COVID-19 Infection Risk & Vaccine Allocation Models: AI/ML & SAS Viya for Nursing Homes

Anzhi Mou, Xiaoyu Zhu, Harvir Virk, Zhaoyu Qiao, Alexandra Allen, and Faculty Advisor Rema Padman.
Heinz College of Information Systems &Public Policy, Carnegie Mellon University, Pittsburgh, PA
Special thanks to SAS Client Representative Manuel Figallo for your unwavering commitment and collaboration.

Executive Summary

Introduction: Roughly 40% of COVID-19 deaths have been among residents and staff of nursing homes and other long-term care facilities in the US—as of December 4, 2020, COVID-19 has infected more than 787,000 people at some 28,000 facilities, resulting in at least 106,000 deaths.[1] Vaccine distribution is underway, but facilities still need supplies to reduce infection spread through the fall surge and beyond.

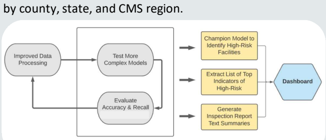
Objective: This project aims to identify Centers for Medicare and Medicaid Services (CMS) nursing homes at high-risk of infection and mortality in order to recommend priority distribution of testing supplies, PPE, and vaccines.

Summary: The team constructed machine learning models to predict nursing home infection and mortality rates based on several datasets containing nursing home COVID-19 infection and mortality data, facility quality ratings, and past inspection report features, as well as systemic factors such as resident demographics, pre-pandemic acuity, and the Social Vulnerability Indexes of their communities to account for the impact of systemic racism and economic inequality on COVID-19 infection and mortality.[2] The team also applied natural language processing techniques to nursing homes' textual inspection reports, demonstrating that they have some predictive potential, and providing a streamlined way for investigators to review past reports. The results were then fed into an interactive SAS Visual Analytic dashboard to display key insights.

Products

Use public data and Artificial Intelligence techniques to:

1. Build a machine learning model to predict/classify Covid-10 infection risk in nursing homes.
2. Stratify facilities by infection and mortality risk to propose a vaccine allocation strategy.
3. Create a web dashboard summarizing important risk indicators and expected nursing home resident vaccine quantities needed by county, state, and CMS region.



**Based on Agile Development Method

References

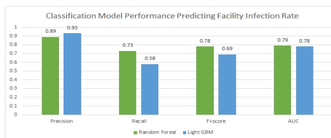
1 The New York Times. (2020, June 27). About 38% of U.S. Coronavirus Deaths Are Linked to Nursing Homes. Retrieved December 18, 2020. https://www.nytimes.com/interactive/2020/us/coronavirus-nursing-homes.html
2 Chidambaram, P., Neuman, T., & Garfield, R. (2020, October 27). Racial and Ethnic Disparities in COVID-19 Cases and Deaths in Nursing Homes. Retrieved December 09, 2020, from https://www.kff.org/coronavirus-covid-19/issue-brief/racial-and-ethnic-disparities-in-covid-19-cases-and-deaths-in-nursing-homes/
3 Centers for Medicare & Medicaid Services (CMS). COVID-19 Nursing Home Dataset, (2020). https://data.cms.gov/Special-Programs-Initiatives-COVID-19-Nursing-Home/COVID-19-Nursing-Home-Dataset/s2uc-8wxp
4 Shaping Long Term Care in America Project at Brown University funded in part by the National Institute on Aging (1P01AG027296). http://ltcfocus.org/1/about-us
5 Centers for Medicare & Medicaid Services (CMS). Provider Information – Nursing Home Compare. (2020). https://data.cms.gov/provider-data/dataset/4pq5-n9py
6 Centers for Medicare & Medicaid Services (CMS). "Full Text of Statements of Deficiencies – October 2020" Note: analysis only used inspection reports from December, 2019 and earlier. https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/CertificationandComplianc/FSQRS
7 Surgo Ventures (2020), Bringing Greater Precision to the COVID-19 Response, Retrieved 11/15/2020, https://precisionforcovid.org/
8 The New York Times, (2020), Coronavirus (Covid-19) Data in the United States, Retrieved 11/15/2020, contained data from 6/5/2020 through 11/15/2020 at the time of download, from https://github.com/nytimes/covid-19-data.
9 Jones, B. and Nagin, D., 2013. A Note on a Stata Plugin for Estimating Group-based Trajectory Models. Sociological Methods & Research, 42(4), pp.608-613.
10 Department of Health & Human Services, & Centers for Medicare & Medicaid Services, (2020, March 15), Tribal Nursing Home & Assisted Living Facility Directory, Retrieved December 19, 2020, from https://www.cms.gov/files/document/2020-directory-tribal-nursing-homes-assisted-living-facilities.pdf
11 Centers for Medicare & Medicaid Services (CMS), Home Health, Hospice, SNF, IRF and LTCH Provider Table, CY 2017, (2017). https://data.cms.gov/Medicare-Hospice/Home-Health-Hospice-SNF-IRF-and-LTCH-Provider-Tabl/44n3-jbm8

Targeting

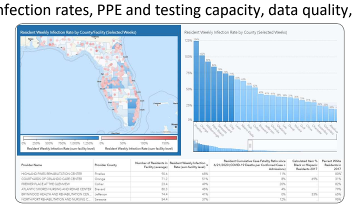
Identify & compare indicators of nursing home COVID-19 infection to target facilities with high infection rates with infection prevention resources such as testing supplies and PPE.

Multiple machine learning models yielded similar recall* levels when using data exclusively from nursing homes to predict which facilities would experience high Covid-19 infection rates:

- *2020 CMS Facility COVID-19: infection, mortality, testing, PPE, staffing*[3]
- *2017 LTCFocus: resident demographics, acuity, and other features*[4]
- *2015-2019 Provider: CMS Quality, Staffing, and Health Inspection Ratings*[5]
- *2019 CMS Inspection Reports: Inspection report key words*[6]
- *2020 COVID-19 Community Social Vulnerability Index*[7]
- *New York Times community infection/mortality*[8]



Descriptive data can still be presented in dashboard to show nursing homes with high recent infection rates, PPE and testing capacity, data quality, and other features.



Sample of Important Features for Predicting Future Infection Rate
From Random Forest & XG Boost

- Last month's resident infection rate
- COVID-19 Community Social Vulnerability
- Community deaths in previous month
- # of Residents (Facility Size)
- % Black Residents
- Total Weighted Health Survey Score

Conclusions/Recommendations: We produced a list of indicators that have some power to predict Covid-19 infection rates in nursing homes. However, the recall scores suggest that there may be limits to how accurately machine learning can predict COVID-19 infection using national-level nursing home data alone. But even linear regression modeling including community infection had difficulties predicting the fall surge, suggesting that the unique nature of the pandemic can be difficult to capture due to novel infection patterns influenced by geography, policy, and human behavior. **We recommend that public health officials use this dashboard tool to target facilities in need of testing/PPE resources and data quality intervention, and to incorporate the import features identified into future predictive time series modeling.**

Impact

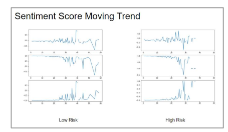
- Equip public health officials to quickly identify facilities/counties experiencing high infection rates and specific PPE/testing resources they need
- Help CMS to more effectively monitor infection/prevention at national level
- Baseline for future modeling to help CMS pre-emptively target facilities for resources

Investigation

Understand nursing home inspection narratives to craft policy changes that address deficiencies predictive of Covid-19 infection/mortality risk.

Natural Language Processing models used on 2019 CMS Inspection text data alone, some predictive power but results limited to mortality before fall surge:

Method	Result /Accuracy
Sentiment Analysis *Target: Mortality/# Bed*	-0.17 Slight negative correlation
BERT *Target: Mortality/# Bed*	0.73
CNN *Target: Mortality/# Bed*	0.69



Integrated custom TextRank key sentence output into dashboard to summarize inspection reports from each facility.

Deficiency Type	Scope-Severity	Key Sentences
Environmental	C Low Severity Widespread Scope	The Administrator and the Director of Maintenance did not provided documentation to indicate that the facility had a comprehensive water management plan in place. The facility did not have a plan that included detailed process flow diagrams, identify dead ends in the water system that were identified during the survey and confirmed by the Director of Maintenance ……..
Pharmacy	D Mild Severity, Isolated Scope	Based on observation, review of facility policy, and staff interviews for medication storage and labeling, the facility failed to accurately label and/or store medications ……..
Quality of Life and Care	G High Severity, Isolated Scope	…The facility failed to ensure the appropriate number of staff were present and assisted with transferring the resident safely out of the bed into the wheelchair and/or failed to ensure safety clips on the mechanical lift were in place to prevent a fall which resulted in an injury The Reportable Event Form dated 1/21/19 at 11:00 AM indicated when the Nursing Supervisor entered the room Resident #1 was observed to be lying on a Hoyer pad on his/her back…….

Conclusions/Recommendations: Sentiment score alone was not a strong predictor of COVID-19 mortality (slight negative correlation), but key words and sentences extracted through ML demonstrated some predictive value. At the same time, the inspection system and reports are both independent factors with a high degree of variability, and the extent to which inspection text alone reflects a facility's mortality or infection prevention capacity is not immediately apparent. **We recommend using custom Text Rank to summarize key sentences from nursing home inspection texts for more efficient review, but not necessarily for predictive modeling.**

Impact

- Allow CMS inspection manager to easily review inspection report summaries rather than having to read full details

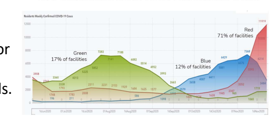
Intervention

Jul Aug Sep Oct Nov Dec

Recommend nursing home vaccine allocation priority levels based on infection and mortality, including anticipated supply needs.

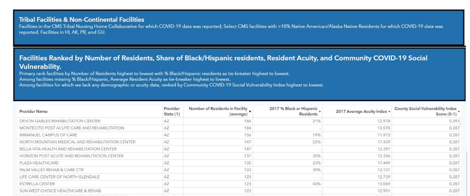
Classified infection groups using Group-Based Trajectory Model (GBTM)[9] for longitudinal clustering on evolving infection rate trends.

TRAJ Classification Shows Varied Past Infection Trends



Created dashboard component to operationalize vaccine allocation strategies, including rankings based on facility size and equity-oriented criteria proof of concept.

Tribal/Non-continental facilities[10,11]
Residents
% Black/Hispanic Residents
High Resident Acuity
High Community COVID Social Vulnerability



Tribal Facilities & Non-Continental Facilities
Facilities in the CMS Tribal Nursing Home Collaborative for which COVID-19 data was reported; Select CMS facilities with >10% Native American/Alaska Native Residents for which COVID-19 data was reported. Facilities in HI, AK, PR, and GU.

Facilities Ranked by Number of Residents, Share of Black/Hispanic residents, Resident Acuity, and Community COVID-19 Social Vulnerability.
Primary rank facilities by Number of Residents highest to lowest with % Black/Hispanic residents as tie-breaker highest to lowest.
Among facilities missing % Black/Hispanic, Average Resident Acuity as tie-breaker highest to lowest.
Among facilities for which we lack any demographic or acuity data, ranked by Community COVID-19 Social Vulnerability Index highest to lowest.

Provider Name	Provider State (1)	Number of Residents in Facility (average)	2017 % Black or Hispanic Residents	2017 Average Acuity Index	County Social Vulnerability Index Score (0-1)
DEVON GABLES REHABILITATION CENTER	AZ	186	21%	12.978	0.391
MONTECITO POST ACUTE CARE AND REHABILITATION	AZ	184	.	13.578	0.287
IMMANUEL CAMPUS OF CARE	AZ	156	19%	11.973	0.287
NORTH MOUNTAIN MEDICAL AND REHABILITATION CENTER	AZ	147	22%	17.439	0.287
BELLA VITA HEALTH AND REHABILITATION CENTER	AZ	147	.	12.391	0.287
HORIZON POST ACUTE AND REHABILITATION CENTER	AZ	137	35%	12.356	0.287
PLAZA HEALTHCARE	AZ	135	23%	17.449	0.287
PALM VALLEY REHAB & CARE CTR	AZ	133	39%	13.121	0.287
LIFE CARE CENTER OF NORTH GLENDALE	AZ	125	.	12.739	0.287
ESTRELLA CENTER	AZ	123	43%	13.069	0.287
SUN WEST CHOICE HEALTHCARE & REHAB	AZ	123	.	12.901	0.287
MESA CHRISTIAN HEALTH AND REHABILITATION CENTER	AZ	120	.	11.927	0.287

Conclusions/Recommendations: Vaccine distribution criteria should be set by public health experts. This dashboard pipeline can operationalize those criteria alongside predictive modeling into an ordered list of facilities at multiple geographic levels. It is the only tool available that joins race/ethnicity and some tribal facility information with COVID-19 data, which is crucial for equitable distribution. **We recommend public health officials build on this tool's framework to factor nursing homes share of Black, Indigenous, and Hispanic residents from vulnerable communities into equitable vaccine distribution decisions.**

Impact:

- Dynamic tool to turn vaccine distribution criteria into ordered list of facilities at multiple geographic levels.
- Provides only resource that incorporates race/ethnicity and some tribal facility impact of COVID-19 in nursing homes