International Journal of Epidemiology, 2022, 1–12
https://doi.org/10.1093/ije/dyac178
Original article

Original article

Limitations introduced by a low participation rate of SARS-CoV-2 seroprevalence data

Olivia Pluss ,[1] Harlan Campbell,[2] Laura Pezzi,[3] Ivonne Morales,[4,5] Yannik Roell,[1] Talia M Quandelacy,[6] Rahul Krishan Arora,[7,8] Emily Boucher,[9,10] Molly M Lamb,[1,6] May Chu,[1] Till Bärnighausen[5] and Thomas Jaenisch,[1,5,6,*]

[1]Center for Global Health, Colorado School of Public Health, University of Colorado, Aurora, CO, USA, [2]Department of Statistics, University of British Columbia, Vancouver, Canada, [3]Unité des Virus Émergents (UVE: Aix-Marseille Univ-IRD 190-Inserm 1207), Marseille, France, [4]Division of Infectious Disease and Tropical Medicine, Center for Infectious Diseases, Heidelberg University Hospital, Heidelberg, Germany, [5]Heidelberg Institute for Global Health, Heidelberg University Hospital, Heidelberg, Germany, [6]Department of Epidemiology, Colorado School of Public Health, University of Colorado, Aurora, CO, USA, [7]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK, [8]Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada, [9]Cumming School of Medicine, University of Calgary, Calgary, AB, Canada and [10]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

*Corresponding author. Department of Epidemiology and Center for Global Health, Colorado School of Public Health, 13199 East Montview Boulevard, Suite 310, Mail Stop A090, Aurora, CO 80045, USA. E-mail: Thomas.Jaenisch@CUAnschutz.edu

Received 23 February 2022; Editorial decision 11 August 2022; Accepted 31 August 2022

Abstract

Background: There has been a large influx of COVID-19 seroprevalence studies, but comparability between the seroprevalence estimates has been an issue because of heterogeneities in testing platforms and study methodology. One potential source of heterogeneity is the response or participation rate.

Methods: We conducted a review of participation rates (PR) in SARS-CoV-2 seroprevalence studies collected by SeroTracker and examined their effect on the validity of study conclusions. PR was calculated as the count of participants for whom the investigators had collected a valid sample, divided by the number of people invited to participate in the study. A multivariable beta generalized linear model with logit link was fitted to determine if the PR of international household and community-based seroprevalence studies was associated with the factors of interest, from 1 December 2019 to 10 March 2021.

Results: We identified 90 papers based on screening and were able to calculate the PR for 35 out of 90 papers (39%), with a median PR of 70% and an interquartile range of 40.92; 61% of the studies did not report PR.

Conclusions: Many SARS-CoV-2 seroprevalence studies do not report PR. It is unclear what the median PR rate would be had a larger portion not had limitations in reporting.

Low participation rates indicate limited representativeness of results. Non-probabilistic sampling frames were associated with higher participation rates but may be less representative. Standardized definitions of participation rate and data reporting necessary for the PR calculations are essential for understanding the representativeness of seroprevalence estimates in the population of interest.

Key words: Response rate, participation rate, COVID-19, SARS-CoV-2, seroprevalence, serology, sero-surveys, validity, representativeness, diagnostic testing

Key Messages

- There is an urgent need for seroprevalence studies to be comparable across geography and over time.
- We carried out a review of anti-SARS-CoV-2 seroprevalence surveys from 1 December 2019 to 10 March 2021, in collaboration with SeroTracker [www.serotracker.com].
- Our results show that out of 706 seroprevalence studies, only 35 (39%) reported participation rates. Participation rates varied between 0.43% and 96.38%, with a mean of 63% and a median of 70%. Only about half of these studies (54%) showed participation rates above 60%.
- Within the 39% that reported participation rates, we found that: (i) probabilistic sampling frames were associated with lower participation rates; and (ii) studies conducted in North America or Europe exhibited lower participation rates relative to studies conducted in Asia.

Introduction

As SARS-CoV-2 has spread globally, there has been considerable effort to understand the prevalence of SARS-CoV-2 antibodies in the population. SARS-CoV-2 seroprevalence studies can help estimate the cumulative burden of infection in each population,[1] help researchers better understand the natural history of the disease[2] and allow a more precise estimate of COVID-19's actual lethality.[3] Therefore, seroprevalence studies represent vital information for public health practitioners and decision makers.[4]

As of December 2021, over 2660 seroprevalence surveys have been published in the literature[5]; for a living systematic review of SARS-CoV-2 seroprevalence studies, see [https://serotracker.com/en/Explore]. When performing a seroprevalence survey, ensuring that the sample of individuals tested is representative of the (target) population of interest is vital for obtaining reliable and unbiased seroprevalence estimates.[6,7] There are three concepts to be reconciled: (i) the (target) population intended to be studied; (ii) the sampling frame to represent this target population; and (iii) the actual sample (of the target population) that was taken. Consequently, this often leads to a substantial difference between the target population that researchers would ideally like to include in a seroprevalence study and the actual population sample that is serologically investigated, the participation rate being the difference between the sample taken and the sampling frame. More concretely, this difference may be caused by: (i) individuals who could not be successfully contacted (e.g. did not answer the phone call); (ii) those who did not consent to participate (e.g. not interested); and (iii) those for whom a valid serological sample was not available (e.g. did not attend the visit). Other reasons for differences between the target population and the sample are not discussed here in more detail. However, they have in common that the results of a given seroprevalence study may not be representative of the target population. If representativeness cannot be assured, seroprevalence estimates may be adjusted by an appropriate weighting of the sample observations, i.e. by post-stratification.[8,9] Sufficient and appropriate adjustment is only possible if the systematic differences between the study sample and the population of interest are known and measured (either specifically or by proxy). This, for example, is difficult for non-responder bias as the information on systematic differences between responders and non-responders is often not available.

The definitions and terminology to indicate who participated in a survey vary in the literature. Due to the lack of standardized definitions, these diverse terms are often used interchangeably to describe different rates: response rate, participation rate and contact rate.[10] These nuances make it imperative for researchers to clearly define their response

and participation rate calculations, as the heterogeneous equations can yield different results.

This paper reviews and analyses seroprevalence studies published during the first year of the COVID-19 pandemic (until 10 March 10 2021), as identified by the SeroTracker project.[5] The aim was to determine participation rates achieved in SARS-CoV-2 seroprevalence studies, following the definition presented by the American Association for Public Opinion Research (AAPOR) (2015): 'the number of respondents who have provided a usable response divided by the total number of initial personal invitations requesting participation', with the intention to motivate further research and enquiry into which factors may be associated with higher or lower seroprevalence study participation rates.

Methods

This study used data compiled by SeroTracker. The SeroTracker study database compiles a wide range of seroprevalence data, including: published and unpublished academic studies in databases (PubMed, medRxiv, bioRxiv, World Health Organization International Clinical Trials Registry Platform) and high-impact medical journals [the *British Medical Journal* (*BMJ*), the *Journal of the American Medical Association* (*JAMA*), the *New England Journal of Medicine* (*NEJM*), the *Lancet*, *Annals of Internal Medicine*]; reports by governments, non-governmental organizations and health systems; and media reports via Google News.[11] In addition, we assessed the comprehensive list of 706 electronic data records (i.e. studies or cohorts) compiled by the SeroTracker living systematic review from 1 December 2019 to 10 March 2021.

We restricted our analysis to seroprevalence studies of the general population rather than of those that obtained samples from more specific target populations (e.g. studies involving only health care workers). Among general population studies, we focused on household and community samples to evaluate seroprevalence studies that reflect the broader population at risk and have comparable study designs and recruitment methods. Because it was not possible to calculate the response rate for blood donor and residual sera cohorts, studies using these samples, as well as reports from news and media, institutional reports or from undefined source types, were also excluded from the analysis. For longitudinal studies with repeat sampling (including multiple sample frame dates), we chose the estimates of the first time point.

Two independent reviewers identified four factors of interest for each study: sampling frame (i), geographical macro-region (ii), sample size (iii) and population invited (iv). Discordant reviews were discussed and resolved. We considered the following.

i. The sampling method for each study was defined as either 'probability-based' or 'non-probability based'. Probability-based studies were those that SeroTracker flagged using either a 'simplified-probability' or 'stratified-probability' sampling frame. Non-probability-based studies included those flagged as having used convenience-based sampling methods.
ii. The geographical macro-regions where the seroprevalence study took place were defined as (a) Europe, (b) North America, (c) Central and South America (referred to as South America from now on), (d) Asia and the Middle East [defined bythe United Nations Statistics Division (UNSD),[12] referred to as Asia from now on]. There were no studies within the included dataset from Africa or Oceania.
iii. The sample size for each study was defined as the number of individuals who provided a valid blood sample used for serological analysis.
iv. The total population invited was defined as the total number of people eligible to be included in the study and invited to participate.

For each study, the participation rate was calculated: the count of people for whom the investigators had a valid serological sample (i.e. the sample size) divided by the count of all people in the original population invited to participate in the study. The denominator represents the total number of people who would participate in the study if all those targeted had been successfully contacted, agreed to participate and provided a blood sample as suggested by the AAPOR standards.[10] If a paper self-reported a participation rate, it was re-calculated according to this definition if the necessary information was available.

We calculated Manski bounds to display the extreme range of potential true seroprevalence rates given non-responder bias. The lower Manski bound represents the scenario where all non-responders tested negative, and the upper Manski bound where all non-responders tested positive (in addition to those negative or positive included in the study). In addition, we calculated 95% confidence intervals (CI) for seroprevalence in studies where 95% CI was not reported by implementing the Clopper–Pearson method.

We employed a multivariable beta generalized linear model (with a logit link) to determine if the participation rate was associated with the four factors of interest (sampling frame, geographical macro-region, sample size and total population invited). Due to highly skewed data, the sample size was transformed into a logarithmic scale. The geographical region reference group was Asia. Studies for

which the participation rate could not be calculated (e.g. the numerator or the denominator or both not available) were excluded from the regression analysis. Predictors were selected a-priori. Statistical analysis was conducted using R software using the betareg package (10.18637/jss.v034.i02).

Results

From 706 records (multiple studies contribute more than one data point based upon cohorts and sampling time points) originally identified by SeroTracker up to 10 March 2021, 93 records were eligible for inclusion; 124 records were excluded due to being from news and media, institutional reports or undefined source types. A further 456 records were excluded due to having a sample frame other than community and household and 33 records that corresponded to multiple cohorts of the same study. We considered each study only once and deleted multiple records of the same study with the intention of a representative dataset. Another three records were further excluded during full-text screening due to the sampling frame being miscategorized in the original dataset.

Among the 90 records deemed eligible (individual studies with the exclusion of multiple cohorts), we calculated the participation rate for 35 studies [39%, interquartile range (IQR) 40.92]. The foremost reason the PR could not be calculated among the 90 eligible studies was that the denominator for PR could not be obtained. One recurring theme within the excluded studies was the difficulty of determining the denominator for household surveys with replacement. For example, when only one volunteer per household was included, the information on how many household members were invited to participate but declined (either or both being the total number of household residents within the sampling frame) was often not provided. Some studies provided participation rates on the household level but seroprevalence rates on the individual level. Some studies reported that when there was a refusal at a household level, the next household in the census tract listing would be surveyed until the recruitment target was reached, without reporting the proportion of non-responder households.

In addition, recruitment using both physical and digital open advertising (which included social media advertising, news articles, staff/students/faculty open e=mail invitation, neighbourhood advertisements and flyers) made it difficult to accurately define the total number of people who would have participated in the study if all those targeted had been successfully contacted. The results of this screening are presented in Figure 1.

The 35 studies included in the final analysis were conducted in 20 different countries, with 20 studies within Europe, four within North America, three within South America and eight within Asia. Participation rates could not be calculated from any study from Oceania and Africa. We present above a breakdown of the 55 studies that could not be included in the final review, which differed from the studies that were included (Table 1). The median sample size for seroprevalence studies included in the analysis was 1659 (range 186–105 651); 19 used a probability-based sampling method and 16 used a non-probability sampling method. The participation rate calculated from 35 studies varied extensively, ranging from 0.43% to 96.38%, with a mean of 63% and a median of 70% (Figure 2). More than half of the studies, 54%, showed participation rates above 60%. Detailed characteristics of the 35 included studies are available in Table 2.

Multivariable regression

Multivariable beta generalized linear regression (with logit link) exhibited that, with all other covariates held, constant non-probability sampling was associated with higher participation rates (parameter estimate of 0.86, 95% CI 0.33, 1.39, P-value = 0.002). Studies conducted in North America and in Europe showed lower participation rates relative to studies conducted in Asia [parameter estimate of -3.52, 95% CI -4.58, -2.47, P-value <0.001 (North America); and -1.38, 95% CI -2.06, -0.70, P-value <0.001 (Europe)] (Table 3). Finally, there was insufficient evidence of any association between log-sample size and participation rate (parameter estimate of 0.00, 95% CI -0.14, 0.14, P-value = 0.98). When we removed the two lowest data points, which appeared to be outliers,[41,42] and re-ran the analysis, we observed the same pattern.

In addition to 95% confidence intervals, we calculated Manski bounds for 35 SARS-CoV-2 seroprevalence studies (see Methods section). Upper Manski bounds varied between 5% and 100% and lower Manski bounds were between 0% and 40% (Figure 3). Visual inspection of the figure suggests the Manski bounds show a heterogenous picture with a slight trend to narrower Manski bounds in the studies with higher seroprevalence estimates from Asia or South America. Conversely, many studies from Europe that exhibit lower seroprevalence estimates exhibit large Manski bounds.

Discussion

During the first year of the COVID-19 pandemic, seroepidemiological data were urgently needed. This motivated researchers around the world to rapidly conduct

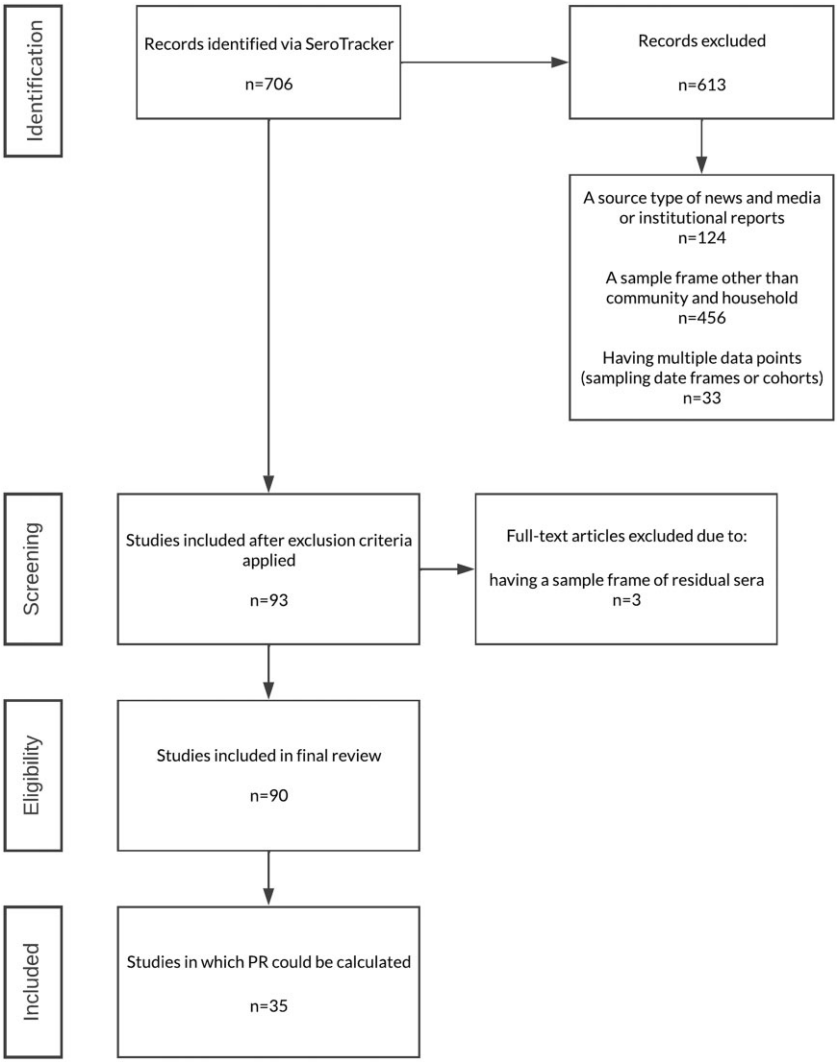

Figure 1 Flowchart of studies selected for inclusion in the analysis

seroprevalence surveys. Unfortunately, the hastiness with which these surveys were designed and conducted may have contributed to low-quality and biased seroprevalence estimates. We examined the availability and variability of studies reporting participation rates to highlight the need for more standardization for seroprevalence studies to be comparable across geography and over time. Our results show that out of 706 seroprevalence studies, only 35 (39%) reported participation rates. Approximately half of these studies (54%) showed participation rates above 60%. Low participation rates can potentially highlight problems with selection bias. Sufficient and appropriate adjustment to counteract bias, specifically selection bias, is only possible if specific factors are known and homogeneously reported.

Bobrovitz *et al.* (2021) applied a modified critical appraisal tool (the Joanna Briggs criteria[48] for bias assessment). As a result, only 36% of seroprevalence studies were considered at low or moderate risk of bias. Reasons for a high risk of bias included: not statistically correcting for demographics or test sensitivity and specificity; using non-probability sampling methods; or using non-representative sampling frames.[11] In another analysis, Chen *et al.* (2021), following their review of the literature (and using a scoring system developed based on a sero-epidemiological protocol from the Consortium for the Standardization of Influenza Sero-epidemiology), reached a similar conclusion: the overall methodological quality of SARS-CoV-2 seroprevalence studies is 'generally low'.[49]

Similar participation/response rate reviews have been conducted for various infectious diseases. For instance, Fritzell *et al.* (2018) conducted a scoping review of dengue, chikungunya and zika seroprevalence studies.[50] They concluded that only 17% of the studies provided response rates ranging from 40% to 100% (mean of 80%). However, Mosha *et al.* (2020), who conducted a review of HIV seroprevalence studies, obtained much higher estimates of

Table 1 Differences between studies where participation rate (PR) could be calculated vs not calculated

	Studies with information on PR ($N = 35$)	Studies without information on PR ($N = 55$)
Continent		
Africa	0	3
Asia	8	15
Europe	20	12
North America	4	15
South America	3	10
Sampling frame		
Convenience	5	10
Entire sample	5	1
Self-referral	4	7
Sequential	0	1
Simplified probability	7	7
Stratified non-probability	1	3
Stratified probability	12	23
Unclear	1	3
Sample size (median, IQR)	1659 (6894; 627–7521)	1545 (3615; 547–4162)
Seroprevalence (mean, SD)	11.47 (14.80)	10.52 (12.56)

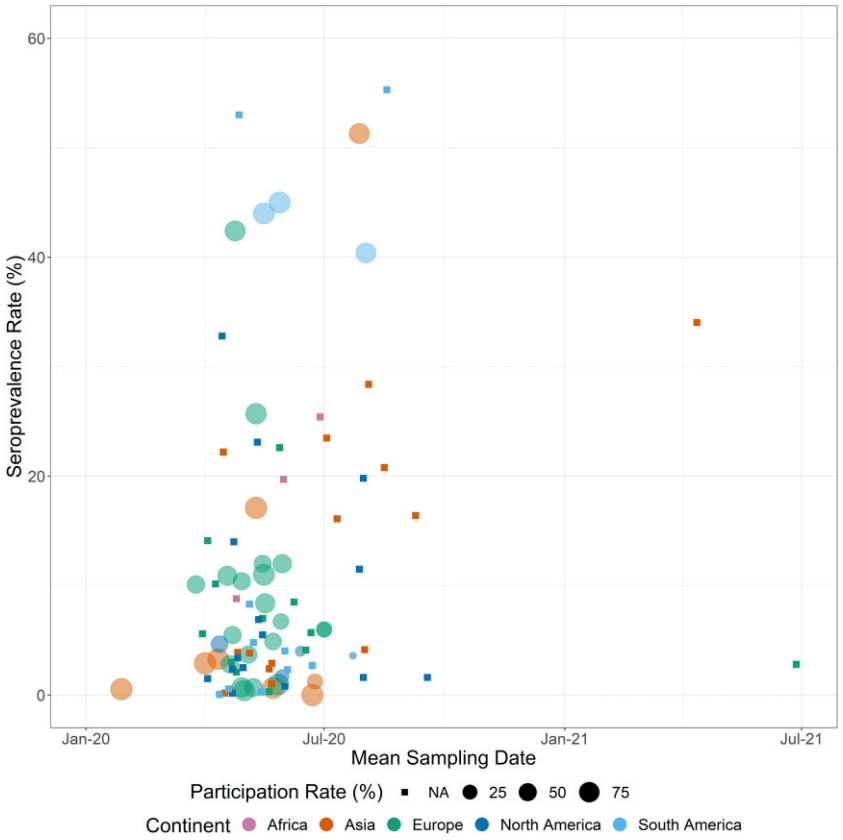

Figure 2 Seroprevalence rate by sampling date with point sizes representing participation rate and colour representing the origin of study. NA indicates studies without a participation rate reported

reported response rates: 21 out of 24 studies (88%) included for review reported a response rate, and these response rates ranged from 32% to 96% (with only two out of 21 studies reporting a response rate of less than 70%).[51]

The representativeness of survey findings in relation to the target population requires high-quality survey estimates and minimizing bias.[52,53] Improving the participation rate of a survey can increase external validity and reduce non-

Table 2 Characteristics of 35 seroprevalence studies were included in the statistical analysis. (Asterix indicates the studies where 95% CI was not reported and was calculated by us)

Author	Sampling start date	Continent	Seroprevalence (%) [95% CI]	Sample size	Probability sampling frame	Participation rate (%)
Ghose A *et al*., 2020[13]	20 July 2020	Asia	51.30 [39.90, 62.40]	1659	Yes	79.4
Hallowell BD *et al*., 1998[14]	28 January 2020	Asia	0.54 [0.01, 2.96] *	186	No	95.4
Ling R *et al*., 2020[15]	25 March 2020	Asia	3.27 [3.02, 3.52]	18391	No	87.9
Murhekar MV *et al*., 2020[16]	11 May 2020	Asia	0.64 [0.30, 0.99]	28000	Yes	92.5
Nawa N *et al*., 2020[17]	14 June 2020	Asia	1.23 [0.17, 2.28]	742	Yes	32.4
Poustchi H *et al*., 2020[18]	17 April 2020	Asia	17.10 [14.60, 19.50]	8902	Yes	95.2
Qutob N *et al*., 2020[19]	15 June 2020	Asia	0.00 [0.00, 0.00]	1319	Yes	94.6
To *et al*., 2020[20]	4 March 2020	Asia	2.88 [1.54, 4.87] *	452	No	96.4
Alessi D *et al*., 2020[21]	23 May 2020	Europe	4.90 [4.33, 5.55]	4987	No	45.5
Aziz NA *et al*., 2020[22]	24 April 2020	Europe	0.97 [0.72, 1.30]	4755	No	87.6
Bognanni A *et al*., 2021[23]	29 April 2020	Europe	0.41 [0.10, 1.38] *	634	No	84.8
Carrat F *et al*., 2020[24]	4 May 2020	Europe	6.72 [6.57, 6.87] *	104001	Yes	36.0
Cito F *et al*., 2020[25]	18 April 2020	Europe	10.90 [8.80, 13.50]	667	No	69.6
Fontanet A *et al*., 2020[26]	28 April 2020	Europe	10.40 [7.92, 13.17] *	552	No	50.3
Knabl L *et al*., 2020[27]	21 April 2020	Europe	42.40 [39.80, 44.70]	1473	No	78.9
Merkely B *et al*., 2020[28]	1 May 2020	Europe	0.68 [0.50, 0.86]	10474	Yes	58.9
Montenegro P *et al*., 2021[29]	21 April 2020	Europe	5.47 [3.44, 8.58]	311	Yes	51.8
Petersen MS *et al*., 2020[30]	27 April 2020	Europe	0.70 [0.32, 1.46] *	1075	Yes	71.7
Pollán M *et al*., 2020[31]	27 April 2020	Europe	3.70 [3.30, 4.00]	51958	Yes	50.7
Roxhed N *et al*., 2020[32]	1 April 2020	Europe	10.10 [7.30, 12.90]	529	Yes	52.9
Royo-Cebrecos C *et al*., 2020[33]	4 May 2020	Europe	11.00 [10.77, 11.23] *	70389	No	92.7
Santos-Hövener C *et al*., 2020[34]	20 May 2020	Europe	12.00 [10.40, 14.00]	2203	Yes	62.3
Stefanelli P *et al*., 2021[35]	5 May 2020	Europe	25.70 [24.60, 26.81] *	6075	No	84.3
Vos ERA *et al*., 2021[36]	31 March 2020	Europe	2.80 [2.10, 3.70]	3207	Yes	53.0
Ward H *et al*., 2020[37]	20 June 2020	Europe	5.96 [5.78, 6.14]	99908	Yes	31.7
Ward H *et al*., 2020[38]	20 June 2020	Europe	6.00 [5.80, 6.10]	105651	Yes	33.5
Weis S *et al*., 2020[39]	12 May 2020	Europe	8.39 [6.33, 10.85] *	620	No	70.2
Wells PM *et al*., 2020[40]	27 April 2020	Europe	12.00 [9.10, 15.20]	431	No	48.2
Feehan AK *et al*., 2020[41]	15 July 2020	North America	3.60 [2.80, 4.40]	2138	Yes	0.4
Mahajan S *et al*., 2021[42]	4 June 2020	North America	4.00 [2.00, 6.00]	567	Yes	5.7
Sood N *et al*., 2020[43]	10 April 2020	North America	4.65 [2.52, 7.07]	865	Yes	44.3
Tang X *et al*., 2021[44]	15 May 2020	North America	1.70 [1.44, 1.98] *	8967	Yes	20.3
da Silva AAM *et al*., 2020[45]	27 July 2020	South America	40.40 [35.60, 45.30]	3156	Yes	77.4
Del Brutto OH *et al*., 2021[46]	1 May 2020	South America	44.00 [38.36, 49.52] *	319	No	90.6
Del Brutto OH *et al*., 2021[47]	25 May 2020	South America	45.00 [41.22, 48.87] *	673	No	92.2

Table 3 Multivariable beta generalized linear regression on participation rate, adjusted for sampling frame, geography and sample size

Covariate	Parameter estimates	95% CI
Intercept	1.42	[0.14, 2.70]
Non-probability sampling vs probability sampling	0.86	[0.33, 1.39]
South America vs Asia	−0.38	[−1.48, 0.71]
Europe vs Asia	−1.38	[−2.06, −0.70]
North America vs Asia	−3.52	[−4.58, −2.47]
Log—sample size	0.00	[−0.14, 0.14]

responder bias, thus providing more representative estimates. The median participation rate of the studies reviewed was 69% (IQR 40.92). However, participation rates could only be calculated in a subset of 39% of eligible studies. In another study, Franceschi *et al.* reported that 59% of COVID-19 population-based seroprevalence studies reported

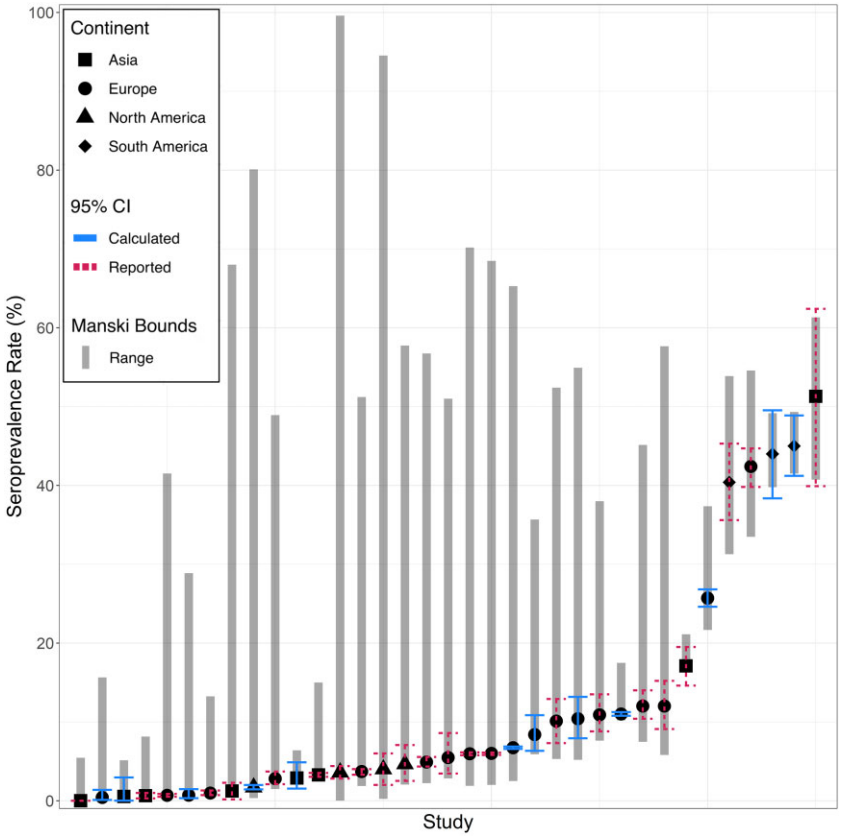

Figure 3 Anti-SARS-CoV-2 seroprevalence (ordered from low to high) in 35 studies, with corresponding 95% confidence intervals (CI) [either reported within the study (dashed line) or calculated (solid line)] and Manski bounds

an 'adequate' response rate of >70%.[54] We estimate the proportion of studies with a participation rate above 70% to be 46%. The small sample size was a limitation of our results and those of Franceschi *et al.* (2021). Franceschi *et al.* examined the overall risk of bias scores of the identified seroprevalence studies; our research further built upon this by examining participation rates more closely. These participation rates are of concern because the potential for selection bias is higher when participation/response rates are lower.[55]

The method of recruitment and the transparency of reporting within the studies impacted on our ability to ascertain and calculate PR. Due to the nature of COVID-19, data collection strategies have been affected, increasing online survey strategies and survey fatigue of participants.[56] For those studies where the PR could not be calculated, we identified two main underlying reasons. These are: (i) recruitment with a replacement on the household level; the number of target households or schools is indicated, without specifying the number of individuals (inhabitants or students);[57,58] and (ii) use of social media or similar for recruitment; when authors did not use a direct method to enrol participants (e-mail, phone call) but instead used advertising on social media, it is impossible to know the number of people who were invited/reached.[41,59–61]

We found notably high or low participation rates in a limited number of papers. Three studies reported a PR above 95%. Two papers[14,20] shared a similar study population: both investigated residents who were evacuated from Hubei (China) in the first months of 2020 and serologically tested after the flights/travel. The imposed quarantine status, and the recent stay in the province considered the epicentre of the SARS-CoV-2 epidemic, likely were factors that could have encouraged invited people to provide blood specimens for testing, being interested to know their serological status. Moreover, the small number of people invited[14,20] potentially enabled higher participation rates (targeted invitations, more time to dedicate to each person to answer questions, making it easier to communicate serology status results to participants after testing). In the third paper, with a PR above 95%,[18] the authors reported that all people invited initially agreed to participate (telephone response rate of 100%) and that only a small part of them rectified their choice (final PR 95.17). This example demonstrates that having a high participation rate is possible even in a relatively large study

population (3709 people from the general population contacted for this survey). Two studies reported a PR below 10%.[41,42] In the study conducted by Feehan *et al.* (2020), the low PR is likely due to the method used for participant recruitment (digital ads targeting 500 000 people). For Mahajan *et al.* (2021), it seems more complicated to identify the reason for such a low PR, despite considerable efforts to maximize the inclusion of harder-to-reach individuals: a dual-frame contact method (landline and cell phone) and several attempts (five) to each randomly selected telephone number, spread over different days of the week.

In this review, we observed that studies that used probability-based sampling frames obtained lower participation rates than those that did not. However, non-probabilistic-based sampling studies are more prone to selection bias.[62] We also observed substantial variation in participation rates by region, with studies conducted in Europe or North America showing on average lower participation rates.

A limitation of this review is the small sample size. We were only able to investigate the relationship between participation rates and four factors of interest (sampling method, macro-region, sample size and population invited). Numerous other aspects could potentially have an impact on participation rates, such as the method of contact and enrolment, amount of information provided to participants and the dissemination method, the type of sampling (self-sampling or sample collection in health facilities), possible remuneration and demographic characteristics of the population sampled, to name a few. The method of contact and enrolment impacts on the representativeness of individuals participating in the study. Studies that rely on online convenience samples with self-selected participants are subject to self-selection and non-coverage bias[63] due to factors such as individuals without access to the internet, age, personality traits and political attitudes. Given the goal of having a representative study, these methods frequently do not reflect the best approach. The relatively small sample size also applies to the publication by Franceschi *et al.* (2021), who examined the overall risk of bias scores of the identified seroprevalence studies; our research further built upon this by examining participation rates more closely. These participation rates are of concern because the potential for selection bias is higher when participation/response rates are lower.[54] Due to the limited sample size, we were unable to include more studies from Africa, which warrants further investigation.

The representativeness of seroprevalence studies depends on whether the sample captures the sample frame, which depends on the sampling strategy (i.e. probability vs non-probability), and sample size, among other factors. We included studies that used a household or community sampling frame, which by design are more representative of the general population than studies that use proxy sampling frames to estimate general population seroprevalence (e.g. blood donors, residual sera). However, the lack of detailed information for each study precluded a more detailed assessment of representativeness.

Researchers should consider these aspects to address potential selection and enrolment bias areas, facilitate participation and report methodologically relevant information. For example, communicating with participants about their responses' importance and value encourages them to feel personally invested in the study[64]; non-response rates are shown to be influenced by a lack of participant investment.[65] In addition, providing survey results allows the opportunity for feedback and dialogue between the participant and researcher, which could help improve the project's efficacy.

Historically, there has been a general lack of consensus regarding best practices for defining and calculating response and participation rates, occasionally leading to overly optimistic rates and estimations.[66] One of the earliest attempts to establish a standard response rate estimation was presented in response to an absence of consensus among colleagues at the University of Illinois Survey Research Laboratory,[67] with other subsequent published attempts.[52] Yet, Spaeth (1992) found in a survey of 38 research organizations that the wording for how the response rate was calculated was not harmonized nor consistent.[68,69] As evidence of the perceived importance of response rates, authors have attempted to determine benchmarks for response rates by examining the average response rate across a body of research.[70,71] In 1998, the American Association for Public Opinion Research (AAPOR) first published a set of standard definitions and formulas, including calculation of response, refusal and contact rates (multiple revised versions have been released[10]). Most recently, in response to the heterogeneity in the quality of seroprevalence studies for SARS-CoV-2, the World Health Organization (WHO) Seroepidemiology Technical Working Group created ROSES-S: Reporting of Seroepidemiologic studies-SARS-CoV-2, a checklist and criteria of items that should be included in study reporting, including a requirement to report the numbers of individuals at each stage in the study as well as giving reasons for non-participation at each stage.[72]

Conclusion

In summary, it is of outstanding importance that authors provide sufficient information to calculate PR and provide clear information on their inclusion and exclusion criteria

for the sampled populations and recruitment methods, including comparative data of responders and non-responders (or participants and non-participants) and clear metrics of the total population initially invited for participation, to determine the extent of potential non-response or non-participation bias in their results. In addition, we emphasize the importance of the scientific community harmonizing and standardizing the definitions for participation and response rates.

Ethics approval

We confirm that such approval is not needed as we did not conduct research on human subjects. All data are publicly available from published studies. We only used aggregate data in the manuscript. This manuscript is the authors' original work and research.

Data availability

The data underlying this article are available in the article provided by SeroTracker [www.serotracker.com].

Author contributions

The authors confirm contributions to the manuscript as follows: R.A., M.C., T.B. and T.J. developed the conceptualization for the manuscript. O.P., H.C. and E.B. carried out the data curation. H.C., Y.R. and M.L. performed the formal analysis. T.J. was responsible for funding acquisition. O.P., H.C., L.P., I.M. and E.B. conducted the research investigation, H.C., L.P., Y.R., T.Q., R.A., M.L. and T.B. the development of the methodology. O.P., M.C. and T.J. functioned as project administration. I.M., T.Q. and E.B. provided computing and data resources. R.A., M.C., T.B. and T.J. conducted the supervision and O.P., H.C., L.P., I.M. and T.Q. were responsible for validation or results, and research outputs. O.P., H.C., Y.R. and T.J. created and presented visualizations. O.P., H.C. and T.J. wrote the original draft. All authors reviewed the results and were critical in reviewing and editing the final draft, including pre- and post-publication stages.

Funding

This work is supported by the ReCoDID study, which has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 825746 and the Canadian Institutes of Health Research, Institute of Genetics (CIHR-IG) under Grant Agreement N. 01886–000. R.A. and E.B. receive funding for the SeroTracker SARS-CoV-2 seroprevalence study evidence synthesis project from the Public Health Agency of Canada through Canada's COVID-19 Immunity Task Force, the World Health Organization Health Emergencies Programme, the Robert Koch Institute and the Canadian Medical Association Joule Innovation Fund. No funding source had any role in the design of this study, its execution, analyses, interpretation of the data or decision to submit results. This manuscript does not necessarily reflect the views of the World Health Organization or any other funder.

Acknowledgements

We would like to acknowledge the SeroTracker team members for database development and data extraction, as well as Judy Chen of McGill University for specific contributions to the database for this project and Christian Cao for his contributions in preparing the data.

Conflict of interest

None declared.

References

1. Chen X, Chen Z, Azman AS *et al.* Serological evidence of human infection with SARS-CoV-2: a systematic review and meta-analysis. *Lancet Glob Health* 2021;**9**:e598–609.
2. Oran DP, Topol EJ. The proportion of SARS-CoV-2 infections that are asymptomatic: a systematic review. *Ann Intern Med* 2021;**174**:655–62.
3. Campbell H, Gustafson P. Inferring the COVID-19 IFR with a simple Bayesian evidence synthesis of seroprevalence study data and imprecise mortality data. *medRxiv* 2021. doi: 10.1101/2021.05.12.21256975, preprint: not peer reviewed.
4. Clapham H, Hay J, Routledge I *et al.* Seroepidemiologic study designs for determining SARS-COV-2 transmission and immunity. *Emerg Infect Dis* 2020;**26**:1978–86.
5. Arora RK, Joseph A, Van Wyk J *et al.* SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. *Lancet Infect Dis* 2021; **21**:e75–76.
6. Draugalis JR, Coons SJ, Plaza CM. Best practices for survey research reports: a synopsis for authors and reviewers. *Am J Pharm Educ* 2008;**72**:11.
7. Elwood JM. Commentary: on representativeness. *Int J Epidemiol* 2013;**42**:1014–15.
8. Brick JM, Kalton G. Handling missing data in survey research. *Stat Methods Med Res* 1996;**5**:215–38.
9. Downes M, Gurrin LC, English DR *et al.* Multilevel regression and poststratification: a modeling approach to estimating population quantities from highly selected survey samples. *Am J Epidemiol* 2018;**187**:1780–90.
10. AAPOR TAA for POR. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 8th edn. 2015. www.aapor.org/AAPOR_Main/media/MainSiteFiles/Standard-Definitions 2015_8thEd.pdf (23 August 2022, date last accessed).
11. Bobrovitz N, Arora RK, Cao C *et al.* Global seroprevalence of SARS-CoV-2 antibodies: a systematic review and meta-analysis. *PLoS One* 2021;**16**:e0252617.
12. UNSD. *Methodology*. https://unstats.un.org/unsd/methodology/m49/ (23 August 2022, date last accessed).
13. Ghose A, Bhattacharya S, Karthikeyan AS *et al.* Community prevalence of antibodies to SARSCoV-2 and correlates of protective immunity in an Indian metropolitan city. *medRxiv* 2020. doi: 10.1101/2020.11.17.20228155, preprint: not peer-reviewed.
14. Hallowell BD, Carlson CM, Jacobs JR *et al.* Severe acute respiratory syndrome coronavirus 2 prevalence, seroprevalence, and exposure among evacuees from Wuhan, China, 2020. *Emerg Infect Dis* 2020;**26**:1998–2004.

15. Ling R, Yu Y, He J *et al.* Seroprevalence and epidemiological characteristics of immunoglobulin M and G antibodies against SARS-CoV-2 in asymptomatic people in Wuhan, China. *medRxiv* 2020. doi:10.1101/2020.06.16.20132423, preprint: not peer-reviewed.
16. Murhekar MV, Bhatnagar T, Selvaraju S *et al.* Prevalence of SARS-CoV-2 infection in India: findings from the national serosurvey, May-June 2020. *Indian J Med Res* 2020;**152**:48–60.
17. Nawa N, Kuramochi J, Sonoda S *et al.* Seroprevalence of SARS-CoV-2 IgG Antibodies in Utsunomiya City, Greater Tokyo, after first pandemic in 2020 (U-CORONA): a household-and population-based study. *medRxiv* 2020. doi:10.1101/2020.07.20.20155945; preprint: not peer-reviewed.
18. Poustchi H, Darvishian M, Mohammadi Z *et al.* SARS-CoV-2 antibody seroprevalence in the general population and high-risk occupational groups across 18 cities in Iran: a population-based cross-sectional study. *Lancet Infect Dis* 2021;**21**:473–81.
19. Qutob N, Awartani F, Salah Z *et al.* Seroprevalence of SARS-CoV-2 in Palestine: a cross-sectional seroepidemiological study. *medRxiv* 2020. doi:10.1101/2020.08.28.20180083, preprint: not peer-reviewed.
20. To KKW, Cheng VCC, Cai JP *et al.* Seroprevalence of SARS-CoV-2 in Hong Kong and in residents evacuated from Hubei province, China: a multicohort study. *Lancet Microbe* 2020;**1**:e111–18.
21. Alessi D, Borré S, Barale A. La prevalenza della risposta immunitaria a SARS-CoV-2 nella popolazione di Borgosesia (VC): una strategia di sorveglianza in fase post-lockdown?[Seroprevalence of anti-SARS-CoV-2 IgG/IgM antibodies in Borgosesia (Piedmont Region, Northern Italy) population: a surveillance strategy in post-lockdown period?]. *Epidemiol Prev* 2020;**44**:5–6.
22. Aziz NA, Corman VM, Echterhoff AK *et al.* Seroprevalence and correlates of SARS-CoV-2 neutralizing antibodies: Results from a population-based study in Bonn, Germany. *medRxiv* 2020. doi:10.1101/2020.08.24.20181206, preprint: not peer-reviewed.
23. Bognanni A, Schiaffino A, Pimpinelli F *et al.* The Conundrum of Giglio Island: unraveling the dynamics of an apparent resistance to COVID-19 – a descriptive study. *medRxiv* 2021. doi:10.1101/2021.01.08.20248948, preprint: not peer-reviewed.
24. Carrat F, de Lamballerie X, Rahib D *et al.* Seroprevalence of SARS-CoV-2 among adults in three regions of France following the lockdown and associated risk factors: a multicohort study. *medRxiv* 2020. doi: 10.1101/2020.09.16.20195693, preprint: not peer-reviewed.
25. Cito F, Amato L, Di Giuseppe A *et al.* A COVID-19 hotspot area: activities and epidemiological findings. *Microorganisms* 2020;**8**:1711.
26. Fontanet A, Grant R, Tondeur L *et al.* SARS-CoV-2 infection in primary schools in northern France: a retrospective cohort study in an area of high transmission *medRxiv* 2020. doi:10.1101/2020.06.25.20140178, preprint: not peer-reviewed.
27. Knabl L, Mitra T, Kimpel J *et al.* High SARS-CoV-2 seroprevalence in children and adults in the Austrian ski resort Ischgl. *medRxiv* 2020. doi:10.1101/2020.08.20.20178533, preprint: not peer-reviewed.
28. Merkely B, Szabó AJ, Kosztin A *et al.*; HUNgarian COronaVirus-19 Epidemiological Research (H-UNCOVER) investigators. Novel coronavirus epidemic in the Hungarian population, a cross-sectional nationwide survey to support the exit policy in Hungary. *GeroScience* 2020;**42**:1063–74.
29. Montenegro P, Brotons C, Serrano J *et al.* Community seroprevalence of COVID-19 in probable and possible cases at primary health care centres in Spain. *Fam Pract* 2021;**38**:153–58.
30. Petersen MS, Strøm M, Christiansen DH *et al.* Seroprevalence of SARS-CoV-2–specific antibodies, Faroe Islands. *Emerg Infect Dis* 2020;**26**:2760–62.
31. Pollán M, Pérez-Gómez B, Pastor-Barriuso R *et al.* Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* 2020;**396**:535–44.
32. Roxhed N, Bendes A, Dale M *et al.* A translational multiplex serology approach to profile the prevalence of anti-SARS-CoV-2 antibodies in home-sampled blood. *medRxiv* 2020. doi:10.1101/2020.07.01.20143966, preprint: not peer-reviewed.
33. Royo-Cebrecos C, Vilanova D, López J *et al.* Mass SARS-CoV-2 serological screening for the Principality of Andorra. *Lancet Reg Health Eur* 2021;**5**:100119.
34. Santos-Hövener C, Neuhauser HK, Rosario AS *et al.*; CoMoLo Study Group. Serology- and PCR-based cumulative incidence of SARS-CoV-2 infection in adults in a successfully contained early hotspot (CoMoLo study), Germany, May to June 2020. *Eurosurveillance* 2020;**25**:2001752.
35. Stefanelli P, Bella A, Fedele G *et al.* Prevalence of SARS-CoV-2 IgG antibodies in an area of northeastern Italy with a high incidence of COVID-19 cases: a population-based study. *Clin Microbiol Infect* 2021;**27**:633.e1–e7.
36. Vos ERA, den HG, Schepp RM *et al.* Nationwide seroprevalence of SARS-CoV-2 and identification of risk factors in the general population of the Netherlands during the first epidemic wave. *J Epidemiol Community Health* 2021;**75**:489–95.
37. Ward H, Cooke G, Atchison CJ *et al.* Declining prevalence of antibody positivity to SARS-CoV-2: a community study of 365,000 adults. *medRxiv* 2020. doi:10.1101/2020.10.26.20219725, preprint: not peer-reviewed.
38. Ward H, Atchison CJ, Whitaker M *et al.* Antibody prevalence for SARS-CoV-2 in England following first peak of the pandemic: REACT2 study in 100,000 adults. *medRxiv* 2020. doi:10.1101/2020.08.12.20173690, preprint: not peer-reviewed.
39. Weis S, Scherag A, Baier M *et al.* Seroprevalence of SARS-CoV-2 antibodies in an entirely PCR-sampled and quarantined community after a COVID-19 outbreak: the CoNAN study. *medRxiv* 2020. doi:10.1101/2020.07.15.20154112, preprint: not peer-reviewed.
40. Wells PM, Doores KJ, Couvreur S *et al.* Estimates of the rate of infection and asymptomatic COVID-19 disease in a population sample from SE England. *J Infect* 2020;**81**:931–36.
41. Feehan AK, Velasco C, Fort D *et al.* Racial and workplace disparities in seroprevalence of SARS-CoV-2 in Baton Rouge, Louisiana, July 15–31, 2020. *medRxiv* 2020, doi:10.1101/2020.08.26.20180968, preprint: not peer-reviewed.
42. Mahajan S, Srinivasan R, Redlich CA *et al.* Seroprevalence of SARS-CoV-2-specific IgG antibodies among adults living in Connecticut: Post-infection prevalence (PIP) study. *Am J Med* 2021;**134**:526–34.e11.

43. Sood N, Simon P, Ebner P *et al.* Seroprevalence of SARS-CoV-2–specific antibodies among adults in Los Angeles County, California, on April 10-11, 2020. *JAMA* 2020;**323**:2425–27.
44. Tang X, Sharma A, Pasic M *et al. COVID Symptoms, Seroprevalence, and Mortality During the First Wave of SARS-CoV-2 in Canada*. Rochester, NY: Social Science Research Network, 2021. Report No.: ID 3752659. https://papers.ssrn.com/abstract=3752659 (23 August 2022, date last accessed).
45. da Silva AAM, Lima-Neto LG, de Azevedo CM *et al.* Population-based seroprevalence of SARS-CoV-2 is more than halfway through the herd immunity threshold in the State of Maranhao, Brazil. *medRxiv* 2020. doi:10.1101/2020.08.28.20180463, preprint: not peer-reviewed.
46. Del Brutto OH, Mera RM, Recalde BY, Costa AF. Social determinants of health and risk of SARS-CoV-2 infection in community-dwelling older adults living in a rural Latin American setting. *J Community Health* 2021;**46**:292–97.
47. Del Brutto OH, Costa AF, Mera RM, Recalde BY, Bustos JA, García HH. SARS-CoV-2 in Rural Latin America. A population-based study in coastal Ecuador. *Clin Infect Dis* 2021;**73**:314–17.
48. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. *Int J Evid Based Healthc* 2015;**13**:147–53.
49. Chen Y, Li L, Zhang Q *et al.* Epidemiology, methodological quality, and reporting characteristics of systematic reviews and meta-analyses on coronavirus disease 2019: a cross-sectional study. *Medicine (Baltimore)* 2021;**100**:e27950.
50. Fritzell C, Rousset D, Adde A, Kazanji M, Van Kerkhove MD, Flamand C. Current challenges and implications for dengue, chikungunya and Zika seroprevalence studies worldwide: a scoping review. *PLoS Negl Trop Dis* 2018;**12**:e0006533.
51. Mosha NR, Aluko OS, Todd J, Machekano R, Young T. Analytical methods used in estimating the prevalence of HIV/AIDS from demographic and cross-sectional surveys with missing data: a systematic review. *BMC Med Res Methodol* 2020;**20**:1–10.
52. Groves RM, Cialdini RB, Couper MP. Understanding the decision to participate in a survey. *Public Opin Q* 1992;**56**:475–95.
53. Ponto J. Understanding and evaluating survey research. *J Adv Pract Oncol* 2015;**6**:168–71.
54. Franceschi VB, Santos AS, Glaeser AB *et al.* Population-based prevalence surveys during the Covid-19 pandemic: a systematic review. *Rev Med Virol* 2021;**31**:e2200.
55. Bootsma-van der Wiel AV, Van Exel E, De Craen AJM *et al.* A high response is not essential to prevent selection bias: results from the Leiden 85-plus study. *J Clin Epidemiol* 2002;**55**:1119–25.
56. De Koning R, Egiz A, Kotecha J *et al.* Survey fatigue during the COVID-19 pandemic: an analysis of neurosurgery survey response rates. *Front Surg* 2021;**8**:690680.
57. Tess BH, Granato CF, Alves MC, *et al.* SARS-CoV-2 seroprevalence in the municipality of São Paulo, Brazil, ten weeks after the first reported case. *medRxiv* 2020. doi:10.1101/2020.06.29.20142331, preprint: not peer-reviewed.
58. Streeck H, Schulte B, Kümmerer BM *et al.* Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *Nat Commun* 2020;**11**:5829.
59. Bendavid E, Mulaney B, Sood N *et al.* Covid-19 antibody seroprevalence in Santa Clara County, California. *Int J Epidemiol* 2021;**50**:410–19.
60. McLaughlin CC, Doll MK, Morrison KT *et al.* High community SARS-CoV-2 antibody seroprevalence in a ski resort community, Blaine County, Idaho, US preliminary results. *medrxiv* 2020. doi:10.1101/2020.07.19.20157198, preprint: not peer-reviewed.
61. Zou J, Bretin A, Gewirtz A, Antibodies to SARS/CoV-2 in arbitrarily-selected Atlanta residents. *medRxiv* 2020. doi:10.1101/2020.05.01.20087478, preprint: not peer-reviewed.
62. Shook-Sa BE, Boyce RM, Aiello AE. Estimation without representation: early severe acute respiratory syndrome coronavirus 2 seroprevalence studies and the path forward. *J Infect Dis* 2020;**222**:1086–89.
63. Mercer AW, Kreuter F, Keeter S, Stuart EA. Theory and practice in nonprobability surveys: parallels between causal inference and survey inference. *Public Opin Q* 2017;**81**:250–71.
64. Clark T. We're over-researched here!' Exploring accounts of research fatigue within qualitative research engagements. *Sociology* 2008;**42**:953–70.
65. VanGeest JB, Johnson TP, Welch VL. Methodologies for improving response rates in surveys of physicians: a systematic review. *Eval Health Prof* 2007;**30**:303–21.
66. Johnson T, Owens L. Survey response rate reporting in the professional literature. In: *58th Annual Meeting of the American Association for Public Opinion Research, May 15-18 2003*. Nashville,TN; 2003.
67. Kviz FJ. Toward a standard definition of response rate. *Public Opin Q* 1977;**41**:265–67.
68. Campanelli PC. *The Impact of the Interviewer: Nonresponse and Response Variance in Social Surveys*. London: London School of Economics and Political Science, 1999.
69. Spaeth MA. Response rates at academic survey research organizations. *Surv Res* 1992;**23**:18–20.
70. Asch DA, Jedrziewski MK, Christakis NA. Response rates to mail surveys published in medical journals. *J Clin Epidemiol* 1997;**50**:1129–36.
71. Sitzia J, Wood N. Response rate in patient satisfaction research: an analysis of 210 published studies. *Int J Qual Health Care* 1998;**10**:311–17.
72. World Health Organization Seroepidemiology Technical Working Group. ROSES-S: Statement from the World Health Organization on the reporting of seroepidemiologic studies for SARS-CoV-2. *Influenza Other Respir Viruses* 2021;**15**:561–68.