. Hence, it is assumed when the population size is small the death rate coefficient is small and when the population size is large the death rate coefficient is large, which will slow down the population growth. Thus, the change in population size can be described as follows:

$$\frac{dP(t)}{dt} = \big(b - nP(t)\big)P(t) \qquad (2)$$

By factoring the constant b, Equation (2) can be rewritten as

$$\frac{dP(t)}{dt} = r\left(1 - \frac{P(t)}{P(m)}\right)P(t). \qquad (3)$$

Here, $r = b$ and $P(m) = b/n$. In this form and by comparison with the mathematical model for the Malthusian

law of population growth (1), the net per capita growth rate, $r(1 - P(t)/P(m))$, is a function of the population size $P(t)$ instead of a constant as in Equation (1). The parameter, $P(m)$, is called the carrying capacity of the human population or, equivalently, the maximum population size. It is noted that the population will grow in size if $P(t) \leq P(m)$ (that is, the net growth rate is nonnegative). Furthermore, when the population is small relative to $P(m)$, the growth rate is large (period of accelerated growth). On the other hand, if the population size is larger and is closer to $P(m)$, the growth rate is still positive but getting smaller (period of decelerated growth). Finally, when t gets larger enough to the point that $P(t) = P(m)$ and $(1 - P(t)/P(m)) = 0$, the growth rate will be zero, and the population size is of constant size given by $P(m)$.

With the initial condition $P(0) = P_0$, the solution of Equation (3) is $P(t) = \frac{P(m)}{1+\left(\frac{P(m)}{P_0}-1\right)e^{-rt}}$ [7].

When $t \to \infty$, we obtain $e^{-tr} \to 0$ and $P(t) \to P(m)$. If we differentiate both sides of Equation (3) with respect to t, we obtain the following equation:

$$\frac{d^2P(t)}{dt^2} = r^2\left(1-\frac{P(t)}{P(m)}\right)\left(1-\frac{2P(t)}{P(m)}\right)P(t) \quad (4)$$

Equation (4) is an expression for the second derivative of $P(t)$, which will give us information on the concavity of the solution curve $P(t)$. In particular, for $P(t) < P(m)$, we observe:

1) When $P(t) < P(m)/2$, we have $d^2P(t) / dt^2 > 0$, which means the solution curve, P(t), is concave.
2) When $P(t) > P(m)/2$, we have $d^2P(t) / dt^2 < 0$, which means the solution curve, P(t), is convex.
3) When $P(t) = P(m)/2$, we have $d^2P(t) / dt^2 = 0$, which corresponds to an inflection point.

Similar to the mathematical model by the Malthusian law, we take the US population of the year 1950 as the initial condition, that is, $P(0) = 152.27 \times 10^6$.

Solving ODEs using MATLAB

MATLAB is a powerful mathematical tool for curve-fitting and solving ODEs. For a system of first order ODEs, an .m file that contains the descriptions of ODEs is first created. Then, by specifying input parameters (parameters in the model, *e.g.*, r, b, and n), initial conditions, and the time interval, where we want to compute the solution, the MATLAB's built-in ODE solver, *i.e.*, ODE45 was used to approximate the solution of the before-mentioned ODE. Besides, MATLAB was also used to visualize the solution and we used ODE45 here. Although MATLAB is powerful, it can only solve first order ODEs. In order to solve higher order ODEs, extra variables were introduced to convert a high order ODE to a system of first order ODEs [7].

Curve Fitting and F-Test

When performing the curve fitting for population growth, the *fminsearch* function in MATLAB is adopted to find the best Logistic and Malthusian growth curve (that is, the best parameters r, b, and n) that can fit the data. *fminsearch* is a Nelder-Mead algorithm to minimize a scalar function of several variables, which requires an initial guess of the unknown parameters. fminsearch was used to find the model parameters to give the best fitting curve via minimizing the sum of squared residuals between the data and the model solution as described by: $min_{r,b,n} J = \sum_{i=1}^{69}\left(y_i^s - y_i^{data}\right)^2$.

In this case, i denotes the year range, y_i^{data} represents the human population data at year i, and y_i^s denotes the model (Malthusian law and Logistic law) population solution at year i. When using the *fminsearch* function, the model parameter values corresponding to the period of 1951 and 1965 were selected as the initial guesses. As to the convergence criteria, we adopt the default options of the *fminsearch* function which used both the termination tolerance on the function value and on the parameter value. To be specific, the *fminsearch* function terminates when the tolerances on the function and parameter values are smaller than or equal to 10^{-4}.

In order to identify the mathematical model that best fits the population data using least squares formulation, we used an F-test. To begin, the null hypothesis, H_0, assumes that $n = 0$ (that is, the nonlinear death rate term in the model using Logistic law is not important, or, equivalently, the mathematical models using the Malthusian and Logistic laws are the same). The F-statistic is then computed from the following formula

$$F = \frac{(SSR_M - SSR_L)/m}{SSR_L/(n-m-1)},$$

where SSR_M and SSR_L are the sum of squared residuals corresponding to the model using Malthusian law and Logistic law, respectively. m is the number of restrictions (in our case, $m = 1$) and n is the number of observations (in our case, $n = 69$). It is noted that $SSR_L < SSR_M$. The question is whether the sum of squared residual associated with the model using the Logistic law is significantly less than the sum of squared residual associated with the model using the Malthusian law. The answer to this question depends on how the F-statistic value computed from the formula given above compares to the F-critical value from the F-table at the required level of significance. If the F-statistic is greater than the critical value at the required level of significance (that is, SSR_L is significantly less than SSR_M), we reject the null hypothesis.

Analysis with different countries

The countries were selected based on their HDI (Human Development Index) which is a measure that accounts for more than incomes. HDI takes many factors into account, such as education, health, and economics, to create a score between 0 and 1 [9]. A score over 0.800 is very high, a score between 0.700 and 0.799 is high, a score between 0.550 and 0.699 is medium, and a score lower than 0.550 is low. We selected ten countries with very high HDI and ten countries

with medium and low HDI. Each country has a generally increasing population and available data from 1950 to 2019.

ACKNOWLEDGEMENTS

I would like to express my special thanks of gratitude to my teacher Dr. Yang who gave me the golden opportunity to spend wonderful time in North Carolina to study. Also special thanks to Guangming Wang, a physics Ph.D. student, and Yahe Yu, a mathematics Ph.D. student at NC State University for offering critical help during my projects.

Last but not the least, I would also like to thank my parents, who helped me a lot in finalizing this opportunity within the limited time.

Received: October 13, 2019
Accepted: August 22, 2020
Published: August 31, 2020

REFERENCES

1. Lassila, Jukka, et al. “Demographic forecasts and fiscal policy rules.” *International Journal of Forecasting*, vol.30, no.4, 2014, pp. 1098-1109.
2. Pierre François Verhulst. “Recherches mathématiques sur la loi d'accroissement de la population.” *Nouvelle mémoire de l'Academie Royale de Sciences et Belle-Lettres de Bruxelles [i.e. Mémoire Series 2]*, vol. 18, 1845, pp. 1–42.
3. Pierre François Verhulst. “Deuxiéme mémoire sur la loi d'accroissement de la population,” *Mémoire de l'Academie Royale des Sciences, des Lettres et de Beaux-Arts de Belgique*, vol. 20, 1847, pp. 1–32.
4. Miranda, Luiz Carlos M. and Carlos A.S. Lima. “On the Logistic Modeling and Forecasting of Evolutionary Processes: Application to Human Population Dynamics.” *Technological Forecasting and Social Change*, vol. 77, no. 5, 9 Feb. 2010, pp. 699–711.
5. Attarian, Adam. “The Official MATLAB Crash Course.” 14 Sep 2008.
6. *US Population by Year*, US Census Bureau, www.multpl.com/united-states-population/table/by-year.
7. Braun, Martin. *Differential Equations and Their Applications: An Introduction to Applied Mathematics*. Springer-Verlag, 1995.
8. Higham, Desmond J., and Nicholas J. Higham. *MATLAB Guide*. SIAM: Society for Industrial and Applied Mathematics, 2005.
9. “Human Development Reports.” *United Nations Development Programme, Human Development Reports*, hdr.undp.org/en/content/human-development-index-hdi.
10. “The Long Term Perspective on Markets.” *Macrotrends*, www.macrotrends.net/.

Sponsorship

Editor's Circle $10,000+

ALLOY THERAPEUTICS

Patron $5,000+

NEW ENGLAND BioLabs® Inc.
enabling technologies in the life sciences

PORTFOLIOS WITH PURPOSE®

Institutional Supporters

HARVARD UNIVERSITY

HARVARD MEDICAL SCHOOL

Tufts UNIVERSITY

Charitable Contributions

We need your help to provide mentorship to young scientists everywhere.

JEI is supported by an entirely volunteer staff, and over 90% of our funds go towards providing educational experiences for students. Our costs include manuscript management fees, web hosting, creation of STEM education resources for teachers, and local outreach programs at our affiliate universities. We provide these services to students and teachers entirely free of any cost, and rely on generous benefactors to support our programs.

A donation of $30 will sponsor one student's scientific mentorship, peer review and publication, a six month scientific experience that in one student's words, 're-energized my curiosity towards science', and 'gave me confidence that I could take an idea I had and turn it into something that I could put out into the world'. **If you would like to donate to JEI, please visit https://emerginginvestigators.org/support, or contact us at questions@emerginginvestigators.org.** Thank you for supporting the next generation of scientists!

'Journal of Emerging Investigators, Inc. is a Section 501(c)(3) public charity organization (EIN: 45-2206379). Your donation to JEI is tax-deductible. '

emerginginvestigators.org