

1 **Wastewater surveillance overcomes socio-economic** 2 **limitations of laboratory-based surveillance when** 3 **monitoring disease transmission: the South African** 4 **experience during the COVID-19 pandemic**

5 **Authors**

6 Gillian Maree ^{1*}, Fiona Els ^{1,2,5}, Yashena Naidoo ¹, Laven Naidoo ¹, Phemelo Mahamuza ¹, Mokgaetji
7 Macheke ², Nkosenhle Ndlovu ², Said Rachida ², Chinwe Iwu-Jaja ², Setshaba Taukobong ²,
8 Sibonginkosi Maposa ², Kathleen O'Reilly ³, Mukhlid Yousif ^{2,4,&}, Kerrigan McCarthy ^{2,4,5,&}

9 **Affiliations**

10 ¹ Gauteng City-Region Observatory (GCRO), a Partnership of the University of Johannesburg, the
11 University of the Witwatersrand, the Gauteng Provincial Government and Organised Local Government in
12 Gauteng (SALGA), Johannesburg, South Africa

13 ² Centre for Vaccines and Immunology, National Institute for Communicable Diseases, Johannesburg,
14 Gauteng, South Africa

15 ³ Centre for Mathematical Modelling of Infectious Diseases, Faculty of Epidemiology and Population
16 Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

17 ⁴ Department of Virology, School of Pathology, University of Witwatersrand, Johannesburg, Gauteng,
18 South Africa

19 ⁵ School of Public Health, University of Witwatersrand, Johannesburg, Gauteng, South Africa

20 & These are joint last authors of the paper

21 * Corresponding author

22 Email: gillian.maree@gcro.ac.za (GM)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

23 **Abstract**

24 Wastewater and environmental surveillance has been promoted as a communicable disease
25 surveillance tool because it overcomes inherent biases in laboratory-based communicable disease
26 surveillance. Yet, little empirical evidence exists to support this notion, and it remains largely an
27 intuitive, though highly plausible hypothesis. Our interdisciplinary uses WES data to show evidence
28 for underreporting of SARS-CoV-2 in the context of measurable and statistically significant
29 associations between economic conditions and SARS-CoV-2 incidence and testing rates. We obtained
30 geolocated, anonymised, laboratory-confirmed SARS-CoV-2 cases, wastewater SARS-CoV-2 viral
31 load data and socio-demographic data for Gauteng Province, South Africa. We spatially located all
32 data to create a single dataset for sewershed catchments served by two large wastewater treatment
33 plants. We conducted epidemiological, persons infected and principal component analysis to explore
34 the relationships between variables. Overall, we demonstrate the co-contributory influences of socio-
35 economic indicators on access to SARS-CoV-2 testing and cumulative incidence, thus reflecting that
36 apparent incidence rates mirror access to testing and socioeconomic considerations rather than true
37 disease epidemiology. These analyses demonstrate how WES provides valuable information to
38 contextualise and interpret laboratory-based epidemiological data. Whilst it is useful to have these
39 associations established for SARS-CoV-2, the implications beyond SARS-CoV-2 are legion for two
40 reasons, namely that biases inherent in clinical surveillance are broadly applicable across pathogens
41 and all pathogens infecting humans will find their way into wastewater albeit in varying quantities.
42 WES should be implemented to strengthen surveillance systems, especially where economic
43 inequalities limit interpretability of conventional surveillance data.

44 **Main Body**

45 **Introduction**

46 Surveillance is a core component of the International Health Regulations, and central to the Global Health
47 Security Agenda[1]. World Health Assembly member states are obliged to detect, assess, notify and report
48 events, and to assess their capacity to do so using the Core Capacities document[2]. During the COVID-19
49 pandemic, which led to an estimated 775 million laboratory-confirmed cases to date[3], testing of
50 individual patient clinical material (usually nasopharyngeal swabs by PCR) and indicators based on
51 these data (including testing rate, incidence rate and proportion testing positive) were the major
52 epidemiological tools used to support monitoring of the pandemic and government decision-making.

53

54 SARS-CoV-2 is transmitted by droplet and airborne transmission. Factors predisposing to
55 transmission include crowded conditions, proximity and duration of contact, particularly in the
56 absence of mask-wearing[4]. The clustering of these factors in low socioeconomic households and
57 communities has been shown to exacerbate disease transmission and lead to higher disease
58 incidence[5,6]. However, from the earliest days of the pandemic, it was observed that countries and
59 regions with poorer socioeconomic status reported fewer cases of SARS-CoV-2[7,8]. Most
60 importantly, a reason for this was limited access to testing, evidenced at a global scale by the African
61 region reporting the fewest SARS-CoV-2 cases across the globe alongside the least number of SARS-
62 CoV-2 tests per capita[8].

63

64 South Africa, a country with a population of over 62 million persons resident in nine provinces and 52
65 health districts, experienced five reported waves of SARS-CoV-2[9] each caused by different genetic
66 variants. Following initial detection, the first (ancestral strain) and second waves (Beta variant) of the
67 pandemic occurred between March-June 2020, and from November 2020 to February 2021[10]. The
68 third and fourth waves occurred from May to September 2021 (Delta variant)[10] and from November

69 2021 to January 2022 (Omicron BA.1 variant)[11]. A subsequent resurgence in cases was responsible
70 for a fifth wave dominated by the Omicron lineage BA.4 and BA.5[10]. The impact of the COVID-19
71 pandemic was devastating. Disease transmission levels across South Africa were high, with
72 seroprevalence data after the third wave of SARS-CoV-2, suggesting that infections were likely 7.8
73 times higher than the number of laboratory-confirmed cases[12]. Excess deaths greatly exceeded
74 reported deaths, indicating that official statistics underestimated the death rate[13]. In spite of active
75 case finding[14] and the availability of a comprehensive laboratory network, these data suggested
76 extensive under diagnosis and under-reporting of cases.

77

78 These observations foregrounded the intrinsic shortcomings of traditional approaches to laboratory-
79 based communicable disease surveillance programmes. Patient factors (such as health care
80 acceptability and accessibility, financial means to procure testing and the presence and severity of
81 symptoms); health system factors (such as clinician propensity to test, test availability, financial
82 support for testing), and laboratory factors (clerical errors, inherent test performance characteristics)
83 mediate laboratory testing and in turn impact the sensitivity, quality and representivity of surveillance
84 data. Traditional case-based surveillance methods thus underestimate disease burden.

85

86 Since the COVID-19 pandemic, wastewater and environmental surveillance (WES) has increasingly
87 been implemented as a complementary surveillance modality that has potential to overcome these
88 limitations[15]. WES has proven utility in supporting polio surveillance by providing highly sensitive
89 data on the presence of poliovirus in communities, material for genomic sequencing and by
90 supporting identification of chains of transmission[16]. These data have greatly enabled polio risk
91 assessments and decision-making regarding public health interventions including the need for
92 vaccination campaigns. During the COVID-19 pandemic, multiple advantages of WES for SARS-
93 CoV-2 became apparent. WES provided first evidence of importation of the virus into new
94 geographical regions[17], heralded the onset of new waves of infection, illustrated disease
95 transmission patterns and allowed inference of relative population burden during endemic phases[18].
96 WES also provided material for genotyping that demonstrated the presence of a broader range of

97 variants than seen clinically[19]. In light of this global experience, the WHO issued updated guidance
98 for countries conducting WES (WHO guidelines), citing the ability of WES to overcome the
99 limitations of laboratory-based surveillance. However, despite these advantages and
100 recommendations, there is a paucity of evidence integrating socioeconomic factors and disease
101 epidemiology based on clinical testing data with wastewater surveillance in order to substantiate the
102 claims that WES overcomes clinical testing limitations. Thus, definitive evidence for the ability of
103 WES to overcome the limitations of laboratory-based surveillance is urgently required. We used
104 clinical, wastewater surveillance, demographic and socioeconomic data in two different socio-
105 economic contexts to explain and quantify the relationship between these variables and burden of
106 SARS-CoV-2 disease, thus substantiating the use of WES as a necessary surveillance tool that enables
107 interpretation of clinical surveillance data.

108 **Methods**

109 **Conceptual framework**

110 We developed a conceptual framework (Fig. 1) to demonstrate the relationships between wastewater
111 concentrations of SARS-CoV-2 and SARS-CoV-2 disease indicators in sewered communities. The
112 population burden of SARS-CoV-2 infections determines the levels of SARS-CoV-2 in wastewater.
113 However, clinical indicators (including incidence rate, testing rate and proportion testing positive)
114 reflect the distribution of testing, health care availability and accessibility. As these factors may be
115 influenced by socio-economic conditions, reported case rates may not accurately reflect the true
116 burden of infections. By determining the inter-relationships between wastewater surveillance data,
117 social determinants of health and reported SARS-CoV-2 cases, the role of WES may be better
118 understood.

119

120 [INSERT FIG. 1]

121 **Fig. 1. Diagram indicating the direction of influence of socio-economic status, population**
122 **structure and mixing on the true burden of SARS-CoV-2 cases, reported burden of SARS-CoV-**
123 **2 cases and levels of SARS-CoV-2 in wastewater. The green arrow indicates the research**
124 **question posed by this work**

125 **Study setting**

126 The study took place in two sewersheds (sewershed D and sewershed O) in different metropolitan
127 areas of Gauteng Province, South Africa. These sewersheds were purposively selected on the basis of
128 geographical representativeness of city populations, namely sewershed D (City of Tshwane), to the
129 north of the Gauteng Province, and sewershed O (City of Ekurhuleni) in the east of Gauteng Province
130 (Fig. 2). Sewershed D is residential with formal housing in the west and central areas, and the
131 metropolitan central business district to the east. A small area of informal housing exists to the far
132 west. Sewershed O is mostly residential with low density, low-rise housing with areas of industrial
133 and manufacturing activity to the north east. Informal settlements and backyard shacks are present in
134 most neighbourhoods in the central areas, whilst a wealthier community of gated estates with a low
135 population density is present in the northernmost section.

136

137 [INSERT FIG. 2]

138 **Fig. 2. Map showing the spatial location of the two sewersheds and locations of health care**
139 **facilities within the Gauteng Province**

140 **Data sources and study period**

141 **Clinical SARS-CoV-2 laboratory testing and data management.**

142 In South Africa public and private laboratories were legally mandated to report all SARS-CoV-2 test
143 results and patient data including residential address via the National Institute for Communicable
144 Diseases (NICD). These data had been geocoded by NICD as part of outbreak response activities and
145 all cases geocoded to Gauteng were extracted. Following ethics review and written approval for this

146 study , anonymised and deidentified geocoded data for positive and negative SARS-CoV-2 PCR test
147 results were supplied for the study period (1 June 2021 (epidemiological week 22 of 2021) to 18
148 March 2022 (epidemiological week 11 of 2022)) and the data was obtained on 23 May 2023.

149

150 **Wastewater SARS-CoV-2 laboratory testing and data management.**

151 Routine SARS-CoV-2 wastewater surveillance data (obtained from laboratory processing of one litre
152 grab-samples collected weekly as previously described[9]) from wastewater treatment plant (WWTP)
153 ‘D’ and ‘O’ were identified for the study period. We obtained wastewater flow rates (in ML per day)
154 from WWTP managers.

155

156 **Quality of life survey- data**

157 The Quality of Life (QoL) survey is a biennial household survey produced by the Gauteng City-
158 Region Observatory (GCRO) which covers a range of topics including demographics, access to
159 services and perceptions of residents. The survey is weighted to ensure representativity to ward level,
160 a geopolitical division of municipalities, developed by the Municipal Demarcation Board and the
161 smallest unit for which demographic and socioeconomic data are provided in South Africa. We
162 obtained data from the 2020/21 (QoL6) survey (n=13,616 respondents) pertaining to 207 and 781
163 respondents from sewersheds D and O respectively[20].

164

165 **Maps of sewershed reticulation methods**

166 Data for the sewage networks and pipelines were obtained from the City of Tshwane and the City of
167 Ekurhuleni, respectively. The City of Tshwane provided a comprehensive dataset including sewershed
168 areas, manholes, and distribution pipelines. The City of Ekurhuleni consulting engineers provided
169 shapefiles of the sewershed. ArcGIS 10.6.2 software was used for all mapping and spatial analysis.

170

171 **GTI hexagon data for population demographics**

172 Population demographic variables were extracted from a 2020/2021 dataset compiled by
173 GeoTerraImage (GTI) comprising population estimates grouped by age cohorts in 400m sided
174 hexagon (0.103755 km²).

175 **Data synthesis and analysis**

176 **General approach**

177 We overlaid sewershed shape files, hexagonal population data, geolocated SARS-CoV-2 cases and
178 geolocated QoL respondents in order to create a dataset comprising these elements for sewersheds D
179 and O. To minimise the effect of the Modifiable Area Unit Problem (MAUP) that is encountered
180 when polygon values require changing because of a change in the shape (zoning effect) or overall area
181 (size effect) of the polygon, we scrutinised our datasets to determine the most viable polygon layer for
182 all spatial datasets to allow for aggregation and comparability without disaggregation of polygon data.
183 Ultimately we manipulated all data to the ward level allowing for comparative analysis over space and
184 time. We conducted these analyses using SPSS software version 29.0.2.0 (20).

185

186 **Data extraction and aggregation to determine population size, socio- 187 economic and epidemiological parameters by sewershed.**

188 To determine population size and age structure within wards and sewersheds, we extracted population
189 data from the GeoTerraImage (GTI) dataset for hexagons whose centroids fell within the sewershed
190 and ward boundaries, and aggregated population data into four broad classes, namely children (0-4
191 years), adolescents (5-19), adults (20-59) and elderly (≥ 60 years). We geolocated QoL respondents
192 within ward and sewershed boundaries and extracted and aggregated relevant socioeconomic and
193 health fields for these individuals. Geolocated positive and negative SARS-CoV-2 PCR test data were
194 aggregated by ward and sewershed level, and used together with population denominator data for each
195 spatial unit to determine overall and weekly incidence, positivity rate and testing rate per 100,000. We

196 created epidemiologic curves using Microsoft Excel 365 using incidence, positivity and testing rates
197 together with SARS-CoV-2 concentration in wastewater for each catchment[21].

198

199 **Principal Component Analysis (PCA)**

200 We conducted principal component analysis (PCA) on the socioeconomic, demographic and clinical
201 variables for each sewershed, using ward-level aggregated data as the unit of analysis to ascertain
202 which variables relate most to epidemiological indicators in each sewershed. Using R (RStudio
203 v4.0.2) and for each sewershed, we generated 1) a correlation coefficient (r) matrix for inter-variable
204 correlation analysis; 2) a Scree plot for ascertaining the dimensions contributing the most to the
205 explained variance as a percentage; 3) a variable loading graph for ascertaining the variables which
206 contribute to the identity of the prominent dimensions; 4) and a biplot for plotting the variables as
207 vectors in 2D space against the two most prominent dimensions.

208

209 **Determination of theoretical infectious case load using mass balance**

210 **equations and comparison with laboratory-confirmed cases by**

211 **epidemiological week**

212 We estimated the theoretical number of persons infected in each sewershed by epidemiological week
213 using mass balance equations from Acheampeong *et al*[22], and Hoffman *et al*[23] as follows:

214

$$215 \text{ Persons infected} = A \times B / C \times D^{23}$$

216

217 where A= RNA per L of wastewater (natural scale); B= estimated flow (L/day) obtained from the
218 wastewater treatment authorities; C= grams of faeces/ person-day, estimated at 128g[24]; D= SARS-
219 CoV-2 RNA per gram faeces, estimated at 2.58×10^8 gene copies shed per day per infected
220 person[23]. We assumed a linear regression to determine the number of persons infected per 100
221 laboratory-confirmed cases by epidemiological week within each sewershed.

222 **Ethical statement**

223 This study was reviewed and approved by the University of the Witwatersrand Human Research
224 Ethics Committee (HREC), M220904. In addition, the National Institute for Communicable Diseases
225 obtained Ethics Approval for essential communicable disease surveillance and outbreak and response
226 activities including SARS-CoV-2 (M210752).

227 **Role of funders**

228 The funders had no role in the study design, collection, analysis and interpretation of data, manuscript
229 writing or journal selection.

230 **Results**

231 **Demographic, socioeconomic and health characteristics**

232 Sewershed O (land area =120 km², population=905,996 persons) has around four times the number of
233 people and just under double the population density of sewershed D (Table 1). Both sewersheds are
234 dominated by younger, working age cohorts, with lower proportions of persons over 60 years of age.
235 The proportion of households earning <USD90, together with the proportion relying on public
236 transport, suggest that households in sewershed O are poorer. As many as 13% of households in
237 sewershed O share sanitation, compared with around 2% in sewershed D. Regarding health care,
238 responses suggested that up to 10% of households in sewershed O vs 4% in sewershed D struggled to
239 access health care during 2020-2021 period. Up to 35% of households in Sewershed O were unable to
240 maintain SARS-CoV-2 non-pharmaceutical measures.

241

242

243 **Table 1: Comparative table of key SARS-COV-2, wastewater surveillance, socioeconomic and**
244 **demographic variables for sewersheds D and O during the study period 1 June 2021**
245 **(epidemiological week 22 of 2021) to 18 March 2022 (epidemiological week 11 of 2022).**
246 **Variable definitions are included.**

	Sewershed D	Sewershed O
Sewershed area (km²)	62·0	120·7
Total population	245,935	905,996
Population density (persons/km²)	3·97	7·51
Number of wards for analysis	11	26
Demographic data (number of people, %)		
0-4 years	21,196 (8·6)	82,230 (9·1)
5-19 years	55,965 (22·8)	204,225 (22·5)
20-59 years	150,798 (61·3)	563,428 (62·2)
≥60 years	17,976 (7·3)	56,111 (6·2)
Wastewater surveillance		
Mean daily flow rate of wastewater through the treatment plant (ML, standard deviation)	38·4 (9·7)	106·2 (9·9)
Median (interquartile range) of SARS-CoV-2 (genome copies/mL)	47·2 (100·5)	46·9 (109·7)
SARS-CoV-2 laboratory surveillance data and indicators		
Testing rate (tests/100 000 population)	23·7	10·6
Total recorded SARS-COV-2 cases during study period	11,026	15,293
Mean positivity rate (% , sd)	17 (12·8)	14 (9·9)
Cumulative incidence rate (cases /100 000 population)	4,483·3	1,688·0
Socio-economic and health indicators from Quality of Life Survey 6 (2020/2021)		
Income below R1600 (sd) <i>Percentage of households that have a combined income less than R1600 per month</i>	21%(4·6)	33%(4·6)
Reliance on public transport (sd) <i>Percentage of respondents that rely on public transport</i>	40%(0·5)	47%(0·5)
Access to shared sanitation (sd) <i>Percentage of respondents who generally use communal toilets or toilets not connected to the sewage system (e.g. portable toilets)</i>	2%(0·2)	13%(0·3)
Quality of Life score (sd) <i>Score out of 100. Calculated with 33 variables, grouped in 7 dimensions</i>	63(9·0)	59(11·0)

Struggled access to health care (sd) <i>Percentage of respondents that struggled to access healthcare from March 2020 to October 2021</i>	4%(0·2)	10%(0·3)
Refused COVID-19 testing (sd) <i>Proportion of respondents who tried but were denied access a COVID-19 test between March 2020 and October 2021</i>	5%(0·2)	3%(0·2)
COVID-19 Index (sd) <i>Percentage respondents unable to implement SARS-CoV-2 preventative measures including social distancing</i>	29(19·0)	35(20·0)

247 (sd)= standard deviation, km=kilometre

248 **SARS-CoV-2 clinical testing, incidence rates and percentage test**
 249 **positive (PTP)**

250 During the study period (1 June 2021 to 18 March 2022), 78% and 60% of positive and negative test
 251 results respectively were successfully geolocated. The SARS-CoV-2 clinical testing rate per 100,000
 252 in sewershed D (23.7) was over twice the rate in sewershed O (10.6). Despite higher absolute numbers
 253 of SARS-CoV-2 cases, sewershed O had a cumulative incidence of laboratory-confirmed SARS-CoV-
 254 2 cases around three times lower than sewershed D (15,293 total cases and 1,688/100,000, vs 11,026
 255 total cases and 4,483/100,000) (Table 1 and Fig. 3).

256 Fig. 3 illustrates changes in testing, incidence and proportion testing positive (PTP) by
 257 epidemiological week during the two waves of infection that occurred during the study period, namely
 258 Delta (between epidemiological weeks 15 and 40 in 2021) and Omicron (epidemiological weeks 46 -
 259 51 in 2021). Whilst all indicators follow the similar trends, rates are lower in sewershed O.

260

261 [INSERT FIG. 3]

262 **Fig. 3. SARS-CoV-2 testing rate (per 100 000 population), incidence rate (per 100 000**
 263 **population), and 4 week moving average proportion test positive (%) by epidemiological week**
 264 **22, 2021 to week 10, 2022, for (a) sewershed D and (b) sewershed O.**

265 Relationship between clinical case data and socioeconomic status

266 Correlation matrices for sewersheds D and O reveal that clinical indicators (PTP, cumulative
267 incidence and testing rates) do not exhibit consistent relationships with demographic, or socio-
268 economic indicators within and between sewersheds (Fig. 4). In sewershed O, *PTP* correlates with
269 indicators associated with poverty (*access to and use of shared sanitation, % of the population with*
270 *income below USD 90*, and *COVID-19 index* (% unable to implement COVID-19 preventive
271 measures)), whilst this relationship is not evident in the more affluent sewershed D. In sewershed D, it
272 appears that PTP and cumulative incidence are correlated with *quality of life, use of public transport*
273 and *proportion of persons with income below USD 90*, suggesting that economic conditions influence
274 access to testing, in turn leading to apparently low cumulative incidence. Taken together, in areas with
275 poorer socio-economic conditions, testing rates and cumulative incidence may not reflect disease
276 burden, whilst PTP better reflects disease risk.

277

278 [INSERT FIG. 4]

279 **Fig.4. Truncated correlation matrices between socio-economic and demographic parameters**
280 **against testing, cumulative incidence and mean proportion test positive for sewersheds D and O,**
281 **annotated as a ‘heat map’ to represents the Spearman’s correlation coefficient (r) between**
282 **socio-economic, demographic variables vs clinical variables (1 June 2021 to 18 March 2022).**

283

284 Scree plots for sewershed D and O demonstrate that 76.7% and 77.5% of data are explained by
285 dimensions 1 and 2 (Fig. S1). The variable loading plot for sewershed D indicates that *access to*
286 *shared sanitation, reliance on non-sewered toilets* and *population groups under 60 years of age*
287 contributed the most to composition of dimensions 1 and 2 (Fig. S2a) whereas in sewershed O,
288 *population group under 60 years of age, testing rate* and *cumulative incidence rate* contributed the
289 most to the composition of dimension 1 and 2 (Fig. S2b). In both sewersheds, testing rate and
290 cumulative incidence contributed to dimensions 1 and 2, but the *proportion test positive* and *refused*
291 *COVID testing* variables contributed minimally.

292

293 The biplots (Fig. 5) demonstrate that SARS-COV-2 testing and incidence rates contribute equivalent
294 influence and spatial distribution on socioeconomic factors for both sewersheds, whilst the *proportion*
295 *test positive* indicator was much less affected by socioeconomic factors. In the biplot for sewershed D,
296 *access to healthcare* and *refused access to COVID testing* are more associated with cumulative
297 incidence and testing rates, while only *refused access to COVID test* is associated with the two
298 clinical variables in sewershed O. In sewershed D, the clustering of the *Quality of Life score* and the
299 *greater than 60 years* age group as well as the clustering of *COVID-19 Index* with *income below*
300 *USD90* indicated these variables are influenced by and can inform one another. In sewershed O,
301 *struggled to access health care* and *Quality of Life score* illustrated similar clustering with the
302 *COVID-19 Index* and *access to public transport*.

303

304 [INSERT Fig. 5]

305 **Fig. 5. PCA biplots displaying socioeconomic and demographic status parameters, cumulative**
306 **incidence rate, testing rate and mean positivity rate within (a) sewershed D and (b) sewershed**
307 **O. In the biplots, the magnitude and colouring of the vectors are related to the variable loading**
308 **scores, while the vector direction and quadrant location is informed by the interrelationship**
309 **between variables and their contribution to dimensions 1 and 2.**

310 **Quantitative SARS-CoV-2 surveillance in wastewater**

311 Wastewater concentrations of SARS-CoV-2 and geolocated laboratory-confirmed SARS-CoV-2 cases
312 in sewersheds D and O were significantly correlated in each sewershed, though less so in sewershed O
313 (Spearman's correlation coefficient 0.723 ($p < 0.001$)), 0.476 ($p = 0.020$) for sewersheds D and O
314 respectively). In each sewershed, higher concentrations and case-loads were observable during the
315 Delta and Omicron waves that occurred during the study period (Fig. 6) and both reached an ebb
316 during weeks 40-42 of 2021. Wastewater concentrations (measured in log genome copies per
317 millilitre) for both sewersheds ranged between 0.5 and 3.5 log copies/mL and were at similar

318 concentrations in the same epidemiological week. Weekly laboratory-confirmed case-counts followed
319 similar trends despite different population sizes in the two catchments.

320

321 [INSERT FIG. 6]

322 **Fig.6. SARS-CoV-2 concentrations in wastewater in log-transformed genome copies per**
323 **millilitre (right axis) and the number of laboratory-confirmed cases (top figures, green bars) or**
324 **incidence per 100,000 persons (bottom figures, blue bars) of SARS-CoV-2 geolocated to a**
325 **residential address in the sewershed by epidemiological week from week 22, 2021 to week 10,**
326 **2022 for sewersheds D (left, figures a and c respectively) and O (right, figures b and d**
327 **respectively).**

328 **Comparison of estimated and actual SARS-CoV-2 case burden.**

329 Regression analysis by sewershed of laboratory-confirmed SARS-CoV-2 cases versus theoretical
330 number of infections indicated that for each 100 reported cases, sewershed O likely had over 63,000
331 infections compared with sewershed D, with 2,700 cases (Fig. 7).

332

333 [INSERT Fig. 7]

334 **Fig.7. SARS-CoV-2 concentrations in wastewater in log-transformed genome copies per**
335 **millilitre (right axis) and the number of laboratory-confirmed cases (top figures, green bars) or**
336 **incidence per 100,000 persons (bottom figures, blue bars) of SARS-CoV-2 geolocated to a**
337 **residential address in the sewershed by epidemiological week from week 22, 2021 to week 10,**
338 **2022 for sewersheds D (left, figures a and c respectively) and O (right, figures b and d**
339 **respectively).**

340 **Discussion**

341 In our transdisciplinary spatial analysis of clinical and environmental data during two large COVID-
342 19 pandemic waves in sewersheds with differing socioeconomic conditions, we observed that despite

343 different population sizes, the concentrations of SARS-CoV-2 in wastewater and the absolute
344 numbers of SARS-CoV-2 cases by epidemiological week were similar. In the light of assumed
345 equivalent excretion rates of SARS-CoV-2 in infected individuals and identical demographic profiles
346 in each sewershed, equivalent wastewater concentrations suggest vast under-reporting of cases in the
347 poorer sewershed. Our socioeconomic analysis demonstrated negative correlations between income
348 and SARS-CoV-2 cumulative incidence and testing rates in the poorer sewershed. Overall, we
349 demonstrated the co-contributory influences of socio-economic indicators on access to SARS-CoV-2
350 testing and cumulative incidence, thus reflecting that apparent incidence rates mirror access to testing
351 and socioeconomic considerations rather than true disease epidemiology. These analyses demonstrate
352 how WES provides valuable information to contextualise and interpret laboratory-based
353 epidemiological data.

354

355 Laboratory-based surveillance systems under-represent the true burden of disease due to a
356 combination of asymptomatic infection, individual and cultural practices regarding health seeking,
357 quality of health care and socio-economic factors that impair access to testing. Furthermore, these
358 same socio-economic factors including education and poverty, are associated with higher SARS-CoV-
359 2 rates[6]. Underreporting of SARS-CoV-2 cases in South Africa was evident through excess
360 mortality reports which indicated over 70,000 excess deaths vs 28,000 reported COVID-19 deaths
361 during 2020[13]. Households in lower income groups, those who rely on public health care, and Black
362 African and Coloured population groups were more likely to have struggled to access healthcare and
363 testing facilities[25]. Our data, demonstrating that cumulative incidence and testing rate in sewershed
364 O was negatively correlated with low income (< USD 90 household income per month) suggest that
365 clinical testing was missing this population segment.

366

367 Few studies have triangulated clinical testing, wastewater surveillance data and socio-economic
368 factors. Using WES data in contrast to clinical testing data, Lancaster *et al*[26] identified specific
369 communities (Black, poor) that were more vulnerable to SARS-CoV-2. Saingam *et al*[27]
370 demonstrated that a machine learning model to predict COVID-19 and post-infectious sequelae is

371 strengthened when including WES data together with socio-demographic parameters. Rogawski
372 McQuade *et al*[28] in Bangladesh, using purposively selected sampling sites across the economic
373 spectrum, demonstrated equivalent SARS-CoV-2 levels in wastewater, but 23 and 70 times the
374 number of clinical tests conducted in the wealthier vs middle and poor income areas. Our
375 interdisciplinary study is the first to use WES data to show evidence for underreporting of SARS-
376 CoV-2 in the context of measurable and statistically significant associations between economic
377 conditions and SARS-CoV-2 incidence and testing rates, something which has been intuitively
378 speculated but not empirically demonstrated[29].

379

380 Several limitations exist in our data collection and interpretation. A small proportion of SARS-CoV-2
381 cases were not successfully geolocated, however non-geolocated cases are more likely to have
382 originated outside the province, or from persons resident in informal settings thus not contributing to
383 wastewater levels of SARS-CoV-2. The absence of wastewater sampling points within wards
384 precluded inclusion of SARS-CoV-2 levels in PCA analyses. Whilst our wastewater data was
385 generated from urban sewersheds, similar findings in non-sewered areas suggest that our findings are
386 generalisable across sewerred and non-sewerred settings[28].

387

388 The addition of WES data to national and global surveillance systems will strengthen sensitivity of
389 event detection for outbreak and pandemic disease, monitoring of endemic disease trends, and will
390 jointly provide material for genomic epidemiology[19]. An evaluation of detection of H5N1 avian
391 influenza by six northern hemisphere surveillance systems in 2010 demonstrated an increase in
392 sensitivity of detection using a combination of data inputs. Authors concluded that the range of
393 surveillance methodologies and variation in system designs created synergy between systems, led to
394 improved data quality and validity, and allowed data to converge on event detection[30]. In a low-
395 middle income country, the need for multiple surveillance systems is even more necessary, as data
396 quality and completeness from single modality surveillance systems may vary, leading to challenges
397 in decision making during a crisis.

398

399 Our findings provide evidence to support intuitive thinking that WES overcomes testing biases
400 particularly in situations with socio-economic disparities and weaker clinical disease surveillance
401 programmes for SARS-CoV-2. Our findings are likely broadly applicable to all communicable disease
402 surveillance programmes, as biases affecting clinical surveillance programmes are not disease-
403 specific, and all pathogens infecting humans are likely to find their way into wastewater albeit in
404 varying quantities. As such, our findings strengthen the case for investment in implementation of
405 WES. Ongoing implementation of WES will allow public health authorities to determine optimal
406 configurations of WES surveillance systems for each pathogen and public health use-case. Further
407 research to determine optimal sample collection, processing and testing methods is needed.

408 Interpretive frameworks or mathematical models will support integration and interpretation of WES
409 data with clinical surveillance data.

410

411 **Data sharing**

412 Clinical and laboratory data were generated at the National Institute for Communicable Diseases. The
413 Quality of Life survey dataset is freely available under the CC BY 4.0 licence from the DataFirst
414 service. For more information please email info@gcro.ac.za. Derived data supporting the findings of
415 this study are available from the corresponding author GM on request.

416 **Acknowledgements**

417 The authors would like to thank the following groups – namely the NICD IT and Surveillance Data
418 Warehouse staff for provision of geolocated SARS-CoV-2 case data; the NICD Epidemiology and
419 Surveillance teams that curated SARS-CoV-2 databases during the pandemic; and the City of
420 Tshwane and City of Ekurhuleni water and sanitation departments for provision of sewershed maps.

421

422 **Author contributions**

423 GM, FE, YN, LN, SR, MY and KM were involved in the conception design of the study. YN, LN,
424 GM and FE curated the data. GM, FE, YN, PM, KO and LN completed the formal analysis. PM and
425 FE completed data visualisation. GM wrote the original draft. MM, NN, SR and MY were involved in
426 laboratory testing. SM did the project administration. MY and KM obtained funding for this study.
427 All authors critically reviewed the manuscript and approved the final version.

428 **Declaration of interests**

429 KO has received funding from the Bill and Melinda Gates Foundation, the World Health Organisation
430 and the Department of Health and Social Care (UK) for research unrelated to this work.

431 **Role of funding source**

432 We acknowledge the Water Research Commission (2020.2021 - 00669), the Bill and Melinda Gates
433 foundation (INV050051 and INV049314) for funding this research. The National Institute for
434 Communicable Diseases, and the Gauteng City Regional Observatory contributed in-kind with staff
435 complements.
436

437

438 **References**

- 439 1. Wolicki SB, Nuzzo JB, Blazes DL, Pitts DL, Iskander JK, Tappero JW. Public Health
440 Surveillance: At the Core of the Global Health Security Agenda. *Health Secur.* 2016;14: 185–
441 188. doi:10.1089/hs.2016.0002
- 442 2. World Health Organisation. Checklist and indicators for monitoring progress in the development
443 of IHR core capacities in states parties, April 2013. UHC Healthier Populations (HEP); 2013.
444 Report No.: WHO/HSE/GCR/2013.2. Available: [https://www.who.int/publications-detail-
445 redirect/who-hse-gcr-2013-2](https://www.who.int/publications-detail-redirect/who-hse-gcr-2013-2)
- 446 3. World Health Organisation. WHO Coronavirus (COVID-19) dashboard > Deaths [Dashboard].
447 2023. Available: <https://data.who.int/dashboards/covid19/cases>
- 448 4. Yazdani S, Heydari M, Foroughi Z, Jabali H. Factors Affecting COVID-19 Transmission and
449 Modelling of Close Contact Tracing Strategies. *Iran J Public Health.* 2021;50: 2121–2131.
450 doi:10.18502/ijph.v50i10.7516
- 451 5. Karmakar M, Lantz PM, Tipirneni R. Association of Social and Demographic Factors With
452 COVID-19 Incidence and Death Rates in the US. *JAMA Netw Open.* 2021;4: e2036462.
453 doi:10.1001/jamanetworkopen.2020.36462
- 454 6. Figueiredo AM de, Figueiredo DCMM de, Gomes LB, Massuda A, Gil-García E, Vianna RP de
455 T, et al. Social determinants of health and COVID-19 infection in Brazil: an analysis of the
456 pandemic. *Rev Bras Enferm.* 2020;73: e20200673. doi:10.1590/0034-7167-2020-0673
- 457 7. Bwire G, Ario AR, Eyu P, Ocom F, Wamala JF, Kusi KA, et al. The COVID-19 pandemic in
458 the African continent. *BMC Med.* 2022;20: 167. doi:10.1186/s12916-022-02367-4

- 459 8. Wamai RG, Hirsch JL, Van Damme W, Alnwick D, Bailey RC, Hodgins S, et al. What Could
460 Explain the Lower COVID-19 Burden in Africa despite Considerable Circulation of the SARS-
461 CoV-2 Virus? *Int J Environ Res Public Health*. 2021;18: 8638. doi:10.3390/ijerph18168638
- 462 9. National Institute for Communicable Diseases. SARS-COV-2 Sequencing Update. Network for
463 Genomic Sequencing. 2023. Available: [https://www.nicd.ac.za/diseases-a-z-index/disease-](https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/sars-cov-2-genomic-surveillance-update/)
464 [index-covid-19/sars-cov-2-genomic-surveillance-update/](https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/sars-cov-2-genomic-surveillance-update/)
- 465 10. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of
466 a SARS-CoV-2 variant of concern in South Africa. *Nature*. 2021;592: 438–443.
467 doi:10.1038/s41586-021-03402-9
- 468 11. Karim SSA, Karim QA. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19
469 pandemic. *Lancet Lond Engl*. 2021;398: 2126–2128. doi:10.1016/S0140-6736(21)02758-6
- 470 12. Mutevedzi PC, Kawonga M, Kwatra G, Moultrie A, Baillie V, Mabena N, et al. Estimated
471 SARS-CoV-2 infection rate and fatality risk in Gauteng Province, South Africa: a population-
472 based seroepidemiological survey. *Int J Epidemiol*. 2022;51: 404–417. doi:10.1093/ije/dyab217
- 473 13. Bradshaw D, Dorrington R, Laubscher R, Groenewald P, Moultrie T. COVID-19 and all-cause
474 mortality in South Africa – the hidden deaths in the first four waves. *South Afr J Sci*. 2022;118.
475 doi:10.17159/sajs.2022/13300
- 476 14. Modisenyane M, Madikezela L, Mngemane S, Ramadan OP, Matlala M, McCarthy K, et al.
477 COVID-19 response in South African communities: Screening, testing, tracing and movement
478 modelling. *S Afr Med J*. 2022;112: 366.
- 479 15. Clark EC, Neumann S, Hopkins S, Kostopoulos A, Hagerman L, Dobbins M. Changes to Public
480 Health Surveillance Methods Due to the COVID-19 Pandemic: Scoping Review. *JMIR Public*
481 *Health Surveill*. 2024;10: e49185. doi:10.2196/49185

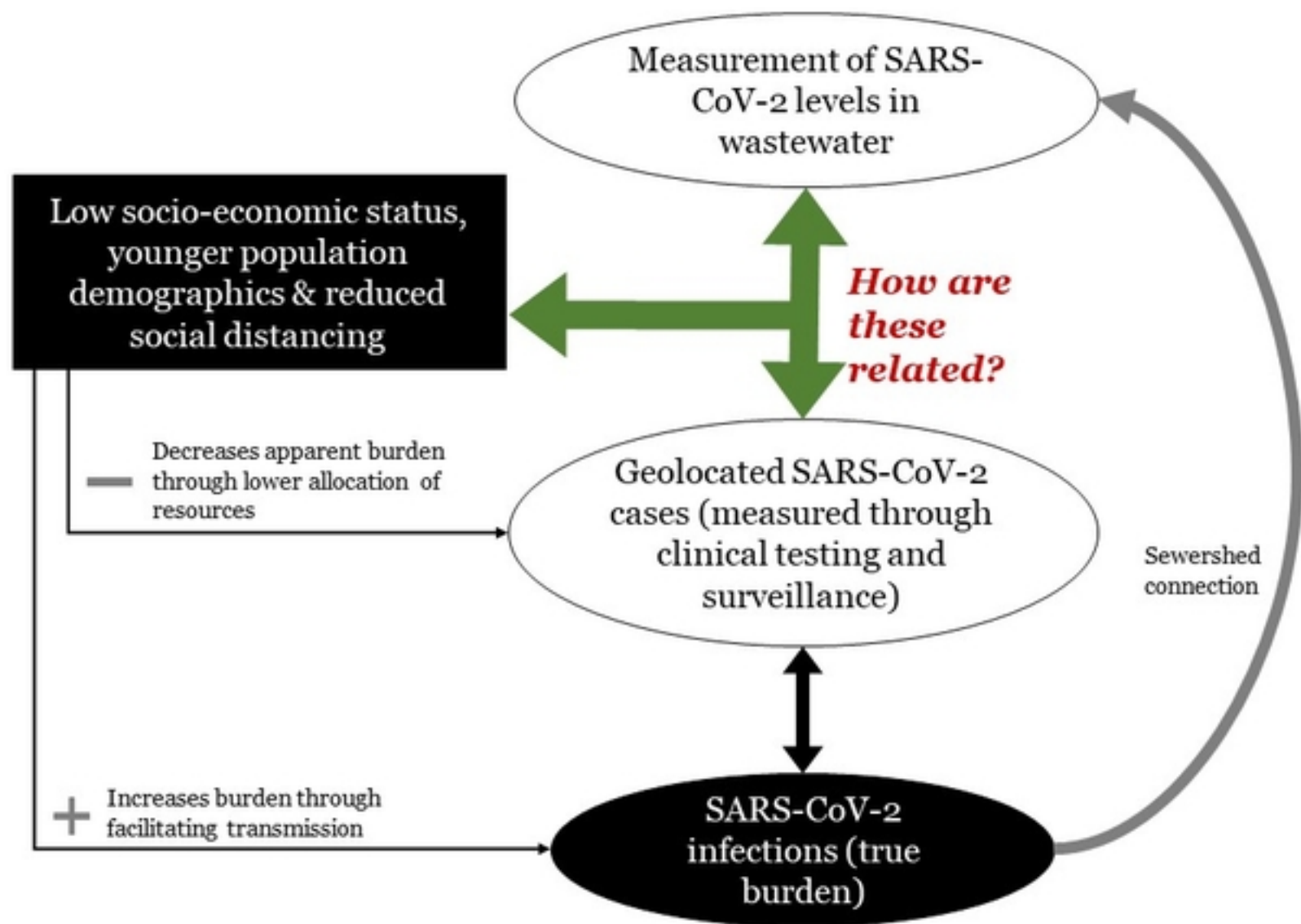
- 482 16. O'Reilly KM, Verity R, Durray E, Asghar H, Sharif S, Zaidi SZ, et al. Population sensitivity of
483 acute flaccid paralysis and environmental surveillance for serotype 1 poliovirus in Pakistan: an
484 observational study. *BMC Infect Dis.* 2018;18: 176. doi:10.1186/s12879-018-3070-4
- 485 17. Ahmed W, Angel N, Edson J, Bibby K, Bivins A, O'Brien JW, et al. First confirmed detection
486 of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater
487 surveillance of COVID-19 in the community. *Sci Total Environ.* 2020;728: 138764.
488 doi:10.1016/j.scitotenv.2020.138764
- 489 18. Smith T, Holm RH, Keith RJ, Amraotkar AR, Alvarado CR, Banecki K, et al. Quantifying the
490 relationship between sub-population wastewater samples and community-wide SARS-CoV-2
491 seroprevalence. *Sci Total Environ.* 2022;853: 158567. doi:10.1016/j.scitotenv.2022.158567
- 492 19. Yousif M, Rachida S, Taukobong S, Ndlovu N, Iwu-Jaja C, Howard W, et al. SARS-CoV-2
493 genomic surveillance in wastewater as a model for monitoring evolution of endemic viruses. *Nat*
494 *Commun.* 2023;14: 6325. doi:10.1038/s41467-023-41369-5
- 495 20. Gauteng City-Region Observatory. Quality of Life Survey 2020-2021, Round 6. [object Object];
496 2021. doi:10.25828/WEMZ-VF31
- 497 21. Iwu-Jaja C, Ndlovu NL, Rachida S, Yousif M, Taukobong S, Macheke M, et al. The role of
498 wastewater-based epidemiology for SARS-CoV-2 in developing countries: Cumulative evidence
499 from South Africa supports sentinel site surveillance to guide public health decision-making. *Sci*
500 *Total Environ.* 2023;903: 165817. doi:10.1016/j.scitotenv.2023.165817
- 501 22. Acheampong E, Husain AA, Dudani H, Nayak AR, Nag A, Meena E, et al. Population infection
502 estimation from wastewater surveillance for SARS-CoV-2 in Nagpur, India during the second
503 pandemic wave. *PLOS ONE.* 2024;19: e0303529. doi:10.1371/journal.pone.0303529

- 504 23. Hoffmann T, Alsing J. Faecal shedding models for SARS-CoV-2 RNA among hospitalised
505 patients and implications for wastewater-based epidemiology. *J R Stat Soc Ser C Appl Stat.*
506 2023;72: 330–345. doi:10.1093/jrsssc/qlad011
- 507 24. Rose C, Parker A, Jefferson B, Cartmell E. The Characterization of Feces and Urine: A Review
508 of the Literature to Inform Advanced Treatment Technology. *Crit Rev Environ Sci Technol.*
509 2015;45: 1827–1879. doi:10.1080/10643389.2014.1000761
- 510 25. Maree G, Fatti CC, Götz G, Hamann C, Parker A. Effects of the COVID-19 pandemic on the
511 Gauteng City-Region: Findings from the GCRO’s Quality of Life Survey 6 (2020/21). 2021.
- 512 26. Lancaster E, Byrd K, Ai Y, Lee J. Socioeconomic status correlations with confirmed COVID-19
513 cases and SARS-CoV-2 wastewater concentrations in small-medium sized communities.
514 *Environ Res.* 2022;215: 114290. doi:10.1016/j.envres.2022.114290
- 515 27. Saingam P, Jain T, Woicik A, Li B, Candry P, Redcorn R, et al. Integrating socio-economic
516 vulnerability factors improves neighborhood-scale wastewater-based epidemiology for public
517 health applications. *Water Res.* 2024;254: 121415. doi:10.1016/j.watres.2024.121415
- 518 28. Rogawski McQuade ET, Blake IM, Brennhofer SA, Islam MO, Sony SSS, Rahman T, et al.
519 Real-time sewage surveillance for SARS-CoV-2 in Dhaka, Bangladesh versus clinical COVID-
520 19 surveillance: a longitudinal environmental surveillance study (December, 2019-December,
521 2021). *Lancet Microbe.* 2023;4: e442–e451. doi:10.1016/S2666-5247(23)00010-1
- 522 29. Layton BA, Kaya D, Kelly C, Williamson KJ, Alegre D, Bachhuber SM, et al. Evaluation of a
523 Wastewater-Based Epidemiological Approach to Estimate the Prevalence of SARS-CoV-2
524 Infections and the Detection of Viral Variants in Disparate Oregon Communities at City and
525 Neighborhood Scales. *Environ Health Perspect.* 2022;130: 067010. doi:10.1289/EHP10289
- 526 30. Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, et al. Evaluation of
527 epidemic intelligence systems integrated in the early alerting and reporting project for the

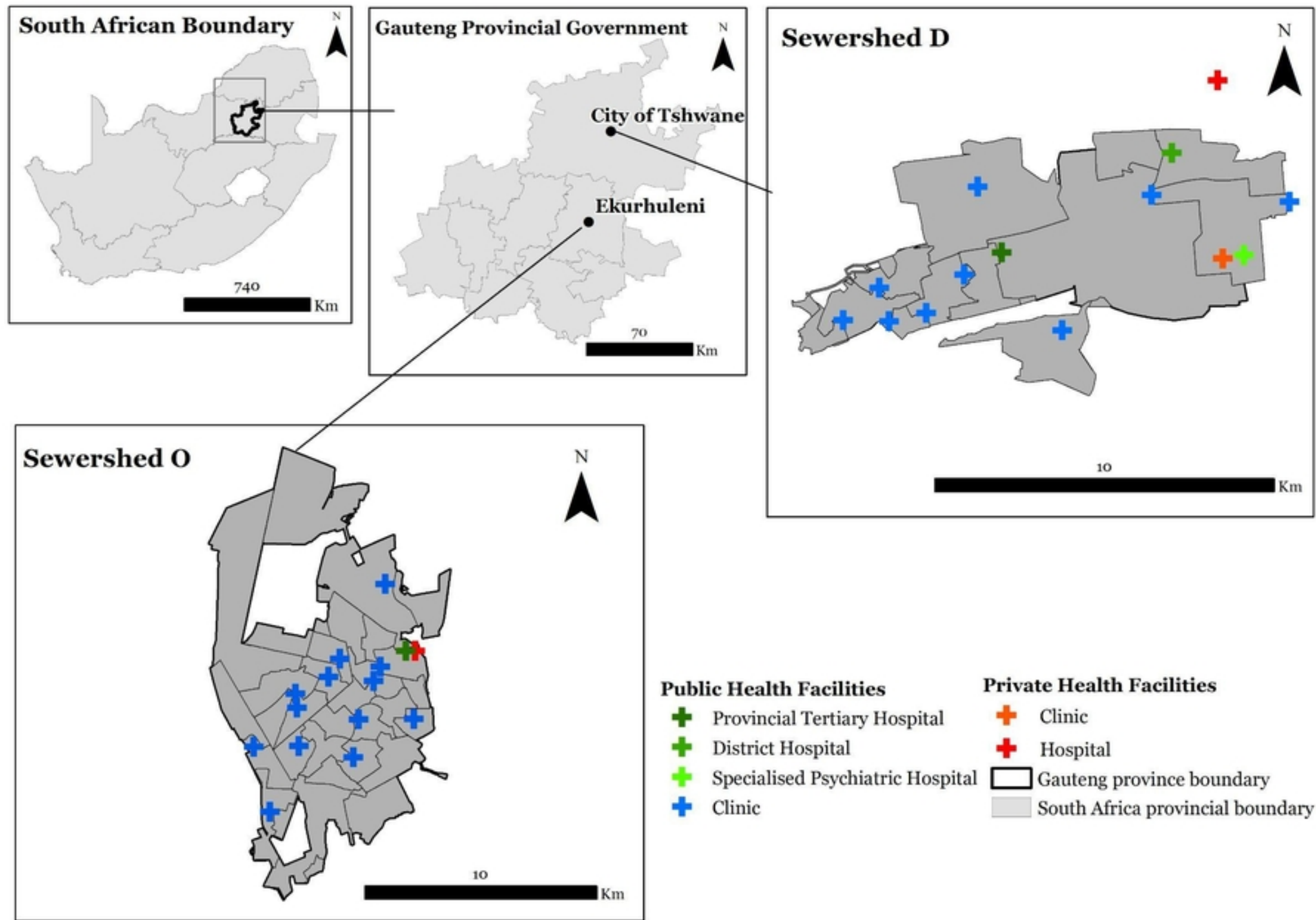
528 detection of A/H5N1 influenza events. PloS One. 2013;8: e57252.

529 doi:10.1371/journal.pone.0057252

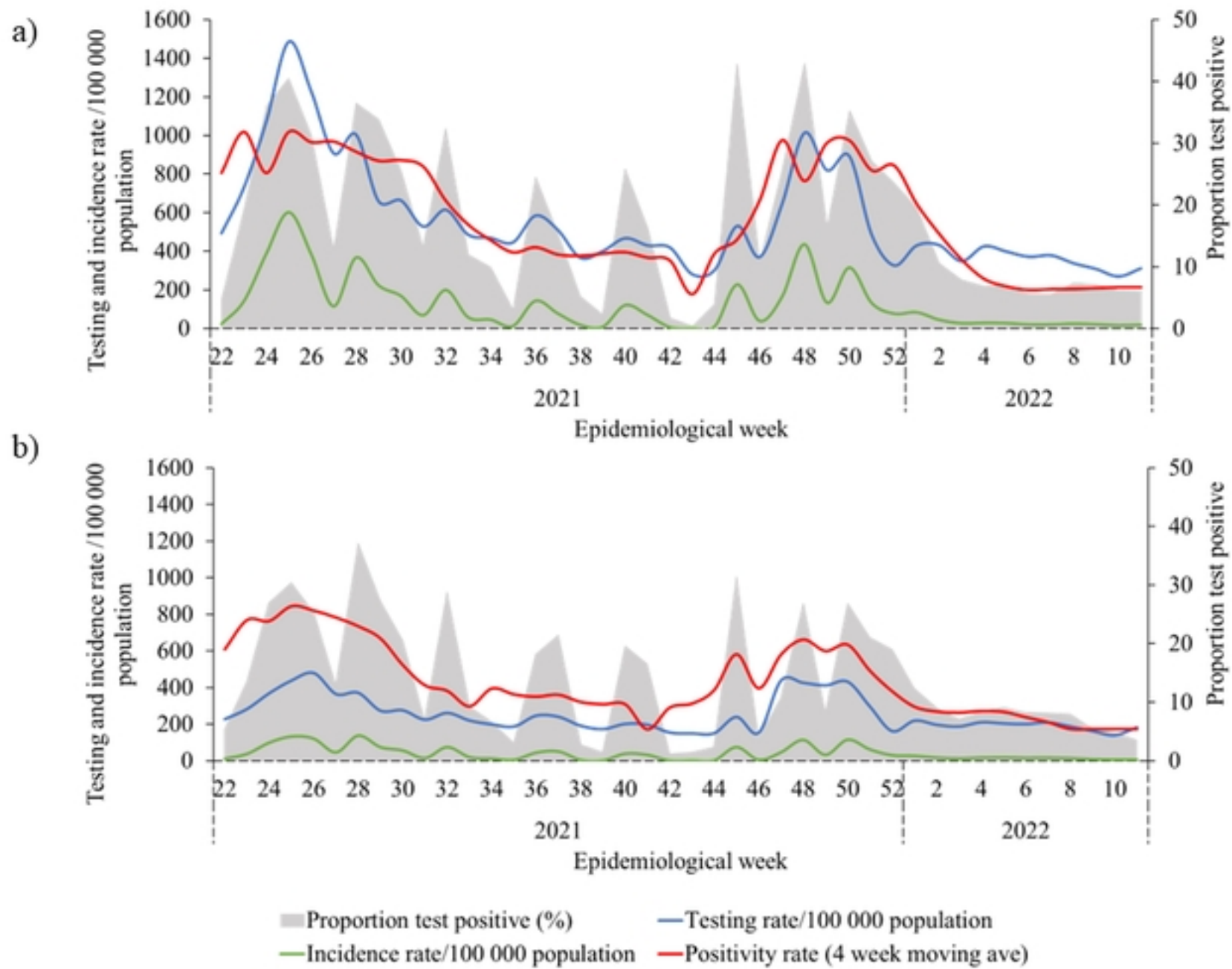
530



Figure



Figure

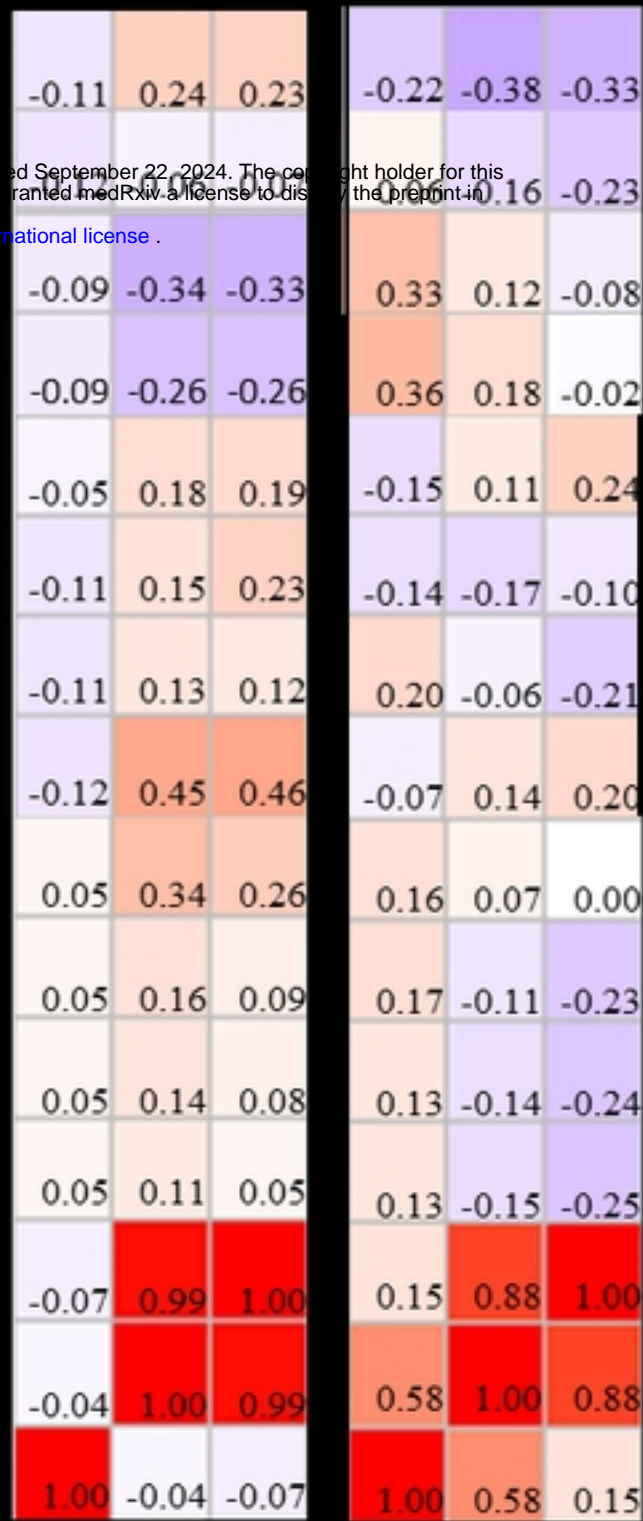


Figure

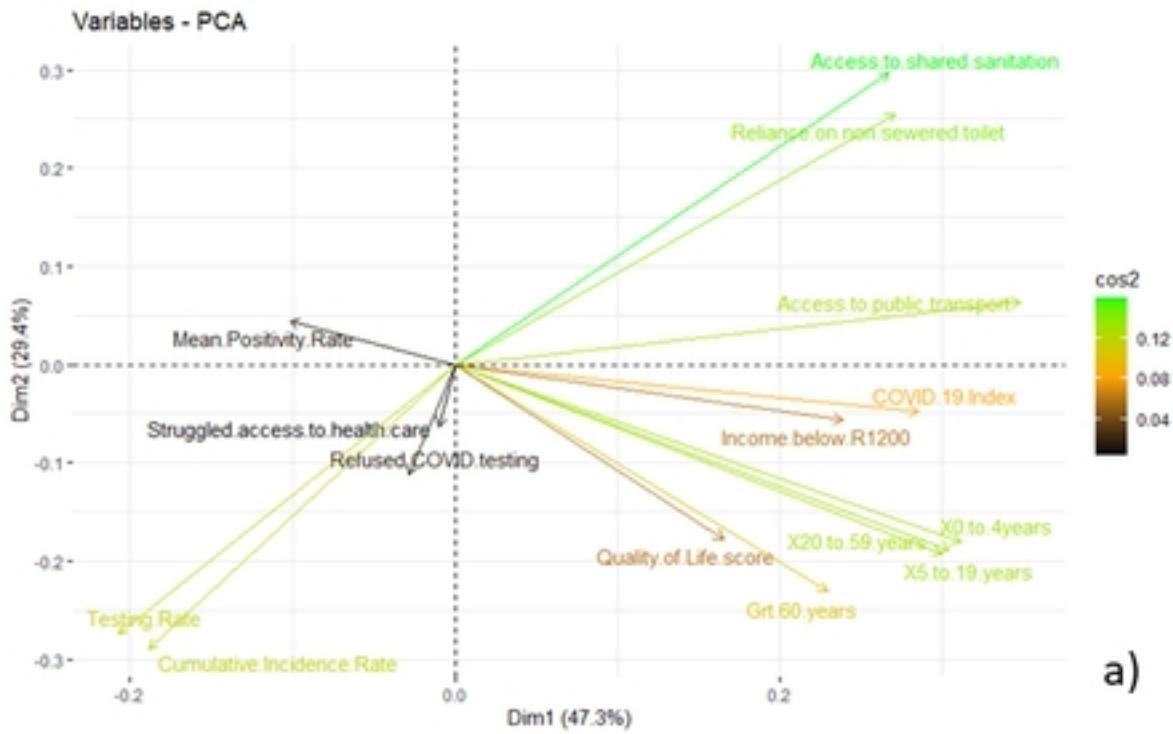
<https://doi.org/10.1101/2024.09.20.24314039>
(which was not certified by peer review)

medRxiv preprint doi: <https://doi.org/10.1101/2024.09.20.24314039>; this version posted September 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

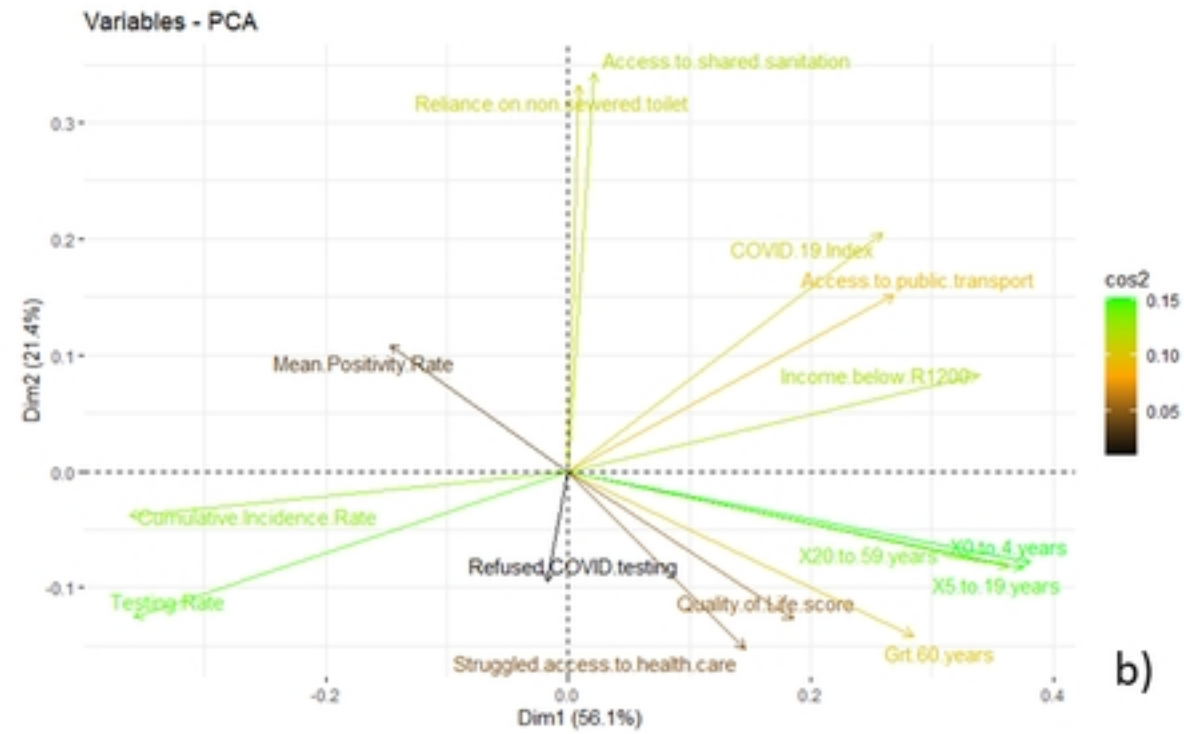
CC-BY-NC 4.0 International license .



Figure

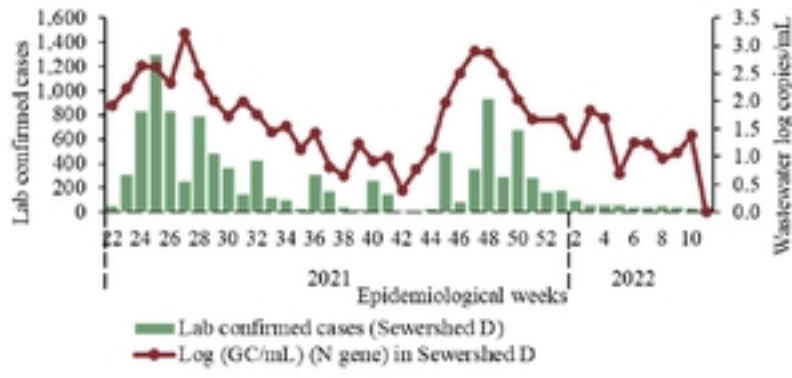


a)

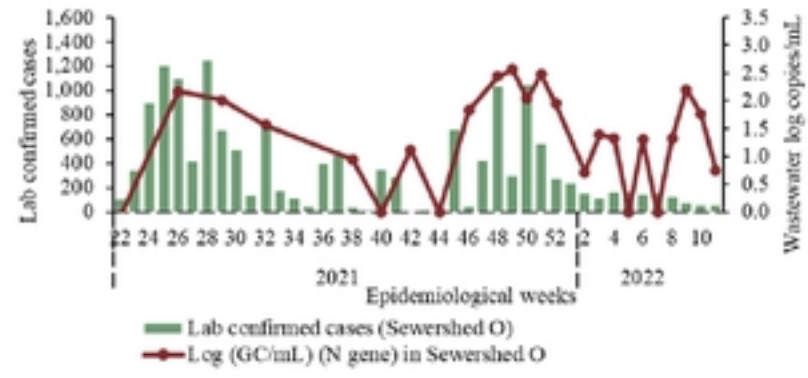


b)

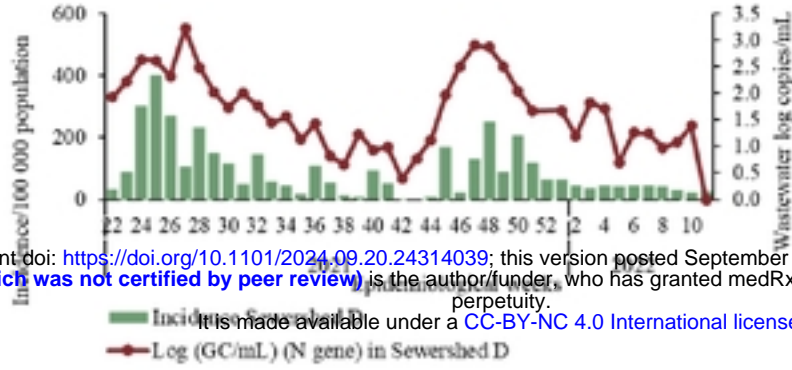
Figure



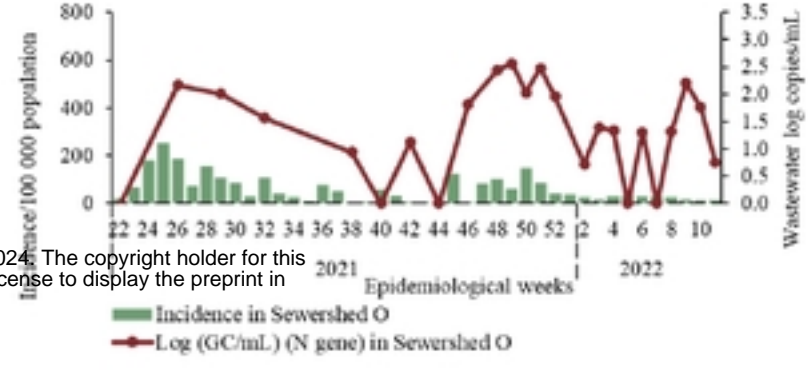
(a)



(b)

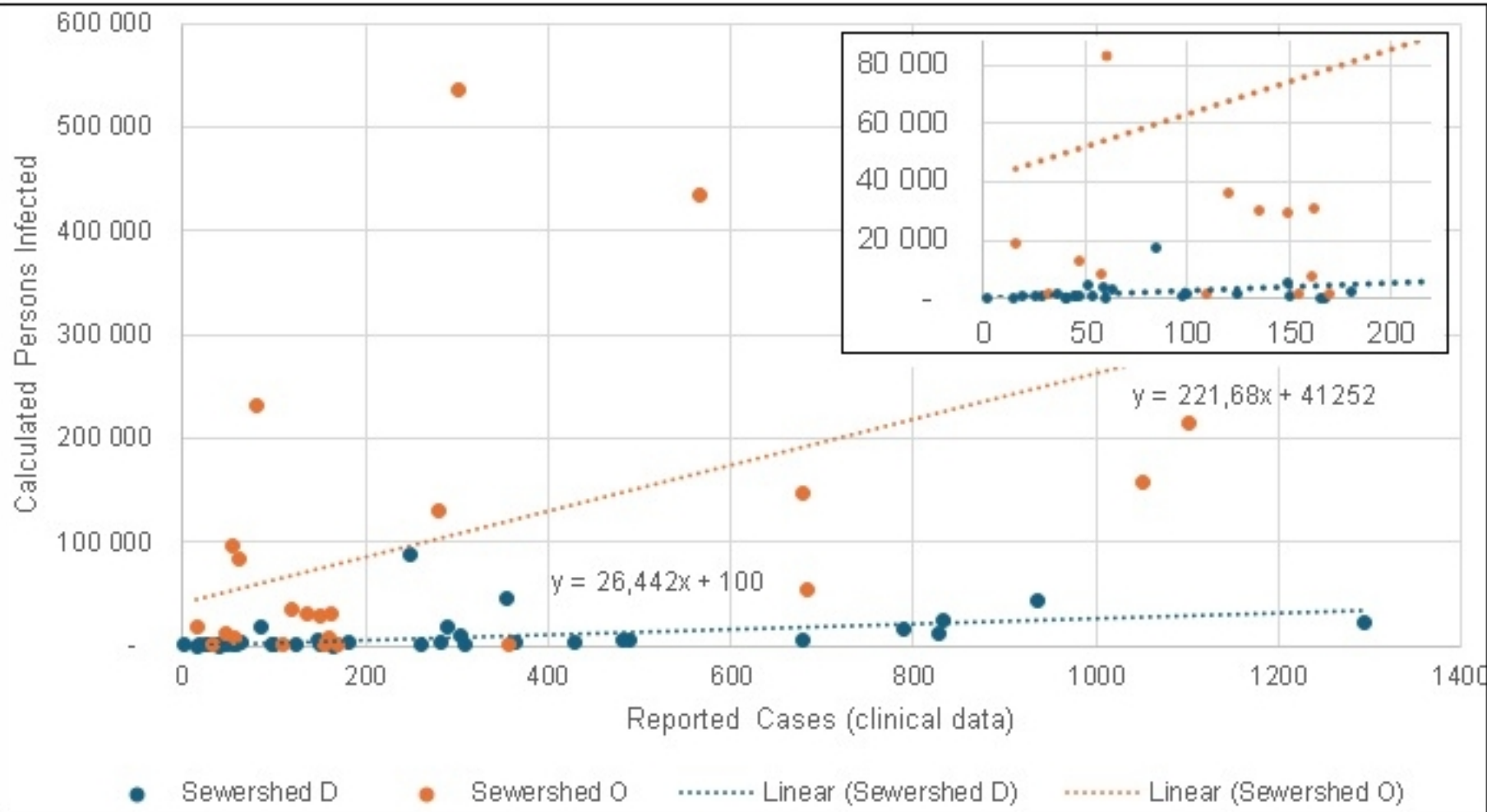


(c)



(d)

Figure 6. SARS-CoV-2 concentrations in wastewater in log-transformed genome copies per millilitre (right axis) and the number of laboratory-confirmed cases (top figures, green bars) or incidence per 100,000 persons (bottom figures, blue bars) of SARS-CoV-2 geolocated to a residential address in the sewershed by epidemiological week from week 22, 2021 to week 10, 2022 for sewersheds D (left, figures a and c respectively) and O (right, figures b and d respectively).



Figure