A new automated national register-based surveillance system for outbreaks in long-term care facilities in Norway detected three times more severe acute respiratory coronavirus virus 2 (SARS-CoV-2) clusters than traditional methods

Kirsten Gravningen PhD1,2,3, Petter Nymark MS1, Torgeir B. Wyller PhD3,4 and Oliver Kacelnik PhD1

1Department of Infection Prevention and Preparedness, Norwegian Institute of Public Health (NIPH), Oslo, Norway, 2Department of Microbiology and Infection Control, Akershus University Hospital, Nordbyhagen, Norway, 3Institute of Clinical Medicine, University of Oslo, Oslo, Norway and 4Department of Geriatric Medicine, Oslo University Hospital, Oslo, Norway

Abstract

Objective: To develop and test a new automated surveillance system that can detect, define and characterize infection clusters, including coronavirus disease 2019 (COVID-19), in long-term care facilities (LTCFs) in Norway by combining existing national register data.

Background: The numerous outbreaks in LTCFs during the COVID-19 pandemic highlighted the need for accurate and timely outbreak surveillance. As traditional methods were inadequate, we used severe acute respiratory coronavirus virus 2 (SARS-CoV-2) as a model to test automated surveillance.

Methods: We conducted a nationwide study using data from the Norwegian preparedness register (Beredt C19) and defined the study population as an open cohort from January 2020 to December 2021. We analyzed clusters (≥ 3 individuals with positive SARS-CoV-2 test ≤ 14 days) by 4-month periods including cluster size, duration and composition, and residents’ mortality associated with clusters.

Results: The study population included 173,907 individuals; 78% employees and 22% residents. Clusters were detected in 427 (43%) of 993 LTCFs. The median cluster size was 4–8 individuals (maximum, 50) by 4-month periods, with a median duration of 9–17 days. Employees represented 60%–82% of cases in clusters and were index cases in 60%–90%. In the last 4-month period of 2020, we detected 107 clusters (915 cases) versus 428 clusters (2,998 cases) in the last period of 2021. The 14-day all-cause mortality rate was higher in resident cases from the clusters. Varying the cluster definitions changed the number of clusters.

Conclusion: Automated national surveillance for SARS-CoV-2 clusters in LTCFs is possible based on existing data sources and provides near real-time detailed information on size, duration, and composition of clusters. Thus, this system can assist in early outbreak detection and improve surveillance.

(Received 16 August 2022; accepted 9 November 2022; electronically published 16 December 2022)
In Norway, ∼35,000 people reside in >900 LTCFs. The median age of residents is >85 years. During the pandemic, guidelines from national authorities recommended that LTCF residents with COVID-19 should not be routinely transferred to hospitals and staff should only be employed in 1 LTCF at a time.

In June 2020, a new real-time population-wide Norwegian preparedness register for COVID-19, Beredt C19, was established as part of the legally mandated responsibilities of the NIPH during epidemics. It combines individual-level data from multiple existing sources to provide authorities with updated knowledge about the pandemic.

Infection surveillance in healthcare is changing from manual data collection by clinical staff to digital systems using unique person identifiers with privacy safeguards that can merge existing individual-level data from various sources.

We sought to determine whether we could use Beredt C19 to develop an automated and more complete system for detecting, defining, and characterizing outbreaks in LTCFs. We evaluated the use of existing data in Beredt C19 (1) to detect and analyze the size, duration, and composition of SARS-CoV-2 clusters among residents and staff in LTCFs in near real-time, (2) to examine residents’ mortality associated with clusters and how the virus variant impacted on these clusters, and (3) to provide recommendations for future monitoring of outbreaks in LTCFs.

**Methods**

**Data sources**

We extracted data from Beredt C19, a national preparedness register containing deidentified individual-level data for the entire Norwegian population from January 2020 as described elsewhere. For this study, we merged data from 6 registers (Table 1).

According to a legal mandate, all confirmed SARS-CoV-2 test results and virus sequencing results are automatically transferred from the microbiology laboratories to the national MSIS laboratory database. The NIPH has established a computer algorithm that will include a positive test result as soon as it is available (updated every 24 hours); thus, a new case will be included in ∼3–4 days from the test date.

**Study population and period**

We defined the study population as an open cohort from January 1, 2020, to December 31, 2021. The resident population was selected from NPR (Table 1). The person had to be alive and registered as a long-term or temporary resident in an LTCF at some point during 2020–2021. Using this definition, the open cohort population of residents was 37,690 distributed across 993 facilities. The employee population was selected from the Aa Register based on registered employment in 1 of 993 facilities during 2020–2021. Altogether, 136,217 employees were included in the open cohort, for a study population of 173,907. We have also described both the study population (individual level) and the LTCFs (institutional level) at study start, using point prevalence estimates on January 1, 2020.

The study period was divided into 4-month periods. For each period, we analyzed cluster characteristics: the index case, size, number, composition, and duration. We calculated vaccine coverage for each period by linking the study population with SYSVAK data. For a second vaccine dose to be valid, it had to be 19 days after the first dose. Two doses were considered full vaccination during this period.

**Variable definitions**

We defined a SARS-CoV-2 case as a person with a confirmed positive SARS-CoV-2 RT-PCR test from MSIS. We defined a SARS-CoV-2 cluster as ≥3 individuals (residents and employees) with SARS-CoV-2–positive test results from MSIS occurring within 14 days of each other that were linked to the same LTCF institutional number in the study period. Clusters could include any combination of residents and employees. An individual could have multiple positive tests and be part of multiple clusters occurring on different dates. The duration of a cluster was the number of days from date of the first positive SARS-CoV-2 test to the date of the last positive test in a cluster. The 14-day all-cause mortality included any instance of death by any cause within 14 days of a positive SARS-CoV-2 test. The 14-day all-cause mortality rate was the number of 14-day all-cause fatalities divided by the number of SARS-CoV-2 cases within a period or cluster. For example, 2 deaths within a cluster with 11 cases would yield a rate of 18%.

---

**Table 1. The Six Registers from Beredt C19 Used in the Study**

<table>
<thead>
<tr>
<th>Register</th>
<th>Description of Register</th>
<th>Data Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>The State Register of Employers and Employees (Aa Register)</td>
<td>Data on all employment relationships in Norway, excluding freelancers and the self-employed: <a href="https://www.nav.no/en/home/employers/nav-state-register-of-employers-and-employees">https://www.nav.no/en/home/employers/nav-state-register-of-employers-and-employees</a></td>
<td>Occupational code, occupation, contract start date, contract end date, organizational number</td>
</tr>
<tr>
<td>Cause of Death Registry (DÅR)</td>
<td>Individual level information about causes of death: <a href="https://www.fhi.no/en/hn/health-registries/cause-of-death-registry/">https://www.fhi.no/en/hn/health-registries/cause-of-death-registry/</a></td>
<td>Date of death</td>
</tr>
<tr>
<td>Norwegian Registry for Primary Health Care (KPR)</td>
<td>Database for residents in LTCFs: <a href="https://helsedata.no/en/forvaltare/norwegian-directorate-of-health/norwegian-registry-for-primary-health-care-kpr/">https://helsedata.no/en/forvaltare/norwegian-directorate-of-health/norwegian-registry-for-primary-health-care-kpr/</a>?</td>
<td>Organization name and number, type of institution, type of stay, date of admission, date of discharge or death</td>
</tr>
<tr>
<td>Norwegian Surveillance System for Communicable Diseases (MSIS) including results from the MSIS Laboratory Database</td>
<td>Laboratory database with information about samples and analysis for COVID-19 and influenza: <a href="https://www.ehelse.no/prosjekt/msis-databasen">https://www.ehelse.no/prosjekt/msis-databasen</a></td>
<td>Positive SARS-CoV-2 PCR test results: test date, virus variant: test date, virus variant sequencing result</td>
</tr>
<tr>
<td>National Population Register (NPR)</td>
<td>Information about the total population in Norway: <a href="https://www.skatteetaten.no/en/person/national-registry/">https://www.skatteetaten.no/en/person/national-registry/</a></td>
<td>Sex, county number, birthdate</td>
</tr>
<tr>
<td>Norwegian Immunization Registry (SYSVAK)</td>
<td>Register for immunization and vaccination status: <a href="https://www.fhi.no/en/hn/health-registries/norwegian-immunisation-registry-sysvak/">https://www.fhi.no/en/hn/health-registries/norwegian-immunisation-registry-sysvak/</a></td>
<td>Date of consultation, vaccine code</td>
</tr>
</tbody>
</table>
Because LTCF residents with COVID-19 were not generally transferred to hospitals during the pandemic, it was not feasible to use hospital admission as a surrogate marker for infection severity. Instead, we calculated the all-cause mortality rate within 14 days of a positive SARS-CoV-2 test among residents. We compared the 14-day all-cause mortality rate in clusters to the same measure in all LTCF-residents in Norway who tested positive in the study period. Although SARS-CoV-2 sequencing was not done for all cases, we used the available sequencing results to calculate the dominant virus variant per week among residents.

Data analyses

We performed descriptive analyses and created figures using R version 4.0.2 software (R Foundation for Statistical Computing, Vienna, Austria). The process of detecting clusters and extracting information about them was automated using R, called automated cluster detection.

Ethics statement

Beredt C19 was established under Section 2–4 of the Health Preparedness Act. The NIPH has conducted a data protection impact assessment of the register. The work included in this study was conducted as part of the mandated work of NIPH. All the data used in this study were stored on secure servers at the NIPH where only the authors employed at the NIPH had access. Details of data protection measures in Beredt C19 are described elsewhere.

Results

The study population at the start of the study on January 1, 2020, included 122,682 individuals in 993 LTCFs in Norway, of whom 79% were employees (88% women) (Table 2). The median age of residents was 88 years and more than two-thirds were women. The mean number of residents in the LTCFs was 34 and the maximum was 244. Furthermore, for a surveillance system to reflect the reality of LTCFs, it must include information on both residents and the staff who work there at any given time.

Clusters

The main body of this study concerns data collected during the first 2 years of the pandemic. Table 3 divides these 2 years into 4-month periods.

Clusters were detected in 427 (43%) of 993 LTCFs. The number of clusters varied, with relatively lower mean numbers in the summers of 2020 and 2021 and increasing numbers in the winter. In the last 4-month period of 2021, with the predominance of the SARS-CoV-2 δ (delta) variant, we detected 4 times as many clusters as in the last period of 2020 with predominance of the α (alpha) variant. The median size of the clusters was 8 in the beginning of 2020 and 5 toward the end of 2021. The median and maximum size of the clusters also remained even, apart from during the second 4-month period of 2020. The median duration of clusters was 14–17 days, apart from the second period of 2020, when it was 8.5 days.

Employees made up 79% of our study population under surveillance at study start January 1, 2020 (Table 2). Employees represented between 60% and 82% of cases depending on the period (Table 3). We found no obvious overrepresentation of either staff or residents in the clusters. We also identified the index case for any cluster. The percentage of clusters for which an employee (rather than a resident) was the index (likely primary) case increased from 60% in the first period of 2020 to ~85% for the rest of the study. The proportion of women in both the resident and employee clusters followed the proportion in the study population.

Effect of modifying the cluster parameters

We tested the effect of changing the minimum requirement for cases to be included as a cluster from 3 to 5 or 10 and the effect of changing the maximum number of days between cases from 14 to 10. Increasing the minimum number of cases to 5 or 10 reduced the total number of clusters from 768 to 392 or 132 and changing the number of LTCFs affected reduced the total from 427 to 275 or 119. Moving from 14 days between cases to 10 days reduced the number of clusters from 768 to 702 and reduced the number of LTCFs affected from 427 to 388.

Vaccination in LTCFs

Starting December 27, 2020, LTCF residents received vaccination against COVID-19, whereas employees were increasingly prioritized for vaccination from early January 2021. In the first 4-month period of 2021, residents attained vaccine coverage of 84%, but coverage was 36% for residents in the clusters (Table 4). Attainment of high vaccine coverage was slower for employees; they did not reach 80% until the second 4-month period of 2021. The coverage within the clusters for employees in the same period was 44%. In the last 4-month period of 2021, when everyone had the opportunity for full vaccination, the difference in vaccine coverage for all residents versus residents in clusters was 5%, whereas this difference was 12% (P < .001) between all employees and employees in clusters.

Cluster visualization

Figure 1 is a visualization of 132 clusters detected with ≥10 cases in 119 LTCFs, the first clusters occurred in week 10 (March 2020).

The figure demonstrates the grouping in terms of time of all the clusters with 10 or more cases in LTCFs. Although the number of clusters varied, the relative contribution of different sized clusters was similar. Furthermore, the number of clusters had already increased greatly through the predominance of the SARS-CoV-2
higher. When the total number of cases was much greater with the $\delta$ variant dominated, when the total number of cases was much greater.

We present findings beyond the main study period through February 21, 2022, including the introduction and dominance of the $\delta$ variant (weeks 31–52 of 2021) and prior to the $\omicron$ variant being dominant in the LTCF population in Norway. This figure can be automated and updated daily. Lastly, it illustrates the long duration of many of these clusters in LTCFs.

**Analysis of LTCF residents**

To contextualize surveillance information on clusters, we designed a system that would set up our clusters to consider all SARS-CoV-2 infections and would additionally include the type of virus variant in the resident study population and the 14-day all-cause mortality rate. We present findings beyond the main study period through February 21, 2022, including the introduction and dominance of the $\omicron$ variant. This part of the analysis included clusters with $\geq$3 cases and excluded all data on employees. In Figure 2, panel A shows the number of SARS-CoV-2 cases for the resident study population and panel B shows the same data for residents who were part of the clusters.

The pattern of positive SARS-CoV-2 tests in our clusters follows closely the results from the resident study population. Cases in the clusters had higher overall 14-day all-cause mortality rate early in the pandemic. Overall, the 14-day all-cause mortality rate was markedly higher among SARS-CoV-2 cases who were part of identified clusters than among cases in the general resident study population. However, in both the clusters and the resident study population we observed a decrease in the mortality rate over the course of the pandemic. This decrease was also pronounced in the period when SARS-CoV-2 $\omicron$ variant dominated, when the total number of cases was much higher.

**Table 3.** SARS-CoV-2 Cases and Clusters With $\geq$3 cases in 4-Month Periods in the Study Population in Long-Term Care Facilities (LTCFs) from March 2020 to December 2021

<table>
<thead>
<tr>
<th>Year and 4-Month Period</th>
<th>Cases</th>
<th>Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total, No.</td>
<td>Employees, No. (%)</td>
</tr>
<tr>
<td>2020 1</td>
<td>441</td>
<td>263 (59.6)</td>
</tr>
<tr>
<td>2020 2</td>
<td>28</td>
<td>21 (75.0)</td>
</tr>
<tr>
<td>2020 3</td>
<td>915</td>
<td>644 (70.4)</td>
</tr>
<tr>
<td>2021 1</td>
<td>649</td>
<td>523 (80.6)</td>
</tr>
<tr>
<td>2021 2</td>
<td>497</td>
<td>390 (78.5)</td>
</tr>
<tr>
<td>2021 3</td>
<td>2,998</td>
<td>2,464 (82.2)</td>
</tr>
</tbody>
</table>

Note. SD, standard deviation; max, maximum.

**Table 4.** Vaccination Coverage (2 doses) of the Study Population (n = 173,907) Divided into Residents and Employees and Vaccine Coverage Within the clusters from January 1, 2020, to December 1, 2021

<table>
<thead>
<tr>
<th>Year and 4-Month Period(s)</th>
<th>Vaccinated Residents, %</th>
<th>Vaccinated Residents in Clusters, %</th>
<th>Vaccinated Employees, %</th>
<th>Vaccinated Employees in Clusters, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020 1b</td>
<td>83.9</td>
<td>59.7</td>
<td>18.7</td>
<td>3.9</td>
</tr>
<tr>
<td>2021 1</td>
<td>90.8</td>
<td>69.7</td>
<td>17.9</td>
<td>4.3</td>
</tr>
<tr>
<td>2021 2</td>
<td>94.8</td>
<td>89.7</td>
<td>90.1</td>
<td>78.6</td>
</tr>
</tbody>
</table>

Discussion

Using COVID-19 as a model, we designed and implemented a system for automated national surveillance of infectious disease clusters in LTCFs in Norway, many of which were previously unreported. Using only data that were already available, we followed and analyzed COVID-19 clusters in LTCFs during 2020–2021. We tracked the size, duration, and composition of clusters as well as the index case for each cluster. Using denominator data representative of the LTCF population in Norway, we calculated the all-cause 14-day mortality rate over time and found that it was associated with the predominant virus variant and was higher among resident cases who were part of the clusters than in cases from the resident study population.

Our new surveillance system has the potential to tackle problems such as underreporting, lack of timeliness, and incompleteness of data. The system detected a 3-fold increase in clusters with $\geq$3 cases and twice the number of cluster cases compared to our traditional method. Data on the index case, cluster composition, and vaccination coverage were complete and updated in the new system, but very few notifications in Vesuv were updated by LTCFs after the outbreak was over. Consequently, the traditional method underestimates the number, size, and duration of SARS-CoV-2 outbreaks in LTCFs. This underestimation may lead to a biased evaluation of the effect of the ICP measures used, which is an important factor in all countries implementing control measures.

Long outbreaks lead to increased staff absences, and increased restrictions should be implemented for prolonged periods every time an outbreak is suspected. Continuous, timely surveillance with cluster detection and the inclusion of both residents and employees enables early outbreak identification and containment, which can decrease the spread within and across facilities and can also reduce outbreak duration, staff absence, and resident morbidity and mortality. The higher all-cause mortality rate in clusters may have been caused by higher viral loads in infected individuals during outbreaks and/or by the reduced quality of care due to significant numbers of staff being absent during outbreaks.

Because our system is based on existing data recorded for other purposes, it does not require extra activity from the LTCFs. Combining laboratory data with individual data on residents and employees in each LTCF, their vaccine status, and residents’ date of death, we were able to automate the rapid detection of SARS-CoV-2 clusters in LTCFs in Norway. The surveillance
Fig. 1. Clusters including ≥10 cases of SARS-CoV-2 in 119 LTCFs in Norway in 2020–2021. The horizontal lines show cluster duration and color indicates the number of cases in each cluster.

Fig. 2. (A) Bar plot showing number of SARS-CoV-2 cases per week from March 2, 2020, to February 21, 2022, by the dominant virus variant for the resident study population. The blue smoothed-line graph shows the 14-day all-cause mortality rate over time. (B) The same data are shown for SARS-CoV-2 cases that were part of the clusters with ≥3 cases.
system can also be used for other agents such as influenza virus.18 Because Beredt C19 includes population-based, deidentified, individual-level data, the registry has stable denominators and is considered representative of the LTCF population in Norway over time. Lastly, the visualization of cluster surveillance may contribute to increased awareness and should be further developed.

When the minimum requirement for cases to be included as a cluster was changed from 3 to 5 (a 60% increase), the number of clusters was reduced by ~50%. However, moving from 14 days between cases to 10 days reduced the number of clusters by <10%. This new system is adaptable enough to react to changes in background noise (eg, community infection rates) and incubation times (eg, new SARS-CoV-2 variants).19 Furthermore, we can easily adjust the parameters for different agents that may be under surveillance.

In Europe, the lack of special surveillance systems and the differences in testing strategies and capacities among countries during the pandemic may have led to underdetection and underreporting of cases in LTCFs.2 The European Centre for Disease Prevention and control (ECDC) recommends a national comprehensive and mandatory LTCF-based surveillance system with cumulative or weekly reporting of cases among residents and staff. The development of electronic reporting through national platforms is considered crucial. Our new system for cluster detection takes this recommendation a step further.

Strengths and limitations

The main strength of our study is the potential for accuracy and timeliness of the new nationwide surveillance system. We included both residents and staff, thereby providing the full picture at little extra cost. The study depended on mandatory, population-based registries, personal identifiers for follow-up of individuals within registries, linkage between registries, and between registries and other data sources.20 It required test results that may have been influenced by local and national testing strategies. Virus variant surveillance depends on genomic sequencing capacity, costs, and which analyses are requested by clinicians. SARS-CoV-2 data are very accurate for the first 2 years of the pandemic due to extensive and free PCR testing. Recent changes in recommendations that PCR be mainly used for clinical cases and not for screening make surveillance data potentially less complete. Furthermore, any delay in the data delivery or changes in data structure will affect the timeliness of analyses. Our method captured clusters on the institutional level and not by ward because this information was not included in the national registers. The clusters may reflect potential outbreaks more accurately in small LTCFs than in larger institutions with a high number of residents and departments; thus, a cutoff for clusters at ≥3 cases may have been too low in large facilities. Lastly, our data were censored at the end of 2022, so the size and duration of outbreaks that continued into 2023 may have been underestimated. This would not be a problem in an established surveillance system.

Future studies should explore how the cluster definition can be adjusted according to the size of the LTCFs, community infection rates, and incubation times of different agents. A system for reporting identification of clusters back to the LTCFs should be established to provide opportunities to use the outcome data from automated surveillance locally.21 Ideally, the implementation of the new system should be linked to scientific evaluations to efficiently collect the necessary information and to examine how the identified clusters correspond to real outbreaks.

We recommend the implementation of automated register-based surveillance of clusters in LTCFs wherever possible prior to the next winter respiratory virus wave in parallel with the development of systems to disseminate this information to individual LTCFs.

Acknowledgments. We thank Hanne-Merete Eriksen-Volle for support and useful advice during the development of the project. We thank all those working in Norwegian LTCFs and with the registers our data are based on. Without your commitment, this study would not have been possible.

Financial support. The study was funded by the Norwegian Institute of Public Health.

Conflicts of interest. All authors report no conflicts of interest relevant to this article.

References


