

Analysis Tutorial Prospectus

Exploring the Evolution of Methicillin-Resistant Staphylococcus aureus (MRSA) Resistance Using Statistical and Genomic Data Analysis in R

1. Research Question:

- How has the rate of methicillin resistance *Staphylococcus aureus* cases evolved over the past decade, and what genetic mutations/MRSA strains are associated with this resistance?
- Are there demographic or geographic factors influencing the development of methicillin resistance in *Staphylococcus aureus*?

2. Objective:

The primary objective of this tutorial is to guide users through the development of an R-based analysis pipeline for investigating the evolution of methicillin resistance in *Staphylococcus aureus* (MRSA). Specifically, the tutorial will:

- Help visualize and analyze the trend of MRSA resistance rates over time.
- Implement statistical modeling to identify demographic and geographic factors associated with resistance development.

The finished R code will allow users to:

- Import and clean resistance-related data.
- Perform statistical analyses to detect significant patterns and trends in resistance rates.
- Create visualizations that show the development of MRSA resistance over time and across different regions or patient demographics.
- Conduct basic genomic analysis on sequence data to identify mutations associated with resistance.

3. Intended Approach/Methods:

The tutorial will follow a clear, step-by-step process that includes:

- **Data Preparation:** Import and clean MRSA resistance data (CSV) and genomic data (FASTA/VCF format) using R. Handle missing data, filter outliers, and format the datasets for analysis.
- **Exploratory Data Analysis (EDA):** Use R packages like ggplot2 to visualize trends in MRSA resistance over time, differences by region or patient demographics, and explore potential correlations.
- **Statistical Modeling:** Apply linear regression or generalized linear models (GLM) to assess the impact of various factors (e.g., region, year, age, sex) on MRSA resistance rates.

- **Geospatial Analysis:** If geographic data is available, implement geospatial visualization to map the distribution of MRSA resistance using packages like ggmap or sf.
- Potential Data:
 - <https://arpsp.cdc.gov/profile/eip/mrsa>
 - <https://catalog.data.gov/dataset/methicillin-resistant-staphylococcus-aureus-mrsa-bloodstream-infections-bsi-in-california--cf5ab>
 - <https://www.cdc.gov/nhsn/hai-report/data-tables-adult/index.html>
 - https://data.cdc.gov/Public-Health-Surveillance/HAICViz-iSA/ssz5-s49e/about_data
 - <https://pubmlst.org/organisms/staphylococcus-aureus>

4. References:

- Liu, C., & Chambers, H. F. (2003). *Methicillin-resistant Staphylococcus aureus: An overview of clinical features and management strategies*. *Clinical Infectious Diseases*, 36(6), 199-206. <https://doi.org/10.1086/368712>
- Boucher, H. W., & Corey, G. R. (2008). *Epidemiology of Methicillin-Resistant Staphylococcus aureus and its Impact on Clinical Outcomes*. *Clinical Infectious Diseases*, 46(5), 1-19. <https://doi.org/10.1086/523340>
- Hidron, A. I., & Patel, J. B. (2011). *Antimicrobial-resistant pathogens associated with healthcare-associated infections: A review of the literature*. *American Journal of Infection Control*, 39(3), 152-160. <https://doi.org/10.1016/j.ajic.2010.04.011>
- Enright, M. C., Day, N. P., Davies, C. E., Peacock, S. J., & Spratt, B. G. (2000). Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *Journal of clinical microbiology*, 38(3), 1008–1015. <https://doi.org/10.1128/JCM.38.3.1008-1015.2000>
- Jolley, K. A., Bray, J. E., & Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome open research*, 3, 124. <https://doi.org/10.12688/wellcomeopenres.14826.1>