Healthcare-associated COVID-19 across 3 pandemic waves: patient characterisation and validation of clinical definitions using genome sequencing

Thomas Demuyser (✉ thomas.demuyser@uzbrussel.be)

UZ Brussel <https://orcid.org/0000-0002-6410-9246>

Lucie Seyler

UZ Brussel

Rhea Buttiens

UZ Brussel

Oriane Soetens

UZ Brussel

Ben Caljon

UZ Brussel

Jessy Praet

Biomerieux

Thomas Seyler

UZ Brussel

Joost Boeckmans

Vrije Universiteit Brussel

Jessy Meert

UZ Brussel

Robin Vanstokstraeten

UZ Brussel

Helena Martini

UZ Brussel

Florence Crombé

UZ Brussel

Denis Piérard

UZ Brussel

Sabine Allard

UZ Brussel

Ingrid Wybo

UZ Brussel

Article

Keywords:

Posted Date: January 28th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1285737/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Worldwide, healthcare-associated SARS-CoV-2 infections are a major problem: they are associated with increased morbidity, mortality, and hospitalization costs. In-depth studies across the pandemic are crucial to understand and prevent transmission in hospital settings. The principal aims of this study were to characterise patients and validate ECDC definitions of healthcare-associated COVID-19 infections.

Methods

We set up a retrospective observational study spanning the first three waves of the COVID-19 pandemic in a Belgian university hospital: it describes the characteristics of COVID-19 patients admitted, with either healthcare- or community-associated infections.

We performed a cluster analysis through epidemiological and viral genome analyses of the healthcare-associated infections, in order to validate the ECDC definitions of healthcare-associated COVID-19 infections.

Results

Between week 10 of 2020 and week 22 of 2021, 168 patients were hospitalized with healthcare-associated COVID-19. The following factors were found more often in symptomatic healthcare- than in community-associated hospitalized patients: older age, increased frailty, smoking habits, and comorbidities.

The genome-based cluster analyses showed that different viral lineages predominated in different timeframes. We observed a good correlation of epidemiological data with genome sequencing results in at least 12 different outbreaks in our hospital, thus validating the ECDC definitions.

Conclusions

This in-depth characterization sheds new light on the problem of healthcare-associated COVID-19 infections, in particular on patients' characteristics, epidemiology, and cluster dynamics. Even though epidemiological evaluation of nosocomial infections is vital, management of nosocomial outbreaks can undoubtedly benefit from genome sequencing analyses to reinforce their strategy.

Introduction

The end of 2019 saw the emergence of a novel severe acute respiratory syndrome - coronavirus 2 (SARS-CoV-2), causing coronavirus disease 2019 (COVID-19). A third distinct epidemic wave was ending in Belgium by July 2021^{1,2}. Because of the intrinsic viral properties of SARS-CoV-2 and increasing pressure

on healthcare facilities since the beginning of the pandemic, healthcare-associated COVID-19 infections (HAIs) have been of major concern.

During the first wave of the pandemic, approximately 10 to 15% of hospitalized COVID-19 cases were HAI^{3,4}. Front-line healthcare workers also have an increased risk of acquiring COVID-19 by a factor of 3.4, compared to the general population⁵. A hospital-wide screening in the UK showed that 49% of healthcare workers who tested positive for SARS-CoV-2 were asymptomatic; 40% had mild symptoms; this contributes to the risk of in-hospital SARS-CoV-2 transmission⁶. Furthermore, even vaccinated individuals can become infected and transfer the virus to others.

However, HAIs in the first wave might not be representative of later phases of the pandemic. Indeed, initial HAIs may have been partly attributable to incorrect isolation procedures, indistinct use of shared healthcare equipment, movements of infected personnel, and insufficient knowledge and awareness on viral transmission properties⁷. Knowledge on the virus' characteristics then improved dramatically. It seemed that the rate of HAI decreased, thanks to adequate responses from infection control teams; shortages of personnel protective equipment were not an issue anymore. Indeed, vaccines and non-pharmaceutical interventions have had dramatic effects on viral transmission⁸.

In the meantime however, the SARS-CoV-2 virus evolved from early 2020. Some variants of concern (VOCs) are now related to more severe infections, vaccine escape and are associated with increased transmissibility⁹. Indeed, transmissibility of the B.1.1.7 SARS-CoV-2 strain (the Alpha variant) is estimated to be 1.56 times higher than the previously dominant wild-type variants¹⁰ and viral loads in B.1.1.7 infected patients tend to be higher¹¹. Today, the global spread of other VOCs, including the B.1.617.2 (Delta) and B.1.1.529 (Omicron) variants, highlight the importance of continuous epidemiological surveillance¹². Many articles have described community-associated COVID-19 (CAI) patients, from the beginning of the pandemic onwards^{3,13}. Fewer have focused on HAIs, despite the risks associated with HAIs including a higher burden of comorbidities, such as malignancies and renal impairment¹⁴. To the best of our knowledge no thorough investigation of HAIs across multiple waves of the pandemic has been conducted.

Another challenge is the correct identification of HAI. Definitions of HAIs have been proposed by several national and international healthcare organizations. In the current manuscript we will use the guidelines provided by the European Centre for Disease Control (ECDC) (Table 1)¹⁵. Genomic sequencing can also provide valuable information to support the epidemiological ECDC HAI definition¹⁶. Combining the epidemiological and genome sequencing approaches could even prove very useful in root cause analysis and outbreak investigations¹⁷.

We therefore decided to describe our HAIs in terms of patients' demographics and clinical data; we then combined ECDC definitions with genomic cluster analyses, to provide an in-depth examination of the HAIs at our tertiary care centre, across the pandemic.

Methods

Timing and Setting

Our study was conducted from week 10 of 2020 until week 22 of 2021, at our academic hospital ‘Universitair Ziekenhuis Brussel’, a 721-bed Belgian tertiary care centre. The hospital has a maximum capacity of 132 “low care COVID-19 beds” and 36 “intensive care COVID-19 beds” during epidemic peaks. The study was done using the hospital-wide severe acute respiratory syndrome surveillance database (SARI Registry) and was approved by our hospital’s ethical committee (EC-2021-176).

Definitions

We used the ECDC case source definitions of HAI, as described in table 1.

Table 1: ECDC case source definitions of COVID-19¹⁵.

Different categories of community (CAI) - or healthcare (HAI) - associated COVID-19 infections, with increasing likelihood of healthcare-associated infection from top to bottom. A “sample” is defined as a positive laboratory test result, confirming an acute COVID-19 infection.

Case source COVID-19	Definition
CAI	Symptoms or sample present on admission or with onset on day 1 or 2 after admission (or on days 3-7 with a strong suspicion of community transmission).
Indeterminate HAI	Symptom onset or sample on day 3-7 after admission, with insufficient information on the source of infection to assign to another category.
Probable HAI	Symptoms onset or sample on day 8-14 after admission (or on days 3-7 and a strong suspicion of healthcare transmission).
Definite HAI	Symptom onset or sample on day >14 after admission.

A 'symptomatic' infection is present in our study when one or more of the World Health Organization defined criteria for SARI is/are present at COVID-19 diagnosis (fever $\geq 37.8^{\circ}\text{C}$, cough, shortness of breath)

Study design

Part A. We performed a retrospective observational study of hospitalised HAIs: our first objective was to compare symptomatic hospitalised HAIs with CAIs; a second objective was to show all cases (including healthcare workers) on epicurves.

Part B. We performed whole genome sequencing (WGS) and cluster analyses with all available samples (including non-hospitalised healthcare workers). Our primary objective here was to describe the different

clusters on the basis of the genetic analyses of the viruses; our second objective was to compare the ECDC source definitions with the sequencing analyses, as a validation tool for the ECDC definitions.

Inclusion of patients and healthcare workers

All included subjects (Figure 1) had a registered polymerase chain reaction (PCR)-confirmed COVID-19 infection.

1. CAI (used as 'controls' for the case-control study).

We extracted demographic, clinical, and laboratory data from a random sample of all hospitalized CAIs (when hospitalized for >24 hours) from our SARI Registry in an anonymized manner.

2. HAI (whether indeterminate, probably or definite).

Hospitalised patients with HAI (used as 'cases' for the case-control study). Their demographic, clinical, and laboratory data were also extracted from our SARI Registry in an anonymized manner.

3. Healthcare workers employed by our hospital and diagnosed with COVID-19

Those subjects included medical and non-medical staff employed in our hospital. The included HAI healthcare workers were not hospitalized. Their data, albeit less detailed, were carefully recorded by the infection control department.

Laboratory: inclusion of samples for genetic sequencing and cluster analyses

Nasopharyngeal samples from HAIs and healthcare workers were systematically stored at -80°C. Samples with sufficiently high (Ct value $\leq$ 25) viral load and remaining sample volume were included in the WGS analysis.

We adapted a SARS-CoV-2 WGS protocol¹⁸. Amplicon libraries were sequenced using MinION flow cells (Oxford Nanopore Technologies, Oxford, UK). Genomes were assembled with reference-based assembly and an in-house bioinformatic pipeline with 300× minimum coverage cut-off for any region of the genome. Consensus fasta sequences were generated using the tools from the artic network¹⁹. A custom scheme using primers of 1200 base pairs was used²⁰. Lineages were assigned using the command-line version (3.1.5) of pangolin²¹. Gene sequences were uploaded onto the Global Initiative on Sharing All Influenza Data (GISAID)'s open access EpiCoV platform (accession numbers in appendix 1)²².

WGS data were processed with the SARS-CoV-2 plug-in of BioNumerics v.7.6.3 (Applied Maths, Biomérieux, Sint-Martens-Latem, Belgium). The subsequences of the Wuhan-Hu-1 (NC 04551219) reference genome were used as reference sequences for a BLAST search²³. After the extraction, these subsequences were screened for single nucleotide polymorphisms (SNPs). Seven entries with an incomplete SNP character set were excluded from further analysis. Next, a similarity matrix was calculated based on the 86 remaining SNP experiments and minimal spanning trees (MSTs) were

constructed. SNP distances were represented in the trees. Forty-eight reference sequences of the circulating VOCs at that moment (B.1.1.7, P.1, B.1.351) were downloaded from the NCBI website and added to the MSTs. Clusters are defined as genomes with $\leq$ two SNPs difference and are marked with a contour. The National Health Service of the United Kingdom defines a nosocomial cluster of COVID-19 as the occurrence of two or more cases of COVID-19 infection in a single setting (e.g. a single ward), where at least one case has become symptomatic or detected on screening at least eight days post hospital admission²⁴.

The hospital wards were anonymized. The numbers correspond to the floor on which a ward is found.

Epidemiological and statistical analyses

For Part A, we described cases over time in epicurves using STATA scripts and Excel software, by week of diagnosis. All HAIs (patients and healthcare workers) and CAIs were included in the representations.

We then described patients' demographics, clinical, and laboratory data and compared continuous and categorical variables between HAI and CAI patients. The median values and interquartile ranges were depicted for continuous data, and univariate statistical analysis was performed applying an unpaired t-test or Mann-Whitney test (depending on normal distribution of data). Categorical data are expressed in absolute numbers and percentages. Univariate statistical analysis was performed by a Fisher's exact test. A p-value < 0.05 was considered as statistically significant. Associative data analysis was performed through a multiple logistic regression method, using parameters shown to be statistically significant in the univariate analysis. Statistical analysis was performed using GraphPad Prism 8.4.3 software (California, USA).

Results

An overview of all inclusions (figure 1): throughout the epidemic waves, a total of 1185 PCR-confirmed COVID-19 patients were admitted to our hospital for at least 24h, of whom 168 were considered as HAI; of those, 95/168 (57%) had at least one SARI symptom.

During the first epidemic wave 43/411 (10.5%) of hospitalized COVID-19 patients were classified as possible HAI; this rose to 59/479 (12.3%) and 66/295 (22.4%) of all hospitalized COVID-19 patients during waves two and three respectively. Amongst adult hospitalised COVID-19 patients, 39/406 (9.6%) were symptomatic HAI. This percentage then dropped during wave two and three respectively, to 8.9% (41/461) and 6.1% (15/246).

We ended up with 95 symptomatic adult HAIs and 191 randomly selected adult symptomatic CAIs (from the 1017 CAIs from the SARI Registry). Those patients were then included in the case-control studies.

Part A. Characteristics of symptomatic hospitalised HAI patients, and comparison with CAI patients

Table 2 shows the patients' demographics, pre-existing co-morbidities, clinical and laboratory parameters of the 95 symptomatic HAIs and 191 symptomatic CAIs. In total 286 adult hospitalized COVID-19 patients (95 HAI + 191 CAI) were included in the comparative study, reported as univariate analysis in the same table.

Amongst symptomatic HAI patients, the gender distribution was 54.8% males for 45.2% females, and the median age distribution was 76.0 years. Compared to CAI patients, HAI patients were significantly older (median age 76.0 vs. 64.0 years) ($P < 0.0001$) and smoked more ($P = 0.0164$); HAI patients had a lower BMI (HAI: 24.5 kg/m² vs. CAI: 26.9 kg/m²) ($P = 0.0025$) and were frailer ($P < 0.0001$). The following comorbidities were more frequent in HAI vs. CAI patients: i.e. anaemia ($P = 0.0066$), cancer ($P < 0.0001$), heart disease ($P = 0.0001$), liver disease ($P = 0.0443$) and renal disease ($P = 0.0483$), while others did not differ, amongst which in particular asthma ($P = 0.074$), and diabetes mellitus ($P = 0.4917$).

At time of COVID-19 diagnosis, the following symptoms were significantly more frequent in CAI patients compared to HAI patients: i.e. cough ($P = 0.0101$), shortness of breath ($P < 0.0001$), headache ($P = 0.0012$), sore throat ($P = 0.0089$), malaise ($P = 0.0142$) and ageusia ($P = 0.0117$).

Although the ICU admission rate was not significantly different amongst HAI (32.3%) compared to CAI patients (21.5%) ($P = 0.0578$); the number of deaths was significantly higher in HAI patients: 33.3% versus 15.7% ($P = 0.0011$). The median length of hospital stay from COVID-19 diagnosis was also significantly longer amongst HAI (13.0 days) than CAI (9.0 days) patients ($P = 0.001$).

Thrombocyte values were significantly higher in HAI patients ($232.0 \times 10^6/\mu\text{L}$) compared to CAI patients ($182.0 \times 10^6/\mu\text{L}$) ($P = 0.0011$), but all were in the physiological range. D-dimer values were above the physiological cut-off of 500 ng/mL in both patient groups and significantly higher in HAI patients (1293.0 ng/mL) compared to CAI patients (785.0 ng/mL) ($P = 0.0013$). Absolute leucocyte numbers ($P = 0.3334$) and neutrophil / lymphocyte ratios ($P = 0.5008$) did not statistically differ between groups. Although both acute phase proteins were above physiological thresholds, ferritin levels did not differ significantly between HAI and CAI patients ($P = 0.1786$) whereas C-reactive protein (CRP) was significantly higher in CAI patients (73.9 mg/L) compared to HAI patients (32.6 mg/L) ($P < 0.0001$).

Table 2: Hospitalized patients' demographics, clinical and laboratory data on symptomatic healthcare-associated versus community-associated COVID-19 (UZ Brussel, 2020-21).

Univariate statistical analysis of continuous data, represented as median (IQR), with unpaired t-test or Mann-Whitney test (depending on normal distribution of data). Univariate statistical analysis of categorical data was done with Fisher's exact test. $P < 0.05$ (marked in bold) was considered as statistically significant. (BMI – body mass index, ICU – intensive care unit, CRP – C-reactive protein)

Patient demographics		Healthcare-associated (N = 95)		%	Community-associated (N = 191)		%	p-value
Gender								0.703
	Male		51	54.8		111	58.1	
	Female		42	45.2		80	41.9	
Age (years)		76.0 (65.5 - 83.5)			64.0 (53.0 - 80.0)			< 0.0001
Smoking	Never		27	41.5		74	62.2	0.0164
	Former		27	41.5		36	30.3	
	Current		11	16.9		9	7.6	
BMI (kg / m²)		24.5 (22.5 - 29.1)			26.9 (24.1 - 30.9)			0.0025
Frailty								< 0.0001
	1		1	1.1		18	9.5	
	2		3	3.2		40	21.2	
	3		11	11.8		46	24.3	
	4		20	21.5		35	18.5	
	5		33	35.5		30	15.9	
	6		11	11.8		12	6.3	
	7		11	11.8		6	3.2	
	8		3	3.2		1	0.5	
	9		0	0.0		1	0.5	
Hospitalization								
ICU admission during stay			30	32.3		41	21.5	0.0578
Deceased			31	33.3		30	15.7	0.0011
Stay from COVID-19 diagnosis (days)		13.0 (8.0 - 21.5)			9.0 (5.0 - 14.3)			0.001
Symptoms at COVID-19 diagnosis								
Fever			58	65.2		125	67.6	0.7843
Chills			15	29.4		45	36.9	0.3849
Cough			45	55.6		125	72.3	0.0101
Shortness of breath			40	47.6		139	75.1	< 0.0001
Tachypnoea			54	61.4		95	52.8	0.1932
Headache			7	13.5		48	37.8	0.0012
Sore throat			2	4.7		26	23.0	0.0089
Coryza			4	10.3		15	15.3	0.5971
Malaise			15	31.3		57	53.8	0.0142
Diarrhoea			16	22.9		47	29.6	0.3374
Nausea			15	23.8		29	19.3	0.4636
Vomiting			11	16.9		25	16.0	0.8442
Ageusia			1	2.6		17	20.0	0.0117
Anosmia			3	7.5		15	17.2	0.1782
Laboratory parameters at COVID-19 diagnosis								
Trombocytes (x 10^6 / µL)		232.0 (173.5 - 316.0)			182.0 (156.8 - 235.0)			0.0011
Leucocytes (x 10^3 / µL)		5.8 (4.8 - 9.1)			6.7 (4.8 - 9.9)			0.3334
Neutrophil / lymphocyte ratio		4.8 (2.5 - 8.4)			5.0 (2.8 - 9.7)			0.5008
D-dimers (ng / mL)		1293.0 (764.3 - 2191.0)			785.0 (492.0 - 1481.0)			0.0013
Ferritin (µg / L)		514.0 (245.0 - 1060.0)			645.0 (326.0 - 1269.0)			0.1786
CRP (mg / L)		32.6 (12.4 - 68.8)			73.9 (32.9 - 153.9)			< 0.0001
Comorbidities of COVID-19 patients								
Anaemia			19	20.4		16	8.5	0.0066
Asthma			4	4.3		21	11.1	0.074
Cancer			27	29.0		9	4.7	< 0.0001
Hyperthyroidism			51	54.8		86	45.3	0.1635
Dementia			12	13.0		16	8.5	0.2888
Diabetes			31	33.3		55	28.8	0.4917
Heart disease			40	43.0		39	20.4	0.0001
Liver disease			7	7.5		4	2.1	0.0443
Neuromuscular disease			10	10.8		11	5.8	0.1509
Renal disease			23	24.7		28	14.7	0.0483

Multiple logistic regression was performed. When demographic and laboratory parameters were found to be significantly different between HAI and CAI patients in the univariate analysis (Table 2), we looked at whether those factors were associated with each other or outcome measures (being an 'HAI' or a 'CAI'). Symptoms and comorbidities were not included in the analysis because of their strong relatedness.

We report our findings in Table 3. The odds ratios reflect the effect of that parameter on the probability that a patient has a HAI. A patient's frailty index, CRP, and thrombocyte levels at COVID-19 diagnosis

seem to be significant at predicting a HAI.

Table 3: Multiple logistic regression modelling of HAI versus CAI patients.

Variable	Odds ratio	95% CI
Age (years)	0.9947	0.9495 to 1.040
Smoking	1.777	0.6752 to 5.210
BMI (kg / m ²)	0.9949	0.8848 to 1.126
Frailty	2.634	1.659 to 4.673
CRP (mg / L)	0.9725	0.9577 to 0.9839
Trombocytes (x 10 ⁶ / μ L)	1.022	1.011 to 1.036
D-dimers (ng / mL)	1	0.9996 to 1.000

Part 2. Description of clusters and validation of the ECDC source definitions for HAI

As explained in Figure 1, subjects from all the different subsets of COVID-19 cases were included in the cluster analysis: hospitalized CAI and HAI patients, as well as not hospitalized healthcare workers of our hospital.

Figure 2/A shows the epicurve of HAI (different colours according to the ECDC source definitions), CAI and healthcare workers. In figure 2/ B we used a MST to visualize genetic clustering of a subset of HAIs (for whom a genetic sample was available). When comparing the ECDC definitions of our 95 adult symptomatic HAIs with the sequencing data, we found that 92.5% of the 'definite', 93.8% of 'probable' and 81.3% of 'indeterminate' HAIs (according to ECDC definition), belonged to a COVID-19 HAI genome cluster. Of the healthcare worker-related HAIs, 84% could be attributed to a HAI genome cluster.

In figure 3, we included hospital wards as different metadata in a similar epicurve and MST. Clustering of genomes has made it possible to detect outbreaks spanning different wards and locate these clusters at different timepoints across the full length of the pandemic. Please note that not all cases depicted in the epicurve (Figure 3/A) are included in the genome analysis of panel B. Also of note, details of the genomes are described in appendix one. A comparison of epidemiological and genome data on HAI is provided in appendix two.

Figure 4 gives an overview of the genomes we sequenced. Panel A depicts the incidence of the different SARS-CoV-2 lineages across the pandemic. During the first and second waves, HAIs could be attributed to viral genomes closely related to the reference Wuhan-Hu-1 strain (B.1, B.1.1, B.1.160, B.1.1.44, B.1.177, B.1.177.77, B.1.1.269, B.1.221, B.38). During the third wave, the B.1.1.7 VOC predominated all HAI cases.

Discussion

1. Burden of HAI across three COVID-19 waves

In our study of COVID-19 HAI, we report numbers of HAI that are comparable to other centres^{3,4}. Interestingly, our study spans the first three waves of the COVID-19 pandemic in Belgium; the number and

percentages of HAI in our centre were stable across the first two waves, and increased towards the third wave, despite more control measures in place by that time of the pandemic. This may reflect more infectious variants as well as more systematic screening of all hospital admissions. Indeed, the percentage of symptomatic HAI patients with time was lowest during the third wave, probably reflecting vaccination rollout, earlier diagnoses and more exhaustive testing across the hospital.

2. Characteristics of HAI vs CAI

We then went on to describe symptomatic HAI and compared them to CAI. Again, there may have been some selection bias reflected in our results; for example, patients with a HAI are expected to be in poorer health before infection, since they were already hospitalised.

In the univariate analysis, some factors seemed to support this selection of sicker patients: our HAI patients were older, had higher frailty scores, more pronounced smoking habits and more comorbidities. Higher percentage of malignancies, kidney disease and older age were also observed in a British study on HAI versus CAI COVID-19¹⁴. With increased age, frailty, comorbidities, and initial reason for hospitalization in mind, it might not come as a surprise that HAI patients had a longer hospital stay from the time of COVID-19 diagnosis. Furthermore, more HAI patients were admitted to the ICU during their stay, compared to CAI. Finally, the mortality was also significantly higher in HAI, compared to CAI patients. Again, these differences are probably due to the selected population, rather than a causal link with HAI per se. Other studies had similar findings which confirm the extent of the problem of HAIs in hospitalized patients⁷.

Apparent differences in the presence or absence of symptoms between HAI and CAI patients might be due to the timing of COVID-19 diagnoses. Indeed, as HAI might be diagnosed at an earlier stage of the disease, because of timeliness of disease detection and laboratory testing – especially at later waves of COVID-19, certain symptoms might not yet have been present in some HAI cases. Similarly for laboratory data, in particular the significantly lower CRP levels in HAI patients which may be indicative of an earlier stage of COVID-19¹³. Other confounders for some laboratory findings, for e.g. D-dimers, are the underlying pathologies of hospitalized patients.

In the multivariate analysis frailty (+), thrombocytes (+) and CRP (-) were significantly associated with HAI. These may become important parameters to take into account when trying to decide if a patient has acquired his/her infection in hospital. Of note, BMI was not significantly associated with having a HAI, compared to a CAI, but we must stress we did not assess associations with severe outcomes – so our results cannot draw any conclusion on the link between BMI amongst HAI and severe COVID-19 disease.

3. Cluster analysis

In part B of the results' section, we report correlations between clinical criteria for the diagnosis of an HAI and genetic sequencing data. When considering bias and limitations, it is true that we were only able to sequence a proportion of the HAI samples.

Despite this, our sequencing analyses allowed us to validate the ECDC definitions for HAI. Sequencing can therefore nicely complement descriptive analyses to describe clusters and seems to be a great tool to better understand COVID-19 transmission within hospitals. Indeed, due to the remaining uncertainties around the incubation period, pre-symptomatic transmission and asymptomatic infections, a definite international consensus on the definition of a HAI is yet to be defined.

We described 12 clusters involving HAI. Some of the clusters we described were large, stressing once again the extreme infectiousness of this infectious agent. In some, different wards and floors were affected by the same cluster (cluster 10): movement of (undetected) infected patients and healthcare workers across wards might have caused transmission. Indeed, looking at figure 2, many clusters have at least a healthcare worker as part of the cluster, thereby suggesting that healthcare workers were involved in nosocomial transmission. A narrative review by Abbas and co-workers highlights the important implications of SARS-CoV-2 transmission to and from healthcare workers²⁵. While protection of healthcare workers is key in prevention of nosocomial COVID-19 outbreaks, preventive measures are predominantly focused on the use of personal protective equipment, such as masks. It seems vital that other preventive measures are also well established, such as physical distancing, appropriate workload, adequate training, population-wide vaccination, etc.²⁵. Our findings support this fully.

With regard to cluster 10, it is not so surprising that the viral strain circulating in this cluster was of the B.1.1.7 lineage. Indeed, we now know that the transmissibility of that SARS-CoV-2 strain is estimated to be 1.56 times higher than the previously reported strains¹⁰. It is also very interesting to see that while several SARS-CoV-2 lineages were circulating in HAI patients earlier in the pandemic, 100% of the isolates from HAI samples were of the B.1.1.7 from the beginning of 2021 onwards.

Conclusion

This study is an in-depth analysis of HAI in a university hospital in Brussels, Belgium, across all past waves of the COVID-19 pandemic (at the time of submission).

It appears that HAI patients tend to be older, frailer, and have more comorbidities. We conducted a genomic cluster analysis of our HAIs, and were able to validate the ECDC clinical criteria to identify HAIs.

Even if no infection control system will prevent HAIs entirely, we therefore suggest combining the ECDC HAI criteria with a timely and automated tracking system with genomic cluster analysis in real-time. Such a system coupled with a hospital-wide alert system, on a background of repeated screening and surveillance, could prevent or limit such large outbreaks in the future. This will become increasingly important with yet other variants emerging.

With those elements in hand, we should urge local and national policy makers to invest more in infection control and HAI surveillance.

Declarations

Contributors

TD, LS, DP, SA, and IW designed the study and wrote the study protocol. TD, RB, OS, BC, and RV aided in sample selection and genomic sequencing. TD, LS, and JM did the clinical data extraction. TD, LS JP, TS, JB, HM, FC, and IW performed data analysis and visualization. TD drafted the first version of the manuscript; LS worked on the second version. All authors had full access to all data, aided in finalizing the text and shared final responsibility for the submission for publication.

Declaration of interests

No funding was asked for this study. The authors declare no competing interests.

Data sharing

Data of patients included in this manuscript are considered sensitive and will not be shared. The study methods and statistical analyses are described in detail in the methods section. Virus genomic data are shared on GISAID.

References

1. COVID-19 situation update for the EU/EEA, as of 23 July 2021.
<https://www.ecdc.europa.eu/en/cases-2019-ncov-eueea>.
2. COVID-19 - Epidemiologische situatie | Coronavirus Covid-19. <https://covid-19.sciensano.be/nl/covid-19-epidemiologische-situatie>.
3. Wang, D. *et al.* Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA - Journal of the American Medical Association* **323**, 1061–1069 (2020).
4. Marago, I. & Minen, I. Hospital-Acquired COVID-19 Infection – The Magnitude of the Problem. *SSRN Electronic Journal* (2020) doi:10.2139/ssrn.3622387.
5. Nguyen, L. H. *et al.* Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *The Lancet Public Health* **5**, e475–e483 (2020).
6. Rivett, L. *et al.* Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. *eLife* **9**, (2020).
7. Barranco, R., du Tremoul, L. V. B. & Ventura, F. Hospital-acquired sars-cov-2 infections in patients: Inevitable conditions or medical malpractice? *International Journal of Environmental Research and Public Health* **18**, 1–9 (2021).
8. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261 (2020).

9. Planas, D. *et al.* Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* 2021 596:7871 **596**, 276–280 (2021).
10. Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021).
11. Frampton, D. *et al.* Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *The Lancet Infectious Diseases* (2021) doi:10.1016/s1473-3099(21)00170-5.
12. Singanayagam, A. *et al.* Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. *The Lancet Infectious Diseases* (2021) doi:10.1016/S1473-3099(21)00648-4.
13. Demeester, S. *et al.* Routine haematology parameters in COVID-19 patients and clinical outcome: A Belgian single-centre study. *International Journal of Laboratory Hematology* (2020) doi:10.1111/ijlh.13313.
14. Khonyongwa, K. *et al.* Incidence and outcomes of healthcare-associated COVID-19 infections: significance of delayed diagnosis and correlation with staff absence. *Journal of Hospital Infection* **106**, 663–672 (2020).
15. Surveillance definitions for COVID-19. <https://www.ecdc.europa.eu/en/covid-19/surveillance/surveillance-definitions>.
16. Meredith, L. *et al.* Rapid implementation of real-time SARS-CoV-2 sequencing to investigate healthcare-associated COVID-19 infections. (2020) doi:10.1101/2020.05.08.20095687.
17. Paltansing, S. *et al.* Transmission of SARS-CoV-2 among healthcare workers and patients in a teaching hospital in the Netherlands confirmed by whole-genome sequencing. *Journal of Hospital Infection* **110**, 178–183 (2021).
18. SARS-CoV2 genome sequencing protocol (1200bp amplicon “midnight” primer set, using Nanopore Rapid kit). <https://www.protocols.io/view/sars-cov2-genome-sequencing-protocol-1200bp-amplicon-btsrnn6>.
19. Artic Network. <https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>.
20. Freed, N. E., Vlková, M., Faisal, M. B. & Silander, O. K. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biology Methods and Protocols* **5**, (2020).
21. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology* 2020 5:11 **5**, 1403–1407 (2020).
22. GISAID - European Union. <https://www.gisaid.org/phylogenetics/european-union/>.
23. Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, co - Nucleotide - NCBI. <https://www.ncbi.nlm.nih.gov/nuccore/1798174254>.
24. COVID-19: epidemiological definitions of outbreaks and clusters in particular settings - GOV.UK. <https://www.gov.uk/government/publications/covid-19-epidemiological-definitions-of-outbreaks->

and-clusters/covid-19-epidemiological-definitions-of-outbreaks-and-clusters-in-particular-settings.

25. Abbas, M. *et al.* Nosocomial transmission and outbreaks of coronavirus disease 2019: the need to protect both patients and healthcare workers. Antimicrobial Resistance and Infection Control vol. 10 (2021).

Figures

Figure 1

Legend not included with this version

Figure 2

Legend not included with this version

Figure 3

Legend not included with this version

Figure 4

Legend not included with this version

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarydata1.docx](#)
- [Supplementarydata2.docx](#)