United States distribution of patients at risk for complications related to COVID-19

Renae Smith-Ray, Erin E. Roberts, Devonee E. Littleton, Tanya Singh, Thomas Sandberg, Michael Taitel

Submitted to: JMIR Public Health and Surveillance
on: April 24, 2020

Table of Contents

United States distribution of patients at risk for complications related to COVID-19

Renae Smith-RayBSc, MA, PhD, ; Erin E. RobertsMPH, BSc, ; Devonee E. LittletonBSc, MSc, ; Tanya SinghBSc, MPH, ; Thomas SandbergBSc, MGIS, ; Michael TaitelBSc, PhD,

Corresponding Author:
Renae Smith-RayBSc, MA, PhD,
Phone: +18472268552
Email: renae.smith-ray@walgreens.com

Abstract

Background: The COVID-19 virus has spread exponentially across the United States. Older adults with underlying health conditions are at especially high risk of developing life-threatening complications if infected. Most ICU admissions and non-ICU hospitalizations have been among patients with at least one underlying health condition

Objective: This study developed a model to estimate the risk status of patients of a nationwide pharmacy chain in the US and to identify the geographic distribution of patients who are at the highest risk of severe COVID-19 complications.

Methods: A risk model was developed using a training test split approach to identify patients who are at high-risk of developing serious complications from COVID-19. Adult patients (age 18+) were identified from the Walgreens pharmacy electronic data warehouse. Patients were considered eligible to contribute data to the model if they had at least one prescription filled at a Walgreens location between October 27, 2019 and March 25, 2020. Risk parameters included age, whether the patient is being treated for a serious or chronic condition, and urban density classification. Parameters were differentially weighted based on their association with severe complications reported in earlier cases. An at-risk rate per 1000 population was calculated at the county level, and ESRI ArcMap was used to depict rate of patients at high risk for severe complications from COVID-19. Real-time COVID-19 cases captured by the Johns Hopkins University Center for Systems Science and Engineering (CSSE) was layered in the risk map to show where cases exist relative to the high risk populations.

Results: Of the 29,824,409 adults included in this study, the average age is 55 years old, 15% have at least one specialty medication, and the average patient has 2 to 3 comorbidities. Nearly 20% of patients have the greatest risk score, and an additional 26.58% of patients are considered high risk with a scores of 8 - 10. Age accounts for 53% of a patient's total risk, followed by the number of comorbidities (30%), inferred COPD, Hypertension, or Diabetes (14%), and urban density classification (4%).

Conclusions: This risk model utilizes data from approximately 10% of the US population. Currently, this is the most comprehensive US model to estimate and depict county-level prognosis of COVID-19 infection. This study shows that there are counties across the US whose residents are at high risk of developing severe complications from COVID-19. Our county-level risk estimates may be used alongside other data sets to improve the accuracy of anticipated healthcare resource needs. The model can also aid in proactive planning and preparations among employers that are deemed critical, such as pharmacies and grocery stores to prevent the spread of COVID-19 within their facilities.

(JMIR Preprints 24/04/2020:19606)
DOI: https://doi.org/10.2196/preprints.19606

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

United States distribution of patients at risk for complications related to COVID-19

Renae L. Smith-Ray, PhD, MA[a], Erin E. Roberts, MPH[a], Devonee E. Littleton, MS[b], Tanya Singh, MPH[a], Thomas Sandberg, MGIS[b], and Michael Taitel, PhD[a]
[a] Walgreens Center for Health and Wellbeing Research, Walgreen Co., Deerfield, IL 60015
[b] Healthcare Planning and Research, Walgreen Co., Deerfield, IL 60015

Corresponding Author: Renae L. Smith-Ray
Email: renae.smith-ray@walgreens.com
ORCID IDs:
Renae L. Smith-Ray: https://orcid.org/0000-0001-9642-7727

Abstract

Background: The COVID-19 virus has spread exponentially across the United States. Older adults with underlying health conditions are at especially high risk of developing life-threatening complications if infected. Most ICU admissions and non-ICU hospitalizations have been among patients with at least one underlying health condition.

Objective: This study developed a model to estimate the risk status of patients of a nationwide pharmacy chain in the US and to identify the geographic distribution of patients who are at the highest risk of severe COVID-19 complications.

Methods: A risk model was developed using a training test split approach to identify patients who are at high-risk of developing serious complications from COVID-19. Adult patients (age 18+) were identified from the Walgreens pharmacy electronic data warehouse. Patients were considered eligible to contribute data to the model if they had at least one prescription filled at a Walgreens location between October 27, 2019 and March 25, 2020. Risk parameters included age, whether the patient is being treated for a serious or chronic condition, and urban density classification. Parameters were differentially weighted based on their association with severe complications reported in earlier cases. An at-risk rate per 1000 population was calculated at the county level, and ESRI ArcMap was used to depict rate of patients at high risk for severe complications from COVID-19. Real-time COVID-19 cases captured by the Johns Hopkins University Center for Systems Science and Engineering (CSSE) was layered in the risk map to show where cases exist relative to the high risk populations.

Results: Of the 29,824,409 adults included in this study, the average age is 55 years old, 15% have at least one specialty medication, and the average patient has 2 to 3 comorbidities. Nearly 20% of patients have the greatest risk score, and an additional 26.58% of patients are considered high risk with a scores of 8 - 10. Age accounts for 53% of a patient's total risk, followed by the number of comorbidities (30%), inferred COPD, Hypertension, or Diabetes (14%), and urban density classification (4%).

Conclusions: This risk model utilizes data from approximately 10% of the US population. Currently, this is the most comprehensive US model to estimate and depict county-level prognosis of COVID-19 infection. This study shows that there are counties across the US whose residents are at high risk of developing severe complications from COVID-19. Our county-level risk estimates may be used alongside other data sets to improve the accuracy of anticipated healthcare resource needs. The ~~model~~ interactive map can also aid in proactive planning and preparations among employers that are deemed critical, such as pharmacies and grocery stores to prevent the spread of COVID-19 within their facilities.

Keywords: COVID-19, modeling, chronic conditions, older adults

Introduction

The first case of COVID-19 was detected in the United States (US) on January 20, 2020 [1]. The spread of the virus increased exponentially across the US during the subsequent two months, with large outbreaks occurring in urban localities including New York City, the San Francisco Bay area, Detroit, and New Orleans [2].

The Centers for Disease Control and Prevention (CDC) analyzed data from lab-confirmed COVID-19 cases in the US from February 12-March 28, 2020. This analysis found that individuals of older age and with underlying health conditions are at higher risk of developing life-threatening complications from COVID-19 [3]. Among COVID-19 patients, 38% had one or more underlying health conditions and the rates of hospitalization among these patients was disproportionately high. The majority of ICU admissions (78%) and non-ICU hospitalizations (71%) were patients with at least one underlying health condition.

Efforts to reduce mortality due to COVID-19 should include identifying and protecting patients who are the highest risk of developing severe complications from the disease. The purpose of this study was to develop a risk model to estimate the risk status for patients of a nationwide pharmacy chain in the US and to identify the geographic distribution of patients who are at the highest risk of severe COVID-19 complications.

Methods

Data Inputs and Sources: Pharmacy Data

Adult patients (age 18+) were identified from the Walgreens electronic data warehouse. Patients were considered eligible to contribute data to the model if they had at least one prescription filled at a Walgreens location between October 27, 2019 and March 25, 2020. Eligible patients were assigned a risk score based on the sum of each patients' risk parameters including: an inferred diagnosis of a serious chronic condition based on a prescription fill within this period for certain specialty medications (Appendix A) or an inferred diagnosis of a chronic condition that is deemed to put the patient at high-risk for severe COVID-19 complications based on a prescription fill to treat these conditions (Appendix B), prescription fills which infer diagnosis of other chronic conditions,

age, and urban density classification. Ethical approval was received by the Advarra Institutional Review Board (protocol number 35300).

Our team assigned a risk value to each parameter based on the findings from recent COVID-19 studies [3, 4]. The risk score algorithm weighted parameters based on their association with complications from COVID-19 infection such as hospitalization and/or death. Parameters shown to be associated with the greatest risk of severe COVID-19 complications were assigned the highest value possible regardless of the presence of other risk factors. The highest risk parameters included a prescription fill for one of the high-risk specialty medications within the study period and/or age 80 or older.

Prescription fills to treat high-risk chronic conditions and other chronic conditions not deemed high-risk were assigned a value based on hazard ratios published in the European Respiratory Journal [5]. Patients with specific underlying health conditions are at high risk of developing severe complications from COVID-19 [3]. The risk score for patients with chronic lung disease, diabetes mellitus, and cardiovascular disease was weighted higher than the risk to patients being treated for other chronic conditions that do not fall into one of these three disease states. Baseline risk is determined by the number of medications the patient is on and whether that medication is for treatment of any chronic condition. Patients treated with medication for one or more of the three high-risk conditions in addition to being treated with additional chronic condition medications received a cumulative value for each category. For instance, a patient being treated for chronic lung disease, diabetes mellitus, and one additional high-risk maintenance medication would receive the following values for these conditions (2.681 + 1.586 + 2.592 = 6.459).

Compounding evidence shows that the risk of developing severe complications from COVID-19 increases exponentially with age; therefore, the risk score was weighted more heavily for patients of increasing age. Observational evidence shows that the spread of COVID-19 occurs most rapidly in urban areas. For this reason, we weighted patients who live in densely populated urban areas with the greatest risk, followed by those in less dense urban, suburban, and rural settings. Counties

categorized as rural contain a population density of < 400 people per square mile, suburban encompasses population density between 400 – 5,000 people per square mile, less dense urban includes counties with 5,000 – 12,500 people per square mile, and urban encompasses population density over 12,500 people per square mile. Population data were acquired from Popstats 2019 (Syergos Technologies, Inc.).

The risk model was developed using a training test split approach. The model was tested and validated using data for patients residing in one state, Georgia, and applied to the full US study cohort. Once cumulative risk values were calculated for each patient, the values were transformed to a maximum risk score of 10 to aid with interpretation using the formula:

$$Risk\,Score = \left(\frac{Patient\,cumulative\,risk\,value}{Highest\,cumulative\,risk\,value} \right) \times 10 \ .$$

Data Inputs and Sources: COVID-19 Surveillance Data

Real-time data of COVID-19 cases captured by the Johns Hopkins University Center for Systems Science and Engineering (CSSE) was layered in the risk map to show where cases exist relative to the populations identified as being at high risk of severe complications from COVID-19 [2].

Model Validation

Model was compared with current trends in COVID-19 cases. Without the availability of confirmed cases, the predictive value of this model is unknown [6].

Mapping

ESRI ArcMap was used to depict the presence of patients identified as being at high risk for severe complications from COVID-19 and real-time COVID-19 cases. The at-risk rate per 1000 population are provided at the county level. County populations with less than 100 residents or less than 10 patients were excluded from the dataset/map. The combined view shows where cases exist relative to the populations identified as high risk. Additionally, Walgreens store and clinic locations are seen with a zoomed in view. The ArcGIS Online platform was used to distribute this map publicly beginning April 16, 2020.

Results

The study included 29,824,409 adults filling at least one specialty or maintenance medication during the study period. Table 1 shows the model inputs and parameters. Using a training test split approach, the model was tested and validated on 623,972 patients residing in Georgia and applied to the full US study cohort (N=29,824,409).

The average age of patients is 50 years old, and the average patient has 2 to 3 comorbidities. Nearly 20% of patients have the greatest risk score, and 26.58% of patients are considered high risk with a score of at least 8 (Table 2). Age accounts for 52% of a patient's total risk, followed by the number of inferred comorbidities (29%), inferred COPD, Hypertension, or Diabetes (16%), and urban density classification (4%).

The risk assigned is weighted the most heavily for adults who are age 80+ (maximum value assigned), followed by adults ages 65-79 [(7 + Age / 100)], ages 50-64 [$(1 + Age / 100)^3$], and ages 18-49 [$(1 + Age / 100)^2$].

Patient addresses were used to depict the distribution of risk status across the US. These data were then compiled to depict a county-level risk status for each county for which we had sufficient data. A county-level at-risk rate was calculated per 1,000 residents. The highest county-level risk category ranged from 265.1-375.0 high-risk residents per 1,000. Eight risk ranges were assembled and color coded onto a county-level US map. The real-time Johns Hopkins University CCSE COVID-19 cases data are layered on top of the county-level risk status to facilitate a visual depiction of the presence of cases in relation to the county-level risk of residents at risk of suffering severe complications from COVID-19 [2]. At the time of publication, the map depicts numerous counties, principally in less densely populated regions of the US that have a high rate of vulnerable residents but have not yet had large numbers of COVID-19 cases. The interactive map depicting the US distribution of patients at risk for complications related to COVID-19 is publically available for viewing [7]. The county-level risk rates are recalculated and refreshed weekly; whereas, the Johns Hopkins University CCSE case numbers are uploaded in real-time.

Discussion

This study shows that there are counties across that US whose residents are at high risk of developing severe complications from COVID-19, many of which had not yet seen many COVID-19 cases when the ~~model~~ interactive map was released. For instance, transmission rates may differ among rural and urban areas, but it is often the case that rural counties have higher risk-statuses and less access to healthcare resources. In the case that disease transmission becomes rampant in a rural county with a high-risk status, healthcare resources may become depleted quickly if a disproportionate number of its residents experience severe complications from the disease.

The risk model utilizes data from approximately 10% of the US population. At the time of publication, this is the most comprehensive US model to depict county-level prognosis of COVID-19 infection [8]. Decaprio et al modeled rates of COVID-19 related pneumonia and hospital admission using 1.5 million records from Medicare claims data from 2015-2016 [9]. Unlike medical claims data, our pharmacy claims data is accessible at a near real-time rate and is which likely improves the precision of the model. Moreover, our data includes US adults age 18 and over, making our population estimates more broad and generalizable.

With the core data, Walgreens was able to implement proactive community outreach by pharmacists to high-risk patients to assure they had sufficient supply of their medications without having to leave their homes by offering home delivery. The pharmacists also inquired about patients' wellbeing during the pandemic and shelter in place orders, and they referred patients to community services as needed. Additionally, by publically sharing de-identified, county-level risk distributions, Walgreens and other organizations are able to plan and respond as COVID-19 begins to spread to areas that previously experienced little impact.

More importantly, our ~~model~~ interactive map will serve to inform public officials and healthcare leaders where there are highly vulnerable pockets of the population so that they may proactively prepare for the possibility of a disproportionately high number patients with severe complications due to COVID-19. Many of these high-risk populations are in rural areas that have limited access to advanced healthcare services such as hospital with respirators. Other ~~models~~ maps have depicted

the availability of healthcare resources, such as intensive care unit beds, available compared to those that will be required in the event of a regional COVID-19 outbreak [10]. Our county-level risk estimates may be used alongside data sets such as that produced by Moghadas et al. to improve the accuracy of anticipated healthcare resource needs.

Our ~~model~~ interactive map will also aid in proactive planning and preparations among employers that are deemed critical, such as pharmacies and grocery stores to prevent the spread of COVID-19 within their facilities. At the time of publication, the interactive map showed that it is relatively uncommon to see a county with a low rate of patients at risk for complications related to COVID-19, but with a high rate of COVID-19 cases. This may be evidence of the differential presentation of SARS-CoV-2 in individuals who are younger and have few comorbidities as compared to their counterparts.

Limitations

There is potential bias in the data source as it only includes Americans who have access to healthcare and can afford to purchase medication. The model would likely be strengthened if it represented less-advantaged individuals who are uninsured or underinsured as well as those who are financially unable to afford their medications. Moreover, since our model relied on pharmacy data, not medical claims data, patient diagnoses were assumed based on the pharmaceutical treatment regimen. Finally, the model could not be externally validated because we did not have access to patient-level COVID-19 case data, which limited our ability to calculate the sensitivity and specificity of the risk model.

While the ~~model~~ interactive map will be useful for multiple purposes, it is for informational purposes only and is not intended to provide medical advice or discourage social distancing or other health-related recommendations. Walgreens will take reasonable steps to timely update this map with the latest available information, COVID-19 is a novel virus and its spread is rapid and unpredictable. We encourage everyone to visit the CDC's Coronavirus (COVID-19) webpage for the latest information and recommendations [11]. We encourage the public to contact their healthcare provider to address any concerns and before taking any personal action in response to the information

provided in the model or map.

Acknowledgements

Author Contributions

RLSR, EER, TS, and MT designed and performed the research; EER wrote the code and developed the weighting for the risk model; RLSR, TS, and MT reviewed the model output and worked with EER to refine the model parameters; DEL developed the map including the upload of both data sources; TS reviewed and performed quality checks of the interactive map; RLSR, EER, DEL, TS, and MT wrote the paper.

Conflicts of Interest

All authors are employees of Walgreen Co.

Abbreviations

JMIR: Journal of Medical Internet Research

References

References

1. Holshue ML, DeBolt C, Lindquist S, et al. First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med.* 2020;382(10):929-936.
2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious diseases.* 2020.
3. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 - United States, February 12-March 28, 2020. *MMWR Morbidity and mortality weekly report.* 2020;69(13):382-386.
4. Wu C, Chen X, Cai Y, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA internal medicine.* 2020.
5. Guan WJ, Liang WH, Zhao Y, et al. Comorbidity and its impact on 1590 patients with Covid-19 in China: A Nationwide Analysis. *The European respiratory journal.* 2020.
6. Russell RE, Katz RA, Richgels KL, Walsh DP, Grant EH. A Framework for Modeling Emerging Diseases to Inform Management. *Emerging infectious diseases.* 2017;23(1):1-6.
7. United States distribution of patients at risk for complications related to COVID-19. 2020; https://walgreens.maps.arcgis.com/apps/webappviewer/index.html?id=cf0649aa9f9f4603a3abb8b0e6732535. Accessed June 5, 2020.
8. Wynants L, Van Calster B, Bonten MMJ, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *Bmj.* 2020;369:m1328.
9. DeCaprio D, Gartner, J, Burgess, T, Kothari, S, Sayed, S, and McCall, CJ. Building a COVID-19 Vulnerability Index. *arXiv.* 2020;arXiv:2003.07347v2.
10. Moghadas SM, Shoukat A, Fitzpatrick MC, et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proceedings of the National Academy of Sciences of the United States of America.* 2020.
11. Coronavirus (COVID-19). 2020;

https://www.cdc.gov/coronavirus/2019-nCoV/index.html. Accessed June 5, 2020.

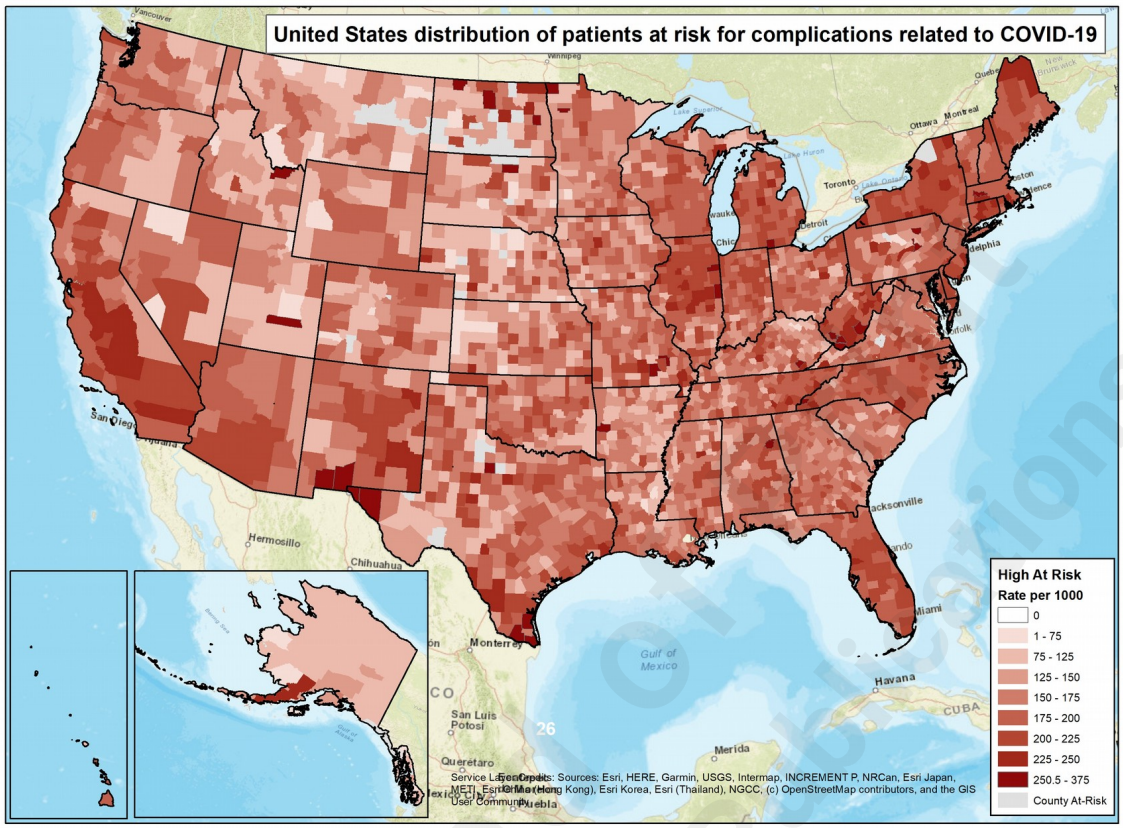
Table 1. Model Inputs and Values

Risk Factor	Risk Value
Baseline Risk	
Maintenance medications for a non-high-risk chronic condition	1.789
Maintenance medications for a high-risk chronic condition	2.289
Maintenance medications (2+)	2.592
Known Disease States For Risk	
Specific specialty medications	maximum
Chronic lung disease medications	2.681
Diabetes mellitus medications	1.586
Cardiovascular disease medications	1.575
Age Related Risk	
18 ≤ Age ≤ 50	$(1+Age/100)^2$
50 < Age ≤ 65	$(1+Age/100)^3$
Age ≥ 65	$(7+Age/100)$
≥ 80 years old	maximum
Urban density classification Risk	
Urban	1
Less dense urban	0.75
Suburban	0.5
Rural	0

Table 2. Risk Category Summary

Risk Category	Number of Patients	Percent
≤ 2	767,538	2.57%
≤ 3	6,521,371	21.87%
≤ 4	4,570,375	15.32%
≤ 5	2,974,637	9.97%
≤ 6	2,373,027	7.96%
≤ 7	2,338,846	7.84%
≤ 8	1,671,698	5.61%
≤ 9	440,574	1.48%
<10	13,892	0.05%
=10	8,152,505	27.33%

Figure 1. United States distribution of patients at risk for complications related to COVID-19



Supplementary Files

Figures

Multimedia Appendixes