



HAL
open science

Estimating the number of probable new SARS-CoV-2 infections among tested subjects from the number of confirmed cases

Y. M. Diarra, P. M. Wimba, P. B. Katchunga, J Bengehya, B Miganda, M Oyimangirwe, L Tshilolo, S. M. Ahuka, Jean Iwaz, Jean-François Étard, et al.

► To cite this version:

Y. M. Diarra, P. M. Wimba, P. B. Katchunga, J Bengehya, B Miganda, et al.. Estimating the number of probable new SARS-CoV-2 infections among tested subjects from the number of confirmed cases. *BMC Medical Research Methodology*, 2023, 23 (1), pp.272. 10.1186/s12874-023-02077-2 . hal-04382523

HAL Id: hal-04382523

<https://hal.science/hal-04382523>

Submitted on 21 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH

Open Access



Estimating the number of probable new SARS-CoV-2 infections among tested subjects from the number of confirmed cases

YM Diarra^{1,2,3,4*}, PM Wimba^{1,2,5,6,7†}, PB Katchunga^{5,6}, J Bengheya⁹, B Miganda⁸, M Oyimangirwe⁵, L Tshilolo⁹, SM Ahuka^{10,11}, J Iwaz^{1,2,3,4}, JF Étar^{12,13}, R Écochard^{1,2,3,4}, P Vanhems^{1,2,4,7,14} and M Rabilloud^{1,2,3,4}

Abstract

Objectives In most African countries, confirmed COVID-19 case counts underestimate the number of new SARS-CoV-2 infection cases. We propose a multiplying factor to approximate the number of biologically probable new infections from the number of confirmed cases.

Methods Each of the first thousand suspect (or alert) cases recorded in South Kivu (DRC) between 29 March and 29 November 2020 underwent a RT-PCR test and an IgM and IgG serology. A latent class model and a Bayesian inference method were used to estimate (i) the incidence proportion of SARS-CoV-2 infection using RT-PCR and IgM test results, (ii) the prevalence using RT-PCR, IgM and IgG test results; and, (iii) the multiplying factor (ratio of the incidence proportion on the proportion of confirmed –RT-PCR+– cases).

Results Among 933 alert cases with complete data, 218 (23%) were RT-PCR+; 434 (47%) IgM+; 464 (~ 50%) RT-PCR+, IgM+, or both; and 647 (69%) either IgG + or IgM+. The incidence proportion of SARS-CoV-2 infection was estimated at 58% (95% credibility interval: 51.8–64), its prevalence at 72.83% (65.68–77.89), and the multiplying factor at 2.42 (1.95–3.01).

Conclusions In monitoring the pandemic dynamics, the number of biologically probable cases is also useful. The multiplying factor helps approximating it.

Keywords SARS-CoV-2, COVID-19 testing, Reverse transcriptase polymerase chain reaction, Serology, Africa.

†Diarra YM and Wimba PM had equal contributions

*Correspondence:

YM Diarra
diarrayvesmarie@gmail.com

¹Université de Lyon, Lyon, France

²Université Claude Bernard Lyon 1, Villeurbanne, France

³Service de Biostatistique-Bioinformatique, Pôle Santé Publique, Hospices Civils de Lyon, Lyon, France

⁴Laboratoire de Biométrie et Biologie Évolutive, Équipe Biostatistique-Santé, CNRS UMR 5558, Villeurbanne, France

⁵Université Officielle de Bukavu, Democratic Republic of the Congo, Bukavu, Congo

⁶Cliniques Universitaires de Bukavu, Democratic Republic of the Congo, Bukavu, Congo

⁷Centre International de Recherche en Infectiologie (CIRI), INSERM U1111-CNRS UMR 5308, Lyon, France

⁸Bureau Information Sanitaire, Division provinciale de la Santé Sud-Kivu, Democratic Republic of the Congo, Bukavu, Congo

⁹Université Officielle de Mbuji-Mayi (UOM), Mbuji-Mayi, Democratic Republic of the Congo

¹⁰Department of Virology, National Institute for Biomedical Research (INRB), Democratic Republic of the Congo, Kinshasa, Congo

¹¹Service of Microbiology, Department of Medical Biology, Kinshasa teaching School of Medicine, Faculty of Medicine, University of Kinshasa, Democratic Republic of the Congo, Kinshasa, Congo

¹²IRD UMI 233, INSERM U1175, Université de Montpellier, Unité TransVIHMI, Montpellier, France

¹³EpiGreen, Paris, France

¹⁴Service d'Hygiène Hospitalière, Infectiovigilance et Prévention, Hospices Civils de Lyon, Épidémiologie, Lyon, France



Introduction

According to the World Health Organisation (WHO), confirmed cases of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infection are those with positive Nucleic Acid Amplification Test (NAAT) or those having both a positive SARS-CoV-2 Antigen-RDT (Rapid Diagnostic Test) and meeting some clinical criteria [1].

It is now recognized that, in Africa, confirmed case counts do not accurately reflect COVID-19 epidemic dynamics [2]. That underestimation of the number of new infections is illustrated by the contrast between a low number of reported cases and a high prevalence of anti-SARS-CoV-2 antibodies (an indicator of high virus circulation) [3]. In 2020, in South Kivu, a low number of reported confirmed cases (418 for 1526 alert –or suspect– cases) contrasted with a high seroprevalence (41.2%) among 359 healthcare workers [4]. According to some authors, that underestimation had two distinct causes: (1) a small portion of the population was tested; (2) a number of tests may have been negative in infected individuals [5, 6]. For economic or logistic reasons, the first cause is difficult to circumvent, whereas the second seems easier to deal with. In fact, estimates of tests' sensitivity (Se) and specificity (Sp) are available as well as statistical methods that allow estimating the number of cases even with imperfect diagnostic tests [7–9].

In South Kivu, NAAT and antibody tests are currently available. Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) SARS-CoV-2 antigen detection test does not yield false-positive results, but false-negative results are possible because of poor sample quality or sample collection in very early or late stages of infection [10]. According to some authors, the Se of RT-PCR SARS-CoV-2 test (its ability to identify truly diseased individuals) was 55% [7], 68% [8], or 85% [9]; a large variability due to sample quality or sampling delay relative to infection. The antibody tests (i.e., rapid lateral flow tests for immunoglobulin M or G –IgM or IgG) that were available during the first wave of the pandemic could yield false-positive and false-negative results in detecting incident cases [11, 12]. Indeed, according to Kostoulas et al. [8], true positive results (Se) for either IgM or IgG during the first, second, and third week after COVID-19 symptom onset were obtained in 32, 75, and 92% of cases, respectively. However, the latter tests Sp's were obviously lower than that of RT-PCR and ranged from 81 to 100% [7, 13].

The WHO definition of 'confirmed cases' based on RT-PCR results avoids count overestimation because of the very high Sp of RT-PCR (nearly 1). However, that definition gives only a lower limit for the real number of cases because of RT-PCR false-negative results. Serum IgM or IgG detection is not routinely used for case definition

because of these tests' false-negative and false-positive results. Besides, while RT-PCR and IgM tests are suitable to diagnose recent infections (incident cases), IgG test is rather suitable to assess past infections (prevalent cases). Given those elements, it seems useful to propose a way to obtain a better estimate of the number of new cases from the number of RT-PCR-confirmed cases.

This article proposes a method to obtain a factor by which the number of confirmed cases (precisely, RT-PCR+cases) may be multiplied to provide a better estimate of the number of new infections (i.e., the number of biologically probable cases).

In the following, we will: (i) estimate the incidence proportion and the prevalence of SARS-CoV-2 infection (here, the proportion of current or previous cases) in the first thousand COVID-19 alert cases in South Kivu using latent class model and a Bayesian inference method; and, (ii) propose an estimate of the multiplying factor that allows obtaining the number of biologically probable new SARS-CoV-2 infections from the number of confirmed cases.

Materials and methods

The study population

The data were extracted from the SARS-CoV-2 infection surveillance system in South Kivu (DRC) during the 2020 pandemic. The population of South Kivu is approximately 4,800,000 people who live in a 64,492 km² area. The study considered the first thousand alert cases recorded between March 29 and November 29, 2020. An alert case was defined as a person with signs suggestive of COVID-19 (fever, headache, breathing difficulties, asthenia... with or without loss of taste and smell) or a person who has been in contact with a person who tested positive for SARS-Cov-2 infection.

From each alert case, two samples were collected: a nasopharyngeal sample for RT-PCR test and a blood sample for IgM and IgG serology. In general, the recording of an alert case and sample collection took place the same day.

The serological test used 'SARS-CoV-2 IgG/IgM Rapid Test Kit' (Abbexa Ltd, Cambridge, UK). This test detects separately but on the same 'cassette' IgM and IgG antibodies against the virus. The tests (RT-PCR and serological test) were carried out in two centers: Kinshasa (March 29 to June 16, 2020) and Bukavu (June 17 to November 29, 2020).

A confirmed case was defined as an alert case with a positive RT-PCR test.

Statistical analyses

Data presentation

Data presentation used 2 by 2 contingency tables for cross-tabulation of RT-PCR versus IgM test results

(numbers and percentages), then for cross-tabulation of RT-PCR versus IgM and IgG test results (positive when IgM+ or IgG+, negative when IgM- and IgG-). The information given by each cell of these tables depends on the proportion of infection cases, the Se, and the Sp of each test. For example, the number of alert cases positive on test A and test B is the sum of two numbers: (i) the number of true positive results on both tests (i.e., the number of alert cases multiplied by the proportion of infection cases and the Se of each test); and, (ii) the number of false positive results on both tests (i.e., the number of alert cases multiplied by the complement to 100% of the percentage of infected cases and the complement to 100% of the Sp of each test). This information is needed to estimate respectively the incidence proportion and the prevalence of SARS-CoV-2 infection using a latent class model and a Bayesian inference method.

The latent class model

In the latent class model, the infection status is considered unknown and the results of the diagnostic tests are used to estimate the proportion of infected cases and the performance (Se and Sp) of the tests. The model was built with the assumption that the RT-PCR and the antibody test results are independent conditionally on the infection status. In fact, this assumption is plausible because the two types of diagnostic tests (RT-PCR and IgM or IgG serology) have different biological mechanisms.

Two separate latent class models were used; one to estimate the incidence proportion (using RT-PCR and IgM serology) and the other to estimate the prevalence (using RT-PCR and IgM/IgG serology).

The bayesian inference method

With two tests, the information provided by the observed data is not sufficient to estimate the proportion of infection cases and the performance of the tests in terms of Se and Sp. A Bayesian inference method was used to add prior knowledge on the Se of the RT-PCR and the Se and Sp of the serological tests to the observed data [14]. The

Sp of RT-PCR was set to 100%. This implies the use of two latent classes (instead of four without this assumption).

Prior knowledge was extracted from the literature and summarized using prior distributions. Prior information on the performance of the tests was obtained from a search on PubMed with various combinations of keywords “COVID-19”, “diagnosis”, “performance”, “accuracy”, “test”, and “serological”. The retained articles were those that reported the performance of at least one of the tests (RT-PCR, IgG, and IgM). The excluded articles were those where the ‘gold standard’ was an imperfect diagnostic test, those that reported on pre-pandemic sera (to determine serologic test specificities), and those that used clinical or biological criteria to select the population. From the articles selected [7–9, 13], we extracted the smallest lower bound and the largest upper bound of the 95% confidence intervals (CoIs) of each test Se and Sp to derive prior intervals. When no confidence intervals were available, point estimates were used to derive the prior intervals (Table 1). Beta distributions were used as prior distributions with means equal to the centers of the corresponding prior intervals and standard deviations equal to the fourths of their ranges (Table 1). For the proportion of infection cases, a beta distribution with both parameters equal to one was used; this corresponds to a uniform distribution between 0 and 1.

Gibbs sampling was used to obtain a sample of the posterior distribution of each parameter from which were derived a point estimate (median of the posterior distribution) and a 95% credibility interval (CrI, between quantiles 2.5% and 97.5% of the posterior distribution) [14].

Three sets of 60,000 values were sampled from the conditional posterior distribution of each of the three parameters of the model using three different sets of starting values for the parameters. These starting sets were chosen using the centres and the upper and lower bounds of the intervals of the literature data on test performance (Table 1). An interval was formed by the proportions of positive RT-PCR and IgG/IgM serology using the cross table of these two test results. The bounds and centre of this interval were used as a starting set for the proportion of infected people in the data. The convergence of the three Markov chains was evaluated by the Gelman index. The first 10,000 iterations of the three chains allowing to reach convergence were removed. The remaining 50,000 iterations of each of the three chains were put together to give point estimates (medians of the posterior distributions) and 95% credibility intervals (quantiles 2.5% and 97.5% of the posterior distributions) of the parameters.

Estimating the multiplying factor

A sample of the posterior distribution of the multiplying factor was obtained by dividing each value of the

Table 1 Prior knowledge on the sensitivities and specificities of the tests used for the diagnosis of SAR-Cov-2 infection

Test characteristic	Prior interval	Prior parameters of beta distributions	
		Alpha	Beta
RT-PCR sensitivity	50–91%	13.25	5.54
IgM test sensitivity ^a	24–44%	30.18	58.58
IgM test specificity	94–100%	124.48	3.85
IgM and IgG test sensitivity ^b	23–97%	3.61	2.40
IgM and IgG test specificity ^b	81–100%	33.58	3.52

^a Sensitivity of the test in diagnosing recent infections

^b IgM and IgG test: positive result when IgM+ or IgG+, negative result when IgM- and IgG-.

posterior distribution sample of the incidence proportion of SARS-Cov-2 infection by the observed proportion of alert cases with positive RT-PCR test result. A point estimate of the factor and a 95% CrI were extracted from that sample (For more details, please see Additional files 1, 2 and 3).

In this work, qualitative variables were summarized by numbers and percentages in various modalities and quantitative variables by the mean, the standard deviation, the median, the first and third quartile.

All statistical analyses were performed with R software version 3.6.3 (2020-02-29, R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>).

Results

Characteristics of the study population

The median age of the individuals in the alert population was 44 years and the standard deviation was nearly 17 years. This wide variability is also illustrated by a 25-year interquartile range (Q1=33 years and Q3=58 years).

Of the 929 individuals with no missing data for sex or diagnostic test results, (659) 71% were males. Among COVID-19 signs that motivated the alert, fever, dry cough, and fatigue were present in (595) 64%, 166 (18%), and 99 (11%) of alert cases, respectively.

Observed diagnostic test results

Full RT-PCR and IgG/IgM test results were available for 933 out of the 1000 alert cases. According to these results, 218 (23%) cases were RT-PCR+; 434 (47%) IgM+; and 464 (50%) RT-PCR+, IgM+, or both. Besides, of these 933 cases, 647 (69%) were IgM+or IgG+ (Table 2).

Estimates obtained with the latent class model

The incidence proportion of SARS-CoV-2 infection in the first thousand COVID-19 alert cases in South Kivu was estimated at 58% (95% CrI: 51.8–64) and the prevalence at 72.83% (95% CrI: 65.68–77.89). In other words, out of 100 alert subjects 58 were incident cases and about 73 were prevalent cases.

The RT-PCR SARS-CoV-2 antigen test sensitivity was estimated at 41.34% (95% CrI: 36.47–46.57). The IgM test sensitivity was estimated at 71.65% (95% CrI: 66.45–76.45) and its specificity at 96.93% (95% CrI: 92.75–99.11). These values were also observable in the posterior distributions of the parameters (see Additional file 4).

Calculation of the multiplying factor

Of the 1000 alert subjects, 240 were predicted as confirmed cases versus 580 as probable cases. The factor for approximating the number of new biologically probable infections from the number of RT-PCR+subjects was

Table 2 SARS-CoV-2 RT-PCR antigen test and serology results in the first thousand COVID-19 cases seen in South Kivu (DRC) in 2020

	RT-PCR negative	RT-PCR positive	Total
IgM test result			
Negative	469 (50.27) ^a	30 (3.22)	499 (53.49)
Positive	246 (26.37)	188 (20.14)	434 (46.51)
Total	715 (76.64)	218 (23.36)	933 (100)
IgM and IgG test result ^b			
Negative	267 (28.62) ^a	19 (2.04)	286 (30.66)
Positive	448 (48.02)	199 (21.32)	647 (69.34)
Total	715 (76.64)	218 (23.36)	933 (100)

^a Each value is expressed as number (percent of all available values)

^b IgM and IgG test: positive result when IgM+or IgG+, negative result when IgM- and IgG-.

Table 3 Calculation of the factor for approximating the number of new biologically probable infections from the number of RT-PCR+ subjects in South Kivu (DRC).

Parameter	Median (95% credibility interval)
Expected number of RT-PCR+ subjects among 1000 alert cases	240 (203–278)
Number of biologically probable new SARS-CoV-2 infections	580 (518–640)
Multiplying factor ^a	2.42 (1.95–3.01)

^a Factor that allows obtaining the number of biologically probable new SARS-CoV-2 infections from the number of confirmed cases

2.42 (95% CrI: 1.95–3.01) (Table 3). This means that the number of confirmed cases should be multiplied by 2.42 to yield an estimate of the number of new biologically probable infections.

Discussion

This analysis of the first 1000 alert cases in South Kivu showed the strong underestimation of the number of new infections obtained by using the number of confirmed cases. Indeed, whereas RT-PCR test indicated 23% of confirmed cases, the incidence proportion as estimated by the latent class model indicated 58% of new infection cases and the estimated prevalence was 73%; i.e., the number of new infections would have been more than double of the number of confirmed cases. The high specificity of RT-PCR tests precludes overestimation and justifies its use to define confirmed cases. However, we might propose to join to the number of confirmed cases the number of biologically probable cases as obtained using the multiplying factor.

The impact of misclassification is well known in other contexts [15–18] and in infectious diseases [19] and has been taken into account in numerous studies. Either they use a method comparable to the one used here, based on estimating a multiplying factor to obtain the corrected value from the observed value [19], or they obtain the corrected value by a latent class model without giving a multiplying factor that can be used in another context. These methods are currently also used to estimate vaccine effectiveness [20–22]. Several previous studies have estimated the number of biologically probable cases using a latent class model to account for test imperfections [7–9, 23]. They reported comparable results in different geographical contexts and suggested for RT-PCR test higher sensitivity estimates than ours (55 to 85% vs. 36 to 47%). This discrepancy may arise from tests performed sometimes at late stages of the disease in this study context, which is supported by the high proportion of IgG+ results.

Here, adding the information on IgG results would increase the estimate of the proportion of infection cases in the alert cases up to 73%. As the study concerned the 1000 alert cases identified during the first months of the pandemic, it is very probable that most infection cases have occurred during the study period and that the true incidence proportion among these first thousand alert cases lays between 58% (an underestimation due to late diagnoses) and 73% (an overestimation due to the potential inclusion of a small number of infection cases that had occurred before the study period). Logically, after the first year of the pandemic, the incidence proportion should be estimated using only the results of RT-PCR and IgM tests.

The early availability of new biological tests at the beginning of the pandemic justifies the caution in serological result interpretation. In particular, some authors have suspected a cross-reactivity between SARS-CoV-2 and other germ antigens already present in the African context [13, 24, 25]. The progressive improvement of serological tests make them useful complements to RT-PCR test whose sensitivity decreases at late stages of infection. Indeed the latent class model with Bayesian inference changed slightly the estimates of the proportion of infection cases by allowing for the imperfections of the tests. In fact, (i) the proportion of either RT-PCR+ or IgM+ subjects was 50%, whereas the estimate of the incidence proportion given by the model was 58%; and, (ii) the proportion of RT-PCR+, or IgM+, or IgG+ subjects was 71%, whereas the estimate of the prevalence given by the model was 73%. This 58% incidence proportion is also found using the expression proposed by Sempas and Tian [26], but the contribution of the present study was to take into account the uncertainty on the performance

of RT-PCR and quantify the multiplier factor in an African context.

The correction factor estimate allowing to approximate the number of probable new SARS-CoV-2 infections among the tested subjects, was based on the number of confirmed cases and was depended of the performance of the used tests. It can be transposed to other countries that used the same tests, regardless of the strength of the epidemic. The correction factor would have to be recalculated in countries that use other tests.

The method and the multiplication factor can be generalized to other African countries with different screening capacities, resources and epidemiological contexts. However, the multiplication factor will have to be recalculated if the tests used are different.

One limitation of this study may be the extrapolation of the incidence proportion found in the 933 subjects (who had diagnostic test results) to the rest of the first 1000 alerts (i.e. 67 subjects). But there was nothing to contradict the fact that these missing data were not random. Even if this bias exists, it will have little impact because these subjects represented only about 7% of the first 1000 alerts.

Another limitation was the unavailability of the delay between symptom onset and lab tests. Indeed, this delay may affect test results and subsequently the value of the multiplier. Furthermore, very few subjects were asymptomatic in our study, which limits subgroup analysis in this particular population.

Conclusion

The present study confirmed that the incidence proportion of SARS-CoV-2 infection is underestimated when only RT-PCR positive subjects are counted. Thus, when dealing with changes in the dynamics of the pandemic, it would be useful to report the number of biologically probable cases along with the number of confirmed cases. The estimated multiplier may be used to approximate this number from the number of RT-PCR+ cases in the African context. To further clarify the applicability of the proposed multiplier, this study should be supported by another one that aims to estimate the multiplier factor as a function of the time elapsed between the onset of COVID-19 symptoms and the carrying out of the diagnostic tests, and also in asymptomatic subjects.

List of abbreviations

CI	Confidence Intervals
CrI	Credibility Interval
DRC	Democratic Republic of the Congo
NAAT	Nucleic Acid Amplification Test
RDT	Rapid Diagnostic Test
Se	Sensitivity
Sp	Specificity
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-023-02077-2>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4

Acknowledgements

Not applicable.

Authors' contributions

DYM, WP, ER and MR designed the study and wrote drafts of the article. WPBJ, MB, OM, TL and ASM contributed to the data collection. KPB, EJF, VP reviewed the article. JJ contributed to the writing and editing. All authors read and approved the final manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

All methods were carried out in accordance with the relevant guidelines and regulations. This study was approved by the Medical Ethics Committee of the official University of Bukavu (N°Ref: 011/CEM/UOB/2023). Informed consent was obtained from all participants for the use of their data for research purposes at the time of their screening test.

Consent for publication

Not applicable.

Competing interests

None

Received: 30 October 2021 / Accepted: 20 October 2023

Published online: 17 November 2023

References

- WHO COVID-19 Case definition. https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-Surveillance_Case_Definition-2020.2. Last accessed: Sept. 2021.
- Kobia F, Gitaka J. COVID-19: are Africa's diagnostic challenges blunting response effectiveness? *AAS Open Res.* 2020;3:4. <https://doi.org/10.12688/aasopenres.13061.1>.
- Nkuba AN, Makiala SM, Guichet E, Tshiminyi PM, Bazitama YM, Yambayamba MK, Kazenza BM, Kabeya TM, Matungulu EB, Baketana LK, Mitongo NM, Thaurignac G, Leendertz FH, Vanlerberghe V, Pelloquin R, Etard JF, Maman D, Mbala PK, Ayoubia A, Peeters M, Muyembe JT, Delaporte E, Ahuka SM. High prevalence of anti-SARS-CoV-2 antibodies after the first wave of COVID-19 in Kinshasa, Democratic Republic of the Congo: results of a cross-sectional household-based survey. *Clin Infect Dis.* 2021;ciab515. <https://doi.org/10.1093/cid/ciab515>.
- Mukwege D, Byabene AK, Akonkwa EM, Dahma H, Dauby N, Cikwanine Buhendwa JP, Le Coadou A, Montesinos I, Bruyneel M, Cadière GB, Vandenberg O, Van Laethem Y. High SARS-CoV-2 Seroprevalence in Healthcare Workers in Bukavu, Eastern Democratic Republic of Congo. *Am J Trop Med Hyg.* 2021;104:1526–30. <https://doi.org/10.4269/ajtmh.20-1526>.
- Scohy A, Anantharajah A, Bodéus M, Kabamba-Mukadi B, Verroken A, Rodríguez-Villalobos H. Low performance of rapid antigen detection test as frontline testing for COVID-19 diagnosis. *J Clin Virol.* 2020;129:104455. <https://doi.org/10.1016/j.jcv.2020.104455>.
- Russell TW, Golding N, Hellewell J, Abbott S, Wright L, Pearson CAB, van Zandvoort K, Jarvis CI, Gibbs H, Liu Y, Eggo RM, Edmunds WJ, Kucharski AJ, CMMID COVID-19 working group. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and Infections. *BMC Med.* 2020;18:332. <https://doi.org/10.1186/s12916-020-01790-9>.
- Hartnack S, Eusebi P, Kostoulas P. Bayesian latent class models to estimate diagnostic test accuracies of COVID-19 tests. *J Med Virol.* 2021;93:639–40. <https://doi.org/10.1002/jmv.26405>.
- Kostoulas P, Eusebi P, Hartnack S. Diagnostic accuracy estimates for COVID-19 RT-PCR and lateral flow immunoassay tests with bayesian latent class models. *Am J Epidemiol.* 2021;190:1689–95. <https://doi.org/10.1093/aje/kwab093>.
- Symons R, Beath K, Dangis A, Lefever S, Smismans A, De Bruecker Y, Frans J. A statistical framework to estimate diagnostic test performance for COVID-19. *Clin Radiol.* 2021;76:75. e1-75.e3.
- WHO. Laboratory testing for coronavirus Disease 2019 (COVID-19) in suspected human cases. Last accessed: Sept; 2021. <https://apps.who.int/iris/bitstream/handle/10665/331329/WHO-COVID-19-laboratory-2020.4-eng.pdf?sequence=1&isAllowed=y>.
- Fischer PU, Fischer K, Curtis KC, Huang Y, Fetcho N, Goss CW, Weil GJ. Evaluation of Commercial Rapid lateral Flow tests, alone or in combination, for SARS-CoV-2 antibody testing. *Am J Trop Med Hyg.* 2021;105:378–86. <https://doi.org/10.4269/ajtmh.20-1390>.
- Mboumba Bouassa RS, Péré H, Tonen-Wolyec S, Longo JD, Moussa S, Mbopi-Keou FX, Mossoro-Kpinda CD, Grésenguet G, Veyer D, Bélec L. Unexpected high frequency of unspecific reactivities by testing pre-epidemic blood specimens from Europe and Africa with SARS-CoV-2 IgG-IgM antibody rapid tests points to IgM as the Achilles heel. *J Med Virol.* 2021;93:2196–203. <https://doi.org/10.1002/jmv.26628>.
- Tso FY, Lidenge SJ, Peña PB, Clegg AA, Ngowi JR, Mwaeselage J, Ngalamika O, Julius P, West JT, Wood C. High prevalence of pre-existing serological cross-reactivity against severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) in sub-saharan Africa. *Int J Infect Dis.* 2021;102:577–83. <https://doi.org/10.1016/j.ijid.2020.10.104>.
- Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of Disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol.* 1995;141:263–72. <https://doi.org/10.1093/oxfordjournals.aje.a117428>.
- Campbell H, Biloglav Z, Rudan I. Reducing bias from test misclassification in burden of Disease studies: use of test to actual positive ratio—new test parameter. *Croat Med J.* 2008;49(3):402–14. <https://doi.org/10.3325/cmj.2008.3.402>. PMID: 18581619; PMCID: PMC2443625.
- Brenner H, Gefeller O. Use of the positive predictive value to correct for Disease misclassification in epidemiologic studies. *Am J Epidemiol.* 1993;138(11):1007–15. <https://doi.org/10.1093/oxfordjournals.aje.a116805>. PMID: 8256775.
- Walraven CV. A comparison of methods to correct for misclassification bias from administrative database diagnostic codes. *Int J Epidemiol.* 2018;47(2):605–16. <https://doi.org/10.1093/ije/dyx253>. PMID: 29253160.
- Kopec JA. Estimating Disease Prevalence in Administrative Data. *Clin Invest Med.* 2022;45(2):E21–27. <https://doi.org/10.25011/cim.v45i2.38100>. PMID: 35752980.
- Tarafder MR, Carabin H, McGarvey ST, Joseph L, Balolong E Jr, Olveda R. Assessing the impact of misclassification error on an epidemiological association between two helminthic Infections. *PLoS Negl Trop Dis.* 2011;5(3):e995. <https://doi.org/10.1371/journal.pntd.0000995>. PMID: 21468317; PMCID: PMC3066162.
- Endo A, Funk S, Kucharski AJ. Bias correction methods for test-negative designs in the presence of misclassification. *Epidemiol Infect.* 2020;148:e216. PMID: 32895088; PMCID: PMC7522852.
- Eusebi P, Speybroeck N, Hartnack S, Staerk-Østergaard J, Denwood MJ, Kostoulas P. Addressing misclassification bias in vaccine effectiveness studies with an application to Covid-19. *BMC Med Res Methodol.* 2023;23(1):55. <https://doi.org/10.1186/s12874-023-01853-4>. PMID: 36849911; PMCID: PMC9969950.
- Habibzadeh F. Correction of vaccine effectiveness derived from test-negative case-control studies. *BMC Med Res Methodol.* 2023;23(1):137. <https://doi.org/10.1186/s12874-023-01962-0>. PMID: 37301843; PMCID: PMC10257167.

23. Bhattacharyya R, Kundu R, Bhaduri R, Ray D, Beesley LJ, Salvatore M, Mukherjee B. Incorporating false negative tests in epidemiological models for SARS-CoV-2 transmission and reconciling with seroprevalence estimates. *Sci Rep.* 2021;11(1):9748. <https://doi.org/10.1038/s41598-021-89127-1>. Erratum in: *Sci Rep.* 2021;11(1):17221. PMID: 33963259; PMCID: PMC8105357.
24. Emmerich P, Murawski C, Ehmen C, von Possel R, Pekarek N, Oestereich L, Duraffour S, Pahlmann M, Struck N, Eibach D, Krumpkamp R, Amuasi J, Maiga-Ascofaré O, Rakotozandrindrainy R, Asogun D, Ighodalo Y, Kann S, May J, Tannich E, Deschermeier C. Limited specificity of commercially available SARS-CoV-2 IgG ELISAs in serum samples of African origin. *Trop Med Int Health.* 2021;26:621–31. <https://doi.org/10.1111/tmi.13569>.
25. Sagara I, Woodford J, Dicko A, Zeguime A, Doucoure M, Kwan J, Zaidi I, Doritchamou J, Snow-Smith M, Alani N, Renn J, Kosik I, Holly J, Yewdell J, Esposito D, Sadtler K, Duffy P. SARS-CoV-2 seroassay optimization and performance in a population with high background reactivity in Mali. *medRxiv* 2021;2021.03.08.21252784. <https://doi.org/10.1101/2021.03.08.21252784>.
26. Sempos CT, Tian L. Adjusting coronavirus prevalence estimates for Laboratory Test Kit Error. *Am J Epidemiol.* 2021;190:109–15. <https://doi.org/10.1093/aje/kwaa174>. PMID: 32803245; PMCID: PMC7454308.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.