**Spatiotemporal Trends in Group A StreptococcalPharyngitis in the United States**

Madeleine C. Kline, Stephen M. Kissler, Yonatan H. Grad

**ABSTRACT**

Background

Group A *Streptococcus* (GAS) causes many cases of pharyngitis each year in the United States (U.S.) and can result in serious complications. While other common respiratory infections, such as influenza and respiratory syncytial virus, have known geographical epidemic trends in the U.S., the timing and geography of GAS pharyngitis trends has not been elucidated.

Methods

This study utilized outpatient claims data from those with private medical insurance between 2010-2018. We assessed the average number of visits per member for GAS pharyngitis across U.S. census regions and subregions. We evaluated the timing of seasonal GAS pharyngitis peaks in individual states and subregions to characterize yearly patterns of geographic spread.

Results

GAS pharyngitis visits showed distinct trends across regions and over the course of a year.

The South had more visits per member than other regions (yearly average 39.11 visits per 1000 people, 95% CI: 36.21-42.01), and the West had fewer visits per member than other regions (yearly average 17.63 visits per 1000 people, 95% CI: 16.76-18.49). There were more visits in winter months than summer months, and regional differences between the South and other regions were most pronounced in the late summer through early winter, reflecting geographic differences in the timing of the rise in visits. GAS pharyngitis visits peaked earliest in southern states in December to January (Louisiana peak month 0.81, 95% CI: 0.41-1.21) before spreading outwards from the South and peaking on the coasts in March (Rhode Island peak month: 2.64, 95% CI: 2.37-2.91).

Conclusions

Understanding regional differences in the timing of GAS pharyngitis visits is important for predicting the highest burden of severe disease, complications, and sequelae, and for uncovering drivers of disease spread that could help target preventative measures.

**INTRODUCTION**

Group A *Streptococcus* (GAS), also known as *Streptococcus pyogenes,* is a common human bacterial pathogen that causes a diverse spectrum of disease. GAS pharyngitis, known commonly as “strep throat,” is responsible for an estimated 5.2 million outpatient visits and 2.8 million antibiotic prescriptions each year.1 According to the Centers for Disease Control and Prevention (CDC), GAS accounts for 20-30% of pharyngitis cases in children and 5-15% of pharyngitis cases in adults.2 Rarely, GAS pharyngitis can have severe complications (e.g., peritonsillar abscess), or post-infection immune sequelae such as acute rheumatic fever or post-streptococcal glomerulonephritis. In the U.S., treatment with antibiotics is recommended to decrease symptom duration, reduce transmission, and prevent complications such as acute rheumatic fever, and diagnosis with a rapid antigen detection test (RADT) or throat culture is required.2

GAS can also cause invasive disease (iGAS), which is defined as bacteria cultured from a typically sterile body site (e.g. bacteremia, meningitis).3 In contrast to GAS pharyngitis, iGAS can be life threatening and require intensive care. After reported iGAS cases declined by about 25% during the COVID-19 pandemic, there was an increase in iGAS infections in the U.S. in the 2022-2023 season. While the connection between GAS pharyngitis and iGAS is not directly causal, better understanding the epidemiology of GAS pharyngitis could help elucidate drivers of iGAS disease.

Many antibiotic courses dispensed to children in the U.S. are associated with respiratory infections.4 GAS pharyngitis is a major driver of antibiotic prescribing in the U.S., accounting for 5.9% of all outpatient antibiotic prescriptions in children ages 3-9.1 While GAS is largely penicillin-susceptible, resistance has emerged to second-line antibiotics such as macrolides and lincosamides.5 Use of antibiotics can promote drug resistance both in the target pathogen and in other prevalent bacteria via bystander selection.6 Efforts to reduce GAS pharyngitis disease burden include the development of vaccines to protect against GAS, some of which are in clinical trials but are not yet approved.7,8 The contribution of GAS pharyngitis to antibiotic prescribing, the potential for future vaccine rollout, and the recent increase in invasive disease all motivate the need to better understand GAS pharyngitis transmission patterns and geographic distribution in the U.S.

GAS pharyngitis is more common in the winter and spring months,2,9 but little else is known about its seasonality and geography in the U.S. In contrast, other common respiratory conditions have well-characterized spatiotemporal trends: epidemic waves of influenza often start in the southern U.S., and respiratory syncytial virus (RSV) typically peaks earliest in Florida seasonally.10,11 Understanding trends in the timing and location of these important infections improves knowledge of pathogen transmission and spread, and can indicate where and when to expect the most severe disease and complications to target interventions that reduce the burden of disease. The existence of a routinely used molecular diagnostic for GAS pharyngitis means that claims data are a reliable indicator of disease prevalence in the U.S. To address the need for improved characterization of GAS epidemiology in the U.S., we used outpatient claims data from private insurers to measure variation in GAS pharyngitis visits by region and over the course of the year.

**METHODS**

**Study Population and Data Source**

Outpatient claims data were extracted from the Merative (formerly IBM) MarketScan database, which is a convenience sample of 16.6-36.4 million privately-insured individuals (5.1-11.6% of the total U.S. population).12 The sample was restricted to individuals who were continuously enrolled for an entire year between 2010-2018. Information on age group, sex, and state were included. Characteristics of the study population averaged over the 9 years of observation are shown in Table 1.

**Disease Incidence**

Visits prompted by GAS pharyngitis were identified by mapping codes from the *International Classification of Diseases, Clinical Modification* ninth (ICD9) or tenth (ICD10) revision to Clinical Classification Software (CCS) codes (Table S1).13 Using CCS codes allowed for consistent identification of GAS pharyngitis cases across both ICD9 and ICD10. Visits were included if a diagnosis code consistent with GAS pharyngitis was the first or second diagnosis billed for the visit.

Yearly visits per 1,000 people were calculated by dividing the number of visits in that year by the number of people during that year and multiplying by 1,000. Monthly visits per 1,000 people were calculated by dividing the number of visits per month by the number of people over the course of the entire year and multiplying by 1,000. Quarterly visits per 1,000 people, where quarter 1 was January, February and March, quarter 2 was April, May and June, quarter 3 was July, August, and September, and quarter 4 was October, November, and December, were calculated by summing the monthly average visits across the 3 months in that quarter. Visit counts by age and sex strata in the MarketScan data were weighted by the proportion of the actual population represented by that stratum to account for the fact that the age and sex distribution in the dataset may not have represented the state’s true age and sex distribution. State-level population counts were obtained from the 2011-2015 American Community Survey (ACS) and accessed using the tidycensus R package14.

Visits were calculated by state and within regions and subregions of the U.S. according to the Census Regions and Divisions of the United States (Tables S2 and S3). For regional analyses, all states were included other than Hawaii and Alaska, which are not part of the continental U.S. (Figure S1). Yearly visits per 1,000 people by region were calculated by dividing the number of visits in that year by the number of people in the region and multiplying by 1,000. Monthly visits per 1,000 people by state, region, or subregion were calculated by dividing the number of visits in that month by the number of people in that state, region, or subregion over the course of the entire year and multiplying by 1,000. Quarterly visits per 1,000 people by region or subregion were calculated by summing the monthly average visits across the 3 months in that quarter in that region or subregion. Visits by age, sex, and state strata in the MarketScan data for each region or subregion were weighted by the actual population proportion represented by that stratum in that region or subregion. There was no clear increasing or decreasing secular trend in visits across years (Figure S2), and thus visits were averaged across all 9 years of observation and 95% confidence intervals were calculated under the assumption of normally distributed errors.

**Statistical Analyses**

*Regional Significance Testing*

To assess differences in bulk visit rates across regions, yearly visits per member in each region were compared using Welch’s two sample t-test (Figures S2 and S3). To assess differences in the seasonality of GAS visits across regions, visits in each region and month were compared to all other regions in that month using Welch’s two sample t-test (see Figure S5). Statistical significance was determined based on a significance level of 0.05 corrected for multiple hypothesis testing using the Bonferroni correction.

*Seasonal Modeling*

To characterize the seasonality in GAS pharyngitis visit trends by state or region, data from the 9 years of observation were fit to sinusoids using nonlinear least squares regression. Trends were fit according to the following equation:

where is the number of visits per thousand people in state or region, *i* , is the amplitude for state/region *i* (difference between the maximum and minimum monthly visits per 1000 people in a year), is the period (the time required for one full cycle of visits), is the month of observation *j* (taking values from 1-12 with 1 corresponding to January 1st), is the phase (the horizontal shift in months, so that is the month in which the maximum visits per 1000 people occurs), and is the baseline visit rate (i.e., the mean number of visits per 1000 people in location *i*). The period was fixed at 12 months (ω = ). Initial guesses for estimating the amplitude, phase, and offset were chosen and then optimized using the *nls* function in R. The initial amplitude was specified as one half of the distance from the average maximum to minimum number of visits in the entire dataset. The initial phase was specified as the average month during which the maximum number of visits occurred. The initial offset was specified as the mean number of visits throughout the year.

The phase represents the month during which the peak in visits occurred, which is 6 months before and after the month with the minimum number of visits according to sinusoid structure with a 12-month period. We mapped months from the dataset to integers from 1-12 (January-December) during which the visit took place. Because the sinusoidal fitting process can result in phases <0 or >12, calculated phases were corrected to fall within the 0-12 range by dividing all phases by 12 and taking the remainder of this division (applying the modulus function). Taking each integer value to be the first day of the specified month (e.g., 1 is January 1st), the final corrected phases fall between 0-11.99 where 0 is the first day of December and 11.99 is the last day of November. The sinusoidal fitting process can also result in negative amplitudes, which effectively shifts the phase of the corresponding positive-amplitude sinusoid by 6 months. None of the amplitude estimates included in the final analysis were negative.

Confidence regions for the sinusoidal fits were calculated via bootstrapping. 1000 samples were drawn from normal distributions centered around each of the 3 fitted sinusoid parameters (amplitude, phase, and offset) and with standard deviation equal to the standard error estimated by the model for each parameter. Sinusoids were generated using each of the 1000 sets of 3 bootstrapped parameters. Bounds for the 95% confidence regions of the mean visits per 1000 people in each month were then extracted as the 2.5th and 97.5th quantiles across this set of 1000 generated sinusoids.

*School Start Time Correlation*

Data on 2019 school start dates was curated by Pew Research Center.15 This dataset contains a 509-district sample of the >13,000 public school districts in the United States, included the 10 largest school districts in each state where possible, and included the 100 largest school districts in the country. Altogether, this dataset represented and estimated 36% of public-school students in the United States. Data were filtered to only include continental U.S. states (excluding Hawaii and Alaska), so that the final dataset contained 497 school districts. School districts were grouped by subregion and the average school start date in each subregion was determined by taking the mean. Correlation between average school start date and sinusoid phase in each subregion was calculated using Pearson’s correlation coefficient.

**RESULTS**

The four census regions (Midwest Northeast, South, West) differed in the bulk number of GAS visits per 1,000 individuals (Figure S2). The South, and particularly the East South Central and West South Central subregions, had more average yearly visits per 1000 people than other regions (Figure S3). The West, and particularly the Pacific West, had fewer average yearly visits per 1000 people than other regions. These differences were stable over the 9-year observation period. The South had an average of 39.11 (95% CI: 36.21-42.01) visits per 1000 people each year, the Midwest had an average of 29.45 (95% CI: 26.78-32.11) visits per 1000 people each year, the Northeast had an average of 29.32 (95% CI: 27.04-31.61) visits per 1000 people each year, and the West had an average of 17.63 (95% CI: 16.76-18.49) visits per 1000 people each year. At a subregional level, the East South Central region had an average of 48.38 (95% CI: 42.40-53.37) visits per 1000 people per year while the Pacific West had an average of 12.39 (95% CI: 11.57-13.22) visits per 1000 people per year. Visits in the South and the West were both statistically significantly different from all other regions (Figure S2), and only the Northeast-Midwest region comparison was not significant.

Across regions, GAS pharyngitis visits were more common in the winter months, with visits nadiring in the summer months before beginning to rise in early autumn and peaking in the first few months of the year (Figure 1). In the South, the January average was 3.78 (95% CI: 3.36-4.21) visits per 1000 people while in July, the average was 1.80 (95% CI: 1.67-1.93) visits per 1000 people. In the West, the January average was 1.76 (95% CI: 1.62-1.90) visits per 1000 people while in July, it was 0.98 (95% CI: 0.93-1.03). Similarly, at a subregional level, the East South Central region had on average 4.70 (95% CI: 4.13-5.26) visits per 1000 people in January and 1.88 (95% CI: 1.71-2.06) visits per 1000 people in July, while the Pacific West had on average 1.2 (95% CI: 1.09-1.31) visits per 1000 people in January and 0.76 (95% CI: 0.71-0.81) visits per 1000 people in July. The West had fewer visits per 1000 people than the other 3 regions throughout the year. However, the differences in visits per 1000 people in the South, Northeast, and Midwest were more pronounced from July through December (July to September South-Northeast difference: 2.51, 95% CI: 2.25-2.78, October to December South-Northeast difference: 3.9, 95% CI: 3.55-4.25) than from January through June (January to March South-Northeast difference: 2.11, 95% CI: 1.58-2.65, April to June South-Northeast difference: 1.27, 95% CI: 0.93-1.61) (Figure S5, Table S4).

The annual peak in GAS pharyngitis visits occurred earliest in the South and from there radiated outwards to the rest of the country (Figure 2, associated gif). The states with the earliest peaks were in the East South Central (phase: 1.50, 95% CI 1.22-1.79, where 1.0 represents January 1st) and West South Central (phase: 1.49, 95% CI 1.20-1.77) subregions (Figure S7), which peaked in January. Louisiana (phase: 0.81, 95% CI 0.41-1.21) and Mississippi (phase: 0.86, 95% CI 0.43-1.3) peaked particularly early, in December (Table S5). The peak in GAS pharyngitis visits then spread up through the Mountain West (phase: 1.68, 95% CI 1.49-1.88) and peaked latest on the coasts in March (Pacific West phase: 2.35, 95% CI 2.10-2.61; New England phase: 2.62, 95% CI 2.44-2.81).

School start dates correlated well with sinusoid trough dates (Pearson’s correlation *r*  = 0.93)(Figure 3, Figure S10). Schools in the South tend to start earlier, and schools in the Northeast tend to start later. The sinusoid phase represents the timing of the peak in GAS pharyngitis visits and is offset from the trough by 6 months. Sinusoid trough date preceded school start date by an average of 21 days or about 3 weeks (range: 12 days to 32 days). Sinusoid trough date preceded school start date by more in subregions that started school earlier, and more closely aligned with school start date in subregions that started school later (Figure S10).

**DISCUSSION**

Visits for GAS pharyngitis showed distinct spatiotemporal patterns. The South, and particularly the East South Central and West South Central regions, documented more visits than other regions throughout the year, especially from July to December. The Pacific West subregion documented fewer GAS pharyngitis visits throughout the year. Annually, GAS pharyngitis incidence began to rise earliest in the same southern subregions with the highest burden of disease and peaked latest in coastal regions. These results indicate that the burden of GAS pharyngitis disease is not evenly distributed across the country, but rather that southern states document more visits. The yearly trends in visits also indicate that disease burden begins in southern states, particularly in Louisiana and Mississippi, and spreads outwards from there to the rest of the country. This could represent the progression of environmental conditions or behaviors that promote the spread of GAS pharyngitis, or the spread of the pathogen itself as time progresses.

The timing of GAS pharyngitis correlates with school start dates; subregions in the South with earlier troughs and peaks in GAS pharyngitis visits start school earlier than coastal subregions with GAS pharyngitis troughs and peaks later in the year. It is plausible that school initiation leads to closer contact between children, who are the primary affected group, therefore facilitating the spread of respiratory infections. This theory has been posited for other respiratory infections in the past.16 However, the fact that school start times are offset from the peak in GAS pharyngitis visits by about 5 months, and the distance between school start time and phase is not constant in every region may indicate that this correlation is circumstantial and based on the fact that school starts in the summer when respiratory infections spread least effectively.

GAS pharyngitis spatiotemporal patterns are similar to but subtly different from previously described trends in yearly influenza and RSV. Previous work has shown that RSV hospitalizations peak earlier in Florida (November/December) than in other states.11 Environmental factors such as mean vapor pressure, minimum temperature, precipitation and seasonal variation in potential evapotranspiration may account for differences in the timing of RSV peaks in different US states.11 Similarly, yearly epidemic waves of influenza tend to originate in the southern U.S., although not specifically in the East South Central and West South Central subregions.10 Absolute humidity has been suggested as one factor that modifies influenza virus transmission and survival and underlies these geographic patterns of spread. It is possible that some of the same drivers of other respiratory infections may contribute to regional differences in burden and timing of GAS pharyngitis. Further work is needed to elucidate these behavioral and environmental contributors, and how the changing climate will impact them.

GAS differs from influenza and RSV in that it is bacterial, rather than viral. It can be carried asymptomatically in some hosts, indicating that this pathogen’s interaction with the human immune system may differ substantially from common respiratory viruses. In addition, trends in GAS pharyngitis may indicate broader trends in GAS disease, or differences in transmission mechanisms between different GAS clinical syndromes. For example, GAS necrotizing fasciitis, a form of iGAS, was not found to be seasonal like GAS pharyngitis,9 but the geographic variation in iGAS, and its relationship to variation in GAS pharyngitis, should be investigated further.

Our results are largely consistent with other studies that have estimated the burden of disease of GAS pharyngitis from outpatient claims data, and with prior studies of seasonality. A study in the U.S. using claims from multiple nationally representative datasets found 19.1 outpatient visits per 1000 U.S. people ages 0-64 years per year for GAS pharyngitis from 2012-2015.1 A study from Australia found that GAS pharyngitis cases peaked in the spring/winter and in the autumn season in 2001-2002,17 and another study from the U.S. found that peaks in GAS pharyngitis occurred in the winter months between 2010-2019.9

The present study has several limitations. The data capturing GAS-related visits are a convenience sample of privately insured individuals in the United States. Because insurance policies are heterogeneous across states and regions, some states have a much higher proportion of publicly insured or uninsured constituents. This could lead to differences in population characteristics and access to care across states that could bias estimates of GAS prevalence across regions or states. Additionally, because our dataset was restricted to individuals who were continuously enrolled in the same state over the course of the year, it excluded people who changed insurers or moved states frequently, whose health and behavior may differ in important ways from the individuals included in the study. All visits where GAS pharyngitis was the primary or secondary diagnostic code were included, but there may have been differences in how providers across different states or hospital systems bill for this condition, for example, billing for follow-up visits, which could also introduce bias.

In conclusion, the South documented more GAS pharyngitis compared to other regions and experienced a seasonal peak in visits earlier than other regions. The Pacific West had fewer GAS-related visits than other regions, and coastal regions experienced seasonal peaks latest in the year. Understanding these patterns is important for designing effective surveillance programs, preventative interventions such as vaccines, and allocating resources to appropriately prepare for expected disease burden.

**MAIN FIGURES AND TABLES**

**Table 1: Study Population Characteristics**

|  |  |  |
| --- | --- | --- |
| Category | Average Membership | % |
| Total | 2.67e+07 (1.66e+07-3.64e+07) | 100 |
| Sex |  |  |
| Male | 1.29e+07 (8.07e+06-1.78e+07) | 48.49 |
| Female | 1.38e+07 (8.53e+06-1.87e+07) | 51.51 |
| Age Group |  |  |
| 00\_04 | 1.50e+06 (9.21e+05-2.08e+06) | 5.6 |
| 05\_09 | 1.79e+06 (1.07e+06-2.49e+06) | 6.71 |
| 10\_19 | 4.15e+06 (2.46e+06-5.78e+06) | 15.55 |
| 20\_29 | 3.53e+06 (2.41e+06-4.80e+06) | 13.22 |
| 30\_39 | 3.90e+06 (2.48e+06-5.26e+06) | 14.62 |
| 40\_49 | 4.74e+06 (2.81e+06-6.58e+06) | 17.77 |
| 50\_59 | 5.21e+06 (3.16e+06-7.08e+06) | 19.51 |
| 60\_69 | 1.88e+06 (1.29e+06-2.40e+06) | 7.03 |
| Region |  |  |
| Midwest | 6.18e+06 (4.01e+06-8.95e+06) | 23.14 |
| Northeast | 4.92e+06 (2.65e+06-7.09e+06) | 18.44 |
| South | 1.03e+07 (6.78e+06-1.31e+07) | 38.75 |
| West | 5.25e+06 (2.79e+06-7.69e+06) | 19.68 |
| Subregion |  |  |
| East North Central | 4.86e+06 (3.00e+06-7.30e+06) | 18.21 |
| East South Central | 1.70e+06 (1.04e+06-2.10e+06) | 6.37 |
| Middle Atlantic | 3.76e+06 (1.96e+06-5.69e+06) | 14.08 |
| Mountain West | 1.60e+06 (9.45e+05-2.24e+06) | 6 |
| New England | 1.16e+06 (6.47e+05-1.72e+06) | 4.36 |
| Pacific West | 3.65e+06 (1.85e+06-5.45e+06) | 13.68 |
| South Atlantic | 5.30e+06 (3.71e+06-6.44e+06) | 19.85 |
| West North Central | 1.32e+06 (1.00e+06-1.73e+06) | 4.93 |
| West South Central | 3.34e+06 (2.03e+06-4.83e+06) | 12.53 |

**A picture containing screenshot, line

Description automatically generated**

**Figure 1:** Visits patterns by region over the course of the year.The average number of visits per 1000 people over the 9-year observation period for all age groups is plotted for each census region. Points represent individual year values. Shading represents the 95% confidence intervals depicting year-to-year variation.

**A map of the united states

Description automatically generated**

**Figure 2:** Phases of US state trend sinusoidal fits. States are colored by the phases of the sinusoids fit to data from people of all ages in each state. Darker colors indicate earlier peaks, and lighter colors indicate later peaks. 1 corresponds to January 1st , 2 corresponds to February 1st etc, with numbers <1 representing December dates. See corresponding gif in supplementary materials.

****

**Figure 3**: Relationship between subregion school start date and GAS pharyngitis sinusoidal trend. Average school start date is plotted on the x-axis and sinusoid trough date, which represents the point in the year with the lowest number of GAS pharyngitis visits and is offset 6 months from the peak in visits, is plotted on the y-axis. The dashed line represents a linear trend line with shading showing the 95% confidence interval of the linear model. Pearson’s correlation is 0.93. The solid line represents a line with a slope of 1, showing what the trend would be if it were the case that starting school 1 day earlier correlated to a sinusoid trough 1 day earlier.

**REFERENCES**

1. Lewnard, J.A., King, L.M., Fleming-Dutra, K.E., Link-Gelles, R., and Van Beneden, C.A. (2021). Incidence of Pharyngitis, Sinusitis, Acute Otitis Media, and Outpatient Antibiotic Prescribing Preventable by Vaccination Against Group A Streptococcus in the United States. Clinical Infectious Diseases *73*, e47–e58. 10.1093/cid/ciaa529.

2. Pharyngitis (Strep Throat): Information For Clinicians | CDC (2023). https://www.cdc.gov/groupastrep/diseases-hcp/strep-throat.html.

3. Streptococcus Disease, Invasive, Group A (GAS) (Streptococcus pyogenes) 1995 Case Definition | CDC (2022). https://ndc.services.cdc.gov/case-definitions/streptococcus-disease-invasive-group-a-1995/.

4. Kissler, S.M., Wang, B., Mehrotra, A., Barnett, M., and Grad, Y.H. (2023). Impact of Respiratory Infection and Chronic Comorbidities on Early Pediatric Antibiotic Dispensing in the United States. Clinical Infectious Diseases *76*, 382–388. 10.1093/cid/ciac811.

5. Brouwer, S., Rivera-Hernandez, T., Curren, B.F., Harbison-Price, N., De Oliveira, D.M.P., Jespersen, M.G., Davies, M.R., and Walker, M.J. (2023). Pathogenesis, epidemiology and control of Group A Streptococcus infection. Nat Rev Microbiol. 10.1038/s41579-023-00865-7.

6. Tedijanto, C., Olesen, S.W., Grad, Y.H., and Lipsitch, M. (2018). Estimating the proportion of bystander selection for antibiotic resistance among potentially pathogenic bacterial flora. Proceedings of the National Academy of Sciences *115*, E11988–E11995. 10.1073/pnas.1810840115.

7. Dale, J.B., and Walker, M.J. (2020). Update on Group A Streptococcal Vaccine Development. Curr Opin Infect Dis *33*, 244–250. 10.1097/QCO.0000000000000644.

8. Dale, J.B., Penfound, T.A., Chiang, E.Y., and Walton, W.J. (2011). New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococci. Vaccine *29*, 8175–8178. 10.1016/j.vaccine.2011.09.005.

9. Kennis, M., Tagawa, A., Kung, V.M., Montalbano, G., Narvaez, I., Franco-Paredes, C., Vargas Barahona, L., Madinger, N., Shapiro, L., Chastain, D.B., et al. (2022). Seasonal variations and risk factors of Streptococcus pyogenes infection: a multicenter research network study. Therapeutic Advances in Infection *9*, 20499361221132100. 10.1177/20499361221132101.

10. Charu, V., Zeger, S., Gog, J., Bjørnstad, O.N., Kissler, S., Simonsen, L., Grenfell, B.T., and Viboud, C. (2017). Human mobility and the spatial transmission of influenza in the United States. PLOS Computational Biology *13*, e1005382. 10.1371/journal.pcbi.1005382.

11. Pitzer, V.E., Viboud, C., Alonso, W.J., Wilcox, T., Metcalf, C.J., Steiner, C.A., Haynes, A.K., and Grenfell, B.T. (2015). Environmental Drivers of the Spatiotemporal Dynamics of Respiratory Syncytial Virus in the United States. PLOS Pathogens *11*, e1004591. 10.1371/journal.ppat.1004591.

12. Real World Evidence | Merative https://www.merative.com/real-world-evidence.

13. Clinical Classifications Software (CCS) for ICD-10-PCS (beta version) https://hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp.

14. Kyle Walker and Matt Herman (2023). tidycensus: Load US Census Boundary and Attributable Data as "tidyverse’ and ’sf’-Ready Data Frames. R package version 1.4.1, https://walker-data.com/tidycensus/.

15. Drew DeSilver (2019). “Back to school” means anytime from late July to after Labor Day, depending on where in the U.S. you live. Pew Research Center. https://www.pewresearch.org/short-reads/2019/08/14/back-to-school-dates-u-s/.

16. Kissler, S.M., Gog, J.R., Viboud, C., Charu, V., Bjørnstad, O.N., Simonsen, L., and Grenfell, B.T. (2019). Geographic transmission hubs of the 2009 influenza pandemic in the United States. Epidemics *26*, 86–94. 10.1016/j.epidem.2018.10.002.

17. Danchin, M.H., Rogers, S., Kelpie, L., Selvaraj, G., Curtis, N., Carlin, J.B., Nolan, T.M., and Carapetis, J.R. (2007). Burden of acute sore throat and group A streptococcal pharyngitis in school-aged children and their families in Australia. Pediatrics *120*, 950–957. 10.1542/peds.2006-3368.

**SUPPLEMENTAL MATERIALS**

**Supplementary Table 1: CCS to ICD Code Mapping**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ICD Revision | CCS Category | CCS Category Description | ICD Code | ICD Code Description |
| ICD-9 | 126 | Ot up rsp in | 340 | STREP SORE THROAT |
| ICD-10 | RSP006 | Other specified upper respiratory infections | J020 | Streptococcal pharyngitis |

**Supplementary Table 2: Regions with constituent states**

|  |  |
| --- | --- |
| Region | States |
| Northeast | Connecticut, Massachusetts, Maine, New Hampshire, Rhode Island, Vermont,  New Jersey, New York, Pennsylvania |
| Midwest | Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota |
| South | Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma, Texas |
| West | Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Alaska, California, Hawaii, Oregon, Washington" |

**Supplementary Table 3: Subregions with constituent states**

|  |  |
| --- | --- |
| Subregion | States |
| New England | Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont |
| Middle Atlantic | New Jersey, New York, Pennsylvania |
| East North Central | Indiana, Illinois, Michigan, Ohio, Wisconsin |
| West North Central | Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota |
| South Atlantic | Delaware, Washington DC, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia |
| East South Central | Alabama, Kentucky, Mississippi, Tennessee |
| West South Central | Arkansas, Louisiana, Oklahoma, Texas |
| Mountain West | Arizona, Colorado, Idaho, New Mexico, Montana, Utah, Nevada, Wyoming |
| Pacific West | California, Oregon, Washington |

**Supplementary Table 4: Region Difference Comparisons by Quarter**

Quarter 1 = January – March, Quarter 2 = April – June, Quarter 3 = July – September, Quarter 4 = October - December

|  |  |  |
| --- | --- | --- |
| Comparison | Quarter | Absolute Difference in Visits per 1000 People (95% CI) |
| Northeast-Midwest | 1 | 0.14 (-0.51-0.79) |
| Northeast-Midwest | 2 | 0.93 (0.61-1.26) |
| Northeast-Midwest | 3 | 0.41 (0.23-0.59) |
| Northeast-Midwest | 4 | 0.51 (0.25-0.78) |
| South-Midwest | 1 | 1.97 (1.41-2.54) |
| South-Midwest | 2 | 2.2 (1.8-2.6) |
| South-Midwest | 3 | 2.1 (1.92-2.29) |
| South-Midwest | 4 | 3.38 (2.97-3.8) |
| South-Northeast | 1 | 2.11 (1.58-2.65) |
| South-Northeast | 2 | 1.27 (0.93-1.61) |
| South-Northeast | 3 | 2.51 (2.25-2.78) |
| South-Northeast | 4 | 3.9 (3.55-4.25) |
| West-Midwest | 1 | 5.02 (3.98-6.07) |
| West-Midwest | 2 | 2.93 (2.49-3.37) |
| West-Midwest | 3 | 1.34 (1.09-1.59) |
| West-Midwest | 4 | 2.53 (2.27-2.79) |
| West-Northeast | 1 | 4.88 (4.13-5.64) |
| West-Northeast | 2 | 3.87 (3.36-4.37) |
| West-Northeast | 3 | 0.93 (0.76-1.1) |
| West-Northeast | 4 | 2.02 (1.7-2.33) |
| West-South | 1 | 7 (5.98-8.01) |
| West-South | 2 | 5.14 (4.59-5.68) |
| West-South | 3 | 3.44 (3.06-3.83) |
| West-South | 4 | 5.91 (5.36-6.46) |

**Supplementary Table 5:** State sinusoid phases with confidence intervals

|  |  |
| --- | --- |
| State | Phase (95% CI) |
| Alabama | 1.4 (1.09-1.7) |
| Arizona | 1.82 (1.49-2.15) |
| Arkansas | 1.37 (0.99-1.74) |
| California | 2.35 (2.06-2.63) |
| Colorado | 1.63 (1.39-1.88) |
| Connecticut | 2.61 (2.42-2.8) |
| Delaware | 2.41 (2.1-2.73) |
| Florida | 2.17 (1.78-2.56) |
| Georgia | 1.55 (1.2-1.91) |
| Idaho | 1.45 (1.17-1.73) |
| Illinois | 2.22 (1.97-2.47) |
| Indiana | 1.97 (1.74-2.21) |
| Iowa | 2.04 (1.78-2.3) |
| Kansas | 1.9 (1.6-2.2) |
| Kentucky | 1.65 (1.34-1.96) |
| Louisiana | 0.81 (0.41-1.21) |
| Maine | 2.38 (2.12-2.65) |
| Maryland | 2.5 (2.26-2.73) |
| Massachusetts | 2.7 (2.5-2.89) |
| Michigan | 2.42 (2.15-2.69) |
| Minnesota | 2.38 (2.12-2.64) |
| Mississippi | 0.86 (0.43-1.3) |
| Missouri | 1.79 (1.56-2.01) |
| Montana | 1.45 (1.21-1.7) |
| Nebraska | 1.95 (1.58-2.31) |
| Nevada | 1.97 (1.43-2.52) |
| New Hampshire | 2.54 (2.33-2.75) |
| New Jersey | 2.66 (2.44-2.88) |
| New Mexico | 1.65 (1.39-1.91) |
| New York | 2.53 (2.35-2.71) |
| North Carolina | 2.17 (1.91-2.44) |
| North Dakota | 1.98 (1.54-2.43) |
| Ohio | 2.2 (1.95-2.44) |
| Oklahoma | 1.44 (1.11-1.78) |
| Oregon | 2.45 (2.12-2.78) |
| Pennsylvania | 2.53 (2.32-2.74) |
| Rhode Island | 2.64 (2.37-2.91) |
| South Carolina | -- |
| South Dakota | 1.98 (1.64-2.33) |
| Tennessee | 1.67 (1.4-1.95) |
| Texas | 1.57 (1.29-1.85) |
| Utah | 1.69 (1.49-1.88) |
| Vermont | 2.17 (1.79-2.55) |
| Virginia | 2.46 (2.2-2.71) |
| Washington | 2.31 (1.97-2.65) |
| Washington DC | 2.74 (2.22-3.27) |
| West Virginia | 1.88 (1.54-2.22) |
| Wisconsin | 2.44 (2.19-2.69) |
| Wyoming | 1.42 (1.07-1.78) |

**A picture containing line, rectangle

Description automatically generated**

**Supplementary Figure 1**: Average membership in each state over the course of the study period. The dashed red line indicates the predetermined quality threshold of an average of 5,000 members per year.

**A picture containing screenshot, line, diagram, plot

Description automatically generated**

**Figure 2a:** Visits per 1000 people in each region over the 9-year observation period.

**A picture containing screenshot

Description automatically generated**

**Supplementary Figure 2b**: Average visits per 1000 people per year in each region with brackets showing 95% confidence intervals representing year-to-year variability.

**A picture containing screenshot, black

Description automatically generated**

**Supplementary Figure 2c**: Comparisons between each region pair. The y-axis shows the difference in average visits per 1000 people per year. The x-axis shows the negative log p-value from Welch’s two-sample t-test comparing the 9 observations from each region. The dashed line indicates the significance threshold of 0.05 corrected for multiple hypothesis testing with the Bonferroni correction, where points to the right of the line are statistically significant and points to the left are not.

**A picture containing screenshot, line, diagram, plot

Description automatically generated**

**Supplementary Figure 3a:** Visits per 1000 people in each subregion over the 9-year observation period.

**A picture containing screenshot

Description automatically generated**

**Supplementary Figure 3b:** Average visits per 1000 people per year in each subregion with brackets showing 95% confidence intervals representing year-to-year variability.

**A picture containing screenshot, line, colorfulness

Description automatically generated**

**Supplementary Figure 3c:** Comparisons of each subregion pair. The y-axis shows the difference in average visits per 1000 people per year. The x-axis shows the negative log p-value from Welch’s two-sample t-test comparing the 9 observations from each region. The dashed line indicates the significance threshold of 0.05 corrected for multiple hypothesis testing with the Bonferroni correction, where points to the right of the line are statistically significant and points to the left are not.

**A picture containing screenshot, colorfulness, art, pixel

Description automatically generated**

**Supplementary Figure 4a:** Quarterly average visits per 1,000 people by region. Quarter 1 is January-March, Quarter 2 is April-June, Quarter 3 is July-September, Quarter 4 is October-December. Error bars are 95% confidence intervals assuming normally distributed errors.

**A picture containing screenshot, colorfulness, candle, art

Description automatically generated**

**Supplementary Figure 4b:** Quarterly average visits per 1,000 people by subregion. Quarter 1 is January-March, Quarter 2 is April-June, Quarter 3 is July-September, Quarter 4 is October-December. Error bars are 95% confidence intervals assuming normally distributed errors.

**A picture containing screenshot, rectangle, diagram, square

Description automatically generated**

**Supplementary Figure 5a:** Region comparisons by month. Regions were compared via Welch’s two sample t-test and significance was determined based on a significance level of 0.05 corrected using the Bonferroni correction for multiple hypothesis testing.

A screenshot of a computer screen

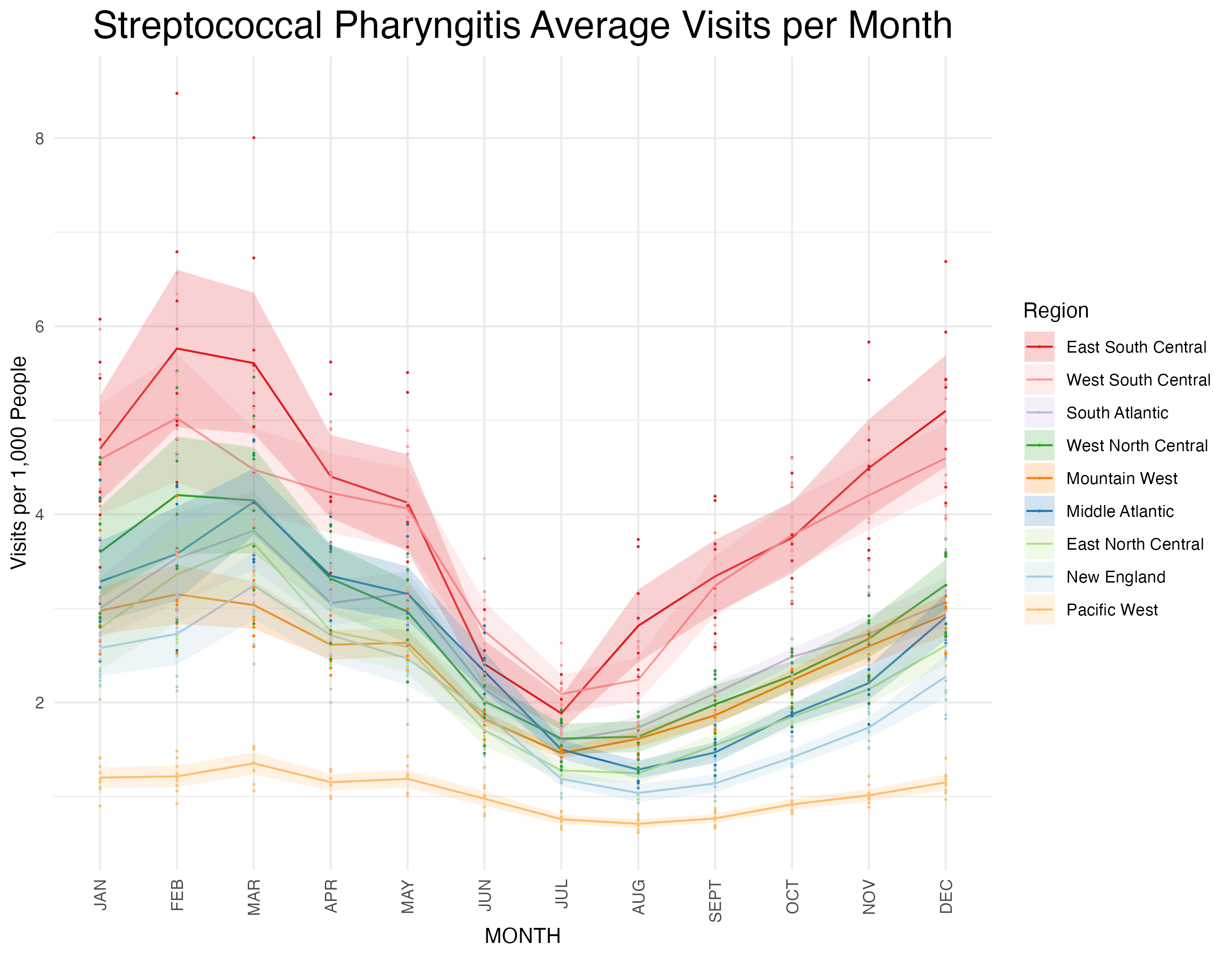
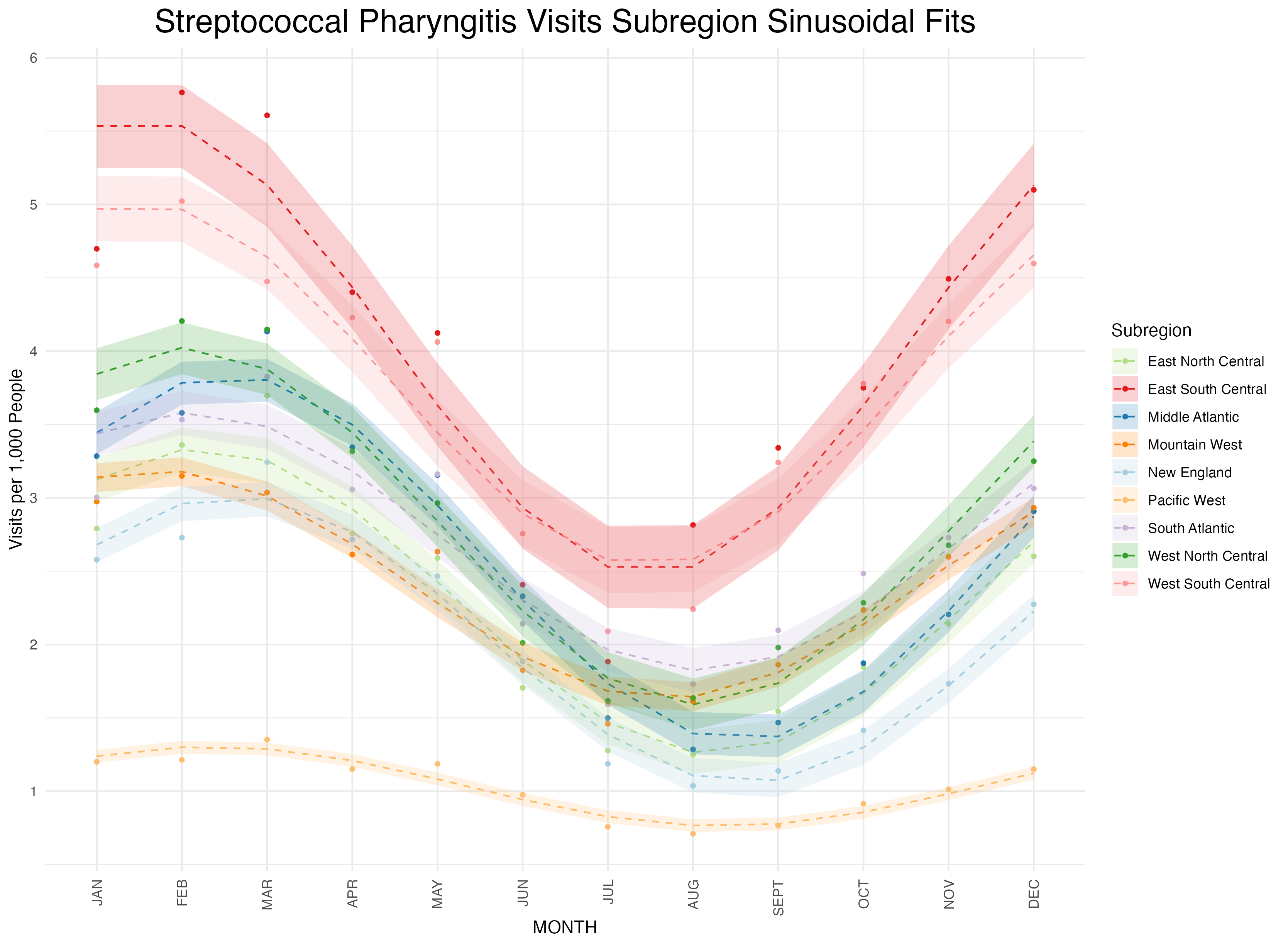
Description automatically generated with low confidence

**Supplementary Figure 5b:** Subregion comparisons by month. Subregions were compared via Welch’s two sample t-test and significance was determined based on a significance level of 0.05 corrected using the Bonferroni correction for multiple hypothesis testing.

**A picture containing screenshot, colorfulness, line, art

Description automatically generated**

**Supplementary Figure 6: Average visits patterns over the course of the year**. The average number of visits per 1000 people in the database over the 9-year observation period for all age groups is plotted for each census subregion. Shading represents the 95% confidence intervals depicting year-to-year variation.



**Supplementary Figure 7a**: Subregion sinusoidal fits. Shading represents 95% confidence intervals assuming normally distributed errors.

A picture containing screenshot

Description automatically generated

**Supplementary Figure 7b:** Subregion sinusoidal fit phases in temporal order. Brackets represent 95% confidence intervals around the phase estimations.

**A picture containing screenshot

Description automatically generated**

**Supplementary Figure 8a:** Region sinusoidal fits. Shading represents 95% confidence intervals assuming normally distributed errors.

A picture containing screenshot, line

Description automatically generated

**Supplementary Figure 8b:** Region sinusoidal fit phases in order. Brackets represent 95% confidence intervals around the phase estimations.

**A picture containing screenshot, square, line, symmetry

Description automatically generated**

**Supplementary Figure 9:** Individual state sinusoidal fits used to generate Figure 2. Points are average GAS pharyngitis visits in that month. Dashed lines represent sinusoid predictions, and shading represents 95% confidence intervals around sinusoid predictions assuming normally distributed errors.

**A picture containing space, screenshot, night, astronomy

Description automatically generatedSupplementary Figure 10**: Subregion sinusoid trough date compared to school start date. Sinusoid trough dates represent the sinusoid phase (peak in annual visits) offset by 6 months. Error bars on sinusoid trough dates are generated from sinusoid standard errors assuming normally distributed errors. School start dates are averaged across the subregion and error bars represent the range of school start dates in that region.