In 2018, Chlamydia, Gonorrhea, and Influenza Were the Top Three Communicable Diseases by Case Count in Toronto*

A Demographic Review

Yixin Guan

01 02 2021

Abstract

In 2018, chlamydia, gonorrhea, and influenza were the top three communicable diseases by case count in Toronto. This paper conducted a demographic review of these three diseases using the data published by Public Health Toronto on the Open Data Toronto website. This paper found that chlamydia and gonorrhea patients who were between 20-39 years old had the most cases compared with other age groups, whereas influenza cases were more evenly distributed among age groups. This paper also found significantly more gonorrhea male patients than gonorrhea female patients in Toronto in 2018.

1 Introduction

Communicable disease refers to the a disease that spreads from one person or animal to another. They can be caused by pathogens such as viruses, bacteria, and fungi (Kandola 2020). Over the past year, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has drastically shown how detrimental one kind of communicable diseases can be to humankind and attracted a lot of public attention. In fact, communicable diseases are no stranger to us. Even in big developed cities like Toronto, people of different ages and sex were still in danger of contracting various communicable diseases. This paper presents a demographic review of the top three communicable diseases by case count in Toronto in 2018. The diseases were chlamydia, gonorrhea, and influenza. This paper found that in Toronto in 2018, patients who were between 20-39 years old have had the most chlamydia and gonorrhea cases compared with other age groups. And more seniors over 60 years had influenza compared with people who had influenza in other age groups. Whereas female and male were some what evenly distributed in the chlamydia and influenza cases, gonorrhea saw more male patients than female patients.

This paper first introduced the source of the communicable diseases dataset used in this study and introduced how the data was collected. It then discussed the limitations and bias of the dataset from both statistical and ethical angle. This paper then indicated the data wrangling made to the dataset and explained the reasons to do so. Using tables and graphs, the paper then drilled down to the age and gender distributions of patients who were infected with chlamydia, gonorrhea and influenza in Toronto in 2018 and ended with a brief discussion of the results and their policy implications.

2 Data

2.1 Data source and data collection

The data is of the communicable diseases in Toronto in 2018 acquired from the Toronto Open Data website (Health 2019) using the R (R Core Team 2020) and the 'opendatatoronto' package (Gelfand 2020). According to the Communicable Disease Surveillance Unit in the City of Toronto, the data were collected by Toronto

*Code and data are available at: LINK.

Public Health (TPH) under the authority of the Health Protection and Promotion Act, Ontario Regulations 559/91, which mandates the notification of all confirmed or suspect reportable diseases to the Medical Officer of Health for the jurisdiction where the patient resides (Communicable Disease Surveillance Unit, n.d.). The data was collected in the following process: Patient information was collected at the point of care. Medical procedures were taken to identify the confirmed and suspect reportable disease patients. Their data with diagnostic results were then sent to the provincial communicable disease information system. The City of Toronto shared the identified data of patients who lived in Toronto as the time of their diagnosis of the communicable disease on the Open Data Toronoto website.

The raw data was structured by the aggregated disease case count by sex (female and male) and age groups of every five years (individuals less than 1 and individuals over 85 has separate age group). The dataset includeed 57 communicable diseases the provincial government mandates to collect (Ontario 2017).

2.2 Data limitation and bias

The data used in this study had bias that could affect our understanding of the actual communicable diseases situation in Toronto in 2018. First, the data of some diseases were not collected in the same time frame s as others. For example, the data of patients with influenza were recorded based on the seasonal year, which spans from September 1, 2018, to August 31, 2019, while the data of patients with other diseases (except malaria and yellow fever) were recorded by the 2018 calendar year. This means that when comparing the case count of influenza against that of the other diseases, we are not comparing them in the same time window. Factors like population and personnel flow in Toronto, general public health measures, public infrastructures may affect the accuracy of the comparison.

Second, missing data may cause statistical bias. According to the data source (Health 2019), due to missing sex data, the total number of cases reported by sex may not correspond with the total number of cases reported for a disease. And as of May 2018, malaria and yellow fever were removed from the communicable disease in Ontario. Moreover, people with stronger immune systems may had recovered without going the see the doctor, in which case, their information would not be documented in the system. Due to the missing data, the actual people infected with communicable diseases will likely be larger than the data used in this paper.

Lastly, the communicable disease was classified by sex but not gender.According to the Canadian Institutes of Health Research, 'sex' refers to a set of biological attributes in humans and animals, whereas 'gender' refers to the socially constructed roles, behaviours, expressions and identities of girls, women, boys, men, and gender diverse people (Health Research 2020). Although it is necessary to study diseases from the lens of patients' biological sex for medical reasons, it is also crucial to recognize the gender diversity in the Ontario population. Trans and/or nonbinary people were not included in the data set and their situation of contracting communicable diseases were not represented in the dataset. Insights gathered from the research of the sex data in this dataset will not reflect the situation of the trans and/or nonbinary people. This may amplify the "data violence" against the trans people (Keyes 2019).

2.3 Data wrangling

In the paper, the 20 age groups were reclassified into five age groups, ie. 0-19, 20-39, 40-59, 60-79, and over 80. The re-classification was conducted for data visualization purposes, although this paper recognize that the more granular the data is, the more accurate the interpretation could be. Moreover, this paper selected the top three diseases by total disease case count for an in-depth demographic review. These diseases selected were chlamydia, gonorrhea, and influenza. The 'tidyverse' package (Wickham et al. 2019), the 'here' package (Müller 2020), and the 'janitor' package (Firke 2021) was used for the data wrangling operations above.

2.4 Data discription

This paper used 'ggplot2' (Wickham 2016) and 'kableExtra'(Zhu 2020) to produce tables and graphs. As shown in (Table 1), there were 13883 cases of chlamydia, ranking the highest among the communicable diseases in Toronto in 2018, followed by gonorrhea (4544 cases) and influenza (2666 cases). (Table 1) also

Table 1: Disease Case Count by Age Group in Toronto in 2018

Disease	Age 0-19	Age 20-39	Age 40-59	Age 60-79	Age 80+	Total
Chlamydia	1839	10440	1487	117	0	13883
Gonorrhea	274	3390	805	74	1	4544
Influenza	485	335	425	718	703	2666

showed the age distribution of these diseases cases. Among chlamydia patients, the 10440 patients falls in the 20-39 age group, making this the largest age group. And no patient over 80 had been infected with chlamydia. Among gonorrhea patients, the age group with the largest disease case count was also 20-39 and only one patient over 80 years old was infected. The top two age groups for patients infected with influenza were the 60-79 age group and the 80+ age group, having 781 and 703 patients. There were 335 patients from the 20-39 age group who were infected with influenza in the 2018 season year, making this the lowest among influenza age groups. (Figure 1) shows the age distribution of the disease case count. Most of the patients with chlamydia and gonorrhea fell in the 20 -39 age group, followed by patients between 0-19 years old and 40-59 years old. More patients between 0-19 than patients between 40-59 were infected with chlamydia, while more patients between 40-59 were infected with gonorrhea compared with those between 0-19. Unlike chlamydia and gonorrhea, influenza had a more prevalent impact across all age groups, as shown in (Figure 1) and (Table 1). Seniors had slightly more influenza cases compared with younger people, with 718 cases for the 60 -79 age group and 703 patients who were over 80. Patients aged between 20-39 had the least influenza cases compared with other age groups.

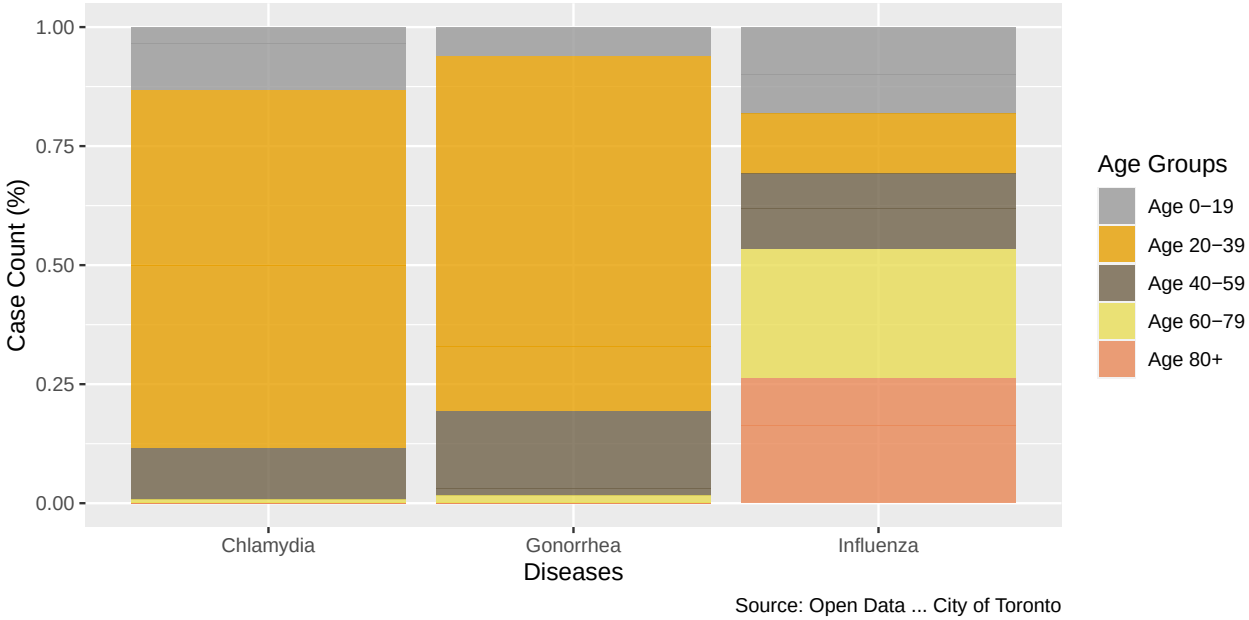

Figure 1: Diseases Case Counts by Age Groups in Toronto in 2018

(Figure 2) and (Table 1) shows the sex distribution among the three diseases. Male and females were relatively evenly distributed in cases of chlamydia and influenza, whereas significantly more male infected with gonorrhea compared with females. (Figure 3) examined the sex distribution of cases across age groups. For chlamydia, female cases were only slightly higher than the male cases in the 20 -39 age group. There were more female patients in the 0-19 age groups and more male cases in the 40-59 age group. For gonorrhea, male cases were significantly higher in the 20-39 and the 40-59 age group. Male and female cases were more evenly distributed among the influenza age groups. However, there were significantly more female influenza patients in the 80+ age group.

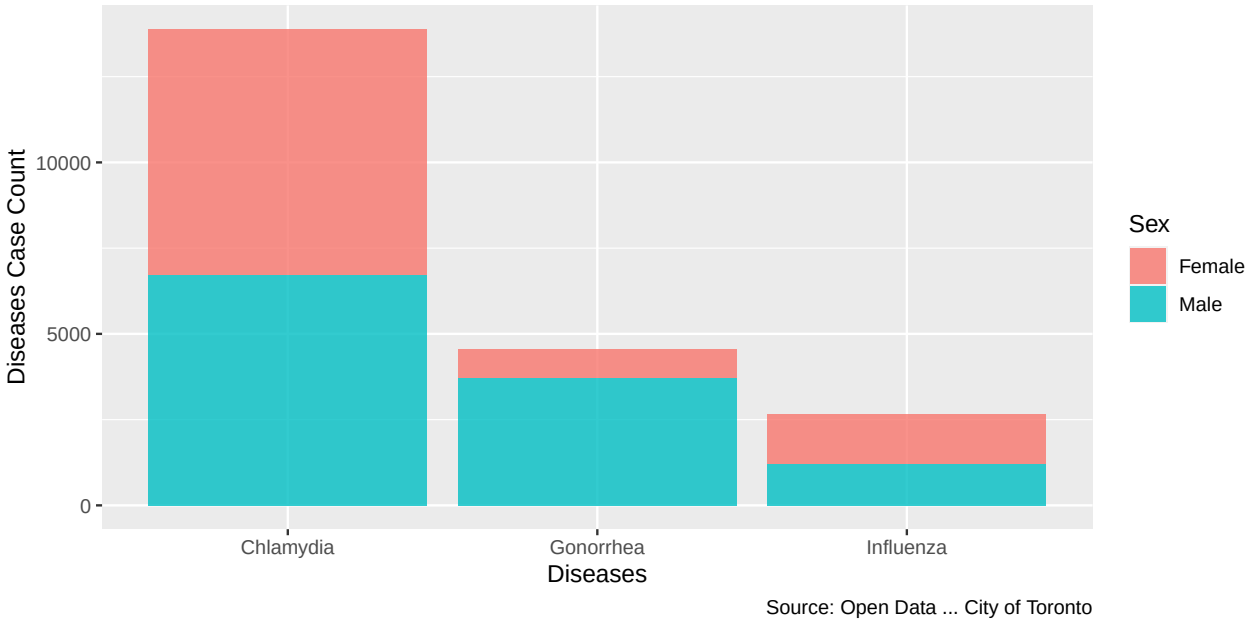

Figure 2: Diseases Case Counts by Sex in Toronto in 2018

Table 2: Disease Case Count by Sex in Toronoto in 2018

Disease	Female	Male
Chlamydia	7168	6715
Gonorrhea	830	3714
Influenza	1450	1216

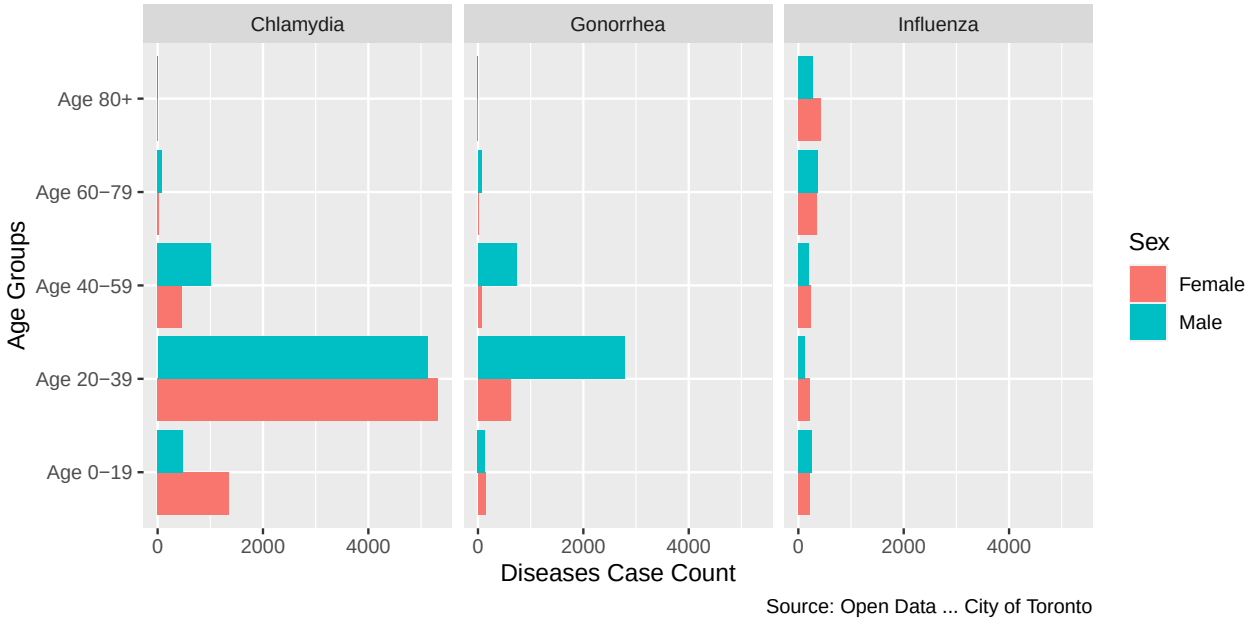

Figure 3: Age and Sex Distributions of Diseases Case Counts in Toronto in 2018

3 Conclusion

In 2018, chlamydia, gonorrhea and influenza were the top three communicable diseases by positive case count in Toronto. The three diseases showed different demographic characteristics. There were slightly more female patients with chlamydia than male patients. Patients between 20 to 39 had the most positive chlamydia cases compared with other age groups. There were significantly more male patients than female patients with gonorrhea. The 20 to 39 age group were also the most populated age group for gonorrhea. Influenza had a more even leveled impact on both sex and all age groups compared with the other two diseases. There were more female influenza patients compared with male patients. Seniors were moderately more infected with influenza than younger age groups.

The results were only descriptions of what happened in Toronto in 2018. This paper did not include any evidence to suggest any causal relationships among sex, age groups, and the communicable disease case count. Further research on these communicable diseases is necessary to obtain a more accurate understanding of the situation and to better inform public health policymakers.

References

Communicable Disease Surveillance Unit, the City of Toronto. n.d. "Communicable Diseases in Toronto – 2018." https://www.toronto.ca/community-people/health-wellness-care/health-inspections-monitoring/communicable-diseases-toronto/.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Health Research, Canadian Institutes of. 2020. *What Is Gender? What Is Sex?* https://cihr-irsc.gc.ca/e/48642.html.

Health, Toronto Public. 2019. "About Annual Summary of Reportable Communicable Diseases." https://www.mdpi.com/2075-4418/10/4/243.

Kandola, Aaron. 2020. "Everything You Need to Know about Communicable Diseases." https://www.medicalnewstoday.com/articles/communicable-diseases.

Keyes, Os. 2019. "Counting the Countless: Why Data Science Is a Profound Threat for Queer People." https://reallifemag.com/counting-the-countless/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Ontario, Public Health. 2017. "List of Reportable Diseases and Unit Contact Information."

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.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.

Zhu, Hao. 2020. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.