**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 an estimated 5.2 million outpatient visits for pharyngitis each year in the United States (U.S.) and can result in serious post-infection immunologic sequelae, such as acute rheumatic fever and post-streptococcal glomerulonephritis. While other common respiratory tract infections, such as influenza and respiratory syncytial virus, tend to follow epidemic patterns of spread geographically across the U.S., the timing and geography of GAS pharyngitis remain unclear.

Methods

This study used 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 the most visits per member (yearly average 39.11 visits per 1000 people, 95% CI: 36.21-42.01), and the West had the fewest (yearly average 17.63 visits per 1000 people, 95% CI: 16.76-18.49). Visits peaked in winter months and nadired in summer months. The South had the earliest start in the rise of visits, with differences between the South and other regions most pronounced in the late summer through early winter. GAS pharyngitis visits peaked earliest in southern states in December to January and latest on the coasts in March. The onset of the rise in GAS pharyngitis visits correlated with, but preceded, average regional school start times.

Conclusions

Private insurance claims showed that the burden and timing of GAS pharyngitis varied across the continental U.S., with highest overall rates and earliest start and peak in cases in the South. Understanding the drivers of these regional differences in the timing of GAS pharyngitis visits will be important for identifying and targeting prevention 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 complications, such as 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. Guidelines advise treatment after diagnosis by a rapid antigen detection test (RADT) or throat culture, rather than by clinical features alone.2,3

GAS pharyngitis is more common in the winter and spring months,2,4 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.5,6 Understanding trends in the timing and location of these important infections aids in defining risks and can indicate where and when to target interventions.

GAS can also cause invasive disease (iGAS), which is defined as bacteria cultured from a typically sterile body site.7 In contrast to the usually mild GAS pharyngitis, iGAS can be life threatening and require intensive care. After reported iGAS cases in the U.S. 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.8 Better understanding of the epidemiology of GAS pharyngitis would enable comparison with iGAS and help elucidate the extent to which and in which ways the two are linked.

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 As use of antibiotics promotes drug resistance not only in the target pathogen but also in other bacteria via bystander selection9, reducing the incidence of GAS pharyngitis will help reduce selective pressure from antibiotics. GAS remains penicillin-susceptible, but resistance has emerged to second-line antibiotics such as macrolides (e.g., azithromycin) and lincosamides (e.g., clindamycin).10 Efforts to reduce GAS pharyngitis disease burden include the development of vaccines to protect against GAS, with candidates now in clinical trials.11,12 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 the temporal and geographic distribution of GAS pharyngitis in the U.S.

To address the need for improved characterization of GAS epidemiology in the U.S., we used outpatient insurance claims data from private insurers to measure variation in visits for GAS pharyngitis 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).13 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).14 Using CCS codes allowed for consistent identification of GAS pharyngitis cases across both ICD9 and ICD10. Visits were included if a diagnosis code indicating 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 enrolled 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 enrolled 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 census-determined 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 census-reported age and sex distribution. State-level population counts were obtained from the 2011-2015 American Community Survey (ACS) and accessed using the tidycensus R package15.

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 continental U.S. states were included (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. 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 = ). Estimates of the amplitude, phase, and offset were 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 to integers by taking the phase modulo 12. Taking each integer value to be the first day of the specified month (e.g., 1 is January 1st), the phases fell between 0-11.99 where 0 is the first day of December and 11.99 is the last day of November.

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.

*Peak and Distance Correlation*

Population-weighted state centroids from 2010 were extracted from the USpopcenters R package.16 Subregion centroids were calculated using a population-weighted average of the state centroids. The reference point was calculated as a population weighted average between the East and West South Central subregions, given their very similar sinusoid phases. Euclidean distance from each subregion centroid to the reference point was calculated and correlated to the sinusoid phase using Pearson’s correlation coefficient.

*School Start Time Correlation*

Data on 2019 school start dates was curated by Pew Research Center.17 This dataset contains a 509-district sample of the >13,000 public school districts in the United States, including the 10 largest school districts in each state where possible, and the 100 largest school districts in the country. Altogether, this dataset represents and estimated 36% of public-school students in the United States. Data were filtered to include only 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. The minimum visit date was calculated by cataloging the month with the fewest visits in each subregion for each of the 9 years of observation, which was either 7 or 8 in all cases, and averaging. Correlation between average school start date and minimum visit date in each subregion was calculated using Pearson’s correlation coefficient. The confidence interval was calculating via bootstrapping; 1000 bootstrapped samples were randomly drawn with replacement from the set of 9 minimum visit months from each year in each subregion, and 1000 bootstrapped samples were also randomly drawn from the vector of school start dates within each subregion. This process created 1000 pairs of vectors of values for each subregion (one for minimum visit, one for school start date) which were then correlated using Pearson’s correlation coefficient.

**RESULTS**

The four census regions (Midwest Northeast, South, West) differed in the number of GAS visits per 1,000 individuals (Figure S2). The South had the most yearly visits per 1000 people at 39.11 (95% CI: 36.21-42.01), and the West the fewest at 17.63 (95% CI: 16.76-18.49. At a subregional level, the East South Central region led all regions with 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 (Figure S3). 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.

At the level of 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 calendar 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 (Figure S4). Throughout the year, the West had fewer visits per 1000 people than the other 3 regions. 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). Similarly, there were more statistically significant differences between regions and subregions in months in the second half of the year (Figure S6).

The annual peak in GAS pharyngitis visits occurred earliest in the South, followed by peaks in adjacent states and through the Mountain West, with the latest peaks in coastal states (Figure 2). Distance from the South was correlated with peak time (r = 0.725, Figure S7).

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 (Figures S8-10), 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 were moderately correlated with minimum visit dates (Pearson’s correlation *r*  = 0.53, 95% CI: -0.04-0.91)(Figure 3, Figure S11). Schools in the South tend to start earlier, and schools in the Northeast tend to start later. Minimum visit dates represent the timing of the uptick in GAS pharyngitis visits in each region. In all subregions, the uptick date preceded the school start date (mean number of days preceding: 36.22, range: 28-47 days; Figure S11).

**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 the fewest 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. The yearly trends in visits also indicate that the annual epidemic spread is first observed in southern states, particularly in Louisiana and Mississippi, followed by the rest of the country roughly in order of increasing distance from the south. This could represent the progression of environmental or social conditions that promote GAS pharyngitis or the spread of the pathogen itself. These findings also indicate that states that have earlier peaks have higher disease burden, which could mean either that the conditions that promote earlier peaks in GAS pharyngitis also lead to more transmission or that states that have environmental or social conditions more amenable to the spread of GAS pharyngitis serve as a focal point of transmission that then expands to the rest of the country.

The reason states in the Pacific West showed markedly fewer visits year-round compared to other states is not clear. Systematic underreporting in private insurance claims could explain these differences, although these may be unlikely to cluster regionally. Other potential reasons include local differences in the distribution of GAS strains and their propensity to cause pharyngitis, fewer susceptible hosts due to increased immunity from prior infection or a cross-protective pathogen, environmental factors or social and behavioral factors that reduce transmissibility. Age-stratified serosurveys, as well as genomic studies that characterize the prevalence and diversity of different strains regionally, could help distinguish among these hypotheses.

The timing of GAS pharyngitis correlated with school start dates. Subregions in the South with earlier upticks in GAS pharyngitis visits also started school earlier than coastal subregions. Given that uptick dates preceded school start dates by over a month, it is unlikely that school attendance initiated transmission in a given season, but it is possible that changes in contact patterns among children associated with the start of school also facilitate the spread of GAS pharyngitis. This theory has been posited for other respiratory infections.18 School initiation is a known driver of respiratory infections that predominantly affect children, as evidenced by the disruption in usual patterns of respiratory infection spread during the COVID-19 pandemic school closures. Further examination of the relationship between rates of GAS pharyngitis and different levels of school closures in different regions could help elucidate the role of school as a nidus of GAS pharyngitis spread.

GAS pharyngitis spatiotemporal patterns were similar to but subtly different from previously described trends in yearly influenza and RSV. RSV hospitalizations peak earlier in Florida (November/December) than in other states, with environmental factors such as mean vapor pressure, minimum temperature, precipitation, and seasonal variation in potential evapotranspiration potentially accounting for timing differences across states.6 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.5 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 clarify these behavioral and environmental contributors, and how the changing climate will impact them.

GAS 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, although this may be due to small sample size,4 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 estimated 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 2001-2002,19 and another study from the U.S. found that peaks in GAS pharyngitis occurred in the winter months between 2010-2019.4

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 graph of different colored lines

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. South Carolina is excluded from the analysis according to data use agreement. 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 (<https://github.com/gradlab/StrepPharyngitis/blob/main/figures/finalgif.gif>).

**A graph with many colored dots

Description automatically generated**

**Figure 3**: Correlation between subregion school start date and GAS pharyngitis minimum visits date. Average school start date is plotted on the x-axis and minimum visit date 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. Points are colored according to their subregion.

**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. Shulman, S.T., Bisno, A.L., Clegg, H.W., Gerber, M.A., Kaplan, E.L., Lee, G., Martin, J.M., and Van Beneden, C. (2012). Clinical Practice Guideline for the Diagnosis and Management of Group A Streptococcal Pharyngitis: 2012 Update by the Infectious Diseases Society of America. Clinical Infectious Diseases *55*, e86–e102. 10.1093/cid/cis629.

4. 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.

5. 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.

6. 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.

7. 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/.

8. Increase in Invasive Group A Strep Infections, 2022-2023 | CDC (2023). https://www.cdc.gov/groupastrep/igas-infections-investigation.html.

9. 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.

10. 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.

11. 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.

12. 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.

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

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

15. 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/.

16. Krieger, N. (2023). USpopcenters: United States Centers of Population (Centroids).

17. 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/.

18. 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.

19. 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**

**All code available at**

**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 graph of state and state

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 screenshot of a graph

Description automatically generated**

**Supplementary Figure 2:** Visits per 1000 people per year in each region. Panel A: Visits per 1000 people in each region over the 9-year observation period. Panel B: Average visits per 1000 people per year in each region with brackets showing 95% confidence intervals representing year-to-year variability. Panel C: 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 graph of different colored lines

Description automatically generated**

**Supplementary Figure 3** Visits per 1000 people per year in each subregion. Panel A: Visits per 1000 people over the 9-year observation period. Panel B **:**Average visits per 1000 people per year in each subregion with brackets showing 95% confidence intervals representing year-to-year variability. Panel C: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 graph with lines and numbers

Description automatically generated

**Supplementary Figure 4: 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.

**A screenshot of a graph

Description automatically generated**

**Supplementary Figure 5:** Quarterly average visits per 1,000 people. 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. Panel A: By region. Panel B: By subregion.

A screenshot of a graph

Description automatically generated

**Supplementary Figure 6:** Monthly statistical comparison. Regions or 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. Panel A: Regions. Panel B: Subregions.

**A map of the united states

Description automatically generated**

**Supplementary Figure 7:** Correlation of Subregion Centroid with GAS Pharyngitis Visit Peak. Panel A shows the population-weighted centroids of each subregion (black dots). The red dot indicates the reference point for distance, which is a weighted average of the centroids of the East and West South Central subregions. Panel B shows sinusoid phase on the x-axis and distance from reference point on the y-axis. Points are colored according to their subregion. The dotted line is a linear regression and shading represents a 95% of the linear model’s predictions.

**A screenshot of a graph

Description automatically generated**

**Supplementary Figure 8**: Subregion sinusoidal fits. Panel A: GAS pharyngitis visit predictions according to subregional sinusoidal fitting. Points represent average visits. Shading represents 95% confidence intervals assuming normally distributed errors. Panel B: Sinusoidal fit phases in temporal order. Brackets represent 95% confidence intervals around the phase estimations.

**A screenshot of a graph

Description automatically generated**

**Supplementary Figure 9:** Region sinusoidal fits. Panel A: GAS pharyngitis visit predictions according to regional sinusoidal fitting. Points represent average visits. Shading represents 95% confidence intervals assuming normally distributed errors. Panel B: Region sinusoidal fit phases in order. Brackets represent 95% confidence intervals around the phase estimations.

**A graph of different colored lines

Description automatically generated**

**Supplementary Figure 10:** 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 screenshot of a graph

Description automatically generatedSupplementary Figure 11**: Subregion minimum visit date compared to school start date. Panel A shows average minimum visit dates for each subregion visit trend with a dashed vertical line. Panel B shows average minimum visit dates with 95% confidence intervals assuming normally distributed errors (red) and school start dates are averaged across the subregion and error bars represent the range of school start dates in that region (blue).