Beyond the Numbers: Dissecting New York City's Leading Causes of Death Across Demographics*

A demographics-based analysis of morbditiy in the Big Apple

Rayan Awad Alim Maria Mangru Reshad

March 12, 2024

We analysed the leading causes of death in New York City, segmented by sex and ethnicity, from 2007 to 2014. Utilizing data derived from NYC death certificates, our research offers insights into public health trends, and the importance of understanding how these causes of death vary across race.

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2....

2 Data

DATA This analysis utilizes data from the New York City Department of Health and Mental Hygiene, specifically provided by the Bureau of Vital Statistics {(**NYCData?**)}. The dataset encompasses records of deaths in New York City since 2007, detailing the leading causes of death categorized by sex and ethnicity. Each entry in this dataset is derived from NYC death certificates which are the official documentation of every death occurring within the city's jurisdiction.

Dataset Context and Broader Implications The investigation into the leading causes of death within New York City holds significant public health importance. It allows for a nuanced

*Code and data are available at: https://github.com/RayanAlim/Analysis-of-Morbidity-in-New-York

understanding of mortality trends, guiding health policy, and intervention strategies tailored to specific demographics and causes. By examining mortality across different sexes and ethnic groups, this analysis contributes to identifying health disparities and targeting efforts to address them.

Variables and Data Examination The dataset comprises several key variables, including: Year: The year of the recorded death. Leading Cause: The primary cause of death as categorized by ICD-10 codes. Sex: The sex of the deceased (Male, Female, or Gender Non-Conforming). Race Ethnicity: The self-reported ethnicity of the deceased, including categories such as Hispanic, White Non-Hispanic, Black Non-Hispanic, Asian and Pacific Islander, and Others. Deaths: The total number of deaths attributed to the leading cause. Death Rate and Age-Adjusted Death Rate: Rates per 100,000 population, providing standardized comparisons across different population sizes and age structures.

Alternative Datasets and Justification While other mortality databases exist, such as the CDC's National Center for Health Statistics, the chosen dataset offers a detailed and localized perspective specific to New York City. This specificity provides a more precise tool for understanding and addressing urban health dynamics, making it more relevant for city-specific policy-making and health intervention strategies.

Data Cleaning and Variable Construction Preliminary data cleaning focused on addressing missing values, ensuring consistency in categorical variables, and suppressing unreliable rates as mentioned. New variables, such as aggregated categories for cause of death or demographic groupings, were constructed to facilitate analysis. For instance, causes of death were grouped into broader categories (e.g., cardiovascular diseases, cancers) to examine trends at a macro level.

Summary Statistics and Relationships Initial exploratory data analysis revealed key trends and disparities in mortality rates across different populations within New York City. We provide summary statistics, including mean death rates and distributions of deaths by cause ethnicity, to provide insights into the health landscape of the city. For example, preliminary findings indicate significant differences in heart disease mortality between ethnic groups, warranting further investigation.

Measurement and Methodological Notes The measurement of mortality and its causes relies on the accurate classification of death certificates, adhering to ICD-10 standards. This classification ensures comparability with other datasets and robustness in identifying health trends. However, it's important to acknowledge potential limitations in cause-of-death reporting and classification, which may impact the analysis. To ensure privacy and reliability, rates based on small numbers (Relative Standard Error, RSE, > 30) and aggregate counts less than 5 have been suppressed. This suppression safeguards against the identification of individuals in rare categories and ensures statistical reliability.

3 Model

3.0.1 Model justification

4 Results

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows…

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows… This suggests…

?@fig-stanareyouokay-2 is a Rhat plot. It shows… This suggests…

References

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.