



Analyzing COVID-19 Mortality Within the Chicagoland Area: Data Limitations and Solutions

MATTHEW BLASER¹, MICHAEL D. CAILAS², JOHN CANAR^{3,4}, BRIAN COOPER^{3,4}, PETER GERACI⁴, KRISTIN OSIECKI^{5,4}, APOSTOLIS SAMBANIS⁴

Key Findings

This study demonstrates that when significant “loss of life” from COVID-19 in Long-Term Care Facilities (LTCF) is not accounted for:

- commonly quoted mortality indicators are likely to be inaccurate;
- the spatial distribution of health outcomes is distorted;
- associations of health outcomes with socioeconomic variables are likely concealed; and
- vulnerability model parameters and their association to health outcomes may be misleading.

The results from this study support the recommendation that public health agencies report health outcomes by accounting for LTCF-related mortality. These findings are valid for the Chicagoland area; however, given that high LTCF-related mortality is widespread on a global scale, these recommendations and findings likely have a broad appeal as well.

Summary

Disseminating reliable information and data is a critical component of an effective risk communication and community engagement (RCCE) strategy to combat any pandemic. During the current public health crisis, many agencies and media outlets are reporting health outcome information based on the overall population of Chicagoland geographic regions. The current study demonstrates that by not accounting for the significant loss of life in Long-Term Care Facilities (LTCF), commonly quoted public health outcome indicators are likely to be inaccurate. Identification of regions with high levels of mortality and infection is a prerequisite for an effective mitigation strategy to protect the public and allocate resources. The common practice for visualizing pandemic information is to rely on overall population loss figures and ratios. The current study demonstrates that by doing so, the spatial distribution of Chicagoland critical areas is likely to be distorted. In the current crisis, inequitable public health outcomes are associated with economic and social factors. Separating Chicagoland mortality into two groups, LTCF and household unit populations, and focusing on the latter, allows us to better discern associations with socioeconomic variables for the general population. This finding has a significant implication on the variable selection and model specification for social vulnerability studies.

AUTHOR AFFILIATIONS

1. United States Environmental Protection Agency, Research Associate under an inter-agency agreement with Oak Ridge Institute for Science and Education.
 2. Environmental and Occupational Health Sciences, School of Public Health, University of Illinois Chicago, Chicago, IL.
 3. United States Environmental Protection Agency Region V.
 4. Health Policy and Administration, School of Public Health, University of Illinois Chicago, Chicago, IL.
 5. University of Minnesota, Rochester, Center for Learning Innovation, Rochester, MN.
- Contact author: Michael D. Cailas at mihalis@uic.edu.

Background

The Census Bureau classifies all people not living in housing units as living in “group quarters.”¹ Subsequently, Health-Related Group Quarters (HRGQ) facilities² is the term used to describe nursing, rehabilitation care, and assisted living centers; in this brief, we will be using the term Long-Term Care Facilities (LTCF) adopted by the agencies of the state of Illinois.³ The segment of the population residing in LTCF has distinct characteristics, such as location and medical needs. LTCF residents are a vulnerable population due to the density of the relatively homogenized group in terms of age, multifaceted health needs, and in many cases, underlying chronic diseases. Reports from various public health agencies across the country have recorded an alarming level of deaths and COVID-19-infected cases.⁴ In the state of Illinois the total deaths among residents in LTCF as of July 17, 2020, was 3,974 (54.8% of the total number of state level fatalities).³ On a global scale a similar trend was discerned and by May 11, 2020, Norway and Spain were reporting death tolls related to LTCF as a percent of the total above 60%.⁵ These numbers corroborate the significance of LTCF-related mortality beyond the borders of Illinois.

In this methodology brief, we will not attempt to analyze the causes of the LTCF disproportionate loss from COVID-19. The main objective of this project is to focus on the implications that the magnitude of LTCF losses has on methods for

analyzing and visualizing the pandemic data and deriving reliable data-driven decision support information. These methods are critical for public health policymakers and government officials, who require practical and reliable information to implement mitigation measures and allocate resources to serve the infected public (e.g., expected hospital bed utilization per region). Location becomes critical for many mitigation measures, and by not accounting for LTCF, the identification of high mortality areas (i.e., clusters) is distorted.

The majority of communications and online portals from public health agencies provide overall population information^{3,6,7} and do not separate COVID-19 health indicators (e.g., COVID-19 mortality per county) in terms of the population living in housing units and the vulnerable residents in LTCF. While many states did not communicate this dimension of the pandemic, the state of Illinois and the Illinois Department of Public Health (IDPH) provided a detailed record of the fatalities in LTCF for each county.³ As a risk communication and community engagement measure, the release of LTCF information can enhance awareness of the LTCF-related issue and justify the need for immediate measures. A drawback of this release is its non-database format, or the lack of a portal with such a database.

The need for reliable and accessible information

Dissemination of reliable* information is a critical component of an effective risk communication and community engagement (RCCE) strategy to combat any pandemic. This need is especially true for the current public health crisis, which is characterized by a massive ‘infodemic’ or over-abundance of information.⁹ Access to reliable data can lead to new insights and methodological breakthroughs that can alleviate the crisis caused by the sudden overload of the health care system. A case in point was the severe acute respiratory syndrome (SARS) outbreak in Singapore during 2003, which led to the creation of a time series model that predicted hospital bed utilization up to three days in advance.¹⁰

Reliable data are also needed to estimate the parameters of deterministic models that describe the spread of infectious disease and predict its abatement. For example, a well known compartmental model in this field describing the number of people infected in a closed population over time, t , is the basic SIR model¹¹ with the following three compartments:

Susceptible (t) → Infected (t) → Recovered (t)

Variations of this 1927 model, with the addition of new compartments and transfer paths (e.g., addition of exposure, asymptomatic, and death compartments), are even used today to model the COVID-19 pandemic.¹² These models become useful only if their parameters are estimated. Estimation relies on reliable data generated, primarily by testing and analyzing the databases provided by public health agencies.

Application of forecasting techniques to estimate the mortality curve is another data-dependent endeavor to develop information useful against the pandemic. In recognition of this importance, the Center for Disease Control (CDC) created a portal “to bring together weekly forecasts for COVID-19 deaths in one place” to “help public health decision-making.”¹³

* Reliability in this context implies that the information (i.e., data) will yield similar results when analyzed by different methods. Consistency and replicability of results over time and space is another characteristic.⁸

Data, platforms and methods

- Socioeconomic data are from the American Community Survey; 2018 release of 5-year estimates. <https://data.census.gov/cedsci/>.
- The major data source for this study is the Medical Examiner (ME) Case Archive of COVID-19-related Deaths.⁶ This archive is organized in a searchable online database format and contains information about “deaths that occurred in Cook County that were under the ME’s jurisdiction.” Some of the recorded cases from this database were not from Cook County and have been removed.
- Data preparation and preliminary analysis was performed with the IBM® SPSS® Modeller 18.2.1. During the data preparation phase, text-mining techniques were applied to identify the LTCF records.
- Geospatial data integration and mapping were performed with ESRI ArcGIS Pro and ArcGIS Online.

The urgency of communicating the implications to public health professionals, as well as the death toll that continues to rise, dictates the methodological approach. For the Chicagoland region, the sheer magnitude of LTCF-related losses precludes the need for simulations to verify any assumptions. Comparisons can demonstrate the basic premise of this brief, namely, that loss of life related to LTCF must be included in standardized reporting and further analysis of the pandemic health outcomes. A similar premise is likely valid for the infection (positive) cases as well, a premise that has not been assessed in this project.

This brief is a point in time assessment of the public health crisis caused by an ongoing pandemic. The progression of this public health crisis compelled us to introduce an ArcGIS “story map” where updated information and much more detailed visualizations are presented. This information is accessible at: <https://pubhealthgis.uic.edu/covid-19-dashboard-maps/>

Results and implications

IMPLICATIONS ON MORTALITY BASED ON RACE

By accounting for the two distinct populations living in LTCF and housing units, the COVID-19 mortality based on race in the Chicagoland area differs from the commonly reported overall population mortality. From Table 1, we can see that the Black household residents of Chicagoland have a relatively high number of COVID-19-related mortalities. This is noteworthy since approximately 30% of the residents in Chicago are Black. For the White and Black population groups, notable losses originate in LTCF. The Latinx HP had nearly the same percent mortality as the White HP; however, at an overall population level, the Latinx mortality rate is 20% less than that of the White population, due to the substantially lower number of Latinx cases in LTCF. By relying on overall population figures only, prevailing health disparities may be obscured.

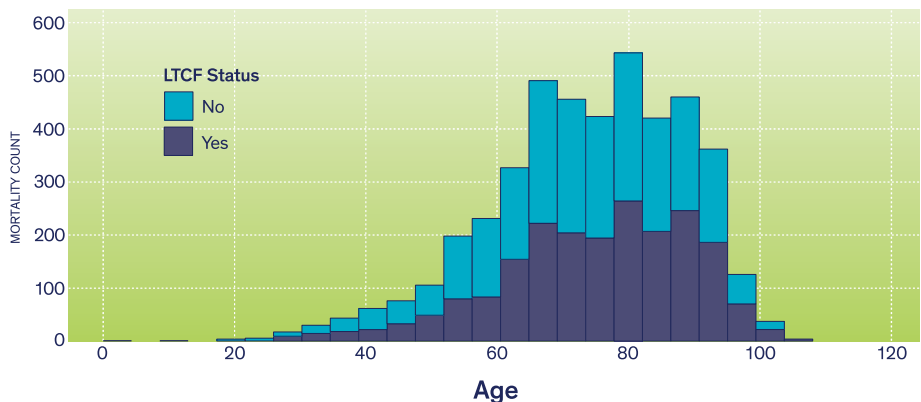
The age distribution of mortality by residency status is presented in Figure 1. The overall pattern conforms to the mortality-age pattern established for the state of Illinois.¹⁴ In addition, this figure underlines the relative magnitude of the loss, especially for older adults above 80 residing in LTCF. Those less than 60 years of age within the LTCF distribution may be attributed to fatalities occurring in rehabilitation centers and specialized clinics, which are included in the LTCF sample.

TABLE 1 Mortality in Chicagoland by race and residency status

		Black	Latinx	White	Other	Total
Household Population	Count	843	688	671	152	2,354
	Row %	35.8	29.2	28.5	6.5	100
	Column %	55.5	78.0	37.7	49.0	52.4
	Total %	18.8	15.3	14.9	3.4	52.4
LTCF Population	Count	677	194	1,111	158	2,140
	Row %	31.6	9.1	51.9	7.4	100
	Column %	44.5	22.0	62.3	51.0	47.6
	Total %	15.1	4.3	24.7	3.5	47.6
Overall Population	Count	1,520	882	1,782	310	4,494
	Row %	33.82	19.63	39.65	6.90	100
	Column %	100	100	100	100	100
	Total %	33.8	19.6	39.7	6.9	100

Data source and notes: Medical Examiner Case Archive of COVID-19-Related Deaths. <https://datacatalog.cookcountyil.gov/>. Date: 07.01.20. LTCF is Long Term Care Facilities.

FIGURE 1: ZIP-Code Level Distribution of Mortality as a Function of Age Within the Long-Term Care Facility (LTCF) and Household Population Groups of Chicagoland Residents (as of July 1, 2020).



VISUALIZING THE IMPLICATIONS ON THE SPATIAL DISTRIBUTION OF MORTALITY

The significance of residency status is likely to have an impact on the spatial distribution of COVID-19-related mortality within the study area. From a geographic point of view, the LTCF and household population groups are distinct. The LTCF mortality is reported as a point (with coordinates) variable, whereas the household population (HP)-related mortality is defined within a polygon surface (HP per ZIP code, per census tract, etc.). To study this effect, we analyzed, at a ZIP code level, the difference between the fatalities in the overall population and those occurring in the HP. The differences were classified into five (5) categories by Jenks' natural breaks (see Figure 2).¹⁵ Under the assumption that the residency status has no implications and that there is no latent factor present (i.e., LTCF), the classification differences would have been minor, falling within Class 1 or 2. The prevalence of the high difference classes, 3 and above (40% of the sample), indicate that the use of the overall population mortality can lead to distorted findings.

The importance of LTCF-related mortality is also verified by directly analyzing the LTCF and HP groups (see Figure 3). Under the no implications assumption, the difference between LTCF and HP mortality per ZIP code would be randomly distributed in the Chicagoland area around the zero horizontal line. Figure 3 demonstrates that ZIP codes with a high LTCF mortality are likely to have a low level of HP-related mortality (i.e., difference is positive and high). As in Figure 2, LTCF-related mortality has a substantial influence on the spatial distribution of mortality at a ZIP code level of aggregation and cannot be ignored. Examining the environmental and socioeconomic reasons for this pattern goes beyond the objectives of this paper and it is the focus of a forthcoming publication.

Attempts to visualize a causal relationship by overlaying point and polygon layers is a common practice widely demonstrated in many recent publications. These maps are based on the mortality points (i.e., coordinates) overlaid on layers depicting the socioeconomic status of geographic areas. Conceptually this is a valid approach; however, it presupposes that each point

FIGURE 2: Difference Between COVID-19 Deaths in the Overall Population and Those Occurring in the Population Residing in Housing Units, by ZIP Code (as of July 1, 2020).

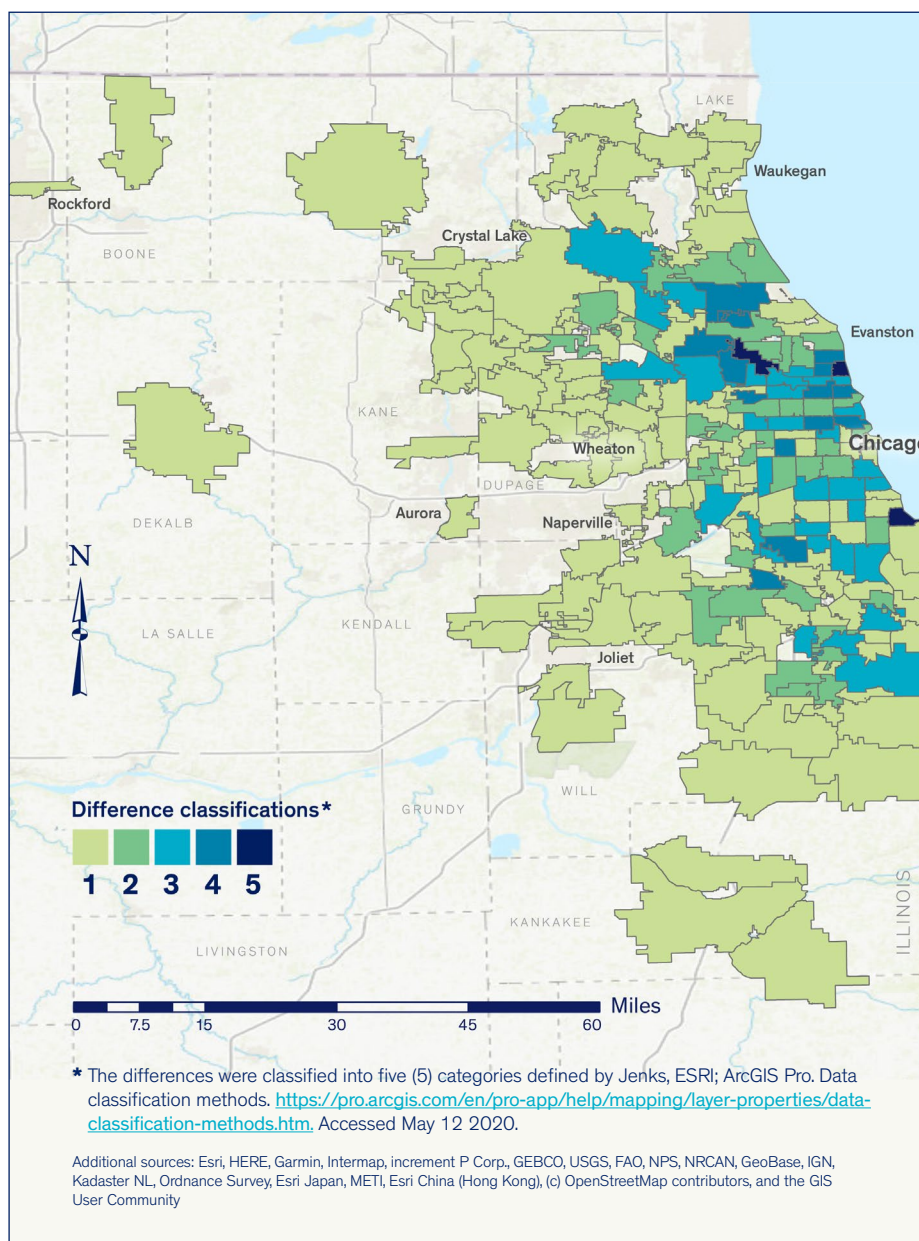
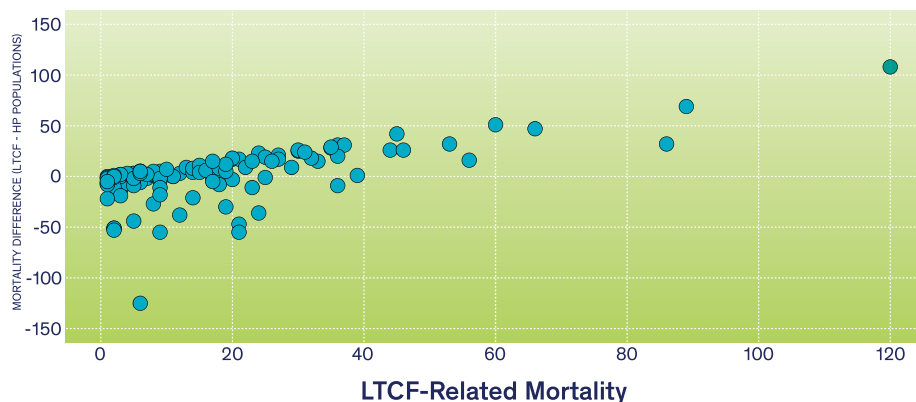


FIGURE 3: Difference of Long-Term Care Facility (LTCF) and Household Population (HP)-Related Mortalities per ZIP Code as a Function of the Number of LTCF Mortalities (as of July 1, 2020).



is a single value. By disregarding the two distinct populations, the visualizations of mortality or infection are distorted since many of these points are, in reality, point mortality “clusters” representing LTCF with multiple observations. These mortality clusters are displayed on maps as single points signifying an individual loss. In reality, they represent something totally different, and in some cases are comprised of clusters with more than 40 deaths per LTCF (e.g., Niles Nursing and Rehab Center).³ By taking into account the LTCF-related mortality clusters, application of spatial autocorrelation tests such as Moran’s I is feasible since “the math for this statistic requires some variation in the variable being analyzed; it cannot solve if all input values are one.”¹⁶ The common approach to avoid this problem is aggregation (e.g., community areas level). Without accounting for LTCF-related mortality, this practice may yield questionable results due to the use of the overall mean, sample size for significance testing, and miss calculation of the distance weights matrix based on the centroid of each aggregation area (i.e., invalid random spatial distribution of deaths assumption).^{16,17}

ANOTHER DIMENSION OF THE CRISIS

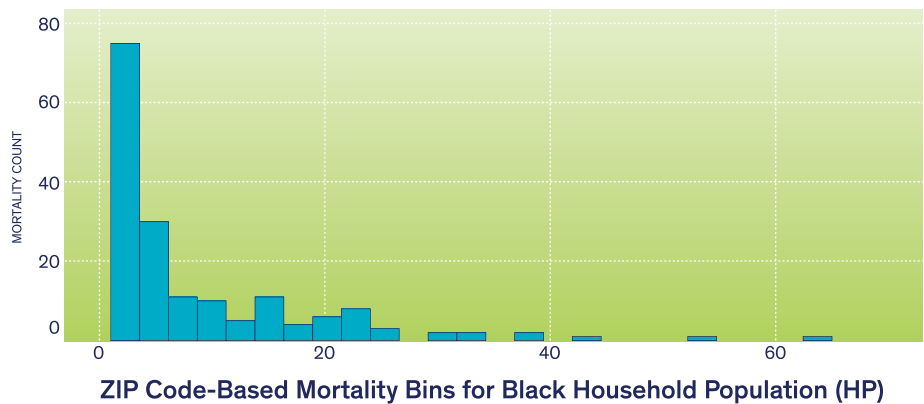
Analyzing the COVID-19 fatalities of household residents by race/ethnicity, as opposed to the overall population, is likely to add another dimension to the health inequity crisis occurring in major metropolitan centers like Chicago. Loss of life per ZIP code is a highly skewed distribution with the majority of ZIP codes registering few fatalities (see Figure 4). To identify patterns and summarize outcome characteristics, the 10 ZIP codes with the highest race-specific mortality were selected for comparison. These cases include ZIP codes with race-specific mortality above the 94th percentile.

HP mortality per race for the top 10 ZIP codes are presented in Table 2. Based on this analysis, the highest death toll for all races occurs in a predominantly Black/Latinx community encompassing the South Lawndale and Little Village communities (i.e., 60623 ZIP code). Comparison of the ZIP code rankings between the overall population and the race-specific columns demonstrates a total discordance for all races. To underline this discordance:

- two of the top ZIP codes in the overall population are not even listed in any of the HP race-specific rankings (i.e., 60714 and 60626);
- the highest overall population mortality ZIP code (60649; 95% Black) with 140 fatalities is only listed for the Black HP with 52 fatalities.

The above findings confirm the importance of LTCF-related mortality for conducting race-specific studies. In addition, they signify the ZIP code-level distinctiveness of the two populations.

FIGURE 4: Distribution of Mortality per ZIP Code for Black Household Population (HP) in Chicagoland (as of July 1, 2020).



White		Black		Latinx		Overall	
ZIP	Mort	ZIP	Mort	ZIP	Mort	ZIP	Mort
60632	19	60628	67	60623	81	60649	140
60623	18	60649	52	60804	50	60623	138
60639	17	60620	49	60639	49	60714	136
60630	15	60619	47	60629	49	60626	110
60638	15	60644	35	60632	40	60625	101
60629	14	60636	35	60647	32	60639	100
60625	14	60617	33	60608	31	60628	91
60804	12	60643	30	60609	24	60453	90
60016	12	60623	29	60618	24	60629	86
60634	11	60651	25	60641	20	60632	85
Avg. =	14.7	Avg. =	40.2	Avg. =	40.0	Avg. =	105.3

Data source and notes: Medical Examiner Case Archive of COVID-19-Related Deaths. <https://datacatalog.cookcountyil.gov/>. Date: 07.01.20

Mort is mortality; Avg. is the average of the ZIP code sample above.

As shown in Table 1, White and Black LTCF-related mortality comprised close to one-half of the overall population mortality for each race. In this section, we will further examine the significance of these numbers. Table 3 underlines the alarming reality for the older population of Chicagoland residing in LTCF. At a comparison level, relying on the overall population numbers conceals that for the White and Black population in Chicagoland, another inequitable public health crisis lies within the LTCF. It is worth noting that the majority of LTCF-related mortalities are not within the top HP-related ZIP codes. The mortality rate picture becomes much more alarming if the population living in “group quarters”¹ is used as a denominator for the LTCF-related mortality rate. LTCF residents belong to this group designation and the US Census Bureau¹ enumerates them. With this modification, the results are presented in the PM.GQ column where PM.GQ is the mortality rate as a percent of the people living in those group quarters. Given that the overall mortality rate for the City of Chicago is approximately 0.1%⁷, the LTCF rates reveal the alarming level of this disparity.

SOCIOECONOMIC CONDITIONS AND HEALTH OUTCOMES

The magnitude of the toll on LTCF residents is likely to distort the association between neighborhood socioeconomic characteristics and disease outcomes (Figures 5 and 6). Regardless of the heteroscedasticity pattern (i.e., in Figure 5, the dispersion of the mortality observations increases with the magnitude of poverty), there is a relatively strong positive association between COVID-19 mortality and the percent of families with income below the poverty level at a ZIP code level (Spearman's rho or $\rho = 0.45$). Such a feature is not discerned in Figure 6 ($\rho = 0.22$). This finding has significant implications on the methodology to develop social vulnerability models for the disease since, in most applications, they rely on a preselected number of variables that are assumed to define vulnerability.¹⁸ Similar patterns have been observed with other socioeconomic variables as well.

White HP		White LTCF			Black HP		Black LTCF		
ZIP	Mort	ZIP	Mort	PM.GQ	ZIP	Mort	ZIP	Mort	PM.GQ
60632	19	60714	62	4.9	60628	67	60649	81	8.4
60623	18	60626	48	1.4	60649	52	60453	38	7.0
60639	17	60614	36	0.9	60620	49	60644	33	3.4
60630	15	60463	35	4.1	60619	47	60626	31	0.9
60638	15	60090	34	5.9	60644	35	60652	28	13.5
60629	14	60068	33	6.4	60636	35	60628	22	3.5
60625	14	60016	32	3.7	60617	33	60430	22	6.8
60804	12	60640	32	1.2	60643	30	60473	21	3.0
60016	12	60706	30	6.9	60623	29	60621	18	2.8
60634	11	60202	29	4.4	60651	25	60625	16	0.7
Avg. =	14.7	Avg. =	37.1	3.8%	Avg. =	40.2	Avg. =	31.0	5.0%

Data source and notes: Medical Examiner Case Archive of COVID-19-Related Deaths. <https://datacatalog.cookcountyil.gov/>. Date: 7.1.20

Mort is mortality; Avg. is the average of the 10 ZIP code sample; PM.GQ is the percent mortality based on the group quarters population per ZIP Code.

HP is Household Population

FIGURE 5: Association of Mortality per ZIP Code and Poverty Level for the Household Chicagoland Population (HP; as of July 1, 2020).

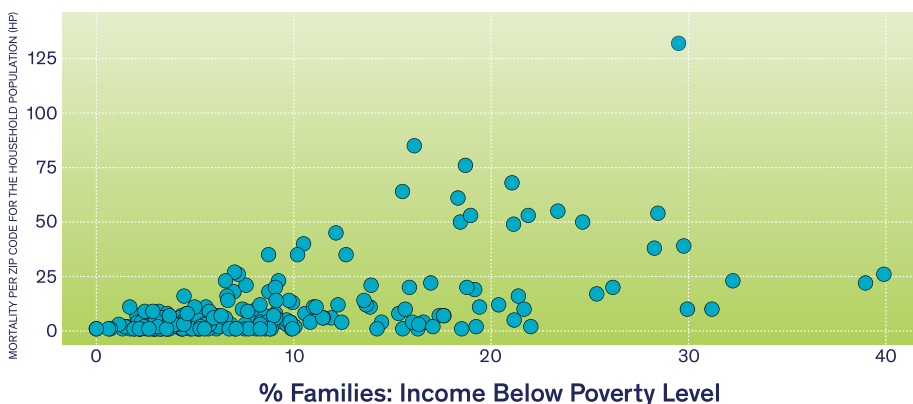
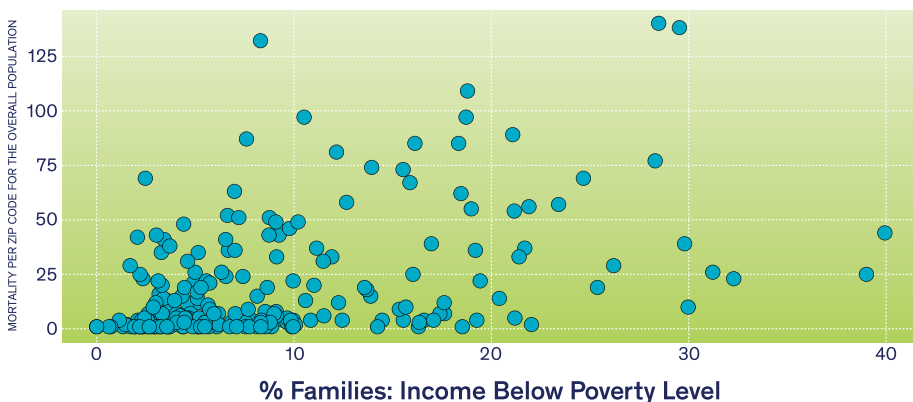


FIGURE 6: Association of Mortality per ZIP Code and Poverty Level for the Overall Chicagoland Population (as of July 1, 2020).



Conclusions and Recommendations

This study demonstrates that when examining COVID-19 mortality it is important to distinguish LTCF-related mortality. Specifically, commonly quoted mortality indicators may be inaccurate and disguise the toll on certain sub-populations, such as was found for the Latinx Chicagoland communities. In addition, the alarming levels of LTCF-related mortality are concealed. Reliably identifying regions with high levels of mortality and infection is a prerequisite for implementing effective mitigation measures and preparedness and response plans to protect the public and allocate resources. The common practice for visualizing pandemic information is to rely on overall population loss figures and ratios; by doing so, the spatial distribution of Chicagoland critical areas is likely to be distorted. Separating mortality into two groups, LTCF and household unit populations, and focusing on the latter, allows us to better discern spatial patterns and critical areas. Without accounting for these two geographically distinct populations, spatial autocorrelation and hot spot analysis methodologies may yield questionable results. From an emergency management

perspective, this separation is critical since the mitigation measures for each population differ. In the current crisis, inequitable public health outcomes are associated with economic and social factors that are likely to exacerbate them. By not accounting for the two distinct populations, these associations are likely to be concealed, limiting the ability to ascertain causal relationships.

The results from this study support a key recommendation that public health agencies report health outcomes by separately accounting for LTCF-related mortality. At a practical level, there is the need to operationalize the LTCF information by organizing it in a relational database format and making it accessible for public and research use from a reliable dissemination portal such as the IDPH.¹⁴

These findings are valid for the Chicagoland area; however, given that high LTCF-related mortality is widespread on a global scale, it is likely that the recommendations and findings have a broad appeal as well.

References

1. U.S. Census Bureau. Group Quarters/Residence Rules. Retrieved from <https://www.census.gov/topics/income-poverty/poverty/guidance/group-quarters.html>. Accessed 12 April 2020.
2. U.S. Census Bureau STUDY SERIES (Survey Methodology #2013-06). Ethnographic Study of the Group Quarters Population in the 2010 Census: Healthcare Facilities (Long Term and Hospice Care). <https://www.census.gov/library/working-papers/2013/adrm/ssm2013-06.html>. Accessed 15 April 2020.
3. Long-Term Care Facility Outbreaks COVID-19. <http://dph.illinois.gov/covid19/long-term-care-facility-outbreaks-covid-19>. Accessed 26 June 2020.
4. McMichael TM, Currie DW, Clark S, et al. Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington. *N Engl J Med*. 2020; 382(21):2005–2011.
5. Coronavirus disease 2019 (COVID-19) in the EU/EEA and the UK – tenth update, 11 June 2020. Stockholm: ECDC; 2020.
6. Medical Examiner Case Archive – COVID-19 Related Deaths. <https://datacatalog.cookcountyil.gov/Public-Safety/Medical-Examiner-Case-Archive-COVID-19-Related-Dea/3trz-enys>. Accessed 26 June 2020.
7. City of Chicago: Daily COVID-19 status report. <https://www.chicago.gov/city/en/sites/covid-19/home/latest-data.html>. Accessed 26 June 2020.
8. International Organization for Standardization (2005): Quality management systems – Fundamentals and vocabulary (ISO 9000:2005).
9. World Health Organization (WHO). Novel Coronavirus (2019-nCoV) Situation Report – 13. 2 February, Geneva; 2020.
10. Earnest A, Chen M, Ng D, Yee L. Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. *BMC health services research*. 2005; 5:36.10.1186/1472-6963-5-36.
11. Kermack WO, McKendrick AG. A Contribution to the Mathematical Theory of Epidemics. *Proc. Roy. Soc. Lond.* 1927; A 115, 700-721, 1927.
12. Sarbaz H A, Rizgar K, Sadegh S. Mathematical modelling for coronavirus disease (COVID-19) in predicting future behaviours and sensitivity analysis *Math. Model. Nat. Phenom.* 2020; 15 33. DOI: 10.1051/mmnp/2020020
13. U.S. Department of Health & Human Services. Centers for Disease Control and Prevention (CDC). Forecasts of Total Deaths; <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html>. Accessed June 25 2020.
14. Illinois Department of Public Health (IDPH): Coronavirus Disease 2019. <http://dph.illinois.gov/covid19>. Accessed June 25 2020.
15. ESRI; ArcGIS Pro. Data classification methods. <https://pro.arcgis.com/en/pro-app/help/mapping/layer-properties/data-classification-methods.htm>. Accessed May 12 2020.
16. ESRI; ArcGIS Pro. Spatial Autocorrelation (Global Moran's I, Spatial Statistics). <https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/spatial-autocorrelation.htm>. Accessed May 17 2020.
17. Goovaerts P. *Geostatistics for Natural Resources Evaluation*. Applied Geostatistics Series, Oxford University Press, New York; 1997.
18. Burton C, Rufat S, Tate E. Social Vulnerability: Conceptual Foundations and Geospatial Modeling. In: Fuchs S, Thaler T, eds. *Vulnerability and resilience to natural hazards*. Cambridge University Press; 2018.

ACKNOWLEDGMENTS

The authors would like to acknowledge the data retrieval, address confirmation and database preparation contribution provided by Mr. Emmanuel Cailas. Many thanks as well to the Director and Deputy Director of the Policy, Practice and Prevention Research Center for their recommendations that improved the quality of this brief.

SUGGESTED CITATION

Blaser M., Cailas M.D., Canar J., Cooper B., Geraci P, Osiecki K., Sambanis A. Analyzing COVID-19 Mortality Within the Chicagoland Area: Data Limitations and Solutions. Research Brief No. 117. Policy, Practice and Prevention Research Center, University of Illinois Chicago. Chicago, IL. July 2020. https://go.uic.edu/P3RC_Publications