Impact of Vaccination Status on ICU numbers*

Krupali Bhavsar

April 06 2022

Abstract

The current global pandemic has brought the world to a standstill since the beginning of 2020. Data was obtained from the Ontario Open Data portal and was studied using tables and graphs. As the pandemic surges on and as the province of Ontario continues going through the motions researching hospital flow pertaining to vaccination status. With better data and sufficient research, adequate information can be relayed about covid-19 patients, vaccination statuses and hospital resources.

Contents

*Code and data are available at: https://github.com/krupalibhavsar/Final-Paper.git

1 Introduction

The current global pandemic has brought the world to a standstill since the beginning of 2020. The spread of SARS- CoV-2 tested both the individual and society as a whole. It is human nature to come together as a community to fight off a threat and that is what we have witnessed these past 20 months. Stay at home orders and mask mandates have aided immensely in the containment of the virus, however with what we know about the nature of this disease and its ability to mutate, vaccines are required to stop the spread. First responders, researchers and manufacturers alike are the reason we are able to imagine the world back at its "natural" state. While health organizations all across the world have been working hard to rid the planet of this disease through vaccine creation and distribution, there has been a rise in a long-standing debate on the effectiveness of vaccines. With this debate comes the inevitable increase and spread of misinformation. Therefore, industry professionals' studies are a reliable and important source of information. Originally, my question focused on how an in depth analysis of covid-19 data could be used to measure the effectiveness of a vaccine from a general perspective. Having combed through the internet and being exposed to all of the data that exists in relation to this topic, I realized that a more precise question would allow for a more detailed analysis. As a result, I began to investigate the correlation between the number of people in the ICU with SARS-CoV-2 and their vaccination status. A large number of news outlets commented on the depletion of resources in hospitals due to the fast-spreading virus during the pandemic. Almost every day, there would be news about hospitals and their intensive care units reaching capacity. Therefore, my research question would aid in understanding how the vaccine would affect the number of people in the Intensive Care Unit and would explore the correlation between vaccination status and admittance into the ICU. My interest in this topic is most probably shared with many others. I became aware of the world's inescapable uncertainty after experiencing the SARS-CoV-2 global pandemic. Many things are out of our control and so the things that are in our control are easier to process and aid us in coping with our anxieties. The purpose of collecting and analyzing data is to find trends that can be applied to other situations. Through this research, I hope to bring some certainty into these uncertain times.

The effects of covid-19 on hospital resources were examined in a number of studies in this area. Prior to understanding how vaccines might have affected the number of patients in the ICU, it is important to understand the initial effects of the pandemic on medical resources. In a study to estimate the depletion of hospital resources as a direct result of Covid-19 in Ontario, Canada it was found that the results would be based on the community following public health measures, (Barrett et al., 2020). If the community were successful in following all public health measures proposed by the government then resources would not be depleted. With inability to follow guidelines resources would be depleted in 14-26 days and as described in the study the 'worst-case scenario' would lead to many patients dying while waiting for resources, (Barrett et al., 2020). There is an important point to remember that this study was conducted in June 2020, long before an adequate vaccination program was underway. Taking into account the beginning stages of the pandemic, we can gain a better understanding of the impact it had on medical resources. We get a clear picture of the strain the global pandemic imposes on health care systems, though there is no definitive proof that a shortage would occur at the time of the study. During the pandemic, a study predicting hospital bed capacity shows a similar pattern. The study itself was conducted to ensure that hospitals are able to plan for capacity earlier to ensure accommodation for all patients, (Deschepper et al., 2020). Additionally, a 2021 study on the effectiveness of the covid-19 vaccine illustrated the success of the vaccine in preventing hospitalization and remaining highly effective during periods of widespread circulation, (Bajema et al., 2021).

2 Data

Data from the Ontario Data Catalogue was obtained to gain a better understanding of how vaccination status affect the number of people that get admitted into the ICU in Ontario, Canada. In this data section 2, a deeper dive into the contents of the data set will be provided. Additionally I will explain how I manipulated the data set to better suit my purposes.

2.1 Data Characteristics

Table 1: Hospital and ICU status

Date	ICU unvac	ICU partially vac	ICU fully vac	hosp unvac	hosp partially vac	hosp fully vac
2021-08-10	22	3	0	23	4	11
2021-08-11	37	5	2	34	7	8
2021-08-12	45	5	2	44	7	9
2021-08-13	52	5	3	65	6	8
2021-08-14	53	4	1	67	6	11
2021-08-15	44	6	1	58	4	14
2021-08-16	48	9	1	55	4	17
2021-08-17	54	6	3	67	8	15
2021-08-18	58	6	5	80	10	16
2021-08-19	55	7	6	76	9	14

All data came from the Ontario Data Catalogue and is titled, "vaccine ICU status.csv." The raw data consists of 7 variables and 74 observation. Daily data was accumulated from August 2021 to October 2021. For everyday represented in the data, information about a patients hospitalization and vaccine status was given. All of the variables in the data set were used and thus each one will be shown in Table 1 and explained in this section 2.1. In Table 1 the first ten rows of the data set are shown. The first variable consists of dates and although this will not be used directly in the model it is important to note that these dates give us an indication of the general vaccine roll out during that given time in Ontario, Canada. Our next three variables, variables 2 to 4, are the total number of people in the ICU separated by their vaccination status consisting of, unvaccinated, partially vaccinated and fully vaccinated. Variables 5 to 7 are the total number of people in hospitals, not in the ICU, separated by their vaccination status consisting of, unvaccinated, partially vaccinated and fully vaccinated. This is the data set as it was in the data catalogue. In the next section, section 2.2, I will present the manipulated data set and explain how it was achieved.

2.2 Data Manipulation

Table 2: Manipulated Data of Hospital and ICU Status

Date	Vaccine Status	ICU	Hospital Non ICU
2021-09-06	partially vaccinated	7	14
2021-10-08	partially vaccinated	11	11
2021-08-31	partially vaccinated	8	19
2021-10-22	unvaccinated	75	114
2021-09-03	partially vaccinated	8	17
2021-08-20	partially vaccinated	7	11
2021-10-07	fully vaccinated	14	49
2021-09-27	unvaccinated	85	104
2021-08-10	fully vaccinated	0	11
2021-08-28	fully vaccinated	8	30

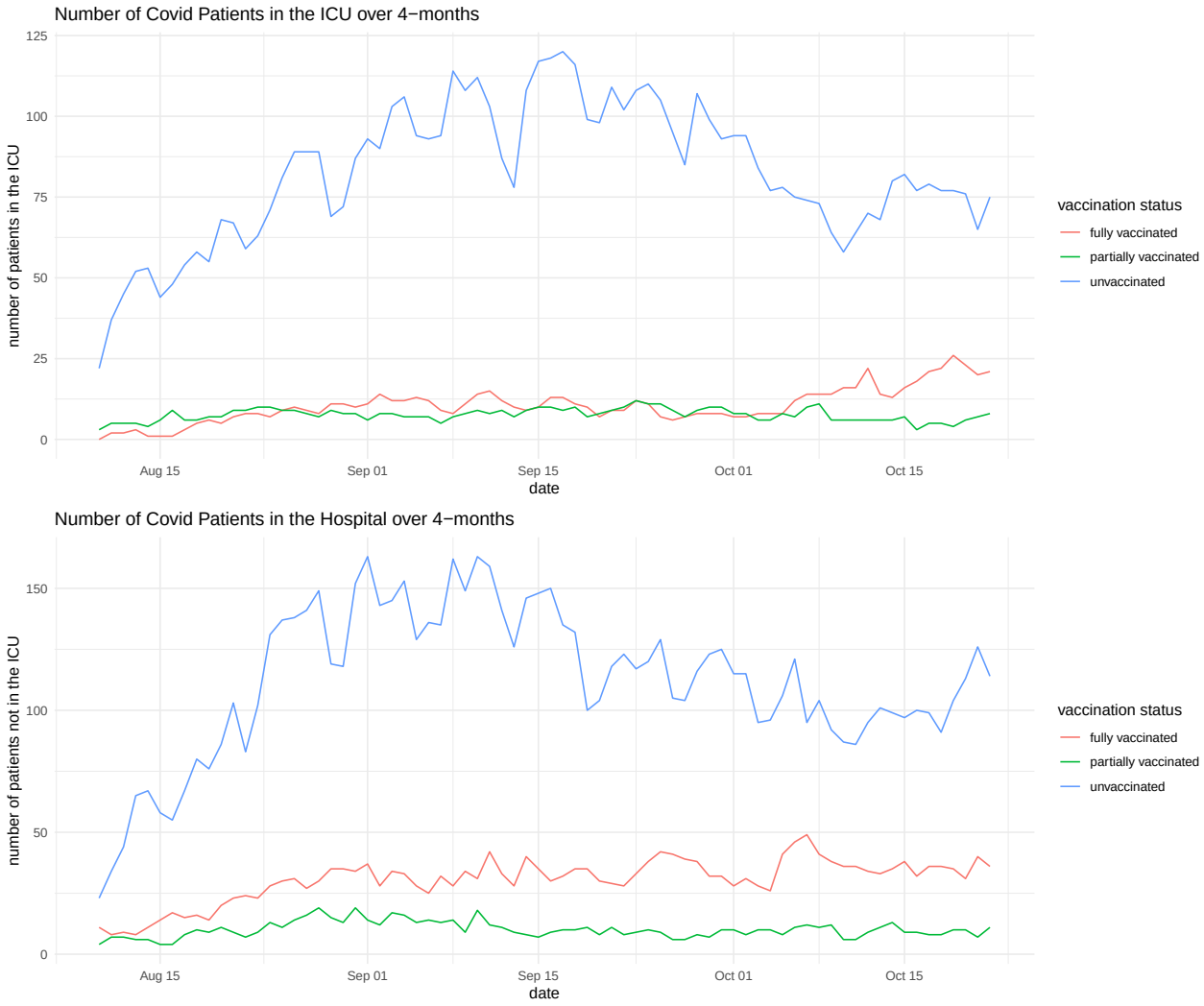
Table 2 shows the first 10 rows of the manipulated data set. Here we still have the first column as date consisting of the date a patient was admitted into the hospital and although this will not be used directly in the model it is important to note that these dates give us an indication of the general vaccine roll out during that given time in Ontario, Canada, as stated earlier. In the second column I added the vaccination status of patients which was deciphered from the original data set. The third column is the total number of

people in the ICU on a specific date separated by their vaccination status, essentially columns 2 to 4 from the original data set were stacked on top of each other. The fourth and last column is the total number of patients in the hospital not in the ICU on a specific date separated by their vaccination status, essentially columns 5 to 7 from the original data set were stacked on top of each other. All of this data was recorded and manipulated using R (R Core Team 2020), tidyverse (Wickham et al. 2019), knitr (Xie, 2021), dplyr (Wickham et al. 2021), tidyr (Wickham 2021) and patchwork (Pedersen 2020).

3 Visualizing the Data and Implications

3.1 Visual One

Here, I am looking to see how vaccination status affected the number of patients in and out of the ICU.

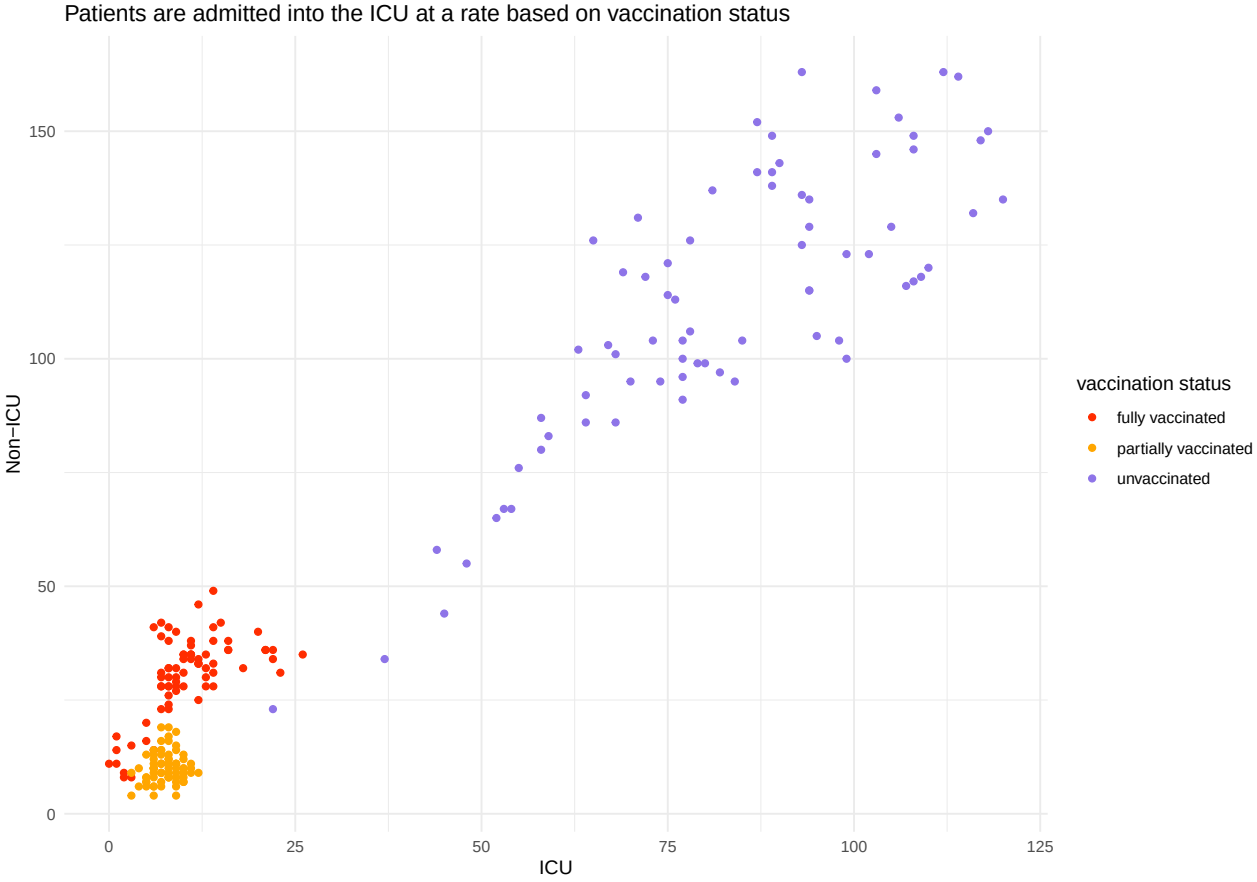


In this graph (Figure 1) we see that the trend in both graphs are very similar. For both graphs we see that the number of unvaccinated patients both in the ICU and in the Hospital outside of the ICU was greater in both graphs, as expected. This tells us that a patients vaccination status does affect the severity of their covid-19 symptoms. Additionally we see that the number of covid patients in the ICU is overall less than the number of covid patients outside of the ICU. This could point to covid-19 patients exhibiting milder symptoms on average however the extent of our data set limits us from making this conclusion. Additionally it seems that more information about population and proportion of people vaccinated would be beneficial to

draw a more detailed and useful conclusion, this will be explored further in section 4.1.

In this first graph we see that there is quite a bit of overlapping between fully vaccinated and partially vaccinated when concerned with patients in the ICU. Then in the second graph we see that the number of partially vaccinated people in the hospital not in the ICU is always lower than the number of patients fully vaccinated in the hospital, not in the ICU. This tells us that although more vaccinated individuals were admitted into the hospital than partially vaccinated, a very similar amount from both groups was later admitted into the ICU. Therefore a greater proportion of partially vaccinated people were admitted into the ICU compared to vaccinated people after having been hospitalized for covid-19.

3.2 Visual Two



We see here in Figure 2 that there is a huge discrepancy between number of people in the hospital separated by vaccination status both inside and outside of the ICU. What is interesting is the the unvaccinated population barely overlaps with the other two groups. This notion was evident in Figure 1 as well however it seems to be shown more clearly here. The unvaccinated population is also the only one out of the three that has a near equal number of people from the hospital admitted into the ICU on a specific day. This means that in the four month duration there was a certain day where every unvaccinated person admitted into the hospital for covid-19 was later admitted into the ICU. For partially and fully vaccinated we immediately see that the number of people in the Hospital not in the ICU is always greater than the number of people in the ICU.

Additionally the hire number of fully vaccinated people both in and outside of the ICU was jarring at first. I then realized that this could be due the the fact that total number of partially vaccinated people in Ontario, Canada was less than the number of fully vaccinated people in the province. Of course this cannot be corroborated with the given dataset alone and thus adds another limitation that will be discussed further in section 4.1.

3.3 Visual Three

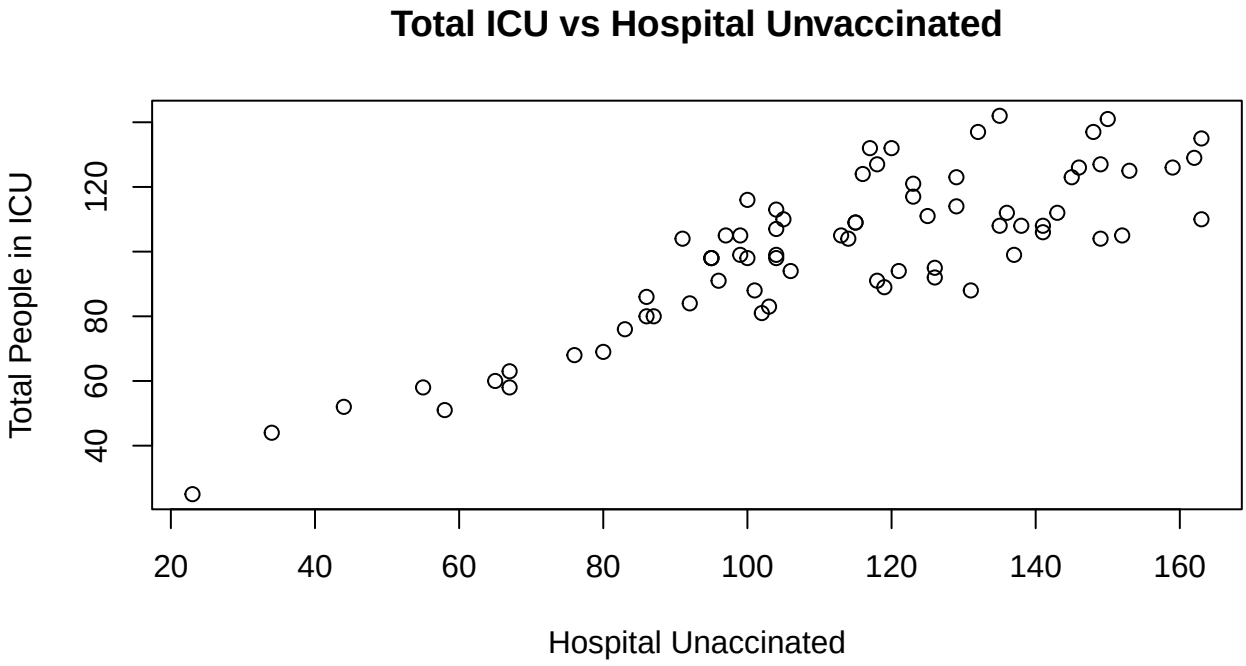


Figure 1: scatter plot of total number of people in the ICU vs number of people hospitalized being unvaccinated

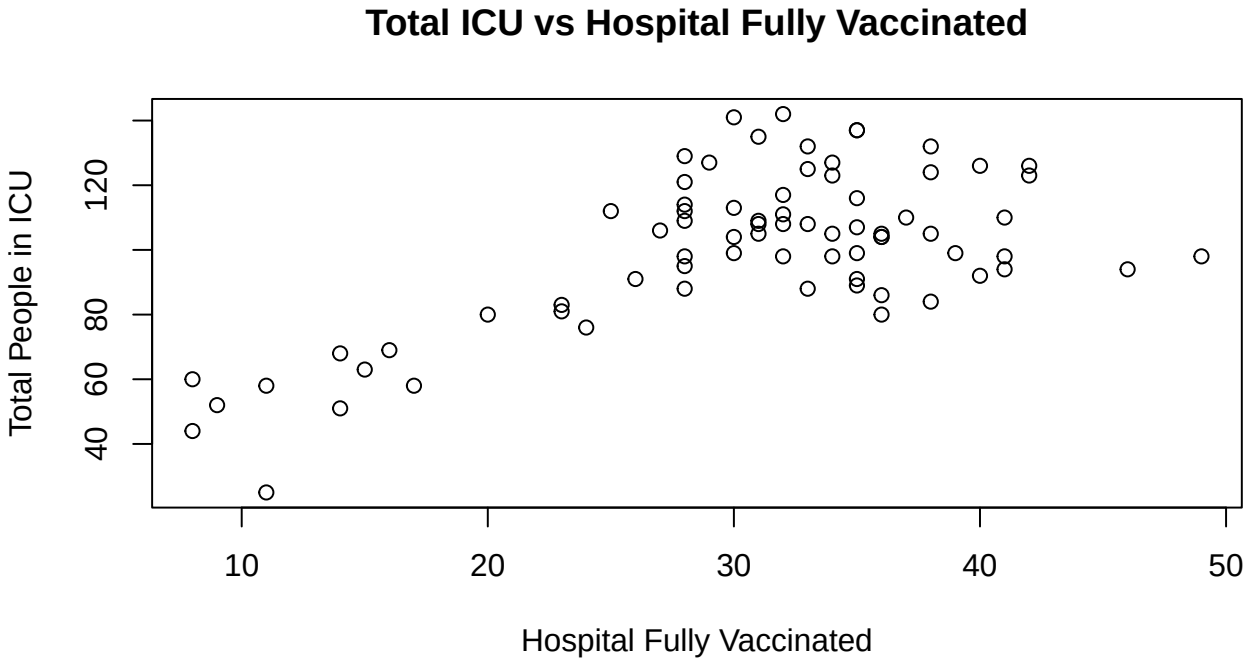


Figure 3 depicts a scatterplot that illustrates the relationship between total number of people in the ICU and total number of unvaccinated individuals both in the hospital and in the ICU. We see a moderate positive linear relationship between both of the variables. As the number of unvaccinated individuals in the hospital goes up the total number of individuals in the ICU follows suit. The variation increases as both values increase.

Figure 4 is a scatter plot that illustrates the relationship between total number of people in the ICU and total number of vaccinated individuals in the hospital. We see a moderate positive linear relationship between both variables here as well however the relationship is considerably weak compared to Figure 3.

With this it seems that we are able to visually see that the number of vaccinated people in the hospital due to contracting covid-19 are less likely to end up in the ICU as depicted by the weak linear relationship. It is evident through both figures that the number of fully vaccinated individuals in the hospital make up a smaller portion of the total number of people in the ICU in comparison with the unvaccinated individuals.

4 Results

Overall, the results obtained from the analysis were an adequate starting point for measuring the efficacy of the covid-19 vaccine and determining what vaccnation status means for the number of patients in the hospital and in the ICU. We were able to identify trend in the given data that showed us how hospital numbers differed based on vaccination status. We were additionaly able to identify the lack of detail in the data set as concrete comparisons could not be made based on the information in the data. We also found that our research question, evidently, could have benefited greatly from a more diverse data set. One that has more information about population and vaccine roll out at the given time. One of the major suggestions points to a need for further investigation due to the many limitations inherent with the data identified through analysis. With a more detailed data set about the vaccines in the province a better analysis can be performed. As a result adequate information can be relayed about covid-19 patients and hospital resources.

4.1 limitations and next steps

The main goal of this study was to bring some certainty into these uncertain times. I wished to better understand the correlation between the number of people in the ICU with SARS-CoV-2 and their vaccination status. Although we gained a better understanding through the various figures created in this paper it seems that limitations in our data calls for some further research and data collection. If information about the general population's vaccine status was given, opposed to only that of those admitted into the hospital, then there would be a clearer understanding of what proportion of people actually got admitted into the ICU based on they vaccination status. Additionally, to further the scope of the research question it would be beneficial to have information about hospital resources. At what rate are the resources depleting? How does the rate of this depletion correlate with the number of covid-19 patients in the hospital separated by their vaccination status? More information about the general population and the state of hospital resources at this given time would aid us in answering these questions. This would then contribute the the greater scope of this research. It would allow us to plan ahead and nalyze trends that would aid the community during this global pandemic. As a society we collect and analyse data in hopes of finding a trend that can be implemented in other situations. By correcting these limitations and implementing these next steps it is my hope that my research is able to bring some clarity to these uncertain times.

5 Appendix

5.1 Datasheet

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
   - The dataset was created to enable analysis of Ontario vaccine data.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
   - The government of Ontario, Canada.
3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*
   - The government of Ontario, Canada.
4. *Any other comments?*
   - No.

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
   - Each row of the dataset represents a group of people in a certain day.
2. *How many instances are there in total (of each type, if appropriate)?*
   - 74
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
   - 3 months are included in the dataset.
4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*
   - Each instance consists of hospital status of individuals based on vaccination status.
5. *Is there a label or target associated with each instance? If so, please provide a description.*
   - Yes, a date.
6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*
   - No.
7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*
   - Yes, through the date.
8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*
   - No.
9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
   - No.
10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any*

restrictions associated with them, as well as links or other access points, as appropriate.
   - Self-contained.
11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*
   - No.
12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*
   - No.
13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*
   - No.
14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*
   - No.
15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*
   - No.
16. *Any other comments?*
   - No.

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*
   - Data was obtained from the Ontario Data Catalogue.
2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*
   - Scraping using R.
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
   - The dataset is not a sample.
4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*
   - Krupali Bhavsar
5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*
   - TBD
6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*
   - No.
7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
   - Third parties.
8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to,*

or otherwise reproduce, the exact language of the notification itself.
   - No.
9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*
   - No.
10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*
   - Consent was not obtained.
11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*
   - No.
12. *Any other comments?*
   - No.

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*
   - Yes cleaning of the data was done
2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*
   - Yes
3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*
   - Yes, R was used
4. *Any other comments?*
   - No

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
   - That information is not available.
2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*
   - No
3. *What (other) tasks could the dataset be used for?*
   - Adding to the given data of several years to study vaccination trends would be interesting.
4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*
   - No.
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
   - No.
6. *Any other comments?*
   - No.

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*

- The dataset is available through GitHub.

2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
   - GitHub
3. *When will the dataset be distributed?*
   - The dataset is available now.
4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
   - No.
5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
   - None that are known.
6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*
   - None that are known.
7. *Any other comments?*
   - No.

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
   - The Government of Ontario, Ontario Data Catalogue
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
   - visit the Ontario Data Catalogue
3. *Is there an erratum? If so, please provide a link or other access point.*
   - Just updated
4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*
   - Yes
5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*
   - No.
6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*
   - Yes, Ontario Data Catalogue.
7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*
   - download dataset from Ontario Data Catalogue.
8. *Any other comments?*
   - No.

6 References

"Covid-19 Vaccine Data in Ontario - Ontario Data Catalogue." COVID-19 Vaccine Data in Ontario - Datasets - Ontario Data Catalogue. Accessed April 6, 2022. https://data.ontario.ca/dataset/covid-19-vaccine-data-in-ontario.

Barrett, K., Khan, Y. A., Mac, S., Ximenes, R., Naimark, D. M. J., & Sander, B. (2020). Estimation of COVID-19– induced depletion of hospital resources in Ontario, Canada. Canadian Medical Association Journal, 192(24). https://doi.org/10.1503/cmaj.200715

Deschepper, M., Eeckloo, K., Malfait, S., Benoit, D., Callens, S., & Vansteelandt, S. (2020). Prediction of hospital bed capacity during the COVID-19 pandemic. https://doi.org/10.21203/rs.3.rs-60176/v1

Bajema, K. L., Dahl, R. M., Prill, M. M., Meites, E., Rodriguez-Barradas, M. C., Marconi, V. C., Beenhouwer, D. O., Brown, S. T., Holodniy, M., Lucero-Obusan, C., Rivera-Dominguez, G., Morones, R. G., Whitmire, A., Goldin, E. B., Evener, S. L., Tremarelli, M., Tong, S., Hall, A. J., Schrag, S. J., ... Tao, Y. (2021). Effectiveness of COVID-19 mrna vaccines against COVID-19–associated hospitalization. MMWR. Morbidity and Mortality Weekly Report, 70(37), 1294–1299. https://doi.org/10.15585/mmwr.mm7037e3

Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

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. 2021. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

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.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.