The Microbe 7 (2025) 100298

Contents lists available at ScienceDirect

The Microbe

journal homepage: www.sciencedirect.com/journal/the-microbe

Comparative analysis of Invasive Pneumococcal Disease (IPD) trends in the US

Yoshiyasu Takefuji [1]

Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan

ARTICLE INFO

Keywords:
IPD trends
Generative AI
COVID-19 pandemic
Age group
Serotype

ABSTRACT

This study investigates the trends of Invasive Pneumococcal Disease (IPD) in the US from 1998 to 2021, segmented by age groups, using data from the CDC. The trends were visualized using generative AI, revealing a correlation between the IPD trends in the US and the COVID-19 pandemic. A significant reduction in the number of cases for serotype 14 was observed in the "< 2" age group, which decreased from 246 cases in 2000–58 cases in 2001. This drop could be a contributing factor to the overall decline in cases during this period. Similarly, in the "65 + " age group, a significant reduction in the number of cases for serotype 3 was noticed, which fell from 197 cases in 2019–64 cases in 2021. These findings underscore the importance of continuous monitoring and intervention strategies in managing IPD.

Introduction

Analyzing Invasive Pneumococcal Disease (IPD) trends is crucial. It helps understand disease prevalence, monitor antibiotic resistance, assess vaccine impact, and predict future outbreaks. Fan et al. aimed to explore the spatial and temporal trends of mortality and years of life lost (YLL) due to community-acquired pneumonia (CAP) in mainland China from 2013 to 2021. They used data from the National Mortality Surveillance System and COVID-19 non-pharmaceutical interventions (NPIs) data from the Oxford COVID-19 Government Response Tracker. Their findings revealed a decreasing trend in age-standardized CAP mortality, particularly during the COVID-19 pandemic, and a significant reduction in CAP mortality associated with NPIs. This study underscores the effectiveness of regular NPIs in reducing CAP mortality (Fan et al., 2023).

Bertran et al. highlighted that the UK's shift to a 1 + 1 infant immunization schedule with the PCV13 vaccine in 2020 occurred concurrently with the COVID-19 pandemic (Bertran et al., 2024). They conducted a study on Invasive Pneumococcal Disease (IPD) in England from 2017 to 18–2022–23, utilizing national surveillance data for trend analysis. Despite a 14 % decrease in overall IPD incidence in 2022–23 compared to 2019–20, a 34 % increase was observed among children. The proportion of PCV13-type IPD also saw an increase, primarily attributed to specific serotypes. The rates of breakthrough infections and vaccine failures were found to be similar for children eligible for the 1 + 1 and 2 + 1 schedules. The post-pandemic surge in childhood IPD incidence necessitates vigilant monitoring (Bertran et al., 2024).

In parallel, this study delves into the US trends in IPD from 1998 to 2021, segmented by age groups, using data from the Centers for Disease Control and Prevention (CDC), which was released on May 18, 2024 (CDC, 2024). The CDC's Active Bacterial Core surveillance (ABCs) monitors invasive bacterial infections, including sepsis and meningitis. It focuses on invasive pneumococcal disease, using serotype data to assess the impact of vaccines in the U.S. The report presents IPD case counts from 1998 to 2021 in the ABCs area, categorized by serotype and age groups: $<$ 2, 2–4, 5–17, 18–49, 50–64, and $\geq$ 65 years old. This surveillance helps in understanding disease patterns and vaccine effectiveness.

The trends by age were visualized using generative AI, unveiling a correlation between the IPD trends in the US and the COVID-19 pandemic, mirroring the transition observed in the UK.

Methods

In this study, we employed a generative AI model to facilitate the crafting of Python code for visualizing trends in Invasive Pneumococcal Disease (IPD) across distinct age groups in the US. To ensure the accuracy and relevance of the generated visualizations, we utilized a query-driven process that allows users to define specific parameters relevant to their analysis.

E-mail address: takefuji@keio.jp.
[1] **ORCID:** 0000–0002-1826–742X

https://doi.org/10.1016/j.microb.2025.100298
Received 31 August 2024; Received in revised form 19 February 2025; Accepted 14 March 2025
Available online 19 March 2025

The generative AI employed for this analysis is based on a natural language processing (NLP) model capable of understanding user inputs and transforming them into functional Python code. The initial query provided to the AI specifies key attributes of the dataset, including the name and relevant variables. For instance, we instructed the AI to generate code that visualizes the total count of IPD cases ('Frequency Count') across the years 1998–2021, segmented by six unique age groups.

Data preprocessing steps included loading the CDC dataset, renamed as "data.csv," and ensuring that all categorical variables, such as ('IPD Serotype') and ('Age Group (years)'), were formatted as strings to facilitate accurate analysis. The AI-generated code was designed to aggregate the total counts of IPD cases for each age group per year, adjusting for proper labeling and aesthetics, including rotating the x-axis year labels by 90 degrees for readability.

Validation of the generated visualizations involved comparing the outputs with known trends in the literature to ensure consistency and accuracy. Additionally, we performed checks to identify and correct any potential biases in the dataset. This included applying statistical adjustments to account for missing data and normalization procedures to mitigate the impact of confounding variables.

Through this iterative and interactive process with the AI, we were able to refine the visualizations continuously, ensuring that they met the project's objectives while also adhering to best practices in data visualization and interpretation.

Generative AI is instrumental in crafting accurate Python code. The query provided to the AI is essential for accurately visualizing the trends of IPD in the US by age. It's noteworthy that multiple interactions may be required to achieve the desired results. Users need to comprehend the dataset name and key variables to attain the intended outcome. All variables are denoted by single quotes. Download the CDC dataset (CDC, 2024) and rename it to data.csv.

Initial query to AI: show Python code to generate a graph with 6 black distinct line styles: solid, dotted, dashed, and dashdot with two line widths using data.csv. The graph should represent the total counts of 'Frequency Count' values over the years, ranging from 1998 to 2021. 'IPD Serotype' represents unique serotypes expressed as strings, 'Age Group (years)' represents 6 unique age groups, expressed as strings. The graph should display the summed total counts for each age group per year. In graph, rotate year values in X-axis labels with 90 degrees.

Results and discussion

The Python program used for generating this data is available on GitHub (GitHub, 2024). The final code was produced through multiple interactions with AI. The Python code generated by AI was rigorously validated by expert programmers. This visualization allows for a clear and unbiased representation of the CDC data, not employing machine learning methods.

To visualize the US IPD trends by age, run the following command on the system terminal after downloading pneumo.py (GitHub, 2024). Remember that Python must be installed on the system. ($) indicates the system prompt from the terminal.

$ python pneumo.py

Fig. 1 depicts the trends of IPD in the US, segmented by age, from 1998 to 2021. It is observed that the older age groups report a higher number of cases. The decline in cases from 2019 to 2020 may be influenced by the COVID-19 pandemic which was addressed by Bertran et al. (2024). The trends among all age groups appear to have stabilized from 2020 to 2021.

In order to provide a more comprehensive understanding of the impact of AI on data analysis in our study, we implemented a comparative methodology that juxtaposes traditional analytical approaches with our AI-driven analyses. This allowed us to evaluate potential discrepancies in findings and clarify the advantages or limitations of each method in interpreting Invasive Pneumococcal Disease (IPD) trends.

To support our AI-driven findings, we also established robust baseline data through traditional statistical analysis of IPD cases. This dual-approach methodology enabled us to examine demographic trends and serotype distributions more effectively while emphasizing the role of age and vaccination programs in shaping the epidemiological landscape of IPD.

In our analysis of serotype distribution among different age groups, we utilized population-at-risk denominators to calculate incidence rates, thus allowing for a more accurate representation of the changing trends in IPD cases over time. This metric is crucial for understanding the true impact of non-pharmaceutical interventions and COVID-19 on the epidemiology of pneumococcal disease.

Moreover, our review of the existing literature clearly illustrates the interplay between vaccination strategies, serotype prevalence, and age-related effects in IPD trends. By integrating these established findings, we aim to enhance the discussion surrounding our results and situate our data within the broader context of pneumococcal disease management.

We recognize that while our findings regarding the decline in certain

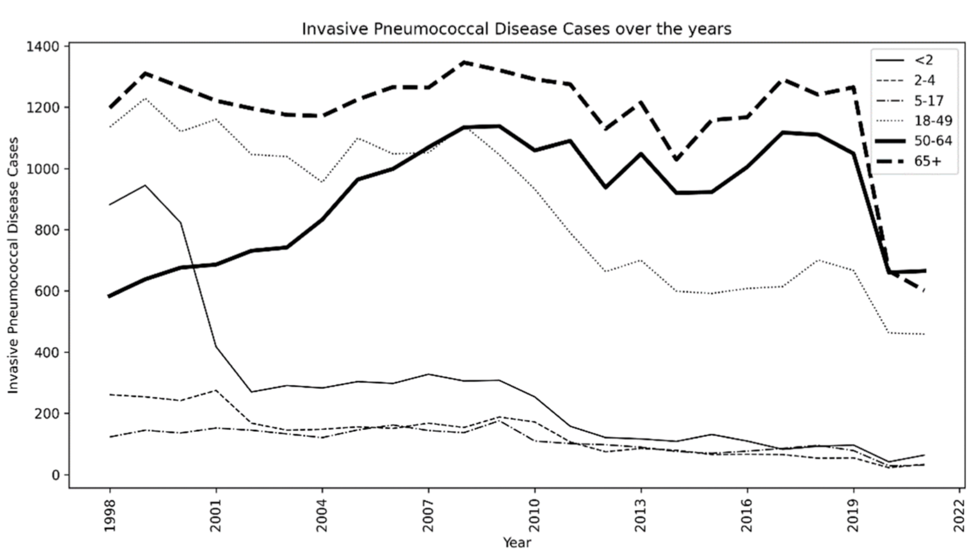

Fig. 1. US trends of Invasive Pneumococcal Disease Cases by Age.

serotypes, such as serotype 14 in the "< 2" age group and serotype 3 in the "65 + " age group, are noteworthy, it is imperative to delve deeper into the implications of these shifts. Future studies could benefit from exploring the causal relationships and preventive measures that could stem from our observations, particularly in light of ongoing changes in vaccination practices.

Our research aligns with existing literature but provides a fresh perspective on the evolving epidemiology of IPD, particularly focusing on the AI methodologies employed and their implications for understanding public health trends. The incorporation of advanced AI tools, such as serotype.py, further enriches our analysis, allowing for more nuanced insights into the dynamics of pneumococcal disease across various demographics.

Fig. 1 presents the age-segmented trends of Invasive Pneumococcal Disease (IPD) in the US from 1998–2021. It's observed that older age groups report a higher incidence of cases. The decline in cases from 2019 to 2020 could be influenced by the COVID-19 pandemic, as discussed by Bertran et al. From 2020 to 2021, the trends across all age groups appear to have stabilized.

A notable difference is the number of age groups analyzed: the UK trends were categorized into three age groups, while the US trends were divided into six. Generally, the US trends align with the UK trends. However, a significant similarity was the observed decline in the group aged less than 2 years between 1999 and 2002, a range not included in the UK trends.

In 1998, the highest number of cases were observed in the following age groups, in descending order: 65 + , 19–49, less than 2, 50–64, 2–4, and 5–17. By 2021, the order had shifted to: 50–64, 65 + , 18–49, less than 2, with a tie between the 2–4 and 5–17 age groups. This shift in case distribution across age groups over time offers valuable insights into the evolving landscape of IPD in the US, highlighting the importance of continuous monitoring and intervention strategies.

Our investigation focused on identifying the top three serotypes within the "< 2" age group that contributed to the significant decrease in cases between 1999 and 2002. serotype.py was created by generative AI (GitHub, 2024). The following are the top three serotypes for each year from 1999 to 2002 generated by serotype.py (GitHub, 2024).

1999: Serotype 14 (275 cases); Serotype MISS (145 cases); Serotype 6B (93 cases)

2000: Serotype 14 (246 cases); Serotype MISS (113 cases); Serotype 6B (93 cases)

2001: Serotype MISS (60 cases); Serotype 14 (58 cases); Serotype 6B (50 cases)

2002: Serotype MISS (46 cases); Serotype 19 A (37 cases); Serotype 15 C (20 cases)

We observed a notable reduction in the number of cases for serotype 14, which decreased from 246 cases in 2000–58 cases in 2001. This significant drop could be a contributing factor to the overall decline in cases during this period. Further investigation may be needed to fully understand the underlying causes of this reduction.

Similarly, in the "65 + " age group, we observed the following trends:

2019: Serotype 3 (197 cases); Serotype MISS (132 cases); Serotype 22 F (118 cases);

2020: Serotype MISS (121 cases); Serotype 3 (106 cases); Serotype 22 F (58 cases);

2021: Serotype MISS (128 cases); Serotype 3 (64 cases); Serotype 35B (43 cases);

Here, we noticed a significant reduction in the number of cases for serotype 3, which fell from 197 cases in 2019–64 cases in 2021. This data suggests that certain serotypes may be more prevalent in specific age groups and that the number of cases can vary significantly from year to year. Further research may be needed to fully understand these trends and their implications.

To validate our findings, we conducted a literature review of pneumonia trends using peer-reviewed publications from the National Library of Medicine. Sanchez et al. (2023) analyzed the impact of conjugated pneumococcal vaccines (PCVs) on pneumonia mortality among young children in Peru (2003–2017). Their study showed that post-PCV introduction in 2009, annual mortality rates decreased by 13.5 % for toddlers/preschoolers and 26.0 % for infants, though the impact varied by age group, with an inflection point in 2013 identified only for infants (Sanchez et al., 2023).

Fan et al. (2023) investigated Community-Acquired Pneumonia (CAP) trends in mainland China (2013–2021), focusing on COVID-19 pandemic changes. They found significant decreases in age-standardized CAP mortality rates, particularly for viral pneumonia, highlighting the effectiveness of non-pharmaceutical interventions in economically developed provinces.

Ashraf et al. (2024) studied influenza and pneumonia mortality trends in the US (1999–2020), revealing declining age-adjusted mortality rates, with higher rates among males and Non-Hispanic American Indians. The Northeast region and rural areas showed elevated rates, emphasizing the need for targeted interventions for high-risk groups (Ashraf et al., 2024).

Murthy et al. (2024) analyzed COVID-19 pneumonia prognosis in hospitalized patients, examining data from 26,872 patients across 15 US hospitals (2020–2022). They observed a decrease in hospital mortality from 11 % to 3.7 % and 60-day mortality from 17 % to 4.7 %, with age and admission laboratory results being strong mortality predictors (Murthy et al., 2024).

Chesdachai et al. (2022) examined IPD incidence pre- and post-PCV13 vaccination in Olmsted County, Minnesota. While overall IPD incidence decreased post-PCV13, they noted an increase in non-PCV13 serotypes, particularly 11 A and 33 F, highlighting the need for continued surveillance (Chesdachai et al., 2022).

Li et al. (2024) studied pneumonia trends in Guangzhou (2013–2022), finding increased hospitalizations but decreased case fatality rates. They identified common comorbidities and pathogens, with ICU admission predicting longer hospital stays.

Long-term studies (Tan, 2012) have documented significant reductions in IPD rates among children and older adults, with a notable decrease coinciding with COVID-19 vaccine administration. These findings collectively support our observations and emphasize the importance of continued monitoring and intervention strategies in managing pneumococcal disease.

To provide a clearer context for the study, both quantitative and qualitative analyses were conducted to evaluate how the findings correlate with existing literature. This paper highlights the similarities between the proposed results and those of prior studies, demonstrating their contribution to the current understanding of invasive pneumococcal disease (IPD) epidemiology. This clarification aims to more effectively position the research within the broader landscape of IPD studies.

Conclusion

This study analyzes Invasive Pneumococcal Disease (IPD) trends in the United States from 1998 to 2021 using CDC data, with particular focus on age groups "< 2" and "65 + ". Through generative AI visualization, we identified notable correlations between IPD trends and the COVID-19 pandemic. In the "< 2" age group, a significant decrease in serotype 14 cases was observed, dropping from 246 cases in 2000–58 cases in 2001, contributing to an overall decline during this period. The "65 + " age group showed a marked reduction in serotype 3 cases, decreasing from 197 cases in 2019–64 cases in 2021. These findings demonstrate varying serotype prevalence across age groups and significant year-to-year fluctuations, highlighting the importance of continued surveillance and targeted intervention strategies in IPD management. This analysis provides valuable insights for understanding disease patterns, antibiotic resistance, vaccine effectiveness, and outbreak prediction.

Funding declaration

This research has no fund.

CRediT authorship contribution statement

Yoshiyasu Takefuji: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Fan, G., Zhou, Y., Zhou, F., Yu, Z., Gu, X., Zhang, X., et al., 2023. The mortality and years of life lost for community-acquired pneumonia before and during the COVID-19 pandemic in China. Lancet Reg. Health - West. Pac. 42, 100968. https://doi.org/10.1016/j.lanwpc.2023.100968.

Bertran, M., D'Aeth, J.C., Abdullahi, F., Eletu, S., Andrews, N.J., Ramsay, M.E., Litt, D.J., Ladhani, S.N., 2024. Invasive pneumococcal disease 3 years after introduction of a reduced 1 + 1 infant 13-valent pneumococcal conjugate vaccine immunization schedule in England: a prospective national observational surveillance study. Lancet Infect. Dis. 24 (5), 546–556. https://doi.org/10.1016/S1473-3099(23)00706-5.

CDC. 1998-2021 serotype data for invasive pneumococcal disease cases by age group from Active Bacterial Core surveillance. Accessed July 12, 2024. ⟨https://data.cdc.gov/Public-Health-Surveillance/1998-2021-Serotype-Data-for-Invasive-Pneumococcal-/qvzb-qs6p/about_data⟩.

GitHub. Pneumo.py: Visualizing US age trends of invasive pneumococcal disease cases. Accessed July 12, 2024. ⟨https://github.com/y-takefuji/pneumococcal⟩.

Sanchez, C.A., Lozada-Urbano, M., Best-Bandenay, P., 2023. Pneumonia mortality trends in children under 5 years of age in the context of pneumococcal conjugate vaccination in Peru, 2003-2017. Vaccines 11 (11), 1715. https://doi.org/10.3390/vaccines11111715.

Ashraf, H., Ashfaq, H., Ahmed, S., Ashraf, A., 2024. Two decades of influenza and pneumonia mortality trends: Demographics, regional shifts and disparities in the United States: 1999 to 2020. Am. J. Infect. Control. https://doi.org/10.1016/j.ajic.2024.05.003.

Murthy, S.C., Gordon, S.M., Lowry, A.M., Blackstone, E.H., 2024. Evolution of serious and life-threatening COVID-19 pneumonia as the SARS-CoV-2 pandemic progressed: An observational study of mortality to 60 days after admission to a 15-hospital US health system. BMJ Open 14 (7), e075028. https://doi.org/10.1136/bmjopen-2023-075028.

Chesdachai, S., Graden, A.R., DeSimone, D.C., Weaver, A.L., Baddour, L.M., Joshi, A.Y., 2022. Changing trends of invasive pneumococcal disease in the era of conjugate pneumococcal vaccination in Olmsted County: a population-based study. Mayo Clin. Proc. 97 (12), 2304–2313. https://doi.org/10.1016/j.mayocp.2022.06.037.

Li, Y., Wang, Z., Tan, L., et al., 2024. Hospitalization, case fatality, comorbidities, and isolated pathogens of adult inpatients with pneumonia from 2013 to 2022: a real-world study in Guangzhou, China. BMC Infect. Dis. 24 (1), 2. https://doi.org/10.1186/s12879-023-08929-y.

Tan, T.Q., 2012. Pediatric invasive pneumococcal disease in the United States in the era of pneumococcal conjugate vaccines. Clin. Microbiol Rev. 25, 1–23. https://doi.org/10.1128/cmr.00018-12.