HHS Public Access

Author manuscript

J Infect Dis. Author manuscript; available in PMC 2024 April 30.

Published in final edited form as:

J Infect Dis. 2022 August 26; 226(3): 546–553. doi:10.1093/infdis/jiac162.

Cluster Transmission Drives Invasive Group A *Streptococcus* Disease Within the United States and Is Focused on Communities Experiencing Disadvantage

Benjamin Metcalf,
Srinivas Nanduri,
Sopio Chochua,
Yuan Li,
Katherine Fleming-Dutra,
Lesley McGee,
Bernard Beall

Respiratory Diseases Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Abstract

Background.—Group A streptococci (GAS), although usually responsible for mild infections, can sometimes spread into normally sterile sites and cause invasive GAS disease (iGAS). Because both the risk of iGAS disease and occurrence of outbreaks are elevated within certain communities, such as those comprising people who inject drugs (PWID) and people experiencing homelessness (PEH), understanding the transmission dynamics of GAS is of major relevance to public health.

Methods.—We used a cluster detection tool to scan genomes of 7552 *Streptococcus pyogenes* isolates acquired through the population-based Active Bacterial Core surveillance (ABCs) during 2015–2018 to identify genomically related clusters representing previously unidentified iGAS outbreaks.

Results.—We found that 64.6% of invasive isolates were included within clusters of at least 4 temporally related isolates. Calculating a cluster odds ratio (COR) for each *emm* type revealed that types vary widely in their propensity to form transmission clusters. By incorporating additional epidemiological metadata for each isolate, we found that *emm* types with a higher proportion of

Correspondence: B. Metcalf, MS, Respiratory Diseases Branch, Centers for Disease Control and Prevention, 1600 Clifton Road NE, MSC H18-1, Atlanta, GA 30333 (ycm6@cdc.gov).

Potential conflicts of interest. All authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

Disclaimer. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://jid.oxfordjournals.org/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

cases occurring among PEH and PWID were associated with higher CORs. Higher CORs were also correlated with *emm* types that are less geographically dispersed.

Conclusions.—Early identification of clusters with implementation of outbreak control measures could result in significant reduction of iGAS.

Keywords

invasive disease; outbreaks; *Streptococcus pyogenes*

Streptococcus pyogenes, or group A *Streptococcus* (GAS), is a Gram-positive beta-hemolytic human pathogen that causes a wide spectrum of diseases that vary markedly in severity [1]. Milder disease manifestations include acute pharyngitis and the highly contagious skin infection impetigo. Together, these 2 common maladies cause hundreds of millions of cases per year globally [2]. Much less frequently, GAS invades normally sterile sites to cause severe infections known as invasive GAS disease (iGAS). Bacteremic cellulitis and bacteremia without a focus are the most common manifestations of iGAS; rarer and more severe syndromes such as necrotizing fasciitis, streptococcal toxic shock, and meningitis also occur [3]. Previous work has shown that risk of iGAS is elevated within residents of long-term care facilities (LTCFs), people experiencing homelessness (PEH), and people who inject drugs (PWID) [4, 5]. Understanding the transmission dynamics of GAS as it spreads within a population and especially within these groups is a major public health concern.

A key component in tracking the transmission of GAS within a community is being able to type each isolate based on specific genomic and phenotypic characteristics. The most common way to categorize GAS is the *emm* genotype-based typing scheme, which identifies alleles (types) based on the variability within the 5′ terminal end of the *emm* virulence gene [6]. *emm* typing has been widely used in elucidating how different strains of GAS spread throughout a population. Although *emm* typing has elucidated the overall dynamics of the transmission of major GAS genetic complexes, it lacks the granularity to reliably identify outbreaks. Because the more common *emm* types often have multiple independent circulating strains that are widely geographically and temporally interspersed, *emm* typing on its own lacks sufficient resolution for outbreak detection. A more discriminating method for identifying outbreaks relies on whole-genome sequencing (WGS) whereby variant alterations observed across the entire genome are used to identify highly related strains.

Although the transmission of *S pyogenes* is local, the patterns of transmission and disease dynamics play out on a much larger scale. Being able to recognize these disease trends across the United States, especially among invasive infections, is critical for guiding public health policy. To that end, the Centers for Disease Control and Prevention (CDC) developed the Active Bacterial Core surveillance (ABCs) program as an active laboratory- and population-based surveillance system for key invasive bacterial pathogens including GAS [3]. The ABCs sites send both a detailed case report form and the bacterial isolate for each invasive disease case occurring within the catchment area. Since 2015, the *Streptococcus* Laboratory has performed WGS all GAS isolates acquired through ABCs [7, 8]. By

combining the epidemiological and genomic data of each isolate, we provide a unique framework to study iGAS spread and host factors that influence transmission.

We aimed to assess (1) the extent to which isolates collected in 2015–2018 from ABCs iGAS cases are part of transmission clusters and (2) whether the propensity to cluster varies across *emm* types. Second, we determined whether cases that are occurring in populations at higher risk of iGAS, such as PEH, PWID, or residents of LTCFs, are more prone to transmission clustering compared with cases from the rest of the population.

METHODS

Isolates

The ABCs conducts active laboratory and population-based surveillance for iGAS infections in 10 states (complete states or selected counties), representing more than 34.4 million persons (https://www.cdc.gov/abcs/reports-findings/survreports/gas18.html). The ABCs defines an iGAS case as illness in a surveillance area resident with isolation of GAS from a normally sterile site, or from a wound culture if accompanied by necrotizing fasciitis or streptococcal toxic shock syndrome. Surveillance staff at sites complete a standardized case report form that includes basic demographic characteristics and risk factors of infection on all cases, and they coordinate collection and transfer of isolates to the CDC's *Streptococcus* Laboratory for characterization. All available iGAS isolates from cases identified in 2015–2018 (7552 of 8680, 87%) were included.

Whole-Genome Sequencing

Group A streptococci chromosomal deoxyribonucleic acid preparation, library construction, and WGS generation for the iGAS isolates acquired through ABCs were performed as previously described [9]. The bioinformatics pipeline describing methods for assignments of *emm* types and other parameters has been described [7, 8].

An acceptable Illumina sequencing run required a Phred quality score ≥Q30, V2 > 75%, V3 > 70%, yield V2 ≥ 6G, and V3 ≥ 10G. Inclusion criteria for genome assemblies specified total contigs ≤250 or, if greater than 250, successful *emm* type and multilocus sequence type calls using the *Streptococcus* Laboratory sequence-based GAS strain characterization pipeline [7].

Transmission Cluster Detection

A schematic describing the cluster detection algorithm is provided in the Supplement (Supplementary Figure 1). The program begins with an initial filtering step that reads in genomic assemblies and calculates their pairwise mash distance using Mash v1.1 [10]. If an isolate does not have at least 2 neighbors with a Mash distance of less than 0.00001, then that isolate is removed from further analysis. Once the initial filtering step is completed, an all-vs-all pairwise single-nucleotide polymorphism (SNP) distance matrix is generated using Nucmer from the MUMmer package v3.9.4 [11]. To improve computational efficiency, this step of the program is run on a high-performance computing system where the pairwise comparisons for each isolate are run as separate jobs. Once all the jobs finish, the values

in the SNP distance matrix are transformed using the equation $e^{-(s/2.3)-1}$, where s is the number of SNPs between 2 genomes. The purpose of this transformation is to give more closely related genomes a higher weight and to invert the values so that genome pairs with fewer SNPs have a higher distance score. The transformed distances are reformatted into an adjacency list and then fed into the clustering tool MCL v14–137, which uses the SNP-based distance graph to identify transmission clusters by using a Markov clustering algorithm [12]. Finally, each isolate belonging to a cluster with a count of at least 4 is annotated with the following epidemiological metadata: patient age, ABC surveillance site, culture date, LTCF residence status (yes/no), PEH status (yes/no), and past or current injection drug use status (yes/no). Due to the strong correlation between PEH and injection drug use, the union of these 2 attributes (defined as either PEH or PWID) was also used in this analysis and is referenced as PEH/PWID status (yes/no). The cluster detection code used in this analysis is publicly available in the GitHub repository (https://github.com/benjamesmetcalf/gas_clusters_code).

Statistical Analysis

The cluster odds ratio (COR) for each *emm* type is the number of cluster-associated isolates over the number of noncluster isolates referenced against all other *emm* types. To illustrate, if a is the number of cluster isolates for the focal *emm* type, b is the number of noncluster isolates for the focal *emm* type, c is the total number of cluster isolates for all other *emm* types, and d is the number of noncluster isolates for all other *emm* types, then COR = $(ad)/(bc)$. Cluster odds ratios were calculated for all *emm* types with at least 10 cases using the "epi.2by2" function in the R "epiR" package v2.0.19 [13]. The Haldane-Anscombe correction was used when calculating odds ratios for *emm* types with no cluster-associated cases. Correlation between COR and epidemiological host attributes was assessed with the Kendall's tau correlation coefficient and visualized as a scatterplot using the "ggplot2" package in R v3.6.1 [13]. To measure the degree to which *emm* types are sporadically emergent and geographically targeted (SEGT), we define a metric called SEGT that uses the Simpson's diversity index to quantify level of surveillance site heterogeneity exhibited by each *emm* type. Lower index values closer to 0 represent *emm* types with lower geographic diversity and a stronger SEGT signal (strong-SEGT), whereas higher diversity scores closer to 1 represent endemic types found across surveillance sites indicating a weaker SEGT pattern (weak-SEGT). The index was calculated with the "diversity" function in the R "vegan" package v2.5.6 using invasive case counts aggregated over *emm* types for each of the 10 GAS surveillance states (California, Colorado, Connecticut, Georgia, Maryland, Minnesota, New Mexico, New York, Oregon, and Tennessee).

RESULTS

Large Proportion of Invasive Group A Streptococci Disease Is Cluster-Associated

Based on genomic relatedness, the algorithm identified 332 transmission clusters across 39 *emm* types (112 *emm* subtypes). It is notable that these cluster-associated isolates represented 64.6% (n = 4881) of the total number of iGAS isolates available from ABCs during 2015–2018 (Supplementary Appendix 2). Transmission clusters had an average size of 15 isolates (ranging from 4 to 372 isolates), and the mean duration from first to most

recent isolate in a cluster was 836 days (range, 31–1448 days). The average pairwise genomic distance for isolates in a cluster ranged from 0.5 to 17.9 SNPs. Weighting the average pairwise distance by the size of the cluster, GAS clusters had an average overall pairwise distance of 7.7 SNPs (unweighted average was 7.2 SNPs). Cluster-associated isolates were from case-patients whose median age was 55 (range, 0–101) years. Across all cluster-associated isolates, 7.4% were recovered from LTCF residents, 14.4% from PEH, 17.1% from PWID, 24.6% from PEH/PWID, and 68.1% from case-patients with none of these 3 risk factors. The impact of clustering on the invasive disease burden within these communities was substantial with cluster-associated isolates accounting for 88.1%, 81.1%, and 73.0% of total cases among PEH, PWID, and residents of LTCF, respectively. Patients living in LTCFs and PEH/PWID both had higher odds of being associated with cluster transmission. Residents of LTCFs were 1.51 times (95% confidence interval [CI], 1.23–1.86) more likely to be cluster-associated than individuals not residing in LTCF, whereas PEH/PWID were 3.22 times (95% CI, 2.79–3.73) more likely to be linked to a transmission cluster versus non-PEH/PWID.

Clusters that contained at least 1 PEH/PWID case had, on average, 32.0% of their cases comprising PEH/PWID isolates, whereas clusters that contained at least 1 LTCF case had, on average, 7.8% of their cases representing LTCF isolates. The relatively low percentages for both groups reveal that clusters that included LTCF residents and PEH/PWID can also spread into the broader general population as well.

emm Types With a Higher Proportion of People Experiencing Homelessness and People Who Inject Drugs Are Associated With a Higher Cluster Odds Ratios

Comparing the proportion of cluster-associated isolates from total invasive cases across *emm* types suggests they diverge widely in their predisposition to cluster (Figure 1). To quantify clustering propensity across GAS lineages, we calculated a COR for all cluster-associated *emm* types (Table 1) where the cluster odds of a focal *emm* type was referenced against the cluster odds of all other *emm* types. The COR analysis confirms that *emm* types vary widely in their tendency to form clusters. Certain *emm* types (eg, 60 and 82) were much more likely to generate clusters compared with background, whereas other *emm* types (eg, 87 and 28) had a much lower clustering propensity.

The wide distribution in COR prompted the question of whether there were underlying features of *emm* types that correlated with this variation in clustering propensity. In particular, we wanted to assess whether LTCF, PEH, or PWID status of the host impacted the propensity of an *emm* type to form transmission clusters. For each type, we calculated the percentage of cases that were isolated from persons experiencing homelessness or individuals who inject drugs and percentage of cases from LTCF residents (Table 1). As shown in Figure 2, *emm* types with a higher proportion of cases from PEH or PWID were positively correlated with COR (Kendall tau = 0.47, $P = 1.51 \times 10^{-5}$) and thus had a higher propensity to generate transmission clusters. Cluster-associated isolates from strong-SEGT *emm* types were more likely to comprise PEH or PWID (Kendall tau = −0.36, $P = .0013$) as well. In contrast, the percentage of LTCF residence did not have any significant association with *emm* type COR.

We were also interested in determining whether *emm* types that were unevenly distributed across our sampling sites had a higher propensity to form clusters: the expectation being that the geographic patchiness of these types was a function of rapid community transmission. To quantify the amount of surveillance site heterogeneity within each *emm* type, we calculated the SEGT index using invasive counts aggregated by geographic site (Supplementary Table 1). The SEGT diversity metric is represented in Figure 2 as the datapoint color gradient, which ranges from dark blue (low diversity and strong-SEGT) to light blue (high diversity and weak-SEGT). As hypothesized, an inverse correlation was observed between site diversity and COR (Kendall tau = −0.30, $P = .0078$) where *emm* types with higher CORs were less likely to be geographically spread across surveillance sites.

The clustering algorithm used in this analysis, MCL, did not require a threshold distance for group designation. This allowed for identification of longer duration clusters because average pairwise distance between outbreak isolates will grow over time. However, without a distance cutoff, some clusters had a relatively high average distance that could reflect unlinked community transmission. Thus, it is reasonable to ask whether the associations among COR, PEH/PWID, and the SEGT metric hold if we restrict the analysis to transmission clusters with an average pairwise distance below a certain threshold. Using an average pairwise distance cutoff of less than or equal 10 SNPs, we identified 266 transmission clusters, 66 fewer than without the constraint, but the significant associations between COR and PEH/PWID (Kendall tau = 0.54, $P = 7.6 \times 10^{-}7$), SEGT and PEH/PWID (Kendall tau = −0.35, $P = .002$), and SEGT and COR (Kendall tau = −0.34, $P = .002$) remained.

DISCUSSION

The large proportion of cluster-associated isolates identified through population-based surveillance suggests that cluster transmission within the community plays a major role in iGAS epidemiology. In previous work, Turner et al [14] performed a retrospective cohort study of 93 patients and identified possible instances of cryptic community transmission. In their investigation, they found 3 clusters, 2 clusters of *emm*1 (2 and 3 isolates each), and a third 4-isolate cluster of *emm*3 in which the patients did not share any identifiable links in healthcare settings. Our results suggest that cluster transmission might be the major driver of iGAS illnesses at the population level. The extensive clustering behavior of iGAS of almost all *emm* types shown in this analysis likely stems from the extremely effective transmission of noninvasive GAS infections. It has been observed that there is a broad overlap between strains that are predominant in pediatric pharyngitis and impetigo and strains causing invasive disease within a given period [15–17].

This investigation also found that strong-SEGT *emm* types were more likely to be associated with iGAS clusters, and these clusters were more likely to have isolates from case-patients who were PEH/PWID. Several studies have documented outbreaks of strong-SEGT *emm* types among PEH and PWID [18–23]. For example, a study performed at a hospital in the United Kingdom [20] compared the *emm* type distributions between PWID treated for GAS bacteremia and controls who do not use drugs. Their analysis revealed that although *emm* types 1 and 89 were commonly found in controls, iGAS cases among PWID were associated

with *emm* types 82 and 83, which were infrequent causes of invasive disease in England and Wales [19]. As others have noted [24], we suspect that these quickly emergent iGAS clusters among populations of younger to middle-aged adults experiencing homelessness and who inject drugs are reflective of noninvasive skin infection reservoirs. Inadequate hygiene, overcrowded living conditions within shelters, and injection drug use could contribute to iGAS cases originating from skin carriage or infection.

The mechanism underlying the clustering of strong-SEGT *emm* types and their association with certain communities may be explained, in part, by their relative inability to cause invasive disease in individuals without these risk factors. By virtue of their prevalence, it is possible that most adults will have been exposed to, and gained some immune protection from, common widely disseminated *emm* types (eg *emm1*, *emm89*, *emm12*, *emm28*). Thus, although they can cause invasive transmission clusters, spreading within a healthy population will most often lead to mild infections with only sporadic cases of invasive disease, resulting in a lower cluster odds ratio. Alternatively, the transmission cluster analysis suggests that strong-SEGT invasive *emm* types are not transmitted efficiently within an immunologically naive healthy population and, instead, mostly spread within networks of individuals who have risk factors and may have had little previous exposure. Their lack of previous exposure in conjunction with certain risk factors such as crowding, sleeping outdoors, and upper extremity skin breakdown create an environment in which strong-SEGT *emm* types can progress rapidly and cause serious disease [22]. It is possible that once the susceptible host population within this network is depleted, the strain is cleared from the community resulting in strong-SEGT *emm* types with disproportionately higher CORs.

The experimental 30-valent GAS vaccine would target broadly distributed, common *emm* types, including *emm1*, *emm3*, *emm28*, and *emm89*, which have large pediatric pharyngitis reservoirs [7] and have been consistently predominant within ABCs for the past 15 years. The reservoir for iGAS cases from these common, broadly distributed *emm* types is thought to be children with pharyngitis or asymptomatic GAS carriage. Thus, by preventing pediatric pharyngitis, the vaccine might be expected to reduce iGAS among nonvaccinated groups within the United States, similar to the reductions in invasive pneumococcal disease seen among adults after introduction of pneumococcal conjugate vaccines in young children [25]. This experimental GAS vaccine, in addition to targeting types that have been predominant during the past few decades, also targets several sporadically emergent types such as *emm49*, *emm82*, and *emm92* that disproportionally affect PWID and PEH. It is conceivable that future vaccination policy could include recommendations for implementation strategies for these individuals at disproportionate risk of invasive GAS.

CONCLUSIONS

In this analysis, we used a cluster detection algorithm to find iGAS outbreaks circulating within ABCs catchment areas using genomic sequence data from isolates acquired from 2015 through 2018. This study found that a large proportion of iGAS isolates were cluster-associated with 65% belonging to a genomically related group of at least 4 isolates. In addition, we found that *emm* types that had a higher proportion of PEH/PWID or were more sporadically emergent and geographically targeted were positively correlated with COR.

Given that 25% of cluster-associated isolates were recovered from PEH/PWID (accounting for 83% of total PEH/PWID cases), a significant reduction of iGAS could be achieved if these transmission clusters were identified early and control efforts were undertaken to limit outbreak progression. Strategies that are effective in controlling outbreaks of GAS among PEH and PWID are needed and could include infection prevention and control measures, antibiotics, and other decolonization measures that can be implemented within facilities such as homeless shelters and facilities serving these populations and within the broader community. In addition, a GAS vaccine could provide critical outbreak control when transmission is occurring within communities of PEH and PWID.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Financial support.

This work was funded by the Centers for Disease Control and Prevention as part of normal responsibilities.

References

1. Walker MJ, Barnett TC, McArthur JD, et al. Disease manifestations and pathogenic mechanisms of group A *Streptococcus*. Clin Microbiol Rev 2014; 27:264–301. [PubMed: 24696436]
2. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. Lancet Infect Dis 2005; 5:685–94. [PubMed: 16253886]
3. Nelson GE, Pondo T, Toews KA, et al. Epidemiology of invasive group A streptococcal infections in the United States, 2005–2012. Clin Infect Dis 2016; 63:478–86. [PubMed: 27105747]
4. Thigpen MC, Richards CL, Lynfield R, et al. Invasive group A streptococcal infection in older adults in long-term care facilities and the community, United States, 1998–2003. Emerg Infect Dis 2007; 13:1852–9. [PubMed: 18258035]
5. Sierra JM, Sánchez F, Castro P, et al. Group A streptococcal infections in injection drug users in Barcelona, Spain: epidemiologic, clinical, and microbiologic analysis of 3 clusters of cases from 2000 to 2003. Medicine (Baltimore) 2006; 85:139–46. [PubMed: 16721256]
6. Beall B, Facklam R, Thompson T. Sequencing emm-specific PCR products for routine and accurate typing of group A streptococci. J Clin Microbiol 1996; 34:953–8. [PubMed: 8815115]
7. Chochua S, Metcalf BJ, Li Z, et al. Population and whole genome sequence based characterization of invasive group A streptococci recovered in the United States during 2015. mBio 2017; 8:e01422–17. [PubMed: 28928212]
8. Li Y, Rivers J, Mathis S, et al. Genomic surveillance of *Streptococcus pyogenes* strains causing invasive disease, United States, 2016–2017. Front Microbiol 2020; 11:1547. [PubMed: 32849323]
9. Metcalf BJ, Chochua S, Gertz RE, et al. Using whole genome sequencing to identify resistance determinants and predict antimicrobial resistance phenotypes for year 2015 invasive pneumococcal disease isolates recovered in the United States. Clin Microbiol Infect 2016; 22:1002.e1–8.
10. Ondov BD, Treangen TJ, Melsted P, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 2016; 17:132. [PubMed: 27323842]
11. Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large genomes. Genome Biol 2004; 5:R12. [PubMed: 14759262]
12. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 2002; 30:1575–84. [PubMed: 11917018]
13. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2020.

14. Turner CE, Bedford L, Brown NM, et al. Community outbreaks of group A *Streptococcus* revealed by genome sequencing. Sci Rep 2017; 7:8554. [PubMed: 28819111]
15. Shulman ST, Tanz RR, Kabat W, et al. Group A streptococcal pharyngitis serotype surveillance in North America, 2000–2002. Clin Infect Dis 2004; 39:325–32. [PubMed: 15306998]
16. Shulman S, Tanz R, Dale J, et al. Seven-year surveillance of North American pediatric group a streptococcal pharyngitis isolates. Clin Infect Dis 2009; 49:78–84. [PubMed: 19480575]
17. Valenciano SJ, Onukwube J, Spiller MW, et al. Invasive group A streptococcal infections among people who inject drugs and people experiencing homelessness in the United States, 2010–2017. Clin Infect Dis 2020; 73:e3718–26.
18. Lamagni TL, Neal S, Keshishian C, et al. Epidemic of severe *Streptococcus pyogenes* infections in injecting drug users in the UK, 2003–2004. Clin Microbiol Infect 2008; 14:1002–9. [PubMed: 19040471]
19. Efstratiou A, Emery M, Lamagni TL, Tanna A, Warner M, George RC. Increasing incidence of group A streptococcal infections amongst injecting drug users in England and Wales. J Med Microbiol 2003; 52(Pt 6):525–6. [PubMed: 12748273]
20. Curtis SJ, Tanna A, Russell HH, et al. Invasive group A streptococcal infection in injecting drug users and non-drug users in a single UK city. J Infect 2007; 54:422–6. [PubMed: 17116332]
21. Kwiatkowska RM, Manley P, Sims B, et al. Outbreak of group A *Streptococcus* emm94.0 affecting people who inject drugs in southwest England, April 2017. Am J Infect Control 2018; 46:238–40. [PubMed: 29031429]
22. Adebanjo T, Mosites E, Beneden CVV, et al. Risk factors for group A *Streptococcus* colonization during an outbreak among people experiencing homelessness in Anchorage, Alaska, 2017. Clin Infect Dis 2018; 67:1784–7. [PubMed: 29788094]
23. Dauby N, Miendje Deyi VY, Delforge V, et al. *Streptococcus pyogenes* infections with limited emm-type diversity in the homeless population of Brussels, 2016–2018. Int J Infect Dis 2019; 81:52–6. [PubMed: 30684741]
24. Fittipaldi N, Beres SB, Olsen RJ, et al. Full-genome dissection of an epidemic of severe invasive disease caused by a hypervirulent, recently emerged clone of group A *Streptococcus*. Am J Pathol 2012; 180:1522–34. [PubMed: 22330677]
25. Pilishvili T, Lexau C, Farley MM, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis 2010; 201:32–41. [PubMed: 19947881]

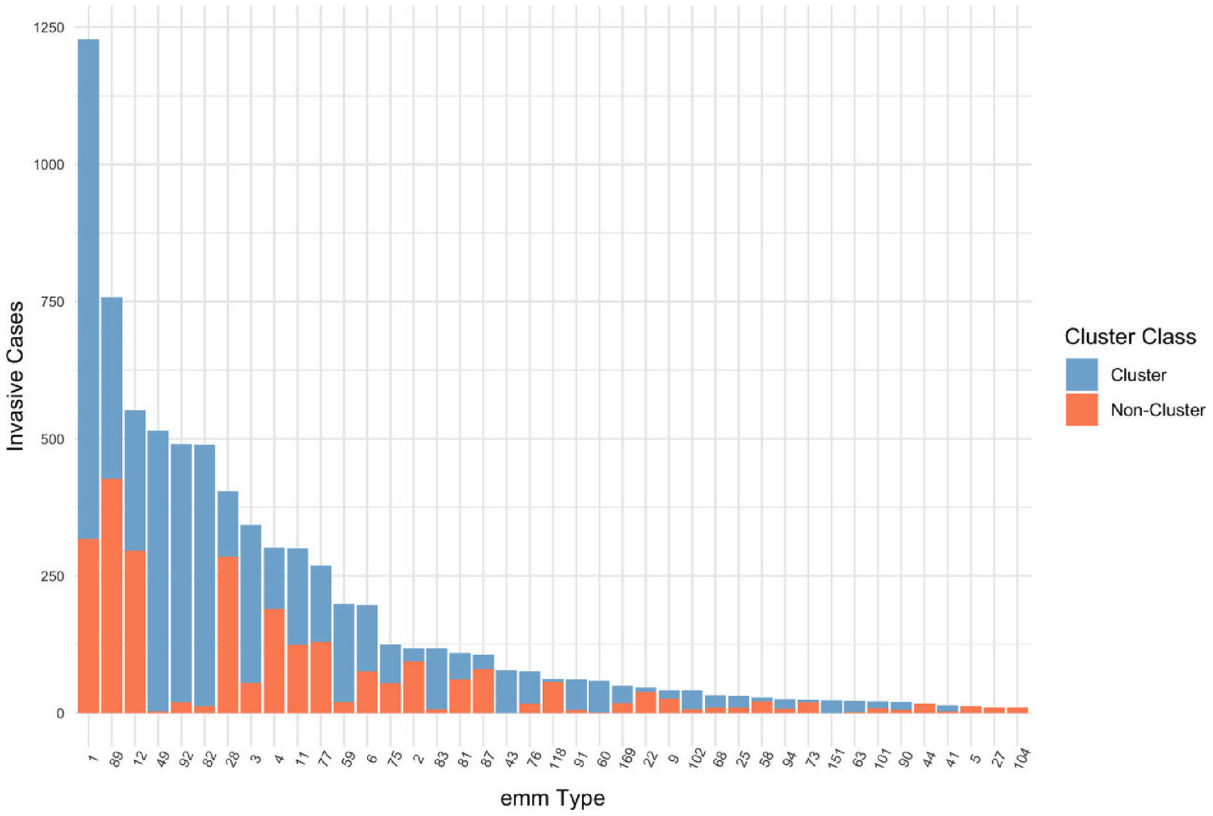

Figure 1.

Proportion of cluster-associated isolates from total invasive cases varies across *emm* types. The stacked bar chart represents *emm* types that contain at least 10 invasive isolates acquired through the Centers for Disease Control and Prevention Active Bacterial Core surveillance program. The graph is ordered by total invasive cases, and each bar comprises 2 categories: cluster-associated (blue) and noncluster-associated isolates (red). The proportion of cluster-associated cases diverges widely across *emm* types: some types, such as *emm49*, are primarily clustered, whereas others (ie, *emm28*) contain mostly nonclustered isolates.

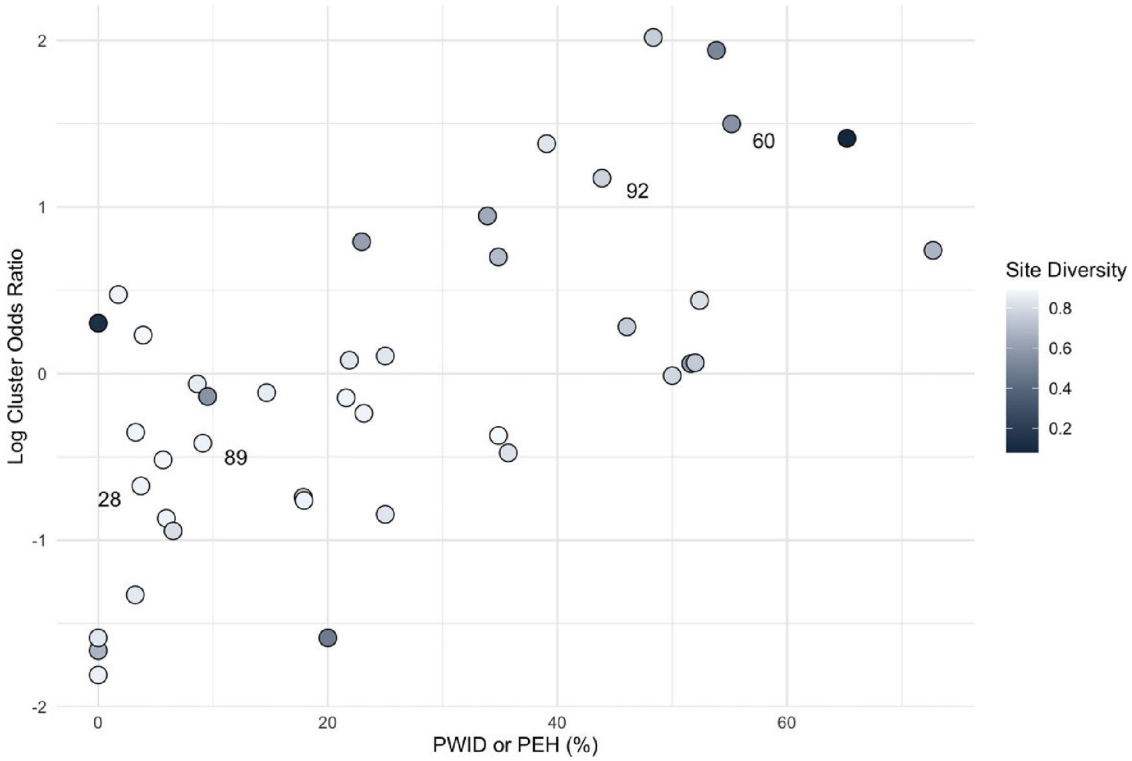

Figure 2.

Correlation between cluster odds ratio and percentage of people experiencing homelessness (PEH) or people who inject drugs (PWID). Plot of the $\log_{10}$ cluster odds ratio against the percentage of PEH or PWID for each *emm* type reveals a positive correlation in which *emm* types with a higher proportion of PEH and PWID have a higher propensity to form transmission clusters. The strength of this correlation was assessed using the Kendall correlation coefficient and was found to have a tau value of 0.47 (P value = 1.51×10^{-5}). The datapoint color gradient represents the sporadically emergent and geographically targeted (SEGT) index value for surveillance site assortment across *emm* types. A darker color indicates the *emm* type was found across fewer surveillance sites (strong-SEGT), whereas a lighter color indicates a larger geographic spread (weak-SEGT). An inverse correlation was identified between site diversity and cluster odds ratio (tau = −0.30, P = .0078). To provide context representative higher COR (*emm*60, *emm*92) and lower COR (*emm*28, *emm*89), *emm* types are labeled on the graph.

Table 1.

Summary of GAS Invasive Disease Cluster Features for Each *emm* Type Ranked by Cluster Odds Ratio[a]

emm Type	Cluster Cases	Total Cases	COR (95% CI)	LTCF (%)	PEH/PWID (%)	SEGT Index
49	512	515	104.22 (33.47–324.55)	2.33	48.35	0.76
43	78	78	87.32 (5.41–1408.9)	1.28	53.85	0.52
60	57	58	31.55 (4.37–227.96)	0	55.17	0.56
151	23	23	25.84 (1.57–425.64)	0	65.22	0.08
82	477	489	24 (13.51–42.65)	6.34	39.06	0.83
92	471	490	14.91 (9.4–23.64)	4.9	43.88	0.78
83	111	118	8.86 (4.12–19.04)	5.08	33.9	0.65
91	56	61	6.19 (2.48–15.47)	6.56	22.95	0.61
63	20	22	5.49 (1.28–23.51)	0	72.73	0.68
59	178	198	5.02 (3.15–7.98)	2.53	34.85	0.7
3	288	343	2.98 (2.23–4)	5.25	1.75	0.87
102	35	42	2.75 (1.22–6.2)	0	52.38	0.82
41	11	14	2.01 (.56–7.21)	42.86	0	0.13
76	59	76	1.91 (1.11–3.28)	2.63	46.05	0.75
1	911	1228	1.7 (1.48–1.96)	2.12	3.91	0.89
90	14	20	1.28 (.49–3.33)	5	25	0.84
68	22	32	1.2 (.57–2.55)	3.13	21.88	0.84
94	17	25	1.16 (.5–2.7)	4	52	0.75
25	21	31	1.15 (.54–2.45)	3.23	51.61	0.59
169	32	50	0.97 (.54–1.74)	12	50	0.79
6	121	197	0.87 (.65–1.16)	3.05	8.63	0.85
11	176	300	0.77 (.61–.97)	15	14.67	0.85
101	12	21	0.73 (.31–1.73)	4.76	9.52	0.57
75	71	125	0.72 (.5–1.02)	5.6	21.6	0.88
77	139	268	0.58 (.45–.74)	6.34	23.13	0.87
12	256	552	0.44 (.37–.53)	5.8	3.26	0.88
81	48	109	0.42 (.29–.62)	1.83	34.86	0.89
89	331	758	0.38 (.33–.45)	21.9	9.1	0.87

emm Type	Cluster Cases	Total Cases	COR (95% CI)	LTCF (%)	PEH/PWID (%)	SEGT Index
9	16	42	0.33 (.18–.62)	0	35.71	0.82
4	111	301	0.3 (.24–.39)	5.65	5.65	0.88
28	120	404	0.21 (.17–.26)	6.93	3.71	0.87
58	7	28	0.18 (.08–.43)	7.14	17.86	0.79
87	26	106	0.17 (.11–.27)	5.66	17.92	0.87
2	24	118	0.14 (.09–.21)	1.69	5.93	0.87
73	5	24	0.14 (.05–.38)	0	25	0.83
22	8	46	0.11 (.05–.24)	4.35	6.52	0.81
118	5	62	0.05 (.02–.12)	3.23	3.23	0.84
27	0	10	0.03 (0–.44)	0	0	0.84
104	0	10	0.03 (0–.44)	10	20	0.48
5	0	12	0.02 (0–.37)	0	0	0.68
44	0	17	0.02 (0–.26)	11.76	0	0.87

Abbreviations: CI, confidence interval; COR, cluster odds ratio; LTCF, long-term care facility; PEH, people experiencing homelessness; PWID, people who inject drugs; SEGT, sporadically emergent and geographically targeted.

[a]The table provides data on the number of cluster cases, total invasive cases, COR with 95% CIs, percentage of LTCF residents among total cases, percentage of PEH/PWID among total cases, and SEGT index for each emm type representing at least 10 isolates. Cluster odds ratios were calculated by taking the ratio of the cluster odds (number of cluster isolates to noncluster isolates) of the focal emm type over cluster odds of all other emm types. The SEGT index was quantified using the Simpson's diversity index in which invasive case counts were aggregated over emm types for each of the 10 GAS surveillance sites (California, Colorado, Connecticut, Georgia, Maryland, Minnesota, New Mexico, New York, Oregon, and Tennessee).