PHUSE US Connect 2024
PP12 - Mining Historical Data and Epidemiological Evidence for Real World Studies
Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

ABSTRACT

Real-world evidence studies always have been used to find definitive or conclusive patterns in a variety of aspects of health and disease, such as epidemiology, disease burden, treatment patterns, efficacy, safety, treatment outcomes, long-term outcomes, and patient-reported outcomes such as satisfaction, medication adherence, quality of life, and patient experience.

One of the main purposes of Real-World Studies is evidence generation for the safety and effectiveness of a drug in larger populations to assess and research if the actual Randomized Controlled Trials could run for a shorter duration based on a robust foundation of evidence. Evidence generation requires a lot of research and analysis to make the right interpretation of the outcome of the experiment.

Our poster would run through a unique technique of mining historical data on pandemics and epidemics across geographies, populations, genotypes, phenotypes, immigration patterns, and pathophysiology, comparing time to novel drug discoveries, management, historical trials, and how public policies were framed to safeguard patients. We will turn the pages of history for time-tracking to discuss how real-world data was utilized for evidence generation and how the research outcomes influenced scientific innovations along with policy making. We will also discuss how Real-world data was gathered and re-presented historically via graphs, visualizations, and how results were communicated to take timely actions and stop the growth of infectious diseases, also how they contributed to public policies and measures. One such example is the valuable contributions of Florence Nightingale's Rose Chart/Coxcomb graph representing unnecessary deaths during the Crimean War because of preventable infections among wounded soldiers. Nightingale was an English social reformer; a statistician and the founder of modern nursing called the "The Lady with the Lamp".

We will finally discuss how Generative AI can be used to parse historical and real-time data from social media platforms to create evidence. We will walk through the evolution of Real-World evidence generation, bolstering the fact that Real World Evidence is the most critical part of Clinical Research. We will conclude with an interactive discussion on how industry and academia can go hand in hand to collaborate in scientific innovations, to help Pharma generate evidence using Real World Data.

METHODS:

Integrating Generative AI Tools To Analyze the Efficacy of Pandemic Response:

1. Model Selection - Identify which LLM (Large Language Model) - GPT4, Gemini (Integrated through Bard), BLOOM, LLaMA, Claude, etc. - is to be used for data collection and parameter definition

•Each LLM has its strengths/weaknesses/abilities (e.g. conceptual understanding and representation vs. computational strength)

•Clear parameters for selection can be defined to place importance on capabilities of priority - for example, access to the internet, or a ranked ability to retrieve certain types of information.

•Features of Interest:

- Multimodality (ability to work with both text and image (Gemini)
- Access to Internet (Bard)
- Accuracy and Computational Ability (GPT 3.5 and 4)
- Reduced Bias/Hallucinations (BLOOM/LLaMA)
- Pattern Recognition & Adaptive Learning (GPT 4 and Gemini)

Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

2. Prompt Engineering - Employ prompt-engineering techniques to refine searches and increase return efficacy.

Potential Methods:

- Self-Ask Prompting
- Meta Prompting
- Chain-of-Thought Prompting
- ReACT
- Symbolic Reasoning and PAL
- Iterative Prompting
- Sequential Prompting
- Self-Consistency Method
- Automatic Reasoning & Tool Use (ART)

*Ask targeted questions ex: duration, geographic location, phenotypes, genotypes, pathogens, vector mode, etc.

3. Fine-Tuning Model - Possibly Fine-Tune LLMs to defined research specifications by preparing and uploading training data (perhaps data collected from prior pandemics/health crises, or for our purpose - a hand-done model analysis of a past pandemic with parameters specified to teach the LLM)

4. Retrieval Augmented Generation - Going back to original references and ensuring content validity/accuracy. This may be supplemented by further LLM queries.

COMPARISON OF LARGE LANGUAGE MODELS

To gauge the strengths/weaknesses of various large language models, we tested a selection of LLMs, by asking them various types of questions relating to epidemiological history/studies. Some of the questions asked, and their respective outputs, by GPT 3.5 and Bard, are shown below as an example of the kinds of responses generally outputted for lines of questioning that were applied. The LLMs tested were Google's Bard, OpenAI's ChatGPT 3.5, Bing Chat (Microsoft Copilot), and Anthropic's Claude. The overarching results were that Bard was very useful in its ability to access the internet and provide resources – links to articles, scholarly studies, and references while ChatGPT was more likely to hallucinate and generate false references/links. However, ChatGPT proved to have a greater memory than Bard, Claude, or Copilot, which kept having to be reminded of the "trains of thought" they were following. Bard was, however, much more capable of analyzing and summarizing historical references and generating comprehensive timeline summaries for epidemiological data. Copilot proved to be useful in handling mathematical data, which showed when performing calculations or statistical summaries, while Bard was greater handy in analyzing graphics/visualizations to summarize them. Claude, however, showed higher response speed, though with lesser specificity in responses, and was more malleable to following a single train of thought. Overall, it seemed that, through preliminary questioning, Bard and GPT 3.5 were the best for research purposes.

Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

List of epidemics in world history	Unfortunately, providing a complete list of all epidemics in world history is virtually impossible, for several reasons:
	
	• **Lack of comprehensive records:** Records of historical outbreaks, especially in ancient times, are often incomplete or inaccurate. Many epidemics may have gone entirely unrecorded.
	• **Varied definitions:** Distinguishing between an epidemic and an endemic disease can be subjective. An outbreak in one region might be considered an epidemic, while similar cases in another area might be seen as endemic.
	• **Continual occurrence:** New epidemics and resurgences of old ones continuously occur, making the list ever-changing.
	
	Therefore, instead of a complete list, I can offer several alternatives to explore historical epidemics:
	
	By Time Period:
	
	• **Ancient Epidemics:**
	  ○ Plague of Athens (430 BC) - possibly typhoid fever
	  ○ Antonine Plague (165-180 AD) - possibly smallpox
	  ○ Plague of Cyprian (249-262 AD) - possibly smallpox
	• **Medieval Epidemics:**
	  ○ Justinian Plague (541-549 AD) - bubonic plague
	  ○ Black Death (1347-1353) - bubonic plague
	  ○ Great Bubonic Plague of London (1665) - bubonic plague
	• **Modern Epidemics:**
	  ○ Spanish Flu (1918-1919) - influenza A virus
	  ○ Asian Flu (1957-1958) - influenza A virus
	  ○ AIDS epidemic (ongoing) - HIV/AIDS
	  ○ COVID-19 pandemic (ongoing) - SARS-CoV-2 virus
	
	By Region:
	
	• **European Epidemics:**
	  ○ Dancing Plague of Strasbourg (1518) - likely mass psychogenic illness
	  ○ Great Famine of Ireland (1845-1852) - potato blight
	• **African Epidemics:**
	  ○ Ebola outbreaks (ongoing) - Ebola virus
	  ○ Malaria (endemic and epidemic outbreaks) - Plasmodium parasite
	• **Asian Epidemics:**
	  ○ Third Cholera Pandemic (1852-1860) - Vibrio cholerae bacterium
	  ○ SARS outbreak (2003) - SARS coronavirus
	
	By Disease:
	
	• Smallpox (eradicated in 1980)
	• Tuberculosis (ongoing) - Mycobacterium tuberculosis bacterium
	• Measles (ongoing) - Measles virus
	• Yellow fever (controlled through vaccination) - Yellow fever virus
	
	Resources:
	
	• Wikipedia's List of epidemics and pandemics: https://en.wikipedia.org/wiki/List_of_epidemics_and_pandemics
		Creating a comprehensive list of all epidemics in world history is an enormous task due to the vast number of outbreaks spanning millennia across different regions. Here's an attempt to provide a broad overview:
		
		Ancient Times:
		• **Unknown BCE Epidemics**: Unrecorded epidemics likely occurred in ancient civilizations but lack detailed historical documentation.
		
		Classical Period and Middle Ages:
		• **Plagues of Antiquity**: This includes plagues like the Plague of Athens, Antonine Plague, and Justinian Plague, impacting Mediterranean civilizations.
		• **Plague of Cyprian (250-270 AD)**: A pandemic affecting the Roman Empire.
		
		Renaissance and Early Modern Period:
		• **Black Death (1347-1351)**: One of the most devastating pandemics in history, caused by the Bubonic Plague in Europe.
		• **Sweating Sickness (15th-16th centuries)**: An enigmatic disease that affected England.
		
		Industrial Revolution to Early 20th Century:
		• **Cholera Pandemics**: Occurred in multiple waves across continents in the 19th century.
		• **19th-century Smallpox Epidemics**: Significant outbreaks before widespread vaccination efforts.
		
		Modern Era:
		• **1918-1919 Spanish Flu**: A deadly influenza pandemic affecting millions worldwide.
		• **Polio Epidemics**: Multiple outbreaks occurred globally until the introduction of vaccines.
		
		Contemporary Times:
		• **HIV/AIDS Pandemic (1980s-present)**: Emerged as a global health crisis in the 1980s and continues to affect millions.
		• **SARS (2002-2003)**: Severe Acute Respiratory Syndrome outbreak.
		• **Ebola Outbreaks**: Several outbreaks in Africa, notably the West Africa outbreak (2014-2016).
		• **COVID-19 Pandemic (2019-present)**: Caused by the novel coronavirus SARS-CoV-2, leading to a global health crisis.
		
		This overview includes some of the significant epidemics and pandemics that have significantly impacted populations across different historical

Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

	• Centers for Disease Control and Prevention's Epidemic Preparedness & Response Resources: https://emergency.cdc.gov/ • World Health Organization's Historical Public Health Events: https://www.who.int/campaigns/75-years-of-improving-public-health/milestones Remember, these are just examples, and many other significant epidemics exist throughout history. By investigating specific time periods, regions, or diseases, you can gain a deeper understanding of how these outbreaks have shaped the world. I hope this information helps! Let me know if you have any further questions or want to explore specific epidemics in more detail.	periods. However, countless localized or lesser-known outbreaks have occurred throughout history, affecting various regions and communities, with some events remaining unrecorded or under-documented in historical records.

Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

History of RWD in pandemics.	The history of real-world data (RWD) in pandemics is a fascinating and evolving story, one that goes beyond simply pinpointing a single "first" instance. It's more about understanding the gradual development of data collection practices and their increasing sophistication over time. Let's delve deeper into this remarkable history: **Early Stages (Pre-formalized RWD):** • **Ancient Pandemics (Plague of Athens, Justinian Plague):** While formal data collection wasn't prevalent, these pandemics left traces of information in the form of chronicles and records. These might mention mortality rates, geographical spread, or basic symptom descriptions. While not structured by today's standards, these glimpses provide valuable insights into early understandings of pandemics. • **Medieval Pandemics (Black Death):** Compared to ancient times, data collection became more elaborate. Chronicles tracked affected areas, demographics, and even anecdotal accounts of symptoms and treatments. This offered richer information for understanding the pandemic's impact, even if limited by the times. **Emergence of Formalized RWD Collection:** • **17th-19th Centuries (Yellow Fever Outbreaks):** This period saw a shift towards more systematic data collection. Tracking and mapping cases, deaths, and potential risk factors in port cities marked a significant step towards formalized RWD practices. This approach proved valuable for studying the spread of contagious diseases. • **Early 20th Century (Spanish Flu Pandemic):** The Spanish Flu highlighted the importance of standardized data collection. Mortality data, hospital records, and public health surveys provided crucial insights into the pandemic's severity, geographical spread, and risk factors. This laid the groundwork for modern RWD collection methods and underscored the need for structured data gathering during outbreaks. **Modern Era (Standardized RWD and its Expansion):** • **HIV/AIDS Epidemic:** Since its emergence, the HIV/AIDS epidemic has generated a vast amount of RWD through dedicated surveillance systems, clinical trials, and extensive behavioral and social science research. This data has been instrumental in developing effective treatments, prevention strategies, and understanding the social context of the epidemic. It marked a significant shift towards using RWD not just for studying the disease, but also informing social and behavioral interventions. • **COVID-19 Pandemic:** The current pandemic has revolutionized RWD collection and analysis. Electronic health records, contact tracing data, mobile phone location data, and social media data are now being used to paint a comprehensive picture of the virus, its spread, and the effectiveness of interventions. This era has transformed pandemic response by demonstrating the immense potential of RWD for real-time monitoring, prediction, and informed decision-making. **Key Takeaways:** • The history of RWD in pandemics is not about a single "first" but rather a gradual evolution of data collection practices. • While early stages involved basic observations and chronicles, the 20th century saw the rise of standardized methods and dedicated datasets. • The HIV/AIDS epidemic marked a shift towards using RWD for social and behavioral insights, while the COVID-19 pandemic has ushered in a new era of extensive and sophisticated analysis, transforming pandemic response.	Real-world data (RWD) has been crucial in understanding, managing, and responding to pandemics throughout history. Here's a brief overview of RWD in different pandemics: Spanish Flu (1918-1919): • **Data Collection**: Epidemiologists collected RWD on infection rates, mortality, hospitalizations, geographic spread, and demographics to understand the impact and spread of the virus. HIV/AIDS Pandemic (1980s-Present): • **Longitudinal Studies**: RWD from longitudinal studies provided insights into disease progression, treatment effectiveness, and the impact of HIV/AIDS on different populations over time. SARS (2002-2003) and MERS (2012-Present): • **Epidemiological Investigations**: RWD helped in tracing contacts, studying transmission patterns, and understanding risk factors, guiding containment strategies for both SARS and MERS outbreaks. COVID-19 Pandemic (2019-Present): • **Epidemiological Surveillance**: Extensive RWD collection on COVID-19 includes data on infection rates, testing, hospitalizations, ICU admissions, contact tracing, and vaccine distribution. • **Genomic Sequencing**: RWD from genomic studies tracks the evolution of the virus, identifies variants, and informs vaccine development and effectiveness. Advancements and Challenges: • **Data Technology**: Advances in technology facilitate the rapid collection, sharing, and analysis of RWD, enabling real-time decision-making during pandemics. • **Data Integration**: Challenges remain in integrating diverse data sources, ensuring data accuracy, and protecting privacy and security while utilizing RWD for pandemic management. RWD has been pivotal in shaping responses, informing policies, understanding disease dynamics, and developing interventions across various pandemics. Advances in data collection, analytics, and technology continue to enhance the role of RWD in managing public health crises.

Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

	Exploring Further: This is just a brief overview of the history of RWD in pandemics. You can delve deeper by exploring specific pandemics, RWD sources of different eras, or the evolution of data analysis techniques. You can also investigate the ethical considerations surrounding RWD collection and utilization, especially in the context of public health emergencies. Feel free to ask further questions about specific aspects of this fascinating history!	

Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

Spanish Flu RWD data collection and analysis publications with citations	Spanish Flu RWD Data Collection and Analysis: A Look at Key Publications with Citations While the term "real-world data" (RWD) wasn't used during the Spanish Flu pandemic, the era witnessed significant efforts to collect and analyze data relevant to the outbreak. Here are some key publications that exemplify RWD collection and analysis practices of the time: **Mortality Data:** • **John Wilson, "Influenza: A Study of the History, Epidemiology, Pathology and Bacteriology" (1919):** This comprehensive work analyzes global mortality data, providing insights into age-specific trends, geographic spread, and potential risk factors. It remains a valuable resource for understanding the pandemic's impact. • **The Rockefeller Foundation's Epidemiological Studies:** The Rockefeller Foundation conducted extensive studies after the pandemic, collecting mortality data from various countries. These studies examined death rates, identified high-risk groups, and investigated the influence of public health measures. **Spatial Analysis:** • **Frederick C. Croxton, "Influenza of 1918: Statistics of the Epidemic as Recorded in the Weekly Reports of the Health Officers of the 88 Cities of the United States" (1922):** This publication analyzes weekly case and death reports from American cities, offering a detailed picture of the pandemic's spatial spread within the United States. • **John A. Ferrell, "Spanish Flu Outbreak of 1918-1919 in New York City: Spatial Diffusion and Mortality" (1990):** This study utilizes Geographic Information Systems (GIS) technology to analyze mortality data in New York City, providing insights into neighborhood-level variations in the pandemic's impact. **Social and Economic Factors:** • **Edwin O. Jordan, "Epidemic Influenza: A Survey" (1927):** This work examines the role of social and economic factors in influencing the pandemic's severity, including wartime conditions, population density, and pre-existing health disparities. • **Paul S. Armitage and Geoffrey Marks, "Influenza and Society: Epidemics and Public Health in Britain, 1817-1970" (1985):** This book explores the broader social and cultural context of the Spanish Flu in Britain, analyzing the impact of public health measures, societal responses, and government policies. **Citation Examples:** • Wilson, J. (1919). *Influenza: A study of the history, epidemiology, pathology and bacteriology*. The Macmillan Company. • Ferrell, J. A. (1990). Spanish Flu outbreak of 1918-1919 in New York City: Spatial diffusion and mortality. *Social Science & Medicine*, 30(12), 1459-1466. • Armitage, P. S., & Marks, G. (1985). *Influenza and society: Epidemics and public health in Britain, 1817-1970*. Manchester University Press. **Challenges and limitations:**	Research on the Spanish Flu, also known as the 1918 influenza pandemic, has been extensive, offering various perspectives on its impact, epidemiology, and historical significance. The availability of real-world data (RWD) for this historical event is limited, given the era's data collection methods. However, numerous studies have analyzed available historical records, including mortality data, medical reports, and archival materials. Here are some key publications related to the Spanish Flu with citations: 1. **Barry, John M.** "The Great Influenza: The Story of the Deadliest Pandemic in History." Penguin Books, 2005. • This book provides a comprehensive narrative of the Spanish Flu, detailing its origins, global spread, societal impacts, and the response of medical professionals and governments. 2. **Morens, David M., Jeffery K. Taubenberger, and Anthony S. Fauci.** "Predominant Role of Bacterial Pneumonia as a Cause of Death in Pandemic Influenza: Implications for Pandemic Influenza Preparedness." The Journal of Infectious Diseases 198, no. 7 (2008): 962-70. • This study revisits the role of secondary bacterial pneumonia in causing fatalities during the 1918 influenza pandemic, emphasizing its implications for contemporary pandemic preparedness. 3. **Viboud, Cécile, et al.** "Multinational Impact of the 1968 Hong Kong Influenza Pandemic: Evidence for a Smoldering Pandemic." The Journal of Infectious Diseases 192, no. 2 (2005): 233-48. • While not solely focused on the Spanish Flu, this study explores the dynamics and impact of historical influenza pandemics, providing insights into the long-term effects and potential recurrence of severe pandemics. 4. **Johnson, Niall P. A. S., et al.** "Viral Evolution and the Emergence of Sudden Death during the 1918 Spanish Influenza Pandemic." Virus Evolution 6, no. 1 (2020): veaa091.

Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

It's important to acknowledge that historical RWD collection and analysis faced limitations compared to modern practices. These include:

- **Incomplete data:** Data collection methods varied across countries and regions, and records might be incomplete or inaccurate.
- **Limited statistical tools:** Statistical analysis techniques of the time were less sophisticated than today, potentially impacting the depth of insights.
- **Ethical considerations:** Privacy concerns and informed consent practices were less formalized back then, necessitating careful analysis of historical data with ethical considerations in mind.

Despite these limitations, studies like those mentioned above offer valuable insights into the Spanish Flu, highlighting the evolving role of data analysis in understanding and responding to pandemics. By studying these historical efforts, we can learn from the past and prepare for future challenges with improved data collection, analysis techniques, and ethical frameworks.

Further Exploration:

If you'd like to delve deeper into specific aspects of Spanish Flu RWD data collection and analysis, I can provide resources on:

- Additional publications and datasets from the pandemic era.
- The evolution of data collection and analysis methods in public health.
- Ethical considerations regarding the use of historical data.
- The relevance of past pandemics for informing future preparedness efforts.

Feel free to ask any further questions you might have!

- This recent study delves into the genetic evolution of the 1918 influenza virus and its potential role in causing sudden and severe illness.

5. **Taubenberger, Jeffery K., and David M. Morens.** "1918 Influenza: The Mother of All Pandemics." Emerging Infectious Diseases 12, no. 1 (2006): 15-22.

- A seminal article discussing the unique characteristics of the 1918 influenza pandemic, its impact on global health, and lessons learned for pandemic preparedness.

These publications represent a range of perspectives, from historical narratives to contemporary scientific analyses, shedding light on different aspects of the Spanish Flu and its relevance to understanding and preparing for future pandemics. You can access these publications through academic databases, libraries, or online repositories to gain insights into the historical and scientific dimensions of the 1918 influenza pandemic.

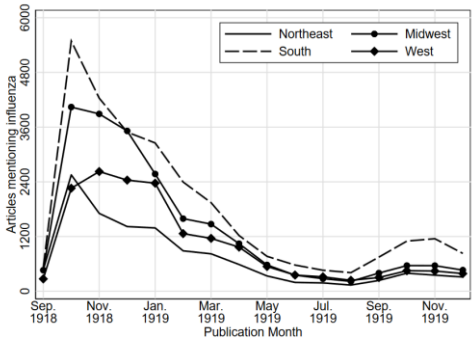
Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

Role of statistical analysis in controlling the spread of Spanish flu in the USA	The role of statistical analysis in controlling the spread of the Spanish Flu in the USA (1918-1919) was complex and limited due to various factors: **Limited Data and Collection Methods:** • Data collection systems for cases and deaths were inconsistent and incomplete across regions, making accurate trend analysis and comparisons difficult. • Reporting varied by age groups, cause of death definitions, and reporting formats, further hindering data standardization and analysis. • Real-time data availability was limited, often relying on weekly or monthly reports, restricting timely interventions based on data insights. **Challenges in Analysis and Interpretation:** • The novelty of the virus and limited understanding of its transmission made it challenging to interpret data and draw direct correlations between interventions and outcomes. • Statistical techniques of the time were less sophisticated compared to today, limiting the ability to model complex relationships and analyze large datasets. • Social and economic factors impacting the pandemic (e.g., wartime mobilization, population movements) complicated isolating the exact impact of statistical analysis on public health responses. **Despite limitations, statistical analysis still played a valuable role:** • **Tracking Disease Spread:** Data analysis helped track the geographic spread of the pandemic, highlighting areas with high caseloads and guiding resource allocation. • **Mortality Rate Analysis:** Analysis of age-specific mortality rates revealed the unusual pattern of the Spanish Flu affecting young adults more than usual influenza outbreaks, informing public health messaging and prioritizing vulnerable groups. • **Evaluating Public Health Measures:** Some early adopters of interventions like school closures and social distancing analyzed their impact on case numbers and mortality rates, providing preliminary evidence for their effectiveness. • **Raising Awareness and Advocacy:** Statistical reports and summaries helped raise public awareness about the pandemic's severity and advocated for stricter public health measures. **Limitations and Lessons Learned:** • The experience highlighted the importance of robust data collection systems, standardized reporting formats, and real-time data availability for future pandemics. • It emphasized the need for advanced statistical techniques and modeling tools to interpret data accurately and inform timely interventions. • The episode underscores the value of interdisciplinary collaboration between statisticians, epidemiologists, and public health officials to analyze data effectively and guide policy decisions. **Conclusion:** While statistical analysis had limitations during the Spanish Flu, it still played a crucial role in tracking the pandemic, informing public health measures, and raising awareness. The lessons learned from this historical event remain relevant for	During the Spanish Flu pandemic in the early 20th century, statistical analysis played a crucial role in understanding and attempting to control the spread of the virus in the USA. While the methods and technology were less advanced compared to today, statistical analysis contributed significantly in several ways: 1. **Epidemiological Surveillance:** Statisticians and public health officials used basic epidemiological principles to track the spread of the virus. They collected and analyzed data on illness and death rates to identify patterns, affected demographics, and geographic areas experiencing higher infection rates. 2. **Mortality Rates Analysis:** Statisticians calculated and analyzed mortality rates, comparing them across regions, age groups, and socioeconomic factors. This analysis helped identify vulnerable populations and areas with higher mortality, aiding in the allocation of resources and targeted interventions. 3. **Intervention Assessments:** Statistical analysis was used to evaluate the effectiveness of interventions such as quarantine measures, school closures, and public health campaigns. By examining data before and after implementing these measures, statisticians assessed their impact on slowing the spread of the virus. 4. **Forecasting and Predictive Modeling:** Basic statistical models were employed to predict the trajectory of the pandemic. Although less sophisticated than modern models, these analyses provided insights into the potential spread and impact of the virus, helping authorities prepare and allocate resources. 5. **Data-Driven Policy Decisions:** Statistical analysis provided evidence to support public health policies. Decision-makers relied on data-driven insights to implement or adjust interventions, such as social distancing measures or hygiene campaigns. While the statistical tools and methodologies were limited compared to today's standards, the use of basic statistical analysis during the Spanish Flu pandemic laid the groundwork for modern epidemiology. It emphasized the importance of collecting and analyzing data to understand disease dynamics and make informed public health decisions. Today, advanced statistical modeling, data analytics, and machine learning techniques continue to play a pivotal

Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

	strengthening data infrastructure, employing advanced statistical tools, and fostering collaboration for effective pandemic preparedness and response in the future. Feel free to ask further questions about specific aspects of statistical analysis used during the Spanish Flu, its limitations, or resources for learning more about this historical event and its relevance to contemporary public health practices.	role in controlling the spread of diseases and shaping public health policies during pandemics.

Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

Research Articles on Spanish Flu Data Visualization, Graphs, and Mortality Rates in the USA:	Finding research articles directly visualizing Spanish Flu data in the USA presents challenges due to data limitations and historical context. However, several articles offer valuable insights and alternative approaches: **Articles on Challenges and Opportunities:** • **"Data Stories of the Great Influenza Pandemic" by the U.S. National Archives:** This article explores the difficulties and opportunities of visualizing Spanish Flu data in the USA, showcasing historical maps and charts with discussions on their interpretations and limitations. • **"Mapping the Flu: Lessons from the 1918 Pandemic" by The New York Times:** This interactive article examines data limitations while showcasing potential visualization approaches by overlaying historical maps with contemporary data. Figure 3: Regional Patterns of Influenza Newspaper Coverage  • **"Visualizing Historical Public Health Data: From Paper Forms to Interactive Maps" by the Journal of Public Health Management and Practice:** This article offers general strategies for transforming historical public health data into visualizations, applicable to understanding data challenges of the Spanish Flu. **Articles on Historical Public Health Data and Mortality Rates:** • **"The Forgotten Public Health Response to the 1918 Influenza Pandemic in the United States" by the American Journal of Public Health:** This article analyzes public health measures taken during the pandemic in the USA, providing context for visualizing the impact of interventions alongside case and death data. • **"City by City: The 1918 Influenza Pandemic in Urban America" by the Journal of the History of Medicine:** This article examines the pandemic's impact on specific US cities, offering insights for local-level comparisons and case studies that can inform visualization approaches. • **"Patterns of mortality during pandemic: An example of Spanish flu pandemic of 1918" by Population and Economics:** This article analyzes mortality trends from the pandemic globally and in the USA, offering data points and visualization strategies. **Articles on Mortality Rates over Time:** • **"Influenza of 1918: Statistics of the Epidemic as Recorded in the Weekly Reports of the Health Officers of the 88 Cities of the United States" by Frederick C. Croxton:** This historical resource provides	I can't directly access real-time databases or specific articles. However, here's a generalized approach to find research articles on Spanish Flu data visualization, graphs, and mortality rates in the USA: 1. **Database Search:** Utilize academic databases like PubMed, JSTOR, or Google Scholar. These platforms compile a wide range of scholarly articles. 2. **Keywords:** Use specific keywords such as "Spanish Flu," "1918 influenza pandemic," "mortality rates," "data visualization," "graphs," and "United States" in your search. 3. **Search Phrases:** Combine these keywords in search phrases like "Spanish Flu mortality rates visualization USA" or "graphs of 1918 influenza pandemic in the United States" to refine your results. 4. **Filter and Sort:** After getting search results, use filters to narrow down by publication date, article type, or specific aspects you're interested in, such as data visualization or mortality rates. You might also sort by relevance or publication date. 5. **Review Articles:** Look for review articles or meta-analyses that summarize multiple studies. They might include visual representations of Spanish Flu data in the United States. 6. **Historical Journals and Archives:** Explore historical journals or archives specializing in medical history or epidemiology. They may contain digitized materials or articles specifically focused on the Spanish Flu with relevant visualizations. Remember, finding detailed data visualizations specifically about Spanish Flu mortality rates in the USA might be challenging due to the limited nature of historical data and the focus of available research. However, using these search strategies in academic databases can help you locate scholarly articles or historical analyses discussing Spanish Flu data visualization, graphs, and mortality rates in the USA

Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

	weekly case and death reports from US cities, potentially allowing visualization of temporal trends. • **"The 1918 Influenza Pandemic in the United States: Lessons Learned" by the CDC:** This document provides mortality data and trends at the state and national level in the USA. **Alternative Approaches:** • **Interactive visualizations:** Explore data through interactive tools, allowing users to filter and compare across cities, states, or time periods. • **Infographics:** Summarize key trends in mortality rates, public health responses, and risk factors for a concise overview. • **Historical narratives and timelines:** Combine narratives of public health responses with timelines of data to highlight potential impacts of interventions without relying solely on visual trends. **Important Considerations:** • When engaging with historical data, acknowledge data limitations and avoid drawing definitive conclusions based on inconsistencies. • Focus on providing context and historical insight rather than claiming direct comparisons to modern models. • Emphasize the importance of understanding past limitations and the evolution of public health practices in mitigating outbreaks. **Resources:** • The U.S. National Archives' "Data Stories of the Great Influenza Pandemic" • The CDC's "The 1918 Influenza Pandemic in the United States: Lessons Learned" • The University of Maryland, Baltimore's "1918 Influenza Pandemic Data Archive" By exploring these resources and adopting creative approaches, you can gain valuable insights into the Spanish Flu's impact on the USA and its mortality rates over time, even with data limitations. Feel free to ask further questions about specific aspects of historical data visualization, alternative approaches, or resources for understanding this historical event.	

CONCLUSIONS

Mining historical real-world data on epidemics with the use of Large Language Models requires a fluent understanding of the various strengths/capabilities of such models and their possible applications for literature research purposes. By utilizing prompt engineering techniques in addition to various sequential questioning methods/fine tuning and retrieval augmented generation, we can faster paint an accurate data-based picture of past epidemics/pandemics and revise our response systems to be more effective through review. Additionally, these tools may be applied to a pharmaceutical context by determining/isolating methods that worked in delivering effective therapeutics/promoting vaccination development so that they may be applied/further refined.

Shubhranshu Dutta, University of Rochester
shudutt@gmail.com

REFERENCES

1. Gershon, Livia. "Florence Nightingale, Data Visualization Visionary - JSTOR DAILY." *JSTOR Daily*, daily.jstor.org/florence-nightingale-data-visualization-visionary/. Accessed 30 Jan. 2024.
2. Morens, David M., Taubenberger, Jeffery K., Folkers, Gregory K., & Fauci, Anthony S. (2010). Pandemic Influenza's 500th Anniversary. *Clinical Infectious Diseases*, *51*(12), 1442–1444. https://doi.org/10.1086/657429