

1 Viral sequencing reveals US healthcare
2 personnel rarely become infected with
3 SARS-CoV-2 through patient contact

4 Katarina M. Braun^{1*}, Gage K. Moreno^{2*}, Ashley Buys³, Max Bobholz², Molly A. Accola³, Laura
5 Anderson³, William M. Rehrauer³, David A. Baker², Nasia Safdar⁴, Alexander J. Lepak⁴, David
6 H. O'Connor^{2,5#}, Thomas C. Friedrich^{1,5#}

7

8 ¹Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI, United
9 States of America

10 ²Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison,
11 WI, United States of America

12 ³University of Wisconsin School of Medicine and Public Health, Madison, WI, United States of
13 America and the William S. Middleton Memorial Veterans Hospital

14 ⁴Department of Medicine, Division of Infectious Diseases, University of Wisconsin School of
15 Medicine and Public Health, Madison, WI

16 ⁵Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, WI,
17 United States of America

18 *These authors contributed equally

19 #These authors contributed equally

20 Summary

21 Background

22 Healthcare personnel (HCP) are at increased risk of infection with the severe acute respiratory
23 coronavirus 2019 virus (SARS-CoV-2). Between 12 March 2020 and 10 January 2021, >1,170
24 HCP tested positive for SARS-CoV-2 at a major academic medical institution in the Upper
25 Midwest of the United States. We aimed to understand the sources of infections in HCP and to
26 evaluate the efficacy of infection control procedures used at this institution to protect HCP from
27 healthcare-associated transmission.

28

29 Methods

30 In this retrospective case series, we used viral genomics to investigate the likely source of
31 SARS-CoV-2 infection in 96 HCP where epidemiological data alone could not be used to rule
32 out healthcare-associated transmission. We obtained limited epidemiological data through
33 informal interviews and review of the electronic health record. We combined viral sequence data
34 and available epidemiological information to infer the most likely source of HCP infection.

35

36 Findings

37 We investigated 32 SARS-CoV-2 infection clusters involving 96 HCP, 140 possible patient
38 contacts, and 1 household contact (total n = 237). Of these, 182 sequences met quality
39 standards and were used for downstream analysis. We found the majority of HCP infections
40 could not be linked to a patient or co-worker and therefore likely occurred in the outside
41 community (58/96; 60.4%). We found a smaller percentage could be traced to a coworker

42 (10/96; 10.4%) or were part of a patient-employee cluster (12/96; 12.5%). Strikingly, the
43 smallest proportion of HCP infections could be clearly traced to a patient source (4/96; 4.2%).

44

45 Interpretation

46 Infection control procedures, consistently followed, offer significant protection to HCP caring for
47 COVID-19 patients in a representative American academic medical institution. Rapid SARS-
48 CoV-2 genome sequencing in healthcare settings can be used retrospectively to reconstruct the
49 likely source of HCP infection when epidemiological data are not available or are inconclusive.
50 Understanding the source of SARS-CoV-2 infection can then be used prospectively to adjust
51 and improve infection control practices and guidelines.

52

53 Funding

54 This project was funded in part through a COVID-19 Response grant from the Wisconsin
55 Partnership Program at the University of Wisconsin School of Medicine and Public Health to
56 T.C.F. and D.H.O. Author N.S. is supported by the National Institute of Allergy and Infectious
57 Diseases Institute (NIAID) Grant 1DP2AI144244-01.

58 Research in context

59 Evidence before this study

60 On 16 January 2021 we searched for “SARS-CoV-2” AND “healthcare workers” AND “viral
61 sequencing” in Google Scholar. This search returned 57 results, and included a number of
62 preprint articles. We found two studies that used viral sequencing to investigate healthcare-

63 associated outbreaks in the Netherlands ¹ and the United Kingdom ². To our knowledge, no
64 study has used viral sequencing to specifically investigate the source of SARS-CoV-2 infections
65 in healthcare workers in the United States. Although we and others have written about the
66 potential utility of sequencing as an infection control asset ³⁻⁶, few have demonstrated the
67 practical application of such efforts.

68

69 Added value of this study

70 Our study suggests infection control measures in place at the institution evaluated in this case
71 series are largely protecting healthcare personnel (HCP) from healthcare-associated SARS-
72 CoV-2 infections. Even so, the majority of healthcare-associated infections we did identify
73 appeared to be linked to HCP-to-HCP spread so additional messaging and guidelines to reduce
74 HCP-to-HCP spread in and out of the workplace may be warranted. In addition, we
75 demonstrated how rapid viral sequencing can be combined with, even limited, epidemiological
76 information to reconstruct healthcare-associated SARS-CoV-2 outbreaks.

77

78 Implications of all the available evidence

79 Healthcare-associated SARS-CoV-2 infections negatively affect HCP, patients, and
80 communities. Infections among HCP add further strain to the healthcare system and put
81 patients and other HCP at risk. We found the majority of HCP infections appeared to be
82 acquired through community exposure so measures to reduce community spread are critical.
83 This further emphasizes the importance of mask-wearing, physical distancing, robust testing
84 programs, and the rapid distribution of vaccines.

85 Introduction

86 Despite the use of personal protective equipment (PPE) and other strategies to mitigate risk,
87 front-line healthcare workers are at increased risk for infection with severe acute respiratory
88 syndrome coronavirus 2 (SARS-CoV-2) compared to the general population ⁷. Healthcare-
89 associated transmission of SARS-CoV-2 poses a serious risk to healthcare workers as well as
90 to other hospital staff and patients ^{8,9}. Here we use rapid viral sequencing and forensic
91 genomics to uncover the likely source of infection in 96 confirmed cases of coronavirus-disease
92 2019 (COVID-19) in healthcare personnel (HCP). We further describe how the results of these
93 investigations informed infection control recommendations within a large academic medical
94 system in the midwestern United States.

95
96 Healthcare-associated SARS-CoV-2 infections negatively affect HCP through direct health
97 impacts, lost wages, and secondary consequences for their close contacts ¹⁰. Healthcare-
98 associated SARS-CoV-2 infections can also negatively impact patient care through staffing
99 shortages, environmental contamination, low morale and other mental health impacts on HCP;
100 each of these can secondarily impact the overall quality of care ^{2,11}.

101
102 The US Centers for Disease Control and Prevention (CDC) have released guidelines for
103 infection prevention for HCP interacting directly with patients with suspected or confirmed
104 infection with SARS-COV-2 ¹². These guidelines include recommendations for the proper use of
105 PPE, hand hygiene, precautions to be taken during aerosol-generating procedures, collection of
106 diagnostic respiratory specimens, environmental infection control practices and many others.
107 These guidelines in addition to institution-specific infection control measures – described in
108 detail in Lepak et al ¹³ – were in place at the institution evaluated here. We posit that these
109 guidelines are generally successful in protecting HCP from SARS-CoV-2 infection in a

110 healthcare setting. In this report, we test this hypothesis using viral sequencing to investigate
111 the likely sources of infection in a series of HCP in the United States.
112
113 Sequencing has been used to explore the origin and the path of spread of various nosocomial
114 *bacterial* pathogens, such as vancomycin resistant-enterococcus (VRE), *Listeria*
115 *monocytogenes*, and *Klebsiella pneumoniae* ^{14–17}. With a few exceptions ^{18–20}, viral sequencing
116 is not currently standard practice for investigating healthcare-associated SARS-CoV-2
117 infections, although we and others have highlighted the potential utility of this approach ^{3–6}. Data
118 suggest SARS-CoV-2 accumulates approximately one fixed mutation every other transmission
119 event ^{21,22}. Therefore, if SARS-CoV-2 is directly transmitted from one individual to another, this
120 “transmission pair” is expected to share identical, or nearly identical (≤ 1 consensus SNV
121 difference), viral sequences. The most parsimonious explanation for identical or near-identical
122 genomes generated from individuals with known contact is that they are transmission pairs ²³. In
123 contrast, two people who became infected at similar times, but from different sources, would be
124 expected to be infected with viruses that differ from each other by two or more nucleotides. This
125 is especially true at this stage of the pandemic in the United States, when transmission rates are
126 high and multiple viruses of distinct genetic lineages cocirculate in many areas ²⁴. By increasing
127 the resolution of inference, rapid viral sequencing can facilitate a targeted approach to examine
128 SARS-CoV-2 nosocomial outbreaks at the level of the individual and the institution, which others
129 have referred to collectively as “precision epidemiology” ²⁵.

130 Results

131 HCP began testing positive for SARS-CoV-2 at a major academic biomedical institution in the
132 American Upper Midwest in early March 2020. From 12 March, 2020 to 10 January, 2021
133 ~1,172 HCP tested positive for SARS-CoV-2 at this center. In collaboration with this institution’s

134 infection control team, we began sequencing viral genomes from residual nasopharyngeal
135 specimens from the individuals involved in these infection clusters. We focused our analyses on
136 HCP infections and infection clusters that were highest risk for nosocomial transmission, as
137 when healthcare-associated transmission could not be ruled out using epidemiological data
138 alone. Each investigation included at least one HCP, all known direct and indirect SARS-CoV-2-
139 positive patient contacts where residual swab was available, and occasionally extended to
140 epidemiologically-linked household contacts. Relevant patient contacts were identified in the
141 Epic electronic medical record using a comprehensive caregiver trace. This function identifies all
142 patient records accessed by a HCP.

143
144 We consider three main potential sources of healthcare-associated infection with SARS-CoV-2:
145 “outside community,” “patient source” (via HCP-patient interactions), and “employee source” (via
146 HCP-HCP interactions). A few HCP infections did not fit neatly into these categories so we have
147 included two additional categories. First, “combined patient and employee cluster”, where a
148 patient-to-HCP transmission event likely initiated a cluster of infections, but we were unable to
149 pinpoint the first HCP to become infected. Second, “inconclusive”, where a consensus
150 sequence was not available, there were no appropriate comparator samples, or epidemiological
151 information were insufficient to interpret the sequence data.

152
153 For us to conclude person A was a likely source of infection for person B, persons A and B must
154 have had known contact with each other, must have been tested within 0-14 days of each other,
155 and must have been infected with identical or near-identical (≤ 1 consensus intrahost single
156 nucleotide variant (iSNV) difference) viruses. In cases where we found the HCP virus diverged
157 from the patient- and employee-contact viruses, we concluded infection most likely occurred
158 outside of the healthcare setting, or in the “outside community”.

159

160 In total, we investigated 32 SARS-CoV-2 infection clusters involving 96 HCP, 140 possible
161 patient contacts, and 1 household contact (total n = 237). This accounted for approximately
162 ~8.2% (96/1,172) of all documented infections in the HCP at this institution. In total, we
163 sequenced 237 samples collected between 15 March and 27 December, 2020. Of these, 182
164 met quality standards (as described in methods) and were used for downstream analysis. We
165 did not find a closely related virus among the epidemiologically linked contacts in 58 HCP
166 infections, so we concluded the most likely source of infection in these cases could be traced to
167 the outside community (58/96; 60.4%). We find a smaller percentage could be traced to a
168 coworker (10/96; 10.4%) or were part of a patient-employee cluster (12/96; 12.5%). Strikingly,
169 the smallest proportion of HCP infections could be clearly traced to a patient source (4/96;
170 4.2%). The remaining HCP infections could not be definitively traced to a single source and
171 were therefore inconclusive (12/96; 12.5%) (**Table 1**). Below, we describe one representative
172 example of three distinct transmission scenarios – outside community-to-HCP, HCP-to-HCP,
173 and patient-to-HCP. A brief overview of all cases included in this study can be found in
174 **Supplementary File 1**.

175

176 In case #20, we compared the viral sequence of a HCP (HCP 20-1), who tested positive in early
177 October, to a patient contact who tested positive eight days prior. A comprehensive caregiver
178 trace of HCP 20-1 revealed a single patient contact with diagnosed COVID-19 (patient 20-A)
179 within the 14 days prior HCP 20-1's symptom onset. HCP 20-1 provided direct care to patient
180 20-A while wearing appropriate PPE and with no reported lapses in PPE. Viral sequencing
181 revealed HCP 20-1 was infected with a virus clustering with the 20G clade whereas patient 20-A
182 was infected with a 20A-clade virus. The sequences of these viruses differed at >20 sites, so we
183 concluded these individuals were unlikely to represent a transmission pair and HCP 20-1 was
184 more likely infected in the outside community (**Figure 1**).

185

186 In case #16, we investigated infections in three HCP who worked in the same department and
187 tested positive in early September (HCP 16-2), mid-September (HCP 16-1), and late September
188 (HCP 16-3). Contact tracing revealed HCP 16-2 worked for two days prior to symptom onset
189 and may have had unmasked contact with HCP 16-1 during overlapping meal breaks. Contact
190 tracing additionally revealed HCP 16-3 had an exposure event lasting >15 minutes in the
191 outside community prior to testing positive. Viral sequencing in this cluster showed HCP 16-1
192 and 16-2 were infected with 20G-clade viruses that were identical at the consensus level, while
193 HCP 16-3 was infected with a genetically dissimilar 20A-clade virus. We therefore concluded
194 HCP 16-2 was a likely source of infection for HCP 16-1, while HCP 16-3 was likely infected
195 elsewhere (**Figure 2**).

196

197 Case #10 involved a HCP (HCP 10-1) who provided care for 15 patients diagnosed with
198 COVID-19 in the 14 days prior to HCP 10-1's symptom onset. HCP 10-1 provided direct care to
199 each of these patients while wearing appropriate PPE with no reported lapses in PPE. We
200 generated consensus sequences from HCP 10-1 and nine patient contacts. There was
201 insufficient viral RNA (vRNA) in the remaining six patient contacts to generate high-quality
202 consensus sequences for comparison. The virus isolated from patient 10-G was identical to the
203 virus from HCP 10-1. Given the known epidemiological association between these two
204 individuals (HCP 10-1 provided direct patient care to patient 10-G), the time separating sample
205 collections (late July & early August), and identical viral sequences, we concluded patient 10-G
206 is a likely source of infection for HCP 10-1 (**Figure 3**).

207

208 The center where we conducted this case series implemented a number of changes to their
209 institutional infection control guidelines based on these sequencing results ¹³. The
210 recommendations for extended reuse of medical grade face masks were clarified and now
211 instruct HCP to consider barrier mask replacement after three days of wear, to inspect the

212 barrier mask prior to each use and to replace if soiled or damaged, and to always ask for PPE
213 when needed. N-95s or powered air-purifying respirators (PAPR) are now universally required
214 on inpatient units housing COVID-19-confirmed and suspected patients. In addition, medical-
215 grade face masks, instead of cloth masks, are now required for HCP in all clinical areas, and not
216 just direct patient care areas. This final recommendation was based on likely HCP-to-HCP
217 transmission involving a HCP who was not directly involved in patient care of COVID-19
218 patients (case #14 in **Supplementary File 1**).

219 Discussion

220 HCP across the hospital are involved in caring for people with COVID-19, whether or not they
221 work on an actual COVID-19 ward. With shifting guidelines and PPE shortages that persist
222 today, it is critical to assess the risk that HCP treating people with known SARS-CoV-2 infection
223 will become infected themselves. Here we used viral genome sequencing to assess the risk that
224 HCP in a large academic medical system treating COVID-19 patients would acquire nosocomial
225 infections. Although others have written about the potential utility of sequencing as an infection
226 control asset within clinical settings, few have demonstrated the practical application of such
227 efforts³⁻⁵. Our results suggest that caring for COVID-19 patients accounted for a minority of
228 HCP infections (n=4), even while evaluating high-risk infections where epidemiological data
229 alone could not be used to rule out healthcare-associated transmission. In contrast, HCP at this
230 institution were much more likely to acquire SARS-CoV-2 from infected coworkers (n=10) or in
231 the outside community (n=58). This result suggests that infection control procedures,
232 consistently followed, offer significant protection to HCP caring for COVID-19 patients in the
233 United States. A similar conclusion was drawn by recent studies evaluating healthcare-
234 associated infections in the Netherlands and in the UK, suggesting this conclusion may hold
235 across healthcare systems^{1,2}.

236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260

Our results suggest that the infection control measures and PPE in place at the institution where we conducted this case series ¹³ can protect against symptomatic SARS-CoV-2 infection, even in settings where the density of individuals infected with SARS-CoV-2 is relatively high. This provides some confidence that PPE and similar infection control measures, when used consistently and correctly, will provide adequate protection in other congregate settings, like schools. The impact of reopening primary schools remains unclear, yet the long-term consequences of remote learning and physical distancing on children, particularly those from disadvantaged backgrounds, are of increasing concern ²⁶. In addition to reducing rates of community spread, it follows that schools should prioritize making PPE and training on its appropriate use available to all staff ²⁷. Similar precautions are needed in other congregate settings as well, such as correctional facilities and factories, to protect inmates and workers.

The effectiveness of hospital- and clinic-like PPE and infection control guidelines used in other congregate settings should similarly be evaluated following implementation. As the cost of viral sequencing continues to drop and sequencing technologies continue to improve, the integration of rapid pathogen sequence information with traditional epidemiological investigation approaches is possible at the level of individual hospitals, clinics, school systems, and correctional facilities. Such systems would allow for the rapid reconstruction of transmission networks. These systems would also inform feedback loops for schools and institutions to continually assess and improve the efficacy of infection control and PPE procedures in protecting HCP, teachers, children, workers, and individuals who are incarcerated. We should work to build such systems now because we can be certain pathogenic viruses will continue to emerge unpredictably.

261 Sampling and contact tracing of nosocomial outbreaks is often coordinated by local hospitals
262 and/or departments of health while expertise in viral sequencing, bioinformatics, and
263 phylogenetics can more often be found in academic laboratories. Successful application of
264 precision epidemiology requires the integration of these areas. This is possible now at academic
265 medical institutions like ours, but presents more of a challenge at smaller, rural, and private
266 patient care centers. Federal support should be provided to help establish and maintain these
267 collaborations in the current pandemic and in anticipation of future outbreaks.

268
269 Importantly, we were only able to evaluate samples for which residual swab was available and
270 for which we were able to generate high-quality consensus sequences. Given this limitation, we
271 were often able to exclude patient contacts and co-workers as likely sources of infection in HCP,
272 but we were rarely able to pinpoint the exact source of infection, especially when it occurred
273 outside of the healthcare setting. In addition, this study only examined SARS-CoV-2 infections
274 in HCP from a single academic medical center so our conclusions may not be broadly
275 generalizable. However, another recent study evaluated healthcare-associated infections in the
276 Netherlands and similarly found no evidence for widespread nosocomial transmission of SARS-
277 CoV-2, suggesting our conclusions may hold across institutions and healthcare systems ¹.
278 Further, we were not able to differentiate between routes of infection (airborne, droplet, contact)
279 with the limited epidemiological data available to us in this study.

280 Here we demonstrated how rapid whole-genome sequencing of current SARS-CoV-2 outbreaks
281 in hospitals can be used retrospectively to reconstruct the likely source of HCP infection and
282 prospectively to adjust and improve infection control practices and guidelines. The approach we
283 describe here need not be limited to investigation of pandemic virus outbreaks. Key concepts
284 from genome sequencing and routine pathogen surveillance can be applied to any nosocomial
285 pathogen and inform changes to infection control practices. Overall, while we do find examples

286 of patient-to-HCP and HCP-to-HCP spread, the majority of HCP infections appeared to be
287 acquired through community exposure, emphasizing the importance of ongoing measures to
288 reduce community spread through mask-wearing, physical distancing, robust testing programs,
289 and rapid distribution of vaccines.

290 **Methods**

291 **Sample approvals and sample selection criteria**

292 From 12 March 2020 to 10 January 2021, ~1,172 HCP tested positive for SARS-CoV-2 at a
293 major academic medical institution in the Upper Midwest. In the case of infections where
294 epidemiological data alone could not rule out healthcare-associated infection, we sequenced
295 viral genomes from all available residual specimens from the individuals involved in these
296 outbreaks. These clusters included at least one HCP, SARS-CoV-2-positive patient contacts,
297 and occasionally extended to epidemiologically-linked hospital staff and household contacts.
298 Relevant patient contacts were identified in the Epic electronic medical record using a
299 comprehensive caregiver trace. This function identifies all patient records accessed by a HCP
300 being traced.

301 **Summary of infection control measures to prevent transmission of** 302 **SARS-CoV-2 at our institution**

303 Detailed descriptions of all infection control measures implemented to prevent transmission of
304 SARS-CoV-2 at the medical institution evaluated here can be found in a recent report by Lepak
305 et al ¹³. Briefly, these guidelines include a universal testing policy for all patients admitted to the
306 hospital, negative air pressure in all locations where SARS-CoV-2 patients are treated, a limit of

307 one visitor or primary support person per patient per day (required to undergo screening prior to
308 entry), establishment of an employee testing site with required employee self-monitoring for
309 signs and symptoms of infection, maintenance of a log of persons entering the room of a
310 confirmed or suspected COVID-19 patient for contact tracing purposes, and establishment of a
311 designated respiratory care unit in the emergency department, among others. PPE guidelines
312 include universal masking and face-shield use with any patient care contact, enhanced PPE
313 requirements for HCP working in the intensive care unit or performing aerosol-generating
314 procedures, and formal donning and doffing protocols and required training for all HCP, among
315 others.

316 vRNA isolation

317 All samples were isolated using a Maxwell isolation instrument and subsequently processed
318 using a modified ARTIC tiled amplicon approach^{28,29}. Nasopharyngeal swabs received in
319 transport medium (VTM) were briefly centrifuged at 14,000 r.p.m. for 30 seconds at room
320 temperature to ensure all residual sample sediments at the bottom of the tube. Viral RNA
321 (vRNA) was extracted from 100 µL of VTM using the Viral Total Nucleic Acid Purification kit
322 (Promega, Madison, WI, USA) on a Maxwell RSC 48 instrument and was eluted in 50 µL of
323 nuclease-free H₂O.

324 Complementary DNA (cDNA) generation

325 Complementary DNA (cDNA) was synthesized using a modified ARTIC Network approach^{28,29}.
326 Briefly, vRNA was reverse transcribed with SuperScript IV Reverse Transcriptase (Invitrogen,
327 Carlsbad, CA, USA) using random hexamers and dNTPs. Reaction conditions were as follows:
328 1 µL of random hexamers and 1 µL of dNTPs were added to 11 µL of sample RNA, heated to
329 65°C for 5 minutes, then cooled to 4°C for 1 minute. Then 7 µL of a master mix (4 µL 5x RT

330 buffer, 1 μ L 0.1M DTT, 1 μ L RNaseOUT RNase Inhibitor, and 1 μ L SSIV RT) was added and
331 incubated at 42°C for 10 minutes, 70°C for 10 minutes, and then 4°C for 1 minute.

332 Multiplex PCR to generate SARS-CoV-2 genomes

333 A SARS-CoV-2-specific multiplex PCR for Nanopore sequencing was performed, similar to
334 amplicon-based approaches as previously described^{28,29}. In short, primers for 96 overlapping
335 amplicons spanning the entire genome with amplicon lengths of 500bp and overlapping by 75 to
336 100bp between the different amplicons were used to generate cDNA. cDNA (2.5 μ L) was
337 amplified in two multiplexed PCR reactions using Q5 Hot-Start DNA High-fidelity Polymerase
338 (New England Biolabs, Ipswich, MA, USA) using the conditions previously described^{28,29}.
339 Samples were amplified through 25 cycles of PCR and each resulting multiplex sample was
340 pooled together before ONT library prep.

341 Library preparation and sequencing

342 Amplified PCR product was purified using a 1:1 concentration of AMPure XP beads (Beckman
343 Coulter, Brea, CA, USA) and eluted in 30 μ L of water. PCR products were quantified using Qubit
344 dsDNA high-sensitivity kit (Invitrogen, USA) and were diluted to a final concentration of 1 ng/ μ L.
345 A total of 5ng for each sample was then made compatible for deep sequencing using the one-
346 pot native ligation protocol with Oxford Nanopore kit SQK-LSK109 and its Native Barcodes
347 (EXP-NBD104 and EXP-NBD114)²⁹. Specifically, samples were end-repaired using the
348 NEBNext Ultra II End Repair/dA-Tailing Module (New England Biolabs, Ipswich, MA, USA).
349 Samples were then barcoded using 2.5 μ L of ONT Native Barcodes and the Ultra II End Repair
350 Module. After barcoding, samples were pooled directly into a 1:1 concentration of AMPure XP
351 beads (Beckman Coulter, Brea, CA, USA) and eluted in 30 μ L of water. Samples were then
352 tagged with ONT sequencing adaptors according to the modified one-pot ligation protocol²⁹. Up

353 to 24 samples were pooled prior to being run on the appropriate flow cell (FLO-MIN106) using
354 the 72hr run script.

355 Processing raw ONT data

356 Sequencing data was processed using the ARTIC bioinformatics pipeline
357 (<https://github.com/artic-network/artic-ncov2019>), with a few modifications. Briefly, we have
358 modified the ARTIC pipeline so that it demultiplexes raw fastq files using qcat as each fastq file
359 is generated by the GridION (<https://github.com/nanoporetech/qcat>). Once a barcode reaches
360 100k reads, it will trigger the rest of the ARTIC bioinformatics workflow which will map to the
361 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 reference (Genbank:
362 MN908947.3) using minimap2. This alignment will then be used to generate consensus
363 sequences and variant calls using medaka (<https://github.com/nanoporetech/medaka>). The
364 entire ONT analysis pipeline is available at [https://github.com/gagekmoreno/SARS-CoV-2-in-](https://github.com/gagekmoreno/SARS-CoV-2-in-Southern-Wisconsin)
365 [Southern-Wisconsin](https://github.com/gagekmoreno/SARS-CoV-2-in-Southern-Wisconsin).

366 Consensus sequence analysis – clade and lineage generation

367 Following the generation of consensus sequences via the above ARTIC pipeline, samples were
368 excluded from downstream analysis if gaps in the consensus sequence totaled $\geq 20\%$ of the
369 genome. Each sample's consensus sequence within a report (or as many as possible), were
370 visually inspected in Geneious Prime (<https://www.geneious.com>) and/or in Nextstrain's
371 Nextclade online tool (<https://clades.nextstrain.org/>). When inspecting in Geneious, we aligned
372 all sequences to a standard reference sequence – an early Wuhan sequence (Genbank:
373 MN908947.3). Using these alignments as well as exported Nextclade CSV files, we were able to
374 quickly identify consensus intrahost single nucleotide variant (iSNV) differences. We additionally

375 used Nextstrain's Nextclade tool to assign clades. We used Pangolin's command-line tool to
376 assign sequences to Pangolin lineages (<https://github.com/cov-lineages/pangolin>).

377

378 As noted in the results, for us to conclude person A was a likely source of infection for person B,
379 persons A and B must have had known contact or epidemiological links, they must have been
380 tested within 0-14 days of each other, and they must be infected with identical or near-identical
381 (≤ 1 consensus iSNV difference) viruses.

382 Consensus sequence analysis – Southeast Wisconsin

383 Phylogenetic tree

384 Wisconsin-centric time-resolved and divergence phylogenetic trees (seen in **Supplementary**
385 **File 1**) were built using the standard Nextstrain tools and scripts ³⁰.

386 Report generation

387 For each sample set, we drafted these results into a PDF report to be shared with the hospital
388 infection control team. These reports were drafted in a standardized format. In each report, we
389 outlined the purpose for the report, the samples involved including any associated metadata, a
390 brief description of the methods used to generate the sequence data, a screen grab of a
391 Nextclade alignment, a table summarizing the iSNV differences, phylogenetic trees if applicable,
392 and finally any overall conclusions that could be drawn regarding the likely and unlikely sources
393 of infection. A summary of each of these reports can be found in **Supplementary File 1**.

394 Data availability

395 Variant identities, alignments, phylogenetic trees, Nextclade CSVs, and Pangolin lineage
396 reports can be found for each case on the GitHub accompanying this manuscript [here](#). An
397 interactive view of the Wisconsin Nextstrain phylogenetic tree can be found [here](#). Figures 1A,
398 2A, and 3A were created with [BioRender.com](#).

399 Study approvals

400 The University of Wisconsin-Madison Institutional Review Board deemed this study quality
401 improvement, rather than research, and considered it exempt from review. Data and metadata
402 were collected as part of routine infection control policy in nosocomial outbreaks and all data
403 were deidentified prior to analysis.

404 Competing interests

405 The authors declare no competing interests.

406 Acknowledgements

407 We gratefully acknowledge Anna Heffron for assisting with sample transport. We also thank all
408 healthcare workers and infection control teams for their ongoing dedication to patient and
409 community health and wellness. This project was funded in part through COVID-19 Response
410 grants from the Wisconsin Partnership Program at the University of Wisconsin School of
411 Medicine and Public Health to T.C.F. and D.H.O. N.S. is supported by the National Institute of
412 Allergy and Infectious Diseases Institute (NIAID) Grant 1DP2AI144244-01.

413

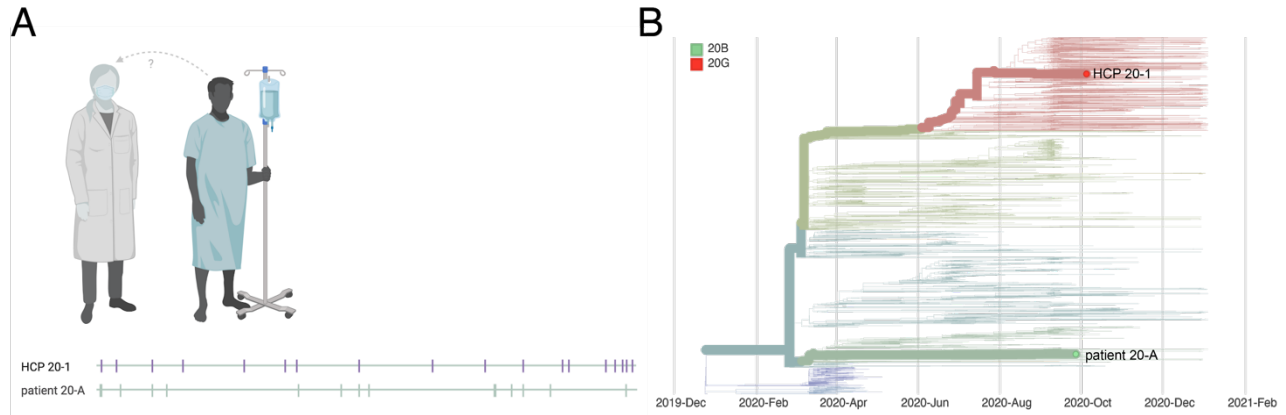
414 References

- 415 1. Sikkema RS, Pas SD, Nieuwenhuijse DF et al. COVID-19 in health-care workers in three
416 hospitals in the south of the Netherlands: a cross-sectional study. *Lancet Infect Dis* 2020;
417 **20**:1273-1280.
- 418 2. Meredith LW, Hamilton WL, Warne B et al. Rapid implementation of SARS-CoV-2
419 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic
420 surveillance study. *Lancet Infect Dis* 2020; **20**:1263-1272.
- 421 3. Deurenberg RH, Bathoorn E, Chlebowicz MA et al. Application of next generation sequencing
422 in clinical microbiology and infection prevention. *J Biotechnol* 2017; **243**:16-24.
- 423 4. Tang P, Croxen MA, Hasan MR, Hsiao WW, Hoang LM. Infection control in the new age of
424 genomic epidemiology. *Am J Infect Control* 2017; **45**:170-179.
- 425 5. Safdar N, Moreno GK, Braun KM, Friedrich TC, O'Connor DH. Using Virus Sequencing to
426 Determine Source of SARS-CoV-2 Transmission for Healthcare Worker. *Emerg Infect Dis* 2020;
427 **26**:2489-2491.
- 428 6. Sikkens JJ, Buis DTP, Peters EJG et al. Serologic Surveillance and Phylogenetic Analysis of
429 SARS-CoV-2 Infection in Hospital Health Care Workers. *medRxiv* 2021; 2021.01.10.21249440.
- 430 7. Nguyen LH, Drew DA, Graham MS et al. Risk of COVID-19 among front-line health-care
431 workers and the general community: a prospective cohort study. *Lancet Public Health* 2020;
432 **5**:e475-e483.
- 433 8. Chou R, Dana T, Buckley DI, Selph S, Fu R, Totten AM. Epidemiology of and Risk Factors for
434 Coronavirus Infection in Health Care Workers: A Living Rapid Review. *Ann Intern Med* 2020;
435 **173**:120-136.
- 436 9. Lai J, Ma S, Wang Y et al. Factors Associated With Mental Health Outcomes Among Health
437 Care Workers Exposed to Coronavirus Disease 2019. *JAMA Netw Open* 2020; **3**:e203976.

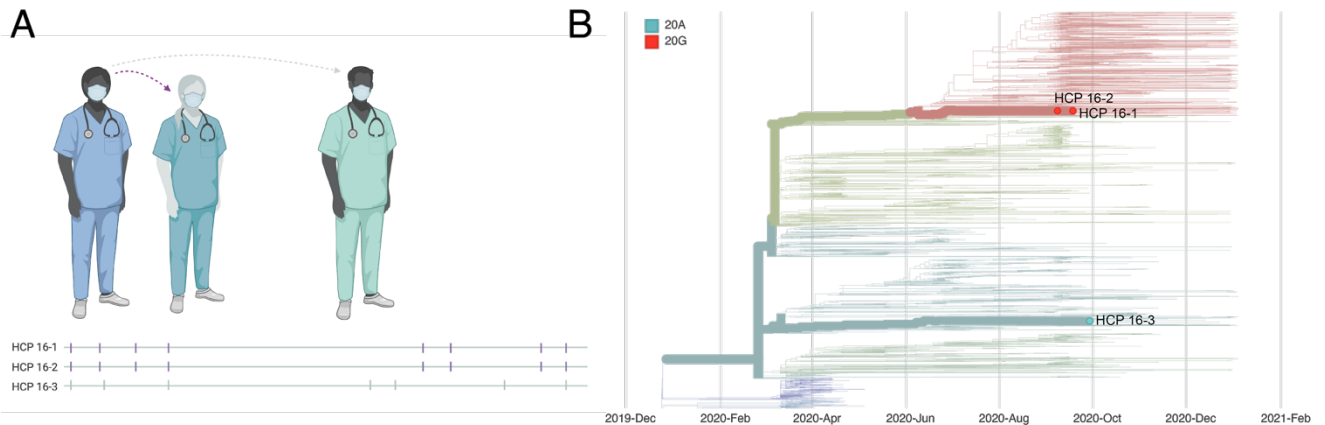
- 438 10. Cheng VC, Wong SC, Yuen KY. Estimating Coronavirus Disease 2019 Infection Risk in
439 Health Care Workers. *JAMA Netw Open* 2020; **3**:e209687.
- 440 11. Arpacioğlu S, Gurler M, Cakiroğlu S. Secondary Traumatization Outcomes and Associated
441 Factors Among the Health Care Workers Exposed to the COVID-19. *Int J Soc Psychiatry* 2020;
442 20764020940742.
- 443 12. CDC. Interim Infection Prevention and Control Recommendations for Healthcare Personnel
444 During the Coronavirus Disease 2019 (COVID-19) Pandemic. 2020;
- 445 13. Lepak AJ, Shirley DK, Buys A, Stevens L, Safdar N. Implementation of infection control
446 measures to prevent healthcare-associated transmission of severe acute respiratory
447 coronavirus virus 2 (SARS-CoV-2). *Infect Control Hosp Epidemiol* 2020; 1-4.
- 448 14. Quainoo S, Coolen JPM, van Hijum SAFT et al. Whole-Genome Sequencing of Bacterial
449 Pathogens: the Future of Nosocomial Outbreak Analysis. *Clin Microbiol Rev* 2017; **30**:1015-
450 1063.
- 451 15. Arnold C. Outbreak Breakthrough: Using Whole-Genome Sequencing to Control Hospital
452 Infection. *Environ Health Perspect* 2015; **123**:A281-6.
- 453 16. Kong LY, Eyre DW, Corbeil J et al. Clostridium difficile: Investigating Transmission Patterns
454 Between Infected and Colonized Patients Using Whole Genome Sequencing. *Clin Infect Dis*
455 2019; **68**:204-209.
- 456 17. Snitkin ES, Zelazny AM, Thomas PJ et al. Tracking a hospital outbreak of carbapenem-
457 resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci Transl Med* 2012;
458 **4**:148ra116.
- 459 18. Houldcroft CJ, Roy S, Morfopoulou S et al. Use of Whole-Genome Sequencing of
460 Adenovirus in Immunocompromised Pediatric Patients to Identify Nosocomial Transmission and
461 Mixed-Genotype Infection. *J Infect Dis* 2018; **218**:1261-1271.

- 462 19. Greninger AL, Zerr DM, Qin X et al. Rapid Metagenomic Next-Generation Sequencing
463 during an Investigation of Hospital-Acquired Human Parainfluenza Virus 3 Infections. *J Clin*
464 *Microbiol* 2017; **55**:177-182.
- 465 20. Houlihan CF, Frampton D, Ferns RB et al. Use of Whole-Genome Sequencing in the
466 Investigation of a Nosocomial Influenza Virus Outbreak. *J Infect Dis* 2018; **218**:1485-1489.
- 467 21. Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. Measurably evolving pathogens in the
468 genomic era. *Trends Ecol Evol* 2015; **30**:306-313.
- 469 22. Rai B, Shukla A, Dwivedi LK. Estimates of serial interval for COVID-19: A systematic review
470 and meta-analysis. *Clin Epidemiol Glob Health* 2021; **9**:157-161.
- 471 23. Bell SM, Hodcroft EB, Müller NF et al. Genomic analysis of COVID-19. Situation report
472 2020-05-15. 2020;
- 473 24. Nextstrain. Genomic epidemiology of novel coronavirus - North America-focused
474 subsampling.
- 475 25. Ladner JT, Grubaugh ND, Pybus OG, Andersen KG. Precision epidemiology for infectious
476 disease control. *Nat Med* 2019; **25**:206-211.
- 477 26. Viner RM, Russell SJ, Croker H et al. School closure and management practices during
478 coronavirus outbreaks including COVID-19: a rapid systematic review. *Lancet Child Adolesc*
479 *Health* 2020; **4**:397-404.
- 480 27. Levinson M, Cevik M, Lipsitch M. Reopening Primary Schools during the Pandemic. *N Engl*
481 *J Med* 2020; **383**:981-985.
- 482 28. Quick J, Grubaugh ND, Pullan ST et al. Multiplex PCR method for MinION and Illumina
483 sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 2017;
484 **12**:1261-1276.
- 485 29. Quick J. nCoV-2019 sequencing protocol. *protocolsio* 2020;
- 486 30. Hadfield J, Megill C, Bell SM et al. Nextstrain: real-time tracking of pathogen evolution.
487 *Bioinformatics* 2018; **34**:4121-4123.

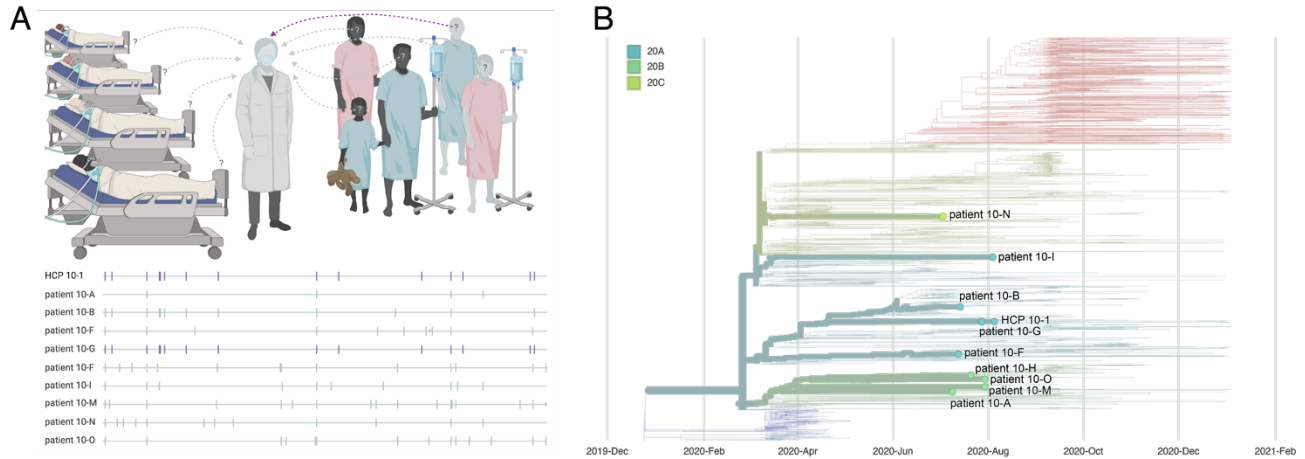
488 **Figures and tables**



489 **Figure 1.** Graphical representation of case #20. A. Virus sequences are aligned against SARS-
490 CoV-2 reference sequence Wuhan-Hu-1 (MN908947.3). Vertical markers denote the location of
491 consensus nucleotide differences between patient viruses and the reference. B. A time-resolved
492 phylogenetic tree built using Nextstrain tools with all Wisconsin sequences available as of 2021-
493 01-15. Viruses involved in this case are denoted with thick branches and labeled tips. Color
494 denotes clade.



495 **Figure 2.** Graphical representation of case #16. A. Virus sequences are aligned against SARS-
496 CoV-2 reference sequence Wuhan-Hu-1 (MN908947.3). Vertical markers denote the location of
497 consensus nucleotide differences between patient viruses and the reference. Purple vertical
498 markers indicate identical virus sequences. B. A time-resolved phylogenetic tree built using
499 Nextstrain tools with all Wisconsin sequences available as of 2021-01-15. Viruses involved in
500 this case are denoted with thick branches and labeled tips. Color denotes clade.



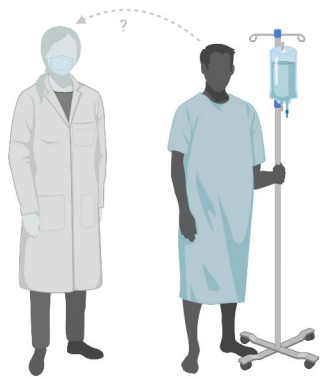
501 **Figure 3.** Graphical representation of case #10. A. Virus sequences are aligned against SARS-
 502 CoV-2 reference sequence Wuhan-Hu-1 (MN908947.3). Vertical markers denote the location of
 503 consensus nucleotide differences between patient viruses and the reference. Purple vertical
 504 markers indicate identical virus sequences. B. A time-resolved phylogenetic tree built using
 505 Nextstrain tools with all Wisconsin sequences available as of 2021-01-15. Viruses involved in
 506 this case are denoted with thick branches and labeled tips. Color denotes clade.

507

508 **Table 1.** Summary of the likely source of infection in the HCP evaluated in this study.

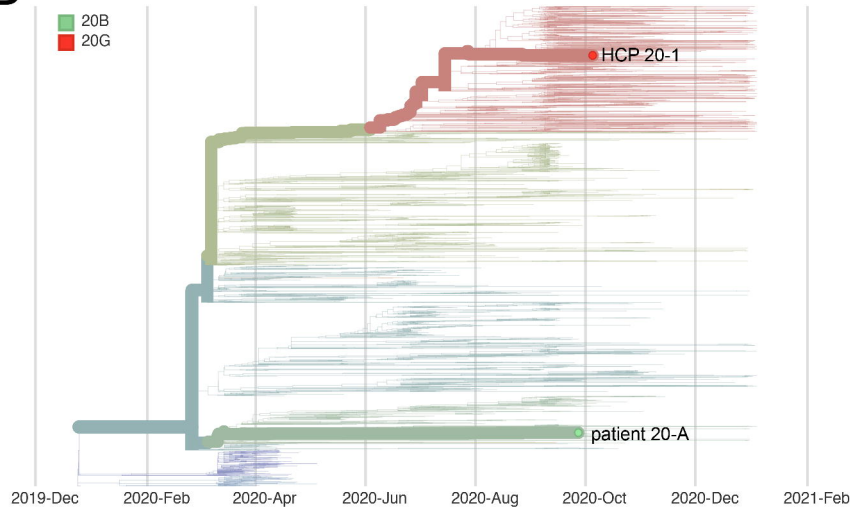
Likely source of infection in HCP	Number of cases
Outside community	58 (60.4%)
Patient source (via employee-patient interactions)	4 (4.2%)
Employee source (via employee-employee interactions)	10 (10.4%)
Combined patient and employee cluster	12 (12.5)
Inconclusive	12 (12.5)
Total	96

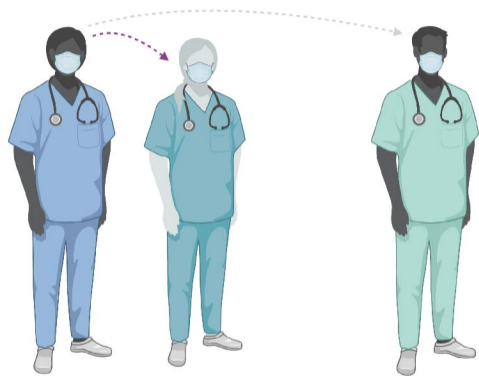
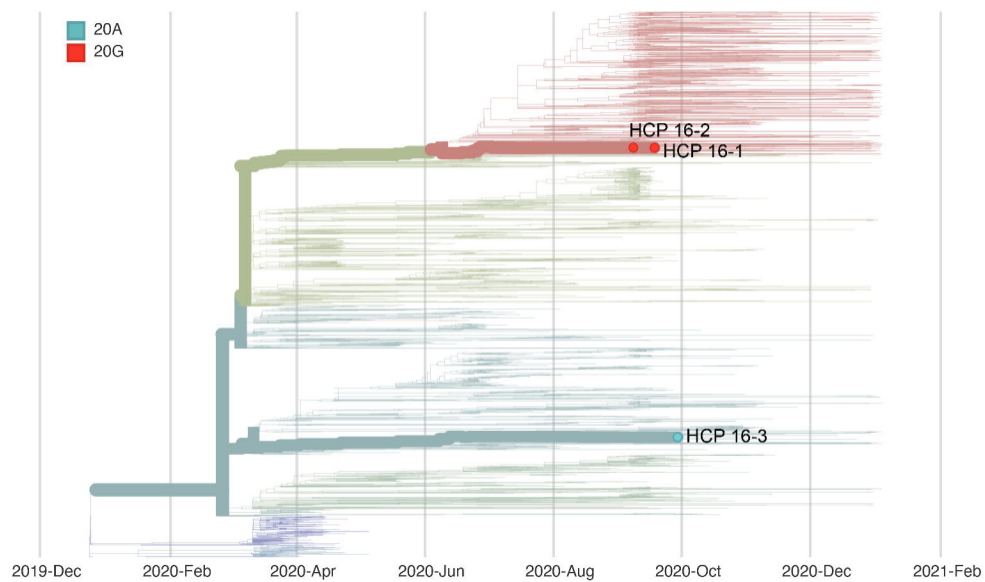
509

A

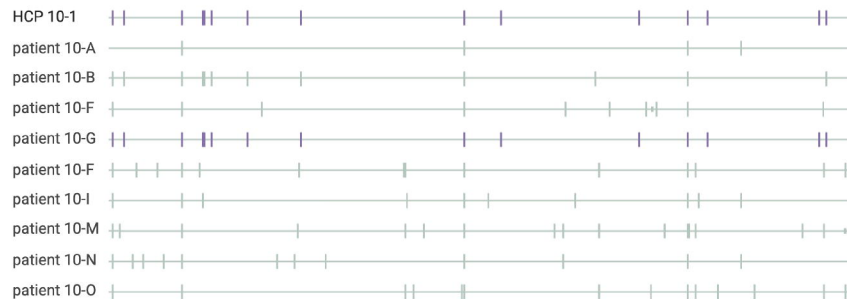
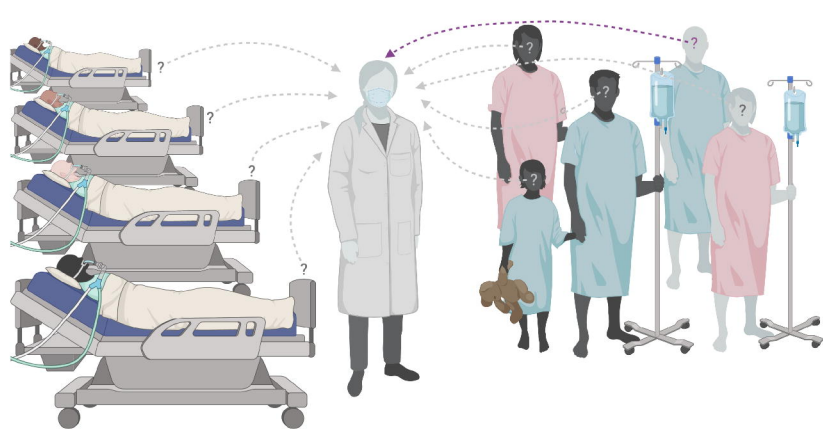
HCP 20-1

patient 20-A

**B**

A**B**

A



B

20A
20B
20C

