


A Typology of COVID-19 Data Gaps and Noise From Long-Term Care Facilities: Approximating the True Numbers

Gerontology & Geriatric Medicine
Volume 8: 1–8
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23337214221079176
journals.sagepub.com/home/ggm


Terry E. Hill¹  and David J. Farrell¹

Abstract

Although there is agreement that COVID-19 has had devastating impacts in long-term care facilities (LTCFs), estimates of cases and deaths have varied widely with little attention to the causes of this variation. We developed a typology of data vulnerabilities and a strategy for approximating the true total of COVID-19 cases and deaths in LTCFs. Based on iterative qualitative consensus, we categorized LTCF reporting vulnerabilities and their potential impacts on accuracy. Concurrently, we compiled one dataset based on LTCF self-reports and one based on confirmatory matching with California's COVID-19 databases, including death certificates. Through March 2021, Alameda County LTCFs reported 6663 COVID-19 cases and 481 deaths. In contrast, our confirmatory matching file includes 5010 cases and 594 deaths, corresponding to 25% fewer cases but 23% more deaths. We argue that the higher (self-report) case total approximates the lower bound of true COVID-19 cases, and the higher (confirmed match) death total approximates the lower bound of true COVID-19 deaths, both of which are higher than state and federal counts. LTCFs other than nursing facilities accounted for 35% of cases and 29% of deaths. Improving the accuracy of COVID-19 figures, particularly across types of LTCFs, would better inform interventions for these vulnerable populations.

Keywords

COVID-19, long-term care, underreporting, public health surveillance, death certificates

Introduction

During the COVID-19 pandemic, discrepancies in data have been common and at times contentious, as in the case of New York's nursing home death toll (James, 2021). The epidemiology community itself has come under criticism. Tarantola and Dasgupta (2021) expressed alarm about the "extreme and invisible heterogeneity that permeates even the most basic measurements," including counts of cases and deaths. Nursing home researchers have relied heavily on COVID-19 data from the Centers for Medicare and Medicaid Services (CMS) while acknowledging that these data understate cases and deaths prior to mandatory reporting in late May 2020 (Shen et al., 2021). Less often emphasized is the other major limitation of CMS data, which is that they are self-reported, unaudited aggregates. Finally, very little is known about the relative impact of COVID-19 and the data quality from other types of long-term care facilities (LTCFs), which house over 800,000 residents nationwide (Harris-Kojetin et al., 2019). In our San

Francisco Bay Area county, our public health surveillance work at ground level throughout the pandemic led us to evaluate an array of LTCF data adequacy issues, with particular attention to undercounts of COVID-19 deaths.

As the COVID-19 pandemic progressed through spring and summer of 2020, the Alameda County Public Health Department enlisted the authors, both long-term care professionals, in its efforts to mitigate pandemic impacts in LTCFs. County officials recognized that disentangling LTCF and non-LTCF community data would enable better targeting of interventions to both groups. In the following months, the challenges to creating a single source of truth about LTCF

¹Alameda County Public Health Department, Oakland, CA, USA

Corresponding Author:

Terry E. Hill, Alameda County Public Health Department, 631 Vernon St, Oakland, CA 94610, USA; thillmd@pacbell.net



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

cases and deaths became abundantly clear. In this report, we offer a description of our methodological approach, elaborate a typology of data vulnerabilities, and offer our best approximations for LTCF cases and deaths. We illustrate the importance of our approach by comparing our numbers with comparable state and national reports and with our county's community-based deaths. In a separate paper (Hill & Farrell, 2022), we focus on the heterogeneity of the pandemic's impacts across the types of LTCFs.

Methods

Our typology of data vulnerabilities and our individual-level surveillance methodology emerged iteratively from our team of public health nurses, epidemiologists, and long-term care experts. Our universe of interest encompassed all COVID-19 cases associated with licensed LTCFs in the Alameda County local health jurisdiction from March 2020 through March 2021. Most of the two-dose vaccination opportunities in our LTCFs occurred in January through March 2021 and led thereafter to a dramatic decrease in infections. We excluded facilities in Berkeley, which has its own public health department, as well as non-licensed residential hotels and boarding houses, which do not have similar reporting requirements. We excluded residents who had been infected elsewhere and then transferred into an LTCF. We included all staff reported by LTCFs as COVID-19 cases; we could not distinguish whether these infections occurred on the job or in the community (Fell et al., 2020).

Master Line List

Facilities are required to report COVID-19 infections to their local health departments using "line lists" that include resident or staff name, demographics, and date of COVID-19 testing. Public health nurses then establish whether an outbreak is present and reinforce mitigation measures. Facilities submit multiple line lists throughout an outbreak as new results become available.

We examined thousands of line lists and other LTCF documents, as well as public health notes in state databases, in order to compile a master line list of all cases. We audited completeness against an internal outbreak tracking file. Quality of submissions varied widely; misspellings and transposed digits were common, and data fields were often left empty. We coded each line on a best-guess basis for COVID-19 test result, month of test, age, and death. We retained and flagged duplicates if they were reported as positive three or more months apart or if staff members were reported as positive by two different facilities at the same time. The master line list proved helpful during the first winter surge for tracking patterns of outbreaks across our local geography and facility types.

Confirmed Match File

Beginning with the line list universe of names, we developed a match file with all cases confirmed as COVID-19 positive in California's COVID-19 database (CaREDIE), California Comprehensive Death File (CCDF), or CalConnect, California's case investigation/contact tracing program. We used Match*Pro for initial probabilistic linkage (version 1.6.5, National Cancer Institute) but enhanced these results with extensive manual searches through CaREDIE, CCDF, and CalConnect. In addition to matching by name and date of birth, we matched on facility identification and dates of test and death if available. The free-text fields in line lists, in addition to newspaper obituaries, were sometimes helpful for resolving uncertain matches. We included several resident cases missing from line lists but identified by CaREDIE as first testing positive in a specific facility. We defined COVID-19 deaths as those with COVID-19 either as an underlying cause of death or a significant contributing condition on the death certificate. We included several deaths occurring in April or May following earlier LTCF infections. CalConnect pointed to another 12 LTCF death certificates with COVID-19, but we excluded these because as of our analysis they had not appeared in CCDF.

For aggregate comparisons, we obtained case and death totals by facility from the California Department of Social Services, which licenses adult residential facilities (ARFs) and residential care facilities for the elderly (RCFEs). For skilled nursing facility (SNF) case and death data, we utilized data from the CMS Nursing Home COVID-19 Public File with a cutoff of April 4, 2021 (CMS, 2021). In our file of Alameda County COVID-19 deaths, we coded infections as LTCF-associated versus community.

This work was done in the course of public health surveillance and is exempt from institutional review as outlined by the Centers for Disease Control and Prevention (CDC) (2010).

Qualitative Results

Table 1 offers our final typology of data vulnerabilities and their potential impacts on reporting, further described below. We define noise as numerous, often bidirectional inaccuracies that may be individually small but that collectively bedevil datasets and diminish signal detection.

Vulnerabilities Specific to Line Lists

Data entry. In some LTCFs, responsibility for completing line lists fell to a staff member with no experience using Microsoft Excel. In our small board-and-care facilities, responsibility often fell to an owner with limited English proficiency. Given hundreds or thousands of fields for manual completion, mistakes were inevitable. The Excel templates had no drop-

Table 1. Typology of data vulnerabilities and their potential impacts on COVID-19 reporting.

COVID-19 Data Vulnerabilities	Impact on Accuracy
Vulnerabilities specific to line lists	
Data entry errors	Source of noise with uncertain impact; induces undercounts when reporting depends on matching across datasets
Multiple, potentially contradictory line list submissions throughout an outbreak	Source of noise with uncertain impact on counts and comparisons
Facilities' inability to learn outcomes after discharge	Large impact in self-report contexts; induces undercounts of hospitalizations and deaths
Vulnerabilities common across datasets	
Prevalence of asymptomatic infections and inadequate testing, particularly early in the pandemic	Large, universal impact; induces undercounts
Unaudited self-reporting by long-term care facilities; aggregate rather than individual level	Source of noise with uncertain impact; potential for undercounts
Absence of national licensing and reporting requirements for LTCFs other than nursing homes	Large impact on understanding pandemic's impacts on older adults; inhibits research; jeopardizes counts and comparisons across jurisdictions
Complexity of long-term care landscape and frequent changes in ownership	Source of noise with uncertain impact; jeopardizes comparisons
Differences in test reporting logistics, for example, across laboratories, among point of care technologies	Source of noise, likely inducing undercounts of uncertain magnitude and jeopardizing comparisons
Uncertainty regarding what constitutes a COVID-19 death	Major source of noise; different definitions jeopardize comparisons
True duplicates to be deleted versus reinfections or infected staff reported by multiple facilities	Potential for noise with uncertain impact; generally overlooked in the literature

downs, no forcing functions, and no interface with census or payroll records, which would have helped avoid errors.

Multiple submissions. Data from multiple line lists submitted during an outbreak were not always captured and accurately summarized in a final file; rather, sequential and final line lists often contradicted each other.

Unknown outcomes. While LTCF staff would know if a resident transferred directly to a hospital, they might not know if a staff member on isolation at home was eventually hospitalized. Similarly, they might not know if a resident or staff member eventually died. Furthermore, if they were aware of a post-discharge death, or even a death in facility, they might not update the line list.

In contrast to these vulnerabilities, we found LTCF reporting of test results to be relatively reliable. While key-stroke errors and contradictory submissions held no consequences to the facility, COVID-19 testing was critical to decision-making by both LTCFs and public health. Once a positive case was reported, public health nurses were tasked with decisions about whether to close the facility to new admissions and, if closed, when to reopen it. False-positive errors occurred, but to our knowledge, positive results were never deliberately over-reported.

Vulnerabilities Common Across Datasets

Inadequate testing. The most significant cause of COVID-19 underreporting in LTCFs, particularly early in the pandemic,

was the prevalence of asymptomatic infections and inadequate testing, a gap discussed below.

Self-reporting. CMS has required SNFs to report an array of aggregate data to the National Healthcare Safety Network (NHSN) since May 2020, with an option to report prior data retrospectively. The use of NHSN's web-based module eliminates many of the key-stroke errors that are prevalent in line lists, but the data are aggregated and reported by facility staff with no provision for auditing. The automated data quality check used by the NHSN is imperfect ([U.S. Office of Inspector General \[U.S. OIG\], 2021a](#)), and we found lingering errors.

Diversity of facility types and capabilities. Apart from SNFs and intermediate care facilities, LTCFs lack federal licensing requirements, and their data infrastructures are meager ([Temkin-Greener et al., 2020](#)). Diversity is marked even within a given state licensing category. California's small board-and care homes and large assisted living facilities operate under the same RCFE license. Alameda County's 6-bed board-and care homes have only 10% of the county's RCFE beds but far outnumber the larger assisted living facilities. Neither are adept at submitting communicable disease reports. Local public health nurses, in direct and repeated contact with LTCFs, are better positioned than state agencies for eliciting accurate reports.

Complexity of LTCFs. The initial lack of familiarity of public health and other agency staff with the LTCF landscape also hindered accurate reporting early in the pandemic ([Levin et al., 2021](#)). Reliance on self-identified facility names, rather than the

name and license number recognized by state agencies, compromised the utility of our internal dashboard, particularly given the frequency of ownership changes in long-term care.

California's continuing care retirement communities (CCRCs), which comprise multiple levels of care, are licensed by two separate state agencies. Because staff often works on both SNF and RCFE levels, the facilities are often uncertain which level is responsible for a staff case, and this uncertainty is incorporated into reports and dashboards. In our final datasets, we combined these SNF and RCFE data into a distinct CCRC category.

Test reporting logistics. Laboratories did not always report results correctly, and interagency reporting was often a challenge. The introduction of COVID-19 point-of-care testing in late 2020 introduced a new source of confusion, both because it required novel reporting pathways and because guidance regarding confirmatory testing changed over time, thus changing the specifications for what constituted a COVID-19 case.

Death determinations. The question of what constitutes a COVID-19 death is beset by yet more uncertainty, particularly in LTCF settings. The line list narrative field often noted that a resident died following COVID-19 infection but also suffered from terminal illness; the death may or may not have been reported as a COVID-19 death to NHSN. California's local health departments have relied on clinical discretion in determining COVID-19 deaths, particularly prior to ready availability of death certificate data.

Duplicates. Duplicates are a common occurrence in databases, usually appropriate for deletion, but true reinfection was not rare in LTCFs even in 2020 (Cavanaugh et al., 2021). From a public health perspective, these cases should not be deleted. In addition, because many staff worked at multiple facilities, two or more facilities often reported a staff member as positive at the same time. When reporting aggregate numbers, for example, county totals, duplicates should be discounted, but when reporting by facility, they should be preserved.

Quantitative Results

As expected, results from our two datasets differ markedly (Table 2). The master line list for licensed LTCFs in the Alameda County jurisdiction contains 6663 unduplicated COVID-19 cases and 481 COVID-19 deaths. Taking the line list names as our universe, we were able to confirm only 5010 cases, 25% fewer. Starting with this universe of names, we confirmed that 594 died with COVID-19 as an underlying cause of death or a significant contributor to death, 23% more.

Several reasons for low case confirmations accounted for at least 1000 unmatched cases. Some test providers submitted morbidity reports without electronic laboratory reports; two of our laboratories failed to submit results properly to CalREDIE; and results captured in other jurisdictions were often not linked back to our outbreaks where the infections occurred. These are known issues being addressed in reconciliation efforts, but any final matching effort will likely fall short of complete.

Deaths may be unknown to LTCFs or misreported, as noted above. Also, LTCFs are rarely privy to death certificate determinations; their clinical judgments may yield comparatively higher or lower numbers. Of 594 deaths in our confirmed match file, 33% were undercounted on line lists. On the other hand, 19% of line list deaths were not confirmed by death certificates and thus overcounted by this definition.

We are left with 6663 cases from the master line list and 594 deaths from the confirmed match file as conservative lower bounds for COVID-19 cases and deaths.

Comparisons of LTCF Deaths with Local Community Deaths

Our LTCF versus community comparison was limited to deaths attributed to the Alameda County local health jurisdiction from March 2020 through April 2021 with COVID-19 on death certificate. Table 3 shows that 47% of these deaths were associated with LTCFs.

Table 2. Self-reported versus confirmed COVID-19 cases and deaths of residents and staff.

	COVID-19 Cases		COVID-19 Deaths	
	Self-Reported Master Line List (%)	Confirmed Match File (%)	Self-Reported Master Line List (%)	Confirmed Match File (%)
Skilled nursing facility (SNF)	4309 (65)	3208 (64)	330 (69)	421 (71)
Residential care facility for the elderly (RCFE)	1667 (25)	1302 (26)	134 (28)	148 (25)
Continuing care retirement community (CCRC)	273 (4)	170 (3)	16 (3)	22 (4)
Adult residential facility (ARF)	207 (3)	164 (3)	— ^b	—
Intermediate care facility (ICF)	152 (2)	118 (2)	—	—
Other ^a	55 (1)	48 (1)		
Total	6663 (100)	5010 (100)	481 (100)	594 (100)

^aOther includes community treatment facilities, mental health rehabilitation facilities, and psychiatric health facilities.

^bValues < 10 are suppressed to protect confidentiality in accordance with state and national confidentiality guidelines.

Figure 1 displays LTCF versus community COVID-19 deaths over time, along with LTCF-associated deaths as a percentage of the monthly total. Through January 2021, prior to the impact of the LTCF vaccination rollout, there were 508 LTCF COVID-19 deaths and 505 community deaths.

Comparisons with State and National Data

We were able to identify 121 adult residential facilities and RCFEs that submitted COVID-19 resident and staff case data

to both our local health department and the California Department of Social Services (CDSS). Compared with our confirmed county data, the CDSS shortfalls were 82% for COVID-19 cases and 77% for deaths (Table 4).

A similar facility-by-facility comparison found that the CMS SNF file contained 91% of cases in our master line list and 86% of deaths in our confirmed match file. Aggregate comparisons with the CMS dataset are vexed by variations in whether facilities retrospectively submitted data for March, April, and early May 2020 (Shen et al., 2021). After

Table 3. COVID-19 as underlying cause or contributing condition on Alameda County death certificates.

	COVID-19 as Underlying Cause of Death (%)	COVID-19 as Contributing Condition (%)	Total Deaths	% of Alameda County Total
Long-term care facility (LTCF)	490 (84)	91 (16)	581 (100)	47
Community (non-LTCF)	615 (92)	52 (8)	667 (100)	53
Total	1105 (89)	143 (11)	1248 (100)	

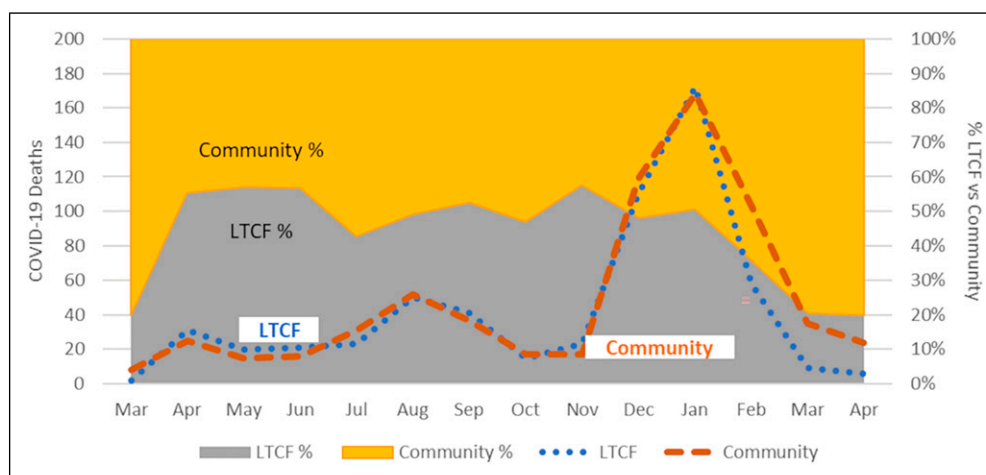


Figure 1. COVID-19 deaths in Alameda County long-term care facilities (LTCF) and community (non-LTCF) and the LTCF versus community percent of monthly totals.

Table 4. Facility-by-facility comparisons of Alameda County cases and deaths with state and federal datasets.

	Resident Cases	Staff Cases	Total Cases	Resident Deaths	Staff Deaths	Total Deaths
ARFs and RCFEs						
Alameda County	963	712	1677	123	— ^c	128
California Department of Social Services (CDSS) ^a	773	584	1357	95	—	99
CDSS as % of Alameda County	80%	82%	81%	77%	—	77%
SNFs						
Alameda County	2496	1665	4161	337	—	342
Centers for Medicare and Medicaid Services (CMS) ^b	2238	1551	3789	289	—	295
CMS as % of Alameda County	90%	93%	91%	86%	—	86%

^aThe CDSS comparison included 121 matched adult residential facilities (ARFs) and residential care facilities for the elderly (RCFEs).

^bThe CMS comparison included 58 matched skilled nursing facilities (SNFs).

^cValues < 10 are suppressed to protect confidentiality in accordance with state and national confidentiality guidelines.

eliminating data from those three months, the CMS counts increased to 96% of county cases but decreased to 69% of county deaths. Line-by-line review revealed erratic variation, that is, noise. Half of individual facility comparisons varied by 10 or more cases, and half varied by 2 or more deaths.

Discussion

Having highly credible data on COVID-19's impact in our LTCFs has helped us mobilize public health and delivery system outreach and support, particularly to assisted living and board-and-care homes. Our typology of data vulnerabilities points to several sources of epidemiological noise that have the potential to push counts either up or down.

The most significant source of noise in COVID-19 death studies, particularly those involving frail older adults, stems from the vagaries of COVID-19 death designation. We relied on death certificates to minimize inconsistency across reporting sources. Federal guidance for certifying COVID-19 deaths on death certificates depends upon "informed medical opinion," consistent with historical precedent (CDC National Center for Health Statistics, 2020). Part I of the death certificate allows for construction of a causal sequence leading to death; part II allows for inclusion of significant conditions contributing to death. As is true for bacterial pneumonia, COVID-19 can be the terminal event for a frail older person who might otherwise live some years longer, hence be recognized in part I, or it might contribute to increasing frailty and thus death from inanition or a later fall (Greco et al., 2021), hence part II. The CDC has found death certificates to be a reasonably accurate foundation for COVID-19 surveillance (Gundlapalli et al., 2021).

Our approach has several additional strengths. Rather than relying on reported aggregate figures, we based our final case count on the LTCFs' individual-level line lists that are used for critical, real-time decision-making by LTCFs and public health nurses. We avoided inappropriate attribution of LTCF residents who had been infected in the community prior to arrival in the facility (Gomolin et al., 2021). Our manual matching process included name, date of birth, LTCF name, and dates of testing and death. We used facility-by-facility comparisons with state and national datasets, excluding missing facilities. Finally, our public health nurses implemented the same data collection techniques across the entire landscape of licensed LTCFs.

In addition to sources of noise, our typology points to two more serious issues. The first is the widely acknowledged, irremediable data gap due to asymptomatic infections and inadequate testing, particularly early in the pandemic. The scale of this underreporting has been recently investigated with excess deaths methodologies for SNFs (U.S. OIG, 2021b) and assisted living (Thomas et al., 2021).

The second issue is the ineluctable underreporting by LTCFs of COVID-19 deaths, many of which occur following

discharge from the facility. Because of our painstaking methodology, we can be confident that half of our county's COVID-19 deaths through January 2021 were associated with LTCFs. We can also be confident that the COVID-19 death numbers that are self-reported by LTCFs are low. The noise level in our relatively small sample of RCFEs and SNFs precludes precise estimates, but the undercount in our jurisdiction and others could be 20% or greater. Our findings deepen concerns about SNF data accuracy and completeness raised by the U.S. OIG (2021a). Underreporting of LTCF deaths has potential implications for resource allocation.

The labor-intensive nature of our individual-level matching approach may explain why it has been rarely used. On a much smaller scale, Telford et al. (2020) retroactively linked test results, LTCF census lists, and case reports from hospitals and medical examiners in studying test strategies and infections in 28 Fulton County LTCFs over the first three months of the pandemic. At that time, more than half of the Fulton County COVID-19 deaths were from LTCFs. Louie et al. (2020) used the death certificate underlying cause of death to distinguish dying with COVID-19 from dying from COVID-19; of their San Francisco decedents through July 14, 2020, 46% resided in SNFs.

Investigators may want to leverage individual-level data—and thus bypass the issue of aggregate self-reporting—using methodologies that are less labor-intensive than those used in our study or in the papers just cited. Claims data, which were used in the OIG's excess-deaths study cited above, can be a good starting point. We are aware of healthcare organizations that used claims data early in the pandemic to identify which of their patients were housed in which RCFEs and SNFs. Matching this starter set of names with state and federal datasets could yield a host of insights, particularly for large private sector organizations or the Veterans Administration. States with all-payer claims databases offer yet richer opportunities to identify these vulnerable populations and match them with COVID-19 datasets.

In our typology, we have attempted to capture the range of limitations that are inherent in attempts to measure COVID-19 cases and deaths in LTCFs. The major limitation of our numerical findings is that we report on a single metropolitan county in California. We intend our typology to be applicable elsewhere, but our specific findings are not generalizable.

Conclusions

Our sustained partnership of public health and long-term care professionals has yielded multiple insights into sources of discrepancies across estimates of COVID-19 cases and deaths. Using individual-level confirmatory matching, we found that through the winter surge ending in January 2021, fully half of the COVID-19 deaths in our local health jurisdiction were associated with LTCFs. Facility-by-facility comparisons revealed case and death undercounts in state and

national datasets. We encourage our long-term care and public health colleagues to devise additional population-based strategies for accurately assessing COVID-19's impact across LTCFs.

Acknowledgments

The authors are grateful to Charlene Harrington for commenting on drafts of this paper and to multiple members of the Alameda County Public Health Department and Health Care Services Agency for their support.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported, in part, by the Stupski Foundation and the Alameda Contra Costa Medical Association.

ORCID iD

Terry E. Hill  <https://orcid.org/0000-0002-6771-8813>

References

- Cavanaugh, A. M., Thoroughman, D., Miranda, H., & Spicer, K. (2021). Suspected recurrent SARS-CoV-2 infections among residents of a skilled nursing facility during a second COVID-19 Outbreak - Kentucky, July-November 2020. *MMWR. Morbidity and Mortality Weekly Report*, 70(8), 273–277. <https://doi.org/10.15585/mmwr.mm7008a3>
- Centers for Disease Control and Prevention. (2010). *Distinguishing public health research and public health nonresearch*. <https://www.cdc.gov/os/integrity/docs/cdc-policy-distinguishing-public-health-research-nonresearch.pdf>
- Centers for Disease Control and Prevention National Center for Health Statistics. (2020). *Vital statistics reporting guidance: guidance for certifying deaths due to COVID-19*. <https://www.cdc.gov/nchs/data/nvss/vsrg/vsrg03-508.pdf>
- Centers for Medicare and Medicaid Services. (2021). *COVID-19 nursing home data*. <https://data.cms.gov/covid-19/covid-19-nursing-home-data>
- Fell, A., Beaudoin, A., D'Heilly, P., Mumm, E., Cole, C., Tourdot, L., & Holzbauer, S. (2020). SARS-CoV-2 exposure and infection among health care personnel - Minnesota, March 6–July 11, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69(43), 1605–1610. <https://doi.org/10.15585/mmwr.mm6943a5>
- Gomolin, I. H., Hartley, D. A., & Polsky, B. (2021). Post-acute care nursing home deaths in the COVID era: Potential for attribution bias. *Journal of the American Geriatrics Society*, 69(6), 1450–1452. <https://doi.org/10.1111/jgs.17132>
- Greco, G. I., Noale, M., Trevisan, C., Zatti, G., Dalla Pozza, M., Lazzarin, M., & Sergi, G. (2021). Increase in frailty in nursing home survivors of coronavirus disease 2019: Comparison with noninfected residents. *Journal of the American Medical Directors Association*, 22(5), 943–947. e943. <https://doi.org/10.1016/j.jamda.2021.02.019>
- Gundlapalli, A. V., Lavery, A. M., Boehmer, T. K., Beach, M. J., Walke, H. T., Sutton, P. D., & Anderson, R. N. (2021). Death certificate-based ICD-10 diagnosis codes for COVID-19 mortality surveillance - United States, January–December 2020. *MMWR. Morbidity and Mortality Weekly Report*, 70(14), 523–527. <https://doi.org/10.15585/mmwr.mm7014e2>
- Harris-Kojetin, L. D., Sengupta, M., Lendon, J. P., Rome, V., Valverde, R., & Caffrey, C. (2019). *Long-term care providers and services users in the United States, 2015–2016*. : National Center for Health Statistics. https://www.cdc.gov/nchs/data/series/sr_03/sr03_43-508.pdf
- Hill, T. E., & Farrell, D. J. (2022). COVID-19 across the landscape of long-term care in Alameda County: Heterogeneity and disparities. *Gerontology and Geriatric Medicine*, 8. <https://doi.org/10.1177/23337214211073419>
- James, L. (2021). *Nursing home response to COVID-19 pandemic*. New York State Office of the Attorney General. <https://wnynnewsnow.com/wp-content/uploads/2021/01/2021-nursinghomesreport-final.pdf>
- Levin, S., Farrell, F., Hill, T., & Howell, N. (2021). Capturing pandemic lessons learned for local health departments and long-term care facilities. *GeriPal Blog*. <https://geripal.org/capturing-pandemic-lessons-learned-for>
- Louie, J. K., Scott, H. M., Lu, W., Chodos, A., DuBois, A., Sturtz, N., & Stoltey, J. (2020). COVID-19-associated deaths in San Francisco: The important role of dementia and atypical presentations in long-term care facilities. *Journal of General Internal Medicine*, 35(11), 3413–3415. <https://doi.org/10.1007/s11606-020-06206-1>
- Shen, K., Loomer, L., Abrams, H., Grabowski, D. C., & Gandhi, A. (2021). Estimates of COVID-19 cases and deaths among nursing home residents not reported in federal data. *JAMA Network Open*, 4(9), Article e2122885. <https://doi.org/10.1001/jamanetworkopen.2021.22885>
- Tarantola, D., & Dasgupta, N. (2021). COVID-19 surveillance data: A primer for epidemiology and data science. *American Journal of Public Health*, 111(4), 614–619. <https://doi.org/10.2105/ajph.2020.306088>
- Telford, C. T., Onwubiko, U., Holland, D. P., Turner, K., Prieto, J., Smith, S., & Shah, S. (2020). Preventing COVID-19 outbreaks in long-term care facilities through preemptive testing of residents and staff members - Fulton County, Georgia, March–May 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69(37), 1296–1299. <https://doi.org/10.15585/mmwr.mm6937a4>
- Temkin-Greener, H., Guo, W., Mao, Y., Cai, X., & Li, Y. (2020). COVID-19 pandemic in assisted living communities: Results from seven states. *Journal of the American Geriatrics Society*, 68(12), 2727–2734. <https://doi.org/10.1111/jgs.16850>

- Thomas, K. S., Zhang, W., Dosa, D. M., Carder, P., Sloane, P., & Zimmerman, S. (2021). Estimation of excess mortality rates among US assisted living residents during the COVID-19 pandemic. *JAMA Network Open*, 4(6), Article e2113411. <https://doi.org/10.1001/jamanetworkopen.2021.13411>
- U.S. Office of Inspector General. (2021a). *CMS's COVID-19 data included required information from the vast majority of nursing homes, but CMS could take actions to improve completeness and accuracy of the data*. <https://oig.hhs.gov/oas/reports/region9/92002005.asp>
- U.S. Office of Inspector General. (2021b). *Data snapshot: COVID-19 had a devastating impact on Medicare beneficiaries in nursing homes during 2020*. <https://oig.hhs.gov/oei/reports/OEI-02-20-00490.pdf>