Serological evidence of human infection with SARS-CoV-2: a systematic review and meta-analysis

Xinhua Chen[1,*], Zhiyuan Chen[1,*], Andrew S. Azman[2,3,4,*], Xiaowei Deng[1], Xinghui Chen[1], Wanying Lu[1], Zeyao Zhao[1], Juan Yang[1], Cecile Viboud[5], Marco Ajelli[6], Daniel T. Leung[7,†], Hongjie Yu[1,†]

*These authors are co-first authors contributed equally to this work.

†These authors are joint senior authors contributed equally to this work.

Corresponding author: Hongjie Yu, Fudan University, School of Public Health, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai 200032, China; E-mail: yhj@fudan.edu.cn

Affiliations:

1. School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai, China
2. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
3. Unit of Population Epidemiology, Division of Primary Care Medicine, Geneva University Hospitals, Switzerland
4. Institute of Global Health, Faculty of Medicine, University of Geneva, Switzerland
5. Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

6. Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN, USA
7. Division of Infectious Diseases, University of Utah School of Medicine, Salt Lake City, UT, USA

Word count (Abstract): 297

Word count (main text): 3,741

Running Title: Seroevidence of human SARS-CoV-2 infections

Keywords: SARS-CoV-2 infections in humans; Serological Evidence

Disclaimer: The views expressed are those of the authors and do not necessarily represent the institutions with which the authors are affiliated.

Abstract

Background: A rapidly increasing number of serological surveys for anti-SARS-CoV-2 antibodies have been reported worldwide. A synthesis of this large corpus of data is needed.

Purpose: To evaluate the quality of serological studies and provide a global picture of seroprevalence across demographic and occupational groups, and to provide guidance for conducting better serosurveys.

Data sources: We searched PubMed, Embase, Web of Science, and 4 pre-print servers for English-language papers published from December 1, 2019 to September 25, 2020.

Study selection: Serological studies evaluating SARS-CoV-2 seroprevalence in humans.

Data extraction: Two investigators independently extracted data from studies.

Data Synthesis: Most of 230 serological studies, representing tests in >1,400,000 individuals, identified were of low quality based on a standardized study quality scale. In the 51 studies of higher quality, high-risk healthcare workers had higher seroprevalence of 17.1% (95% CI: 9.9-24.4%), compared to low-risk healthcare workers and general population of 5.4% (0.7-10.1%) and 5.3% (4.2-6.4%). Seroprevalence varied hugely across WHO regions, with lowest seroprevalence of general population in Western Pacific region (1.7%, 0.0-5.0%). Generally, the young (<20 years) and the old (≥65 years) were less likely to be seropositive compared to middle-aged (20-64 years) populations.

Seroprevalence correlated with clinical COVID-19 reports, with pooled average of 7.7 (range: 2.0 to 23.1) serologically-detected-infections per confirmed COVID-19 case.

Limitations: Some heterogeneity cannot be well explained quantitatively.

Conclusions: The overall quality of seroprevalence studies examined was low. The relatively low seroprevalence among general populations suggest that in most settings, antibody-mediated herd immunity is far from being reached. Given the relatively narrow range of estimates of the ratio of serologically-detected infections to confirmed cases across different locales, reported case counts may help provide insights into the true proportion of the population infected.

Primary Funding source: National Science Fund for Distinguished Young Scholars (PROSPERO: CRD42020198253).

Introduction

The epidemic of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was first reported in Wuhan, Hubei province in December, 2019 and quickly spread globally (1). As of October 13, 2020, more than 37 million COVID-19 cases, including 1,077,799 deaths, had been reported in 235 countries/regions (2, 3). The real number of SARS-CoV-2 infections is undoubtedly much higher than the officially reported cases due to various factors, including the occurrence of asymptomatic infections, variable healthcare-seeking for clinically mild cases, varied testing strategies in different countries, false-negative virologic tests, and incomplete case reporting. Therefore, the reported COVID-19 cases based on clinical identification with virologic confirmation only represents the "tip of iceberg", with a large number of asymptomatic and mild infections in the general population who may only be identified by seroepidemiological studies (4).

Serological studies are a useful tool to estimate the proportion of the population previously infected, to quantify the magnitude of transmission of pathogens, estimate the infection fatality rate (5), assess the impact of interventions (6), and when correlates of protection are available, estimate the degree of population immunity (7, 8). Insights from serological surveillance can be valuable for policymakers and health officials when planning public health decision-making.

94 A large number of serological investigations across the world have been
95 published during the first 9 months of the COVID-19 pandemic, with highly
96 variable estimates of seroprevalence that may largely be due to differences in
97 attack rates, but which also feature heterogeneous sampling strategies and
98 assays used. Two rapid systematic reviews of SARS-CoV-2 seroprevalence were
99 identified (in preprint at the time of writing) but both had a limited scope and
100 did not explore important differences between subpopulations, assess study
101 quality or perform standard meta-analyses techniques (9, 10).
102
103 Here, we conduct a systematic review and meta-analysis to summarize serologic
104 surveys for SARS-CoV-2 infections in humans, to comprehensively evaluate the
105 study designs, laboratory methods, and outcome interpretations for each
106 included serological study, and to estimate the risk of infections by populations
107 with different presumed levels of exposure to SARS-CoV-2. We aim for these
108 results to help inform decision makers and researchers alike as plans are made
109 for the next phases of the global pandemic.

110 Methods

111 Data Sources and Searches

112 According to the Preferred Reporting Items for Systematic Reviews and
113 Meta-Analyses (PRISMA) guideline (http://www.prisma-statement.org/) (11),
114 we performed a systematic literature review from three peer-reviewed databases
115 (PubMed, Embase and Web of Science) and four preprint severs (medRxiv,

116 bioRxiv, SSRN and Wellcome) with predefined search terms **(Appendix Table 1)**.

117

118 **Study Selection and Extraction**

119 Two independent researchers screened titles and abstracts of English-language
120 papers published from December 1, 2019 to September 25, 2020 meeting the
121 following criteria: (1) a report of seroprevalence in either the general population
122 or some other well-defined population of non-COVID clinical cases; (2)
123 conducted after the first reported case in the area and (3) reported the specific
124 assays used. We excluded studies that only reported serologic responses among
125 COVID-19 patients, those using samples with known infection status (e.g.
126 validation studies of assays), and animal experiments. We excluded abstracts of
127 congress meetings or conference proceedings, study protocols, media news,
128 commentaries, reviews or case reports.

129

130 **Data Extraction and Quality Assessment**

131 The full text of included studies after initial screening were scrutinized to assess
132 the overall eligibility based on the inclusion and exclusion criteria by two
133 independent researchers. A third researcher was consulted when the two
134 reviewers disagreed on study assessment. For eligible studies, data were
135 extracted by researchers on the number of participants who provided specimens
136 and the number of these who were seropositive to calculate the seroprevalence.
137 When data were inconsistent between reviewers, they were asked to discuss and

138 revisit the article until reaching a consensus. We developed a scoring system on
139 the basis of a seroepidemiological protocol from Consortium for the
140 Standardization of Influenza Seroepidemiology (CONSISE), a previously
141 published scoring system for seroprevalence studies of zoonotic influenza
142 viruses, and a seroepidemiological protocol developed by World Health
143 Organization (WHO), to further assess the study quality (12-14). We
144 comprehensively assessed study design (representativeness of study
145 participants), laboratory assay (whether internal assay validations or a
146 confirmatory assay was performed) and outcome adjustment (correction for
147 demographics and/or test performance) and based on a scoring scheme, an
148 overall score was determined for each included study **(Appendix Table 6)**. From
149 this score, we classified each study's quality into one of four grades: A, B, C or D.
150 Grade A, the highest quality category, spanned studies with scores ranging from
151 12 to 10, grade B from 9 to 7, grade C from 4 to 6, and grade D from 0 to 3. We
152 only included grade A and grade B studies in the main analysis but provide
153 additional results with all studies, irrespective of grade, in the supplement. We
154 also described the characteristics, laboratory testing method, and primary
155 outcome for each available study in **Appendix Table 2-5**.
156
157 **Data Synthesis and Analysis**
158 Seroprevalence was defined as the prevalence of SARS-CoV-2-specific antibodies
159 at or above a designated antibody titer to define a seropositive result in each

160 original study. For serial cross-sectional studies, we calculated the sum of the
161 total number of participants who provided specimens and total number of
162 seropositive individuals during the whole study period, to avoid repeated
163 inclusion of the same study. Similarly, only data from the first blood collection
164 were analyzed for studies with a longitudinal design. For studies that used
165 multiple serological assays, we used the seropositive results from the assay with
166 the highest sensitivity and specificity (calculated by Youden's index). We also
167 conducted a sensitivity analysis with results from the other assays **(Appendix**
168 **Table 10)**. If a study used a confirmatory assay (e.g. microneutralization assay)
169 to validate the positive and/or equivocal result from the initial screening, we
170 used the results from the confirmatory assay. For those studies reporting
171 multiple isotypes including IgG, we included only IgG in the main analyses as
172 they remain elevated for a longer period post-infection compared to IgM and IgA
173 (15). If seropositivity based only on IgG was not reported separately or
174 seropositivity reported was based only on total antibodies, these results were
175 also included. While many studies did adjust for various factors, we decided to
176 use the crude (unadjusted) estimates in our analyses to ease interpretation
177 across different studies.
178
179 To reduce heterogeneity between individual seroprevalence estimates and to
180 provide more meaningful summary statistics, we stratified eligible studies by
181 WHO regions (i.e. African region, Region of Americas, Eastern Mediterranean

182 Region, European Region, South-East Asia Region, Western Pacific Region) (16).
183 To estimate seroprevalence by different types of exposures, within each WHO
184 region, we categorized all study participants into five groups: 1) close contacts, 2)
185 high-risk healthcare workers, 3) low-risk healthcare workers, 4) general
186 populations, and 5) poorly-defined populations **(Appendix Table 5)**. The
187 poorly-defined population classification represents populations with undefined
188 or unknown exposure to laboratory-confirmed or suspected COVID-19 patients
189 (e.g., blood donors, residual blood samples from labs, patients of other diseases,
190 etc.), as well as study populations that cannot be categorized into the first four
191 study populations due to limited exposure information [e.g., healthcare workers
192 without reporting usage of PPE (personal protective equipment) or
193 COVID-19-related exposures]. Based on a random-effect meta-analysis model, we
194 used inverse variance method to estimate pooled seroprevalence by WHO
195 regions and different sub-populations, combined with the use of
196 Clopper-Pearson method to calculate 95% confidence intervals (17, 18).
197
198 For seroprevalence estimates from the general population, we further explored
199 potential determinants affecting the seroprevalence, such as sex, age, and the
200 reported cumulative incidence of COVID-19 cases for the location. We calculated
201 cumulative incidence of confirmed COVID-19 by dividing the number of cases
202 reported in the same target population as the serosurvey two weeks before the
203 serosurvey mid-point at the location, by the estimated population size.

204 Furthermore, we meta-analyzed the number of serologically-detected infections
205 (the number of individuals with positive SARS-CoV-2 serology) per confirmed
206 case (the number of reported cases in the serologic study's target population)
207 with available epidemiological data included in the articles, the WHO website
208 (19), the COVID-19 dashboard of Johns Hopkins University (2), and local health
209 authorities (20). Unless reported in the article, we used population size estimates
210 from WorldPop (21) or a local statistics bureau (22). Studies were included in the
211 meta-analyses of the ratio of serologically-detected infections per confirmed
212 cases if they reported seroprevalence in the representative general population
213 (non-convenience sample) with population data and confirmed case data
214 available for the same population.
215
216 Variability between studies was determined by the heterogeneity tests with
217 Higgins' I^2 statistic. We explored the reasons for variations among eligible studies
218 and examined whether prevalence of SARS-CoV-2 specific antibodies varied by
219 study location, study quality, and level of exposure by multivariable
220 meta-regression models. All statistical analyses were done using R (version
221 3.6.3), with 'meta' package to conduct the meta-analysis. For all statistical tests,
222 two tailed P-value less than 0.05 were considered statistically significant.
223
224 **Role of the funding source**
225 The funder had no role in study design, data collection, data analysis, data

226 interpretation, or writing of the report. The corresponding author had full access
227 to all the data in the study and had final responsibility for the decision to submit
228 for publication.
229
230 **Results**

231 We identified a total of 8,487 studies after systematically searching multiple data
232 sources with 4,763 coming from peer-reviewed databases, 3,710 from preprint
233 severs and 14 reports from governments or health authorities. After removing
234 duplicates and screening titles and abstracts, 1,535 studies reporting serological
235 evidence of SARS-CoV-2 infections were found. Finally, a total of 230 studies
236 involving 1,445,028 participants were included in our meta-analysis after
237 full-text scrutiny **(Figure 1)**, with the majority of studies conducted in European
238 Region (n=106), Region of the Americas (n=69), and Western Pacific Region
239 (n=34) **(Figure 1, Appendix Figure S9)**. Among 214 studies reporting the exact
240 starting sampling date, seventeen (17/214, 8%) were conducted after more than
241 75% of the total cases (in that country/states/provinces) had been reported as of
242 Sept 25, 2020, most of which (14/17, 82%) were conducted in China **(Appendix**
243 **Figure 2-7)**. For those seventeen studies, the serosurveys were conducted a
244 median of 46 days (range: 12-102) after the epidemic peak. Similarly, fifty-three
245 studies (53/214, 25%) were conducted after more than 50% of total cases had
246 been reported, with a median of 40 (range: 2-113) days after the peak.
247

248 General populations (55/230, 24%) and healthcare workers (both high-risk and
249 low-risk healthcare workers) (78/230, 34%) were sampled with the highest
250 frequency. The age profile of participants varied across studies of different
251 targeted populations **(Appendix Table 2)**. Most studies involved participants
252 aged 18-69 years (222/230 studies) and only seven studies exclusively recruited
253 young people (<18 years), including two papers focusing on children (<10 years),
254 and one exclusively targeting older (>80 years) persons from nursing home
255 **(Appendix Table 2)**.
256
257 The overall quality of studies was low based on our grading system **(Appendix**
258 **Table 6)**. The majority of studies were categorized as Grades C or D **(Appendix**
259 **Figure 1)**, including all but one study of high-risk healthcare workers, and all
260 studies of close contacts. Of the 55 general-population-based studies, 5 achieved
261 Grade A, and 19 achieved Grade B.
262
263 Approximately two-third of studies (149/230, 65%) described serological results
264 from convenience samples, while only thirty studies (30/230, 13%) used
265 multi-stage/stratified random sampling to select study participants. Most studies
266 measured IgG antibodies using chemiluminescence immunoassays (CLIA,
267 92/230, 40%), followed by enzyme-linked immunosorbent assays (ELISA,
268 80/230, 35%), and lateral flow immunoassays (LFIA, 59/230, 26%), with 19%
269 (43/230) of studies using more than one serologic assay. Additionally, nineteen

270 studies utilized a microneutralization assay to detect neutralizing antibodies.
271 Among 165 studies that provided the targeted protein for serological assays, 123
272 (75%) studies used tests targeting the S-protein, and 104 (63%) studies used
273 tests targeting the N-protein. More than half of the studies (126/230, 55%)
274 reported age/sex-specific seroprevalence or corrected their findings for age
275 and/or sex. Thirty-one studies (31/230, 13%) adjusted for sensitivity and
276 specificity of the serological assay(s). Only twenty-eight studies corrected for
277 both demographic factors and test performance.
278
279 Among 51 Grade-A and Grade-B studies, seroprevalence varied across WHO
280 regions and study populations. Generally, high-risk healthcare workers had a
281 higher seroprevalence (17.1%, 95% CI: 9.9-24.4%), compared to low-risk
282 healthcare workers (5.4%, 95% CI: 0.7-10.1%) and general populations (5.3%,
283 95% CI: 4.2-6.4%) **(Figure 3A, Appendix Table 9)**. The seroprevalence of the
284 populations in studies that did not specify exposure was 4.1% (95% CI: 3.0-5.1%)
285 **(Figure 3A, Appendix Table 9)**. Estimates of seroprevalence from the general
286 population was highest in the one study conducted in the African Region (25.4%,
287 95% CI:19.1-31.7%), followed by Eastern Mediterranean Region (2 studies,
288 14.4%, 95% CI:2.2-26.7%), Region of the Americas (9 studies, 5.2%, 95%
289 CI:2.9-7.5%), and European Region (10 studies, 4.2%, 95% CI:2.7-5.8%), with the
290 lowest seroprevalence in studies conducted in the Western Pacific Region (2
291 studies, 1.7%, 95% CI:0.0-5.0%) **(Figure 3B-G, Appendix Table 9)**. Sensitivity

292 analysis for studies using multiple assays showed that few changes were
293 observed in seroprevalence by different WHO regions and populations compared
294 to our main analysis **(Appendix Table 10)**.
295
296 The seroprevalence increased with age for study participants under 65 years,
297 constituting 1.5% (95% CI: 0.0-3.1%), 4.2% (95% CI: 2.1-6.2%) and 5.2% (95%
298 CI: 3.2-7.2%) for study participants younger than 20 years, 20-49 years, and
299 50-64 years among the 51 studies graded as A or B **(Figure 3H)**. A similar
300 tendency was also observed when analyzing by WHO regions **(Appendix Figure**
301 **10B-C).** After further (imperfectly) merging age groups, the relative risk of
302 seropositivity in the young (<20 years) was approximately 20% lower than that
303 of the middle-age group (20-64 years) (RR=0.76, 95%CI: 0.69-0.85), suggesting
304 young individuals may be less susceptible to SARS-CoV-2 infections or may
305 mount a less robust antibody response **(Appendix Table 12)**. The risk of
306 seropositivity in the old (≥ 65 years) was also lower compared to the middle-age
307 group (RR=0.71, 95%CI: 0.54-0.94) **(Appendix Table 12)**.
308
309 Pooled seroprevalence of males (5.0%, 95% CI: 3.4-6.6%) and females (4.4%, 95%
310 CI: 2.7-6.0%) were similar, with 58% (7/12) of studies reporting sex-specific
311 seroprevalence having higher point estimates of seroprevalence in males than
312 females (Figure 3I). Similarly, pooling sex-specific relative risks across studies to
313 adjust for the differences in risk across settings revealed no significant increase

314 in risk of seropositivity in males (RR=1.03, 95% CI: 0.95-1.13) **(Appendix Table**
315 **12),** with similar estimates across WHO regions (**Appendix Table 12**).
316
317 The relationship between reported COVID-19 incidence (confirmed cases
318 reported from public sources) and the number of infections identified through
319 serologic surveys can be useful for trying to understand the evolution of the
320 pandemic without serologic surveillance in each and every locale. For studies of
321 the general population, the cumulative reported incidence of COVID-19
322 correlated with seroprevalence across locations (Pearson correlation coefficient:
323 0.80), though this is largely driven by one study **(Appendix Figure 11)**. For
324 studies including individuals from general populations, the ratio of
325 serologically-detected infections to virologically-confirmed cases varied across
326 locations, with a pooled ratio of 7.7 (95% CI: 5.5-10.9), suggesting that for each
327 virologically confirmed SARS-COV-2 infection, at least 6 infections remained
328 undetected by surveillance systems. This ratio was similar between the Region of
329 the Americas (7.9, 95% CI: 1.7-35.3) and the European Region (7.7, 95%CI:
330 5.8-10.2) **(Figure 4)**.
331
332 **Discussion**
333 With the increasing availability of serological assays for SARS-CoV-2, a large body
334 of literature describing seroprevalence studies in different populations has also
335 emerged. In this study, we examine the quality and results of over two hundred

336 reports of seroprevalence studies from around the globe, both published, and in
337 preprint form. In general, the quality of existing serological studies was low,
338 involving less rigorous sampling strategies, poorly-validated and
339 non-standardized laboratory methods, and lack of statistical correction in
340 analyses. As expected, we found that high-risk healthcare workers had higher
341 prevalence of SARS-CoV-2 specific antibodies than that of low-risk healthcare
342 workers and the general population. Young individuals (<20 years) and older
343 persons (≥65 years) are less likely to be seropositive than those of middle age
344 (20-64 years), while there was not a significant difference in seroprevalence
345 between males and females. Additionally, we find that the ratio of infections per
346 confirmed case is generally on the same order of magnitude across populations
347 with very different surveillance and healthcare systems.
348
349 Representative serosurveys can provide useful snapshots of the infection history
350 of a population. However, very few studies (30/230, 13%) provided
351 representative estimates for their target population. An optimal study design for
352 estimating seroprevalence includes a detailed sampling framework, rigorous
353 sampling methods (i.e. multi-stage, stratified sampling), and adjustments for
354 selection bias and assay performance (12).
355
356 Various detection assays were used for determination of seropositivity. We found
357 large variations in test performance, targeted antigens and immunoglobulin

358 isotypes, and threshold used. Furthermore, more than half of the studies lacked
359 independent validations of the sensitivity and specificity of the diagnostic kits
360 prior to assessment of serosurvey samples to verify their initial results (23-25).
361 Notably, although the WHO established a generic population-based serological
362 study protocol, there is a lack of standardized guidelines and procedures for
363 laboratory testing, which may contribute towards such heterogeneity in both
364 performance and reporting of results. We call on national and international
365 governance bodies to develop standardized antibody testing protocols and
366 reporting practices and create biobanks of reference standards (e.g. monoclonal
367 antibodies), to reduce lab-to-lab variations, thus facilitating the comparability
368 and interpretability amongst seroprevalence studies. Despite the WHO
369 recommendations, the estimates described by many of the population-based
370 serosurveys did not adjust for demographic structure of the target population
371 (14), nor the testing performance (sensitivity and specificity) of the assay, which
372 made the comparison among studies difficult.
373
374 Most of the high-quality serological surveys identified were conducted in the
375 Region of the Americas and the European Region, predominantly on general
376 populations. There were a very limited number of high-quality studies of
377 exposed populations, especially for healthcare workers and close contacts, and
378 studies to address this knowledge gap are needed. For the other four WHO
379 regions examined, there was a paucity of high-quality studies across all

380 populations examined, suggesting that attention should be paid to optimize the
381 design of future sero-epidemiological studies to include good representativeness
382 of samples, standardized laboratory methods, and reasonable adjustments.
383 Higher quality studies provide more accurate measures of disease burden and
384 transmission, to better inform public health efforts against COVID-19.
385
386 We found high seroprevalence among high-risk healthcare workers, defined as
387 those who provided routine medical care to COVID-19 patients without PPE. On
388 the contrary, low-risk healthcare workers, defined as those wearing adequate
389 PPE or those who provided care for non-COVID-19 patients, had seroprevalence
390 estimates closer to (and with overlapping confidence intervals) that of the
391 general population (26). We found a pooled seroprevalence of 5.3% in the
392 general population, suggesting that globally, the number of persons infected thus
393 far is unlikely to approach that satisfying estimates of what it would take to
394 achieve antibody-mediated herd immunity (27-29). The seroprevalence in the
395 general population also varied across WHO regions, indicating different levels of
396 community transmission of SARS-CoV-2 in different locales. In addition, the very
397 limited number of high-quality studies in all but two of the WHO regions
398 underlines our incomplete understanding of COVID-19 burden in much of the
399 world. (30) We also found notable differences in seroprevalence between age
400 groups, with the seroprevalence increasing with age among participants younger
401 than 65 years. We found that the young (<20 years) were less likely to be

402 seropositive compared to middle-aged (20-64 years) individuals, consistent with
403 reports of lower numbers of virally-confirmed cases in children compared to
404 other age groups (31). In some areas, this may have been due to the effects of
405 lockdowns, limiting exposures of school-aged populations, in contrast to
406 adult-aged essential workers who had continued community exposure (32, 33).
407 The prevalence of SARS-CoV-2 specific antibodies and relative risk of being
408 seropositive among older persons (≥65 years) was also low, which may be
409 explained by poorer serologic responses after infection (34), lower rates of
410 infection as a result of biological differences or, perhaps more likely, due to
411 behavior changes leading to reduced frequency of infectious contacts.
412
413 High heterogeneity was observed for all study populations. Although meta
414 regression could partly explain such varied seroprevalence between different
415 study populations with different type of exposures, the internal heterogeneity for
416 the same population should be interpreted carefully. The seroprevalence of the
417 general population is likely a reflection of the duration and intensity of
418 community transmissions. We showed using a regression analysis that higher
419 cumulative incidence of reported cases is associated with higher seroprevalence.
420 For locations where data on the number of confirmed cases were available, we
421 found that the number of infections per confirmed case varied (2.0-23.1) but the
422 order of magnitude was quite consistent across locales. Such evidence indicates
423 the existence of a large number of undetected cases in the community and

424 provides a range of scaling factors for translating the observed confirmed cases
425 into unobserved infections in the community.
426
427 Our study has several limitations. First, although we have performed
428 meta-regression and subgroup analysis to explore heterogeneity of varied
429 seroprevalence for different populations, there are still some factors that we have
430 not taken into account, so that some heterogeneity cannot be well explained
431 quantitatively. Second, misclassification bias may occur due to the limited
432 information on exposures for the study populations, especially for the
433 "Poorly-defined" populations. For some key information (e.g., the use of PPE for
434 healthcare workers) that cannot be extracted from the original articles, we have
435 tried to contact the authors, but the response rate was low. Third, our findings
436 involving cumulative incidence of confirmed cases are limited by the inherently
437 heterogeneous incidence across locations, due to the diversity of testing strategy,
438 case tracing and identification, and public health interventions. Fourth, only raw
439 data without correcting for assay performance or sampling design were used in
440 the analysis to ensure comparability within studies. Fifth, the seroprevalence of
441 IgG was possibly overestimated due to the inability to isolate IgG-based
442 seroprevalence from total antibodies-based seroprevalence in some studies.
443 Lastly, our infection-to-case estimates did not account for age-specific differences
444 in exposure risk and symptom severity.
445

446 In conclusion, the overall quality of the existing seroprevalence studies of
447 SARS-CoV-2 is low and international efforts to standardize study design and
448 assays are urgently needed (14). Pooled estimates of SARS-CoV-2 seroprevalence
449 based on currently available data demonstrate a higher infection risk among
450 close contacts and healthcare workers lacking PPE, while the relatively low
451 prevalence of SARS-CoV-2 specific antibodies among general populations
452 suggests that the majority of examined populations have not been infected.
453 Therefore, antibody-mediated herd immunity is likely far from being reached in
454 most settings, and continuous serological monitoring is necessary to inform
455 public health decision making.
456
457 **Contributors**

458 H.Y. designed and supervised the study. X.C. and Z.C. did the literature search, set
459 up the database and did all statistical analyses. X.C., Z.C., A.S.A., and D.T.L.
460 co-drafted the first version of the article. X.C, Z.C., X.D., W.L, Z.Z, and X.C. helped
461 with checking data and did the figures. D.T.L., A.S.A., J.Y, M.A., C.V. and H.Y.
462 commented on the data and its interpretation, revised the content critically. All
463 authors contributed to review and revision and approved the final manuscript as
464 submitted and agree to be accountable for all aspects of the work.
465
466 **Declaration of interests**

467 H.Y. has received investigator-initiated research funding from Sanofi Pasteur,

468 GlaxoSmithKline, and Yichang HEC Changjiang Pharmaceutical Company; M.A.
469 has received research funding from Seqirus; D.T.L. and A.S.A. has received
470 research funding from the US National Institutes of Health. None of those
471 research funding is related to COVID-19. All other authors report no competing
472 interests.
473

474 **Acknowledgments**

475 We thank Wei Wang, Qianhui Wu, Yongli Zhang, Junbo Chen, Qianli Wang, and
476 Yuxia Liang from the Fudan University. This study was funded by the National
477 Science Fund for Distinguished Young Scholars (grant no. 81525023), Program of
478 Shanghai Academic/Technology Research Leader (grant no. 18XD1400300),
479 National Science and Technology Major project of China (grant no.
480 2018ZX10713001-007, 2017ZX10103009-005, 2018ZX10201001-010), the US
481 National Institutes of Health (R01 AI135115 to D.T.L. and A.S.A.)
482

483 **References**

484 1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus
485 from Patients with Pneumonia in China, 2019. N Engl J Med. 2020.
486 2. CSSE J 2020;Pages https://coronavirus.jhu.edu/map.html on 17 Aug 2020.
487 3. WHO
488 2020;Pages https://www.who.int/emergencies/diseases/novel-coronavir
489 us-2019.
490 4. Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A Novel

491 Coronavirus Emerging in China - Key Questions for Impact Assessment. N
492 Engl J Med. 2020;382(8):692-4.
493 5. Perez-Saez J, Lauer SA, Kaiser L, Regard S, Delaporte E, Guessous I, et al.
494 Serology-informed estimates of SARS-CoV-2 infection fatality risk in
495 Geneva, Switzerland. Lancet Infect Dis. 2020.
496 6. Sughayer MA, Mansour A, Al Nuirat A, Souan L, Ghanem M, Siag M.
497 Covid-19 Seroprevalence rate in healthy blood donors from a community
498 under strict lockdown measures. 2020:2020.06.06.20123919.
499 7. Griffin S. Covid-19: Herd immunity is "unethical and unachievable," say
500 experts after report of 5% seroprevalence in Spain. Bmj.
501 2020;370:m2728.
502 8. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al.
503 Estimates of the severity of coronavirus disease 2019: a model-based
504 analysis. Lancet Infect Dis. 2020.
505 9. Bobrovitz N, Arora RK, Yan T, Rahim H, Duarte N, Boucher E, et al. Lessons
506 from a rapid systematic review of early SARS-CoV-2 serosurveys.
507 2020:2020.05.10.20097451.
508 10. Arora RK, Joseph A, Van Wyk J, Rocco S, Atmaja A, May E, et al. SeroTracker:
509 a global SARS-CoV-2 seroprevalence dashboard. Lancet Infect Dis. 2020.
510 11. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for
511 systematic reviews and meta-analyses: the PRISMA statement. PLoS Med.
512 2009;6(7):e1000097.
513 12. Horby PW, Laurie KL, Cowling BJ, Engelhardt OG, Sturm-Ramirez K,
514 Sanchez JL, et al. CONSISE statement on the reporting of
515 Seroepidemiologic Studies for influenza (ROSES-I statement): an

516 extension of the STROBE statement. Influenza Other Respir Viruses.
517 2017;11(1):2-14.
518 13. Sikkema RS, Freidl GS, de Bruin E, Koopmans M. Weighing serological
519 evidence of human exposure to animal influenza viruses - a literature
520 review. Euro Surveill. 2016;21(44).
521 14. WHO. Population-based age-stratified seroepidemiological investigation
522 protocol for COVID-19 virus infection. 2020.
523 15. Ma H, Zeng W, He H, Zhao D, Jiang D, Zhou P, et al. Serum IgA, IgM, and IgG
524 responses in COVID-19. Cell Mol Immunol. 2020;17(7):773-5.
525 16. Rostami A, Sepidarkish M, Leeflang M, Riahi SM, Nourollahpour Shiadeh M,
526 Esfandyari S, et al. First snap-shot meta-analysis to estimate the
527 prevalence of serum antibodies to SARS-CoV-2 in humans.
528 2020:2020.08.31.20185017.
529 17. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated
530 in the case of the binomial. Biometrika. 1934;26:404-13.
531 18. Newcombe RG. Two-sided confidence intervals for the single proportion:
532 comparison of seven methods. Stat Med. 1998;17(8):857-72.
533 19. WHO
534 2020;Pageshttps://www.who.int/news-room/feature-stories/detail/who
535 -updates-covid-19-dashboard-with-better-data-visualization on 28 Aug
536 2020.
537 20. FOPH
538 2020;Pageshttps://www.bag.admin.ch/bag/en/home/krankheiten/ausbr
539 ueche-epidemien-pandemien/aktuelle-ausbrueche-epidemien/novel-cov
540 /situation-schweiz-und-international.html#-1680104524 on 11 Aug

541 2020.
542 21. WorldPop 2020;Pageshttps://www.worldpop.org/ on 28 Aug 2020.
543 22. Office FS 2020;Pageshttps://www.bfs.admin.ch/bfs/en/home.html on
544 Sept 25 2020.
545 23. Chen Y, Tong X, Wang J, Huang W, Yin S, Huang R, et al. High SARS-CoV-2
546 antibody prevalence among healthcare workers exposed to COVID-19
547 patients. J Infect. 2020.
548 24. Daniel C. Payne PSES-J, MPH1; Gosia Nowak, MPH2; Uzo Chukwuma,
549 MPH2; Jesse R. Geibe, MD2; Robert J. Hawkins, PhD, DNP2; Jeffrey A.
550 Johnson, PhD1; Natalie J. Thornburg, PhD1; Jarad Schiffer, MS1; Zachary
551 Weiner, PhD1; Bettina Bankamp, PhD1; Michael D. Bowen, PhD1; Adam
552 MacNeil, PhD1; Monita R. Patel, PhD1; Eric Deussing, MD2; CDC COVID-19
553 Surge Laboratory Group; Bruce L. Gillingham, MD2. SARS-CoV-2 Infections
554 and Serologic Responses from a Sample of U.S. Navy Service Members —
555 USS Theodore Roosevelt, April 2020. MMWR Morb Mortal Wkly Rep.
556 2020.
557 25. Kelvin Kai-Wang To* VC-CC, Jian-Piao Cai*, Kwok-Hung Chan*, Lin-Lei
558 Chen, Lok-Hin Wong, Charlotte Yee-Ki Choi,, Carol Ho-Yan Fong AC-KN, Lu
559 Lu, Cui-Ting Luo, Jianwen Situ, Tom Wai-Hin Chung, Shuk-Ching Wong,
560 Grace See-Wai Kwan,, Siddharth Sridhar JF-WC, Cecilia Yuen-Man Fan,
561 Vivien W M Chuang, Kin-Hang Kok, Ivan Fan-Ngai Hung, Kwok-Yung Yuen.
562 Seroprevalence of SARS-CoV-2 in Hong Kong and in residents evacuated
563 from Hubei province, China: a multicohort study. Lancet Microbe. 2020.
564 26. Liu M, Cheng SZ, Xu KW, Yang Y, Zhu QT, Zhang H, et al. Use of personal
565 protective equipment against coronavirus disease 2019 by healthcare

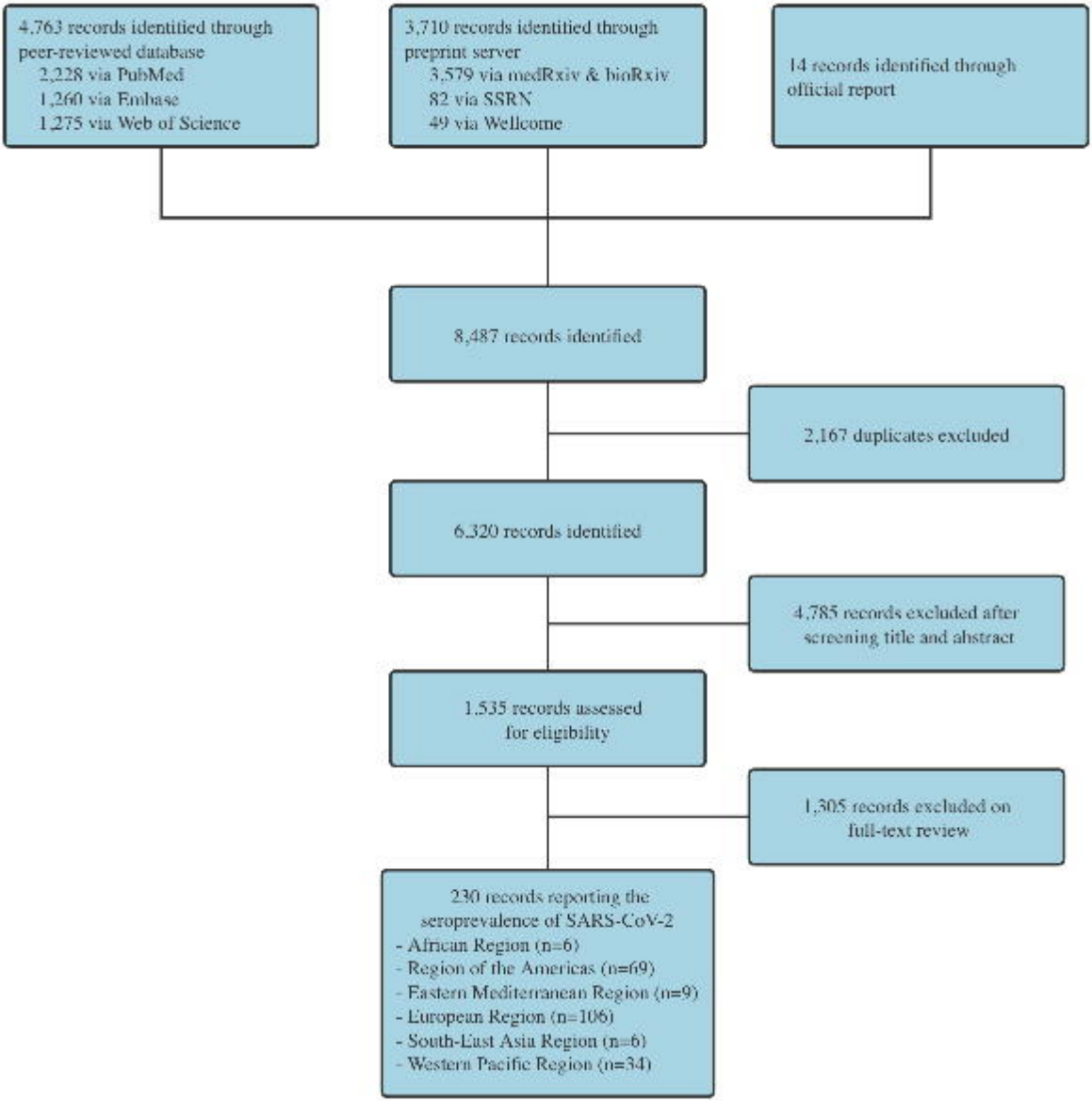
566 professionals in Wuhan, China: cross sectional study. Bmj.
567 2020;369:m2195.
568 27. Stringhini S, Wisniak A, Piumatti G, Azman AS, Lauer SA, Baysson H, et al.
569 Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland
570 (SEROCoV-POP): a population-based study. The Lancet. 2020.
571 28. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA,
572 Pérez-Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a
573 nationwide, population-based seroepidemiological study. Lancet. 2020.
574 29. Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Profiling Early Humoral
575 Response to Diagnose Novel Coronavirus Disease (COVID-19). Clin Infect
576 Dis. 2020.
577 30. Majiya H, Aliyu-Paiko M, Balogu VT, Musa DA, Salihu IM, Kawu AA, et al.
578 Seroprevalence of COVID-19 in Niger State. 2020:2020.08.04.20168112.
579 31. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P,
580 Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. N
581 Engl J Med. 2020;382(24):2302-15.
582 32. Del Fava E, Cimentada J, Perrotta D, Grow A, Rampazzo F, Gil-Clavel S, et al.
583 The differential impact of physical distancing strategies on social contacts
584 relevant for the spread of COVID-19. 2020:2020.05.15.20102657.
585 33. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in
586 contact patterns shape the dynamics of the COVID-19 outbreak in China.
587 Science. 2020.
588 34. Pawelec G, Weng NP. Can an effective SARS-CoV-2 vaccine be developed
589 for the older population? Immun Ageing. 2020;17:8.
590

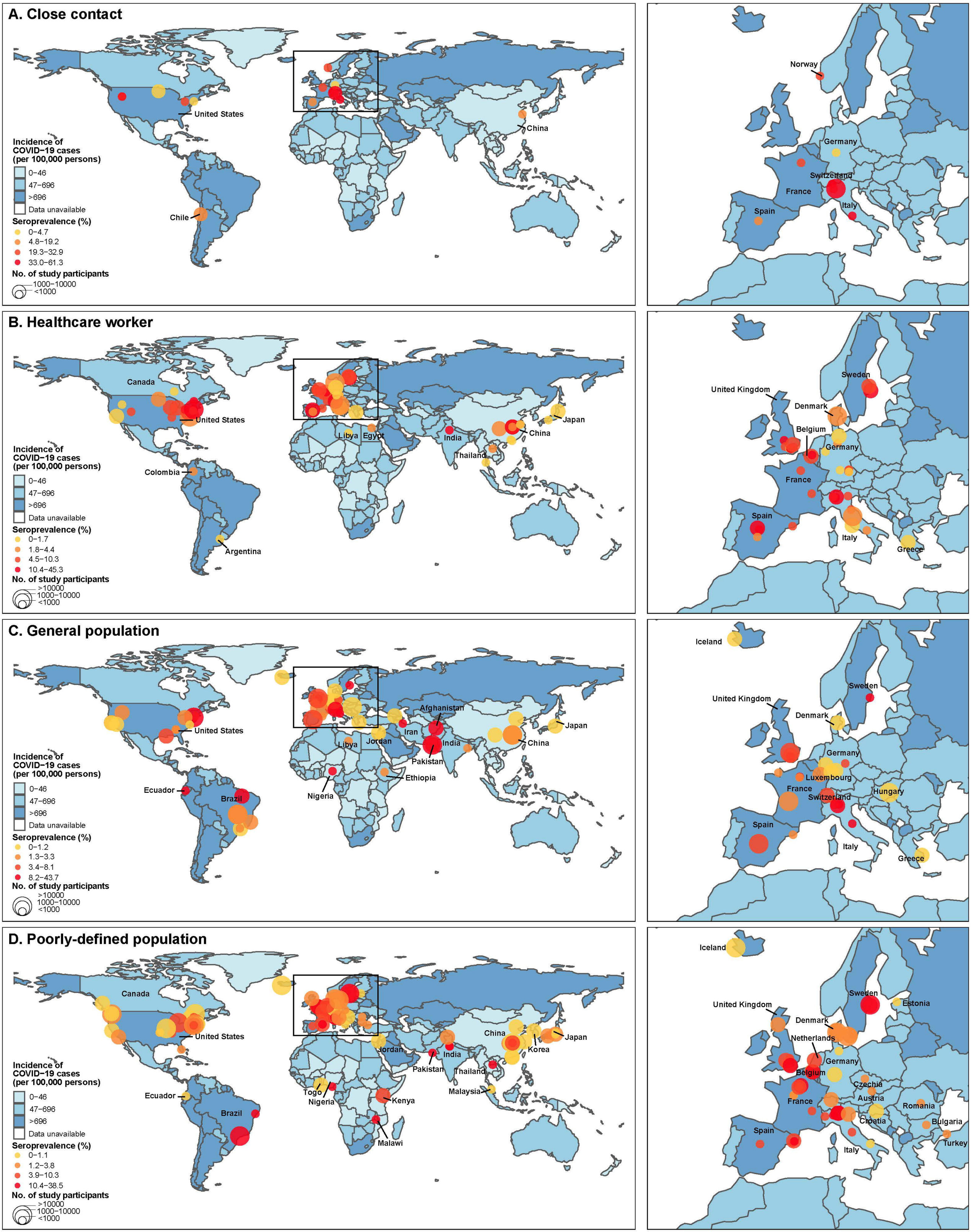
591 **Figure Legends**

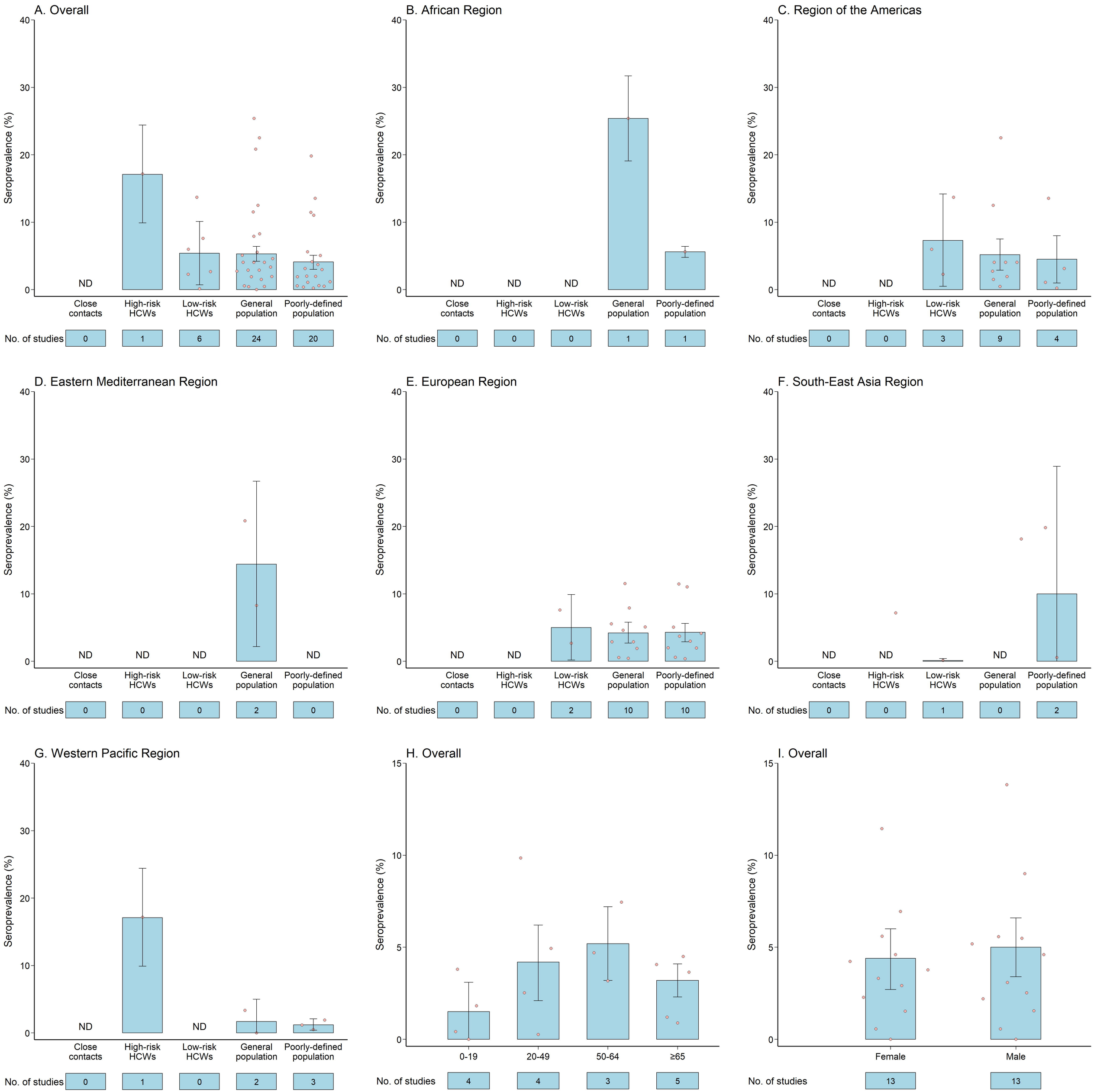
592 Figure 1. Flowchart of the selection of serological studies of SARS-CoV-2 infection,
593 December 2019-September 2020
594
595 Figure 2. Geographical distribution of SARS-CoV-2 serosurveys in humans by
596 study populations, December 2019-September 2020. The color of the map
597 indicates the cumulative incidence of reported cases (19) with darker colors
598 representing higher values. **(A)** Close contact. **(B)** Healthcare workers. **(C)**
599 General population. **(D)** Poorly-defined population
600
601 Figure 3. Estimated seroprevalence by WHO regions and study populations. **(A)**
602 Overall estimated seroprevalence by study populations; **(B)** Estimated
603 seroprevalence in African region; **(C)** Estimated seroprevalence in Region of the
604 Americas; **(D)** Estimated seroprevalence in Eastern Mediterranean region; **(E)**
605 Estimated seroprevalence in European region; **(F)** Estimated seroprevalence in
606 South-East Asia region; **(G)** Estimated seroprevalence in Western Pacific region;
607 **(H)** Overall estimated seroprevalence by age groups; **(I)** Overall estimated
608 seroprevalence by sexes (one study were not shown due to the y-axis limit); The
609 bar represents the pooled estimates. The vertical line represents the 95%CI. Each
610 dot represents the result of one single study.
611
612 Figure 4. Estimated ratio of human infections with SARS-CoV-2 per COVID-19

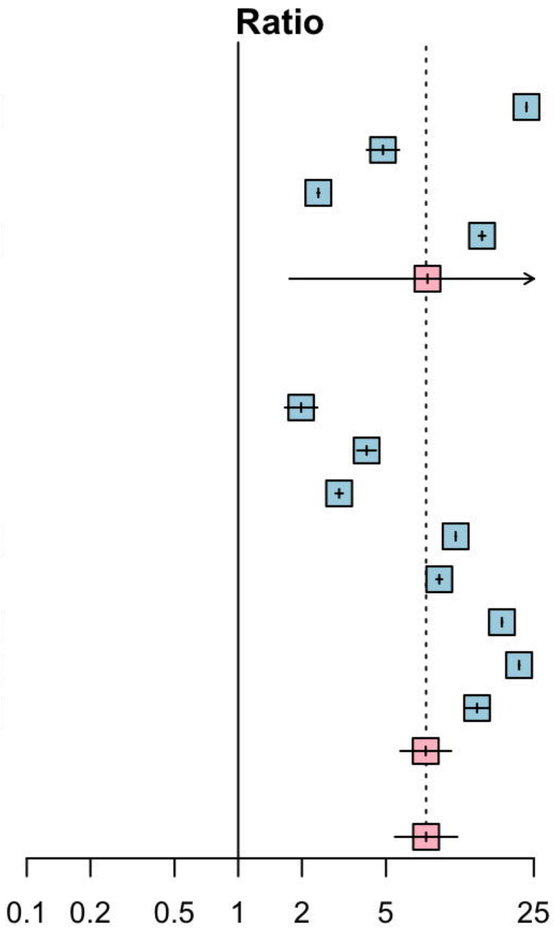
613 confirmed cases. The size of boxes represents the weight for each study. The

614 whisker represents the 95%CI.


4,763 records identified through peer-reviewed database
2,228 via PubMed
1,260 via Embase
1,275 via Web of Science
3,710 records identified through preprint server
3,579 via medRxiv & bioRxiv
82 via SSRN
49 via Wellcome
14 records identified through official report
8,487 records identified
2,167 duplicates excluded
6,320 records identified
4,785 records excluded after screening title and abstract
1,535 records assessed for eligibility
1,305 records excluded on full-text review
230 records reporting the seroprevalence of SARS-CoV-2
- African Region (n=6)
- Region of the Americas (n=69)
- Eastern Mediterranean Region (n=9)
- European Region (n=106)
- South-East Asia Region (n=6)
- Western Pacific Region (n=34)


A. Close contact
United States
China
Chile
Incidence of COVID−19 cases (per 100,000 persons)
0−46
47−696
>696
Data unavailable
Seroprevalence (%)
0−4.7
4.8−19.2
19.3−32.9
33.0−61.3
No. of study participants
1000−10000
<1000
Norway
Germany
Switzerland
France
Italy
Spain
B. Healthcare worker
Canada
United States
Colombia
Argentina
Libya
Egypt
India
Thailand
China
Japan
Incidence of COVID−19 cases (per 100,000 persons)
0−46
47−696
>696
Data unavailable
Seroprevalence (%)
0−1.7
1.8−4.4
4.5−10.3
10.4−45.3
No. of study participants
>10000
1000−10000
<1000
Sweden
United Kingdom
Denmark
Belgium
Germany
France
Spain
Italy
Greece
C. General population
United States
Ecuador
Brazil
Nigeria
Libya
Jordan
Iran
Afghanistan
India
Pakistan
Ethiopia
China
Japan
Incidence of COVID−19 cases (per 100,000 persons)
0−46
47−696
>696
Data unavailable
Seroprevalence (%)
0−1.2
1.3−3.3
3.4−8.1
8.2−43.7
No. of study participants
>10000
1000−10000
<1000
Iceland
Sweden
United Kingdom
Denmark
Germany
Luxembourg
France
Switzerland
Hungary
Spain
Italy
Greece
D. Poorly-defined population
Canada
United States
Ecuador
Brazil
Togo
Nigeria
Kenya
Malawi
Jordan
Pakistan
India
Thailand
Malaysia
China
Korea
Japan
Incidence of COVID−19 cases (per 100,000 persons)
0−46
47−696
>696
Data unavailable
Seroprevalence (%)
0−1.1
1.2−3.8
3.9−10.3
10.4−38.5
No. of study participants
>10000
1000−10000
<1000
Iceland
Sweden
Estonia
United Kingdom
Denmark
Netherlands
Germany
Belgium
Czechia
Austria
France
Romania
Croatia
Bulgaria
Spain
Turkey
Italy


A. Overall
Seroprevalence (%)
ND
Close contacts
High-risk HCWs
Low-risk HCWs
General population
Poorly-defined population
No. of studies
0
1
6
24
20
B. African Region
ND
ND
ND
No. of studies
0
0
0
1
1
C. Region of the Americas
ND
ND
No. of studies
0
0
3
9
4
D. Eastern Mediterranean Region
ND
ND
ND
ND
No. of studies
0
0
0
2
0
E. European Region
ND
ND
No. of studies
0
0
2
10
10
F. South-East Asia Region
ND
ND
ND
No. of studies
0
0
1
0
2
G. Western Pacific Region
ND
ND
No. of studies
0
1
0
2
3
H. Overall
0-19
20-49
50-64
≥65
No. of studies
4
4
3
5
I. Overall
Female
Male
No. of studies
13
13

Location (Author)	Infections	Cases	Ratio	95%CI
Region of the Americas				
Brazil (Hallal et al.)	4111354	178214	23.1	[23.0; 23.2]
USA, California, Marin (Appa et al.)	716	148	4.8	[4.1; 5.8]
USA, Connecticut (Mahajan et al.)	109383	45715	2.4	[2.4; 2.4]
USA, Georgia, DeKalb and Fulton (Biggs et al.)	45166	3176	14.2	[13.7; 14.7]
Overall			7.9	[1.7; 35.3]
European Region				
Denmark, Faroe Islands (Petersen et al.)	365	184	2.0	[1.7; 2.4]
Germany, Kreis Heinsberg, Gangelt (Streeck et al.)	1777	439	4.0	[3.7; 4.5]
Luxembourg (Snoeck et al.)	9816	3270	3.0	[2.9; 3.1]
Spain (Marina et al.)	2137063	200210	10.7	[10.6; 10.7]
Switzerland, Geneva (Stringhini et al.)	37830	4238	8.9	[8.6; 9.2]
UK, England (Office of National Statistics)	2791800	158078	17.7	[17.6; 17.8]
UK, England (Ward et al.)	3360000	157797	21.3	[21.2; 21.4]
UK, Jersey (Government of Jersey)	3300	245	13.5	[11.8; 15.3]
Overall			7.7	[5.8; 10.2]
Overall			7.7	[5.5; 10.9]

Test for overall effect: Z = 11.80 (P < 0.01)