

Over half of men and 28% of women aged 65-70 died in Manaus, Brazil by July, 2020

Adam Zimmerman

December 1, 2020

Background and Overview

What happened in the Amazon and São Paulo during the COVID-19 pandemic?

This exploratory analysis of three epidemiological and one genetic data sets begins by showing that the city of Manaus, and its Northern Brazilian state, Amazonas, were hit harder per capita than the city and state of São Paulo through the second and third quarters (Q2; Q3) of 2020.

From Q2 into Q3, the intense period of infection correlated with the emergence of an antibody prevalence among Manaus blood donors that was approximately threefold greater than antibody prevalence among São Paulo blood donors (Figure 4; Buss *et al.*, 2020). This was an indication of good news and herd immunity for Manaus, heading into Q4. But the path to that antibody prevalence was characterized by a two- to three-fold increase in mortality for Manaus residents, especially those that were male and over 65 (Figure 6; Prowse *et al.*, 2020). Over half of Manaus men aged 65-70 died.

Genetic data contributed one more distinct, undeniable trend for the Amazon. A variant call analysis of 33 SARS-CoV-2 samples from patients across the Northern Brazilian states surrounding Manaus showed that the great majority of single nucleotide polymorphisms (SNPs) displayed transitions from thymine to cytosine (Figure 8).

Methods

Software

Rstudio 4.0.2 and Ubuntu 20.04 powered three novel analyses of public tabular data (Figures 1-6), alongside a variant call analysis based on code by Naupaka Zimmerman (Figures 7-8).

The pre-print data browsing tool `medrxivr` contributed to the initial exploration of tabular data, as did `0xCOVID19` (McGuinness and Schmidt, 2020;

Mahdi *et al.*, 2020). Hadley Wickham wrote and co-wrote several R packages that supported data reading, tidying, filtering, and formatting (Wickham *et al.*, 2020; Wickham, 2020; Wickham and Hester, 2020; Golemund and Wickham, 2011). His invaluable `ggplot2` package rendered Figures 1-6 and Figure 8, with help from `ggthemes`. (Wickham, 2016; Arnold, 2019). Figure 7 is a screenshot from variant call analysis made possible by VCFR (Knaus and Grünwald, 2017).

Tabular data methods

Throughout 2020, Oxford University hosted and made freely accessible large amount of tabular epidemiological COVID-19 data from across the world (Mahdi *et al.*, 2020). This provided the demographics and case rates data of Manaus and São Paulo (Figures 1-3).

Buss and colleagues inferred elevated levels of herd immunity in Manaus from their tabular data, which included antibody prevalence (Figure 4; Buss *et al.*, 2020). Figure 4 collapses multi-week date ranges to a single data point per month. The source data include specific dates, as well as antibody prevalence values before and after adjusting for seroreversion over time. For full details, refer to Table 1 of Buss and colleagues' pre-print paper (Buss *et al.*, 2020).

Prowse and colleagues also reported on herd immunity in the Amazon and Northern Brazil, providing risk factor data, including mortality by age and sex (Figures 4-5; Prowse *et al.*, 2020).

Genetic data methods

Across Q2 and Q3 of 2020, scientists across Northern Brazil and the Amazon collected 33 Illumina NGS samples and posted them as `.fasta` files on the website of the National Center for Biotechnology Information (NCBI) under the project number PR-JNA662684.

Within a Makefile, quality checking scripts demonstrated the high quality of the majority of this data. Next, a trimmomatic script based on code by Data Carpentry removed the shorter, lower quality runs from the analysis. Several more bash scripts based on code by C. Titus Brown and colleagues compared the reads to the NCBI's reference SARS-CoV-2 genome to output a .vcf variant call format file for interpretation in RStudio. Lastly, single nucleotide polymorphisms (SNPs) were checked for quality (Figure 7) and grouped by position and nucleotide substitution (Figure 8).

Results and Discussion

Tabular data analysis

Cities like Manaus track and report mortality on an ongoing basis, setting a baseline for deaths per week. Excess mortality is the observed number of deaths, minus the baseline expected mortality. Mortality is measured in deaths per 1000 people.

Prowse and colleagues reported that Manaus was the most quickly and severely hit city in the hardest hit region of the Amazon/Northern Brazil (Figure 4; Prowse *et al.*, 2020). OxCOVID19 data corroborated this, showing that Manaus was hit much harder per capita than São Paulo Figure 1-3; (OxCOVID19). Buss and colleagues contributed antibody prevalence data consistent with the herd immunity interpretation (Figure 4; Buss *et al.*, 2020).

Prowse and colleagues showed that total female mortality (996) was more than double (2.13x) expected mortality (468) in the oldest age group of patients, over 75 (not pictured; Prowse *et al.*, 2020). In fact, the situation was worse for slightly younger female patients, ages 65-70, who faced a total mortality (283) 2.78x expected mortality (102) (Figure 5; Prowse *et al.*, 2020).

In contrast, total male mortality (1208) were higher in raw numbers, as well as in proportion (3.16x) to expected mortality (382) in the oldest age group of patients, over 75 (not pictured; Prowse *et al.*, 2020). The ratio of expected (149) to observed mortality (510) was most severe at 3.42x in men, ages 65-70 (Figure 6; Prowse *et al.*, 2020). Men in this age group died at 1.80 times the rate of women of the same age group. In percentage terms, 51.0% of men and 28.3% of women ages 65-70 died by the end of Q2 in Manaus.

In summary, living in Manaus, being over 65, and being male correlated with elevated case rates, anti-

body prevalence, and mortality for Brazilians facing COVID-19 in Q2 and Q3 of 2020.

Genetic data analysis

Typical read quality was high across the 33 SARS-CoV-2 samples in the the Northern Brazil NCBI project analyzed 0 (Figure 7).

The majority of single nucleotide polymorphism mutations (SNPs) among Brazilian sample genomes were pyrimidine transitions from thymine to cytosine (Figure 8). Contextualizing this clear trend could be a fruitful direction for further discovery.

Figures

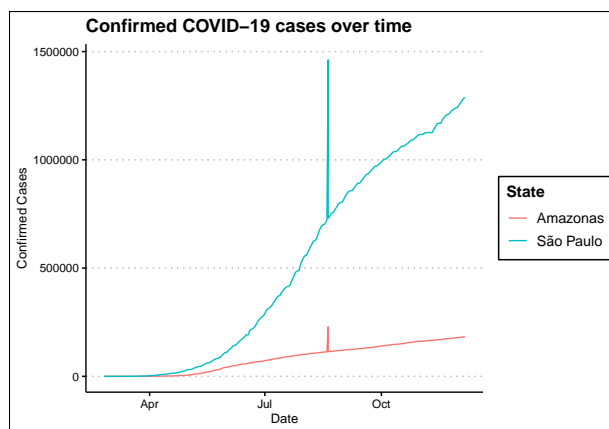


Figure 1: The state of São Paulo saw more total confirmed cases than the state of Amazonas.

Population of Brazil

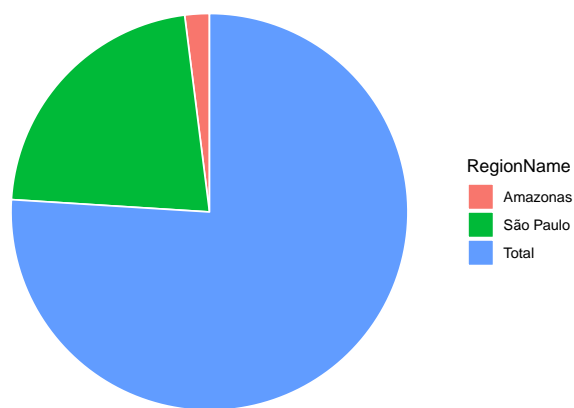


Figure 2: The state of São Paulo is much larger than Amazonas by population at 46 million versus 4 million.

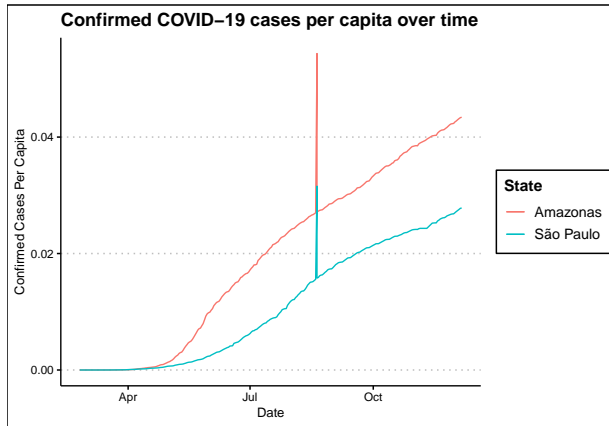


Figure 3: Amazonas saw greater confirmed cases per capita than São Paulo.

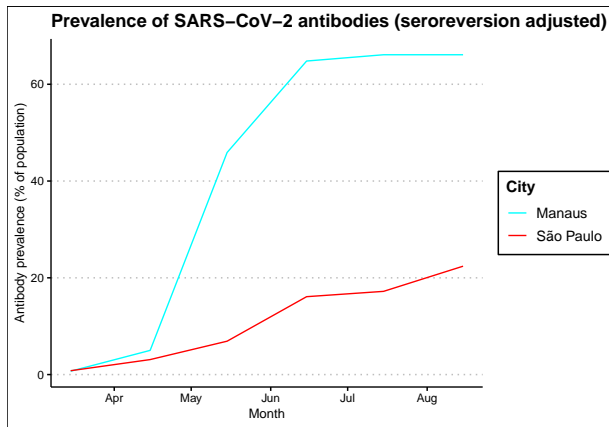


Figure 4: Manaus saw more widespread prevalence of SARS-CoV-2 antibodies vs. São Paulo, up to 66% vs. 22% of the population respectively.

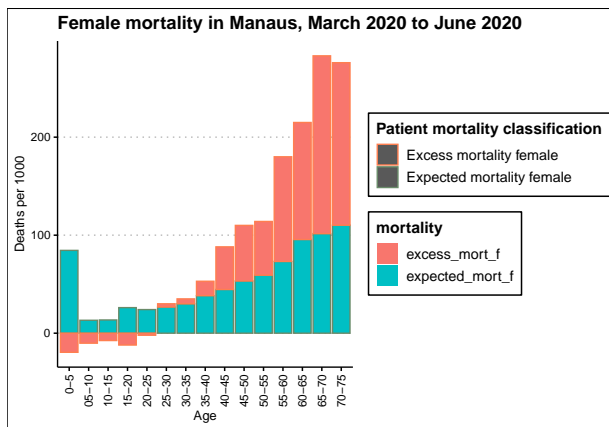


Figure 5: Excess mortality in women peaked at 2.78x times expected mortality in the age group 65-70 years, with a total mortality of 283 deaths per 1000 residents of Manaus.

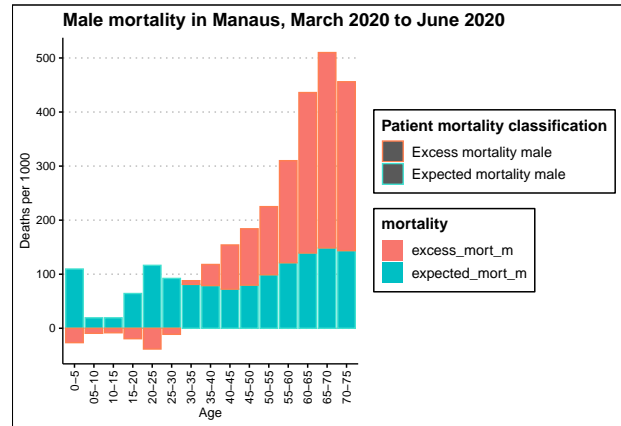


Figure 6: Excess mortality in men peaked at 3.42 times expected mortality in the age group 65-70 years, resulting in a total mortality of 510 deaths per 1000 residents of Manaus. Men in this age group died at 1.80 times the rate of death for women of the same age group.

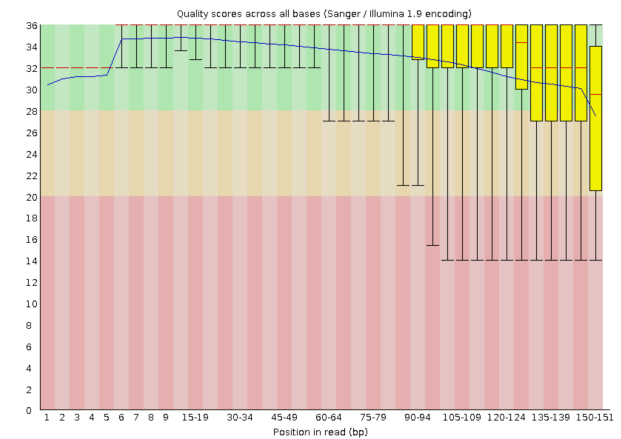


Figure 7: A typical fastq sample of SARS-CoV-2 from Northern Brazil showed high read quality, with most reads scoring above 26.

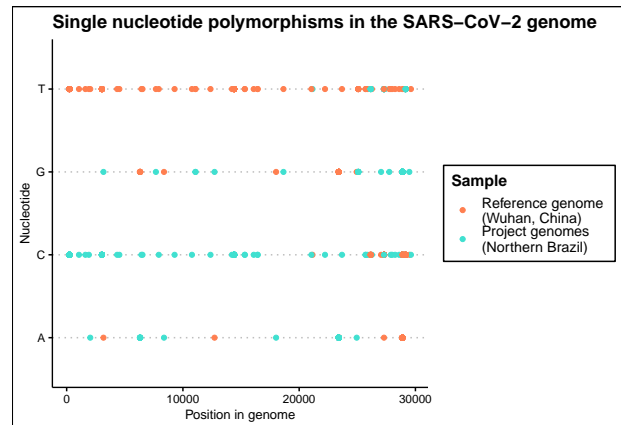


Figure 8: The majority of single nucleotide polymorphisms are located in the 5' region of the SARS-CoV-2 genome.

phism mutations among Brazilian sample genomes were transitions from thymine to cytosine.

Sources Cited

Arnold,J.B. (2019) Ggthemes: Extra themes, scales and geoms for 'ggplot2'.

Buss,L.F. *et al.* (2020) COVID-19 herd immunity in the brazilian amazon. *medRxiv*.

Grolemund,G. and Wickham,H. (2011) Dates and times made easy with lubridate. *Journal of Statistical Software*, **40**, 1–25.

Knaus,B.J. and Grünwald,N.J. (2017) VCFR: A package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, **17**, 44–53.

Mahdi,A. *et al.* (2020) OxCOVID19 database: A multimodal data repository for better understanding the global impact of covid-19. *medRxiv*.

McGuinness,L.A. and Schmidt,L. (2020) Medrxivr: Accessing medRxiv and bioRxiv data in r.

Prowse,T.A.A. *et al.* (2020) Inferred resolution through herd immunity of first covid-19 wave in manaus, brazilian amazon. *medRxiv*.

Wickham,H. (2016) Ggplot2: Elegant graphics for data analysis Springer-Verlag New York.

Wickham,H. (2020) Tidyr: Tidy messy data.

Wickham,H. *et al.* (2020) Dplyr: A grammar of data manipulation.

Wickham,H. and Hester,J. (2020) Readr: Read rectangular text data.