BMJ Open

COVID-19 morbidity in Afghanistan: a nationwide, population-based seroepidemiological study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-060739
Article Type:	Original research
Date Submitted by the Author:	21-Jan-2022
Complete List of Authors:	Saeedzai, Sayed Ataullah; Ministry of Public Health, Monitoring, Evaluation and Health Information System Sahak, Mohammad; World Health Organization, WHE Arifi, Fatima; WHO Aly, Eman; World Health Organisation Regional Office for the Eastern Mediterranean Van Gurp, Margo; KIT Royal Tropical Institute White, Lisa; University of Oxford, Nuffield Department of Medicine; University of Oxford Chen, Siyu; University of Oxford, Discovery, Nuffield Department of Medicine Barakat, Amal; World Health Organisation Regional Office for the Eastern Mediterranean Azim, Giti; Ministry of Public Health Rasoly, Bahara; Ministry of Public Health Safi, Soraya; Ministry of Public Health Flegg, Jennifer; University of Melbourne Ahmed, Nasar; Florida International University, Department of Epidemiology Ahadi, Mohmmad Jamaluddin; Ministry of Public Health Achakzai, Niaz; Ministry of Public Health AbouZeid, Alaa; Cairo University Kasr Alainy Faculty of Medicine
Keywords:	COVID-19, Epidemiology < INFECTIOUS DISEASES, EPIDEMIOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

COVID-19 morbidity in Afghanistan: a nationwide, population-based seroepidemiological study

Authors & affiliations

Sayed Ataullah Saeedzai[1], Mohammad Nadir Sahak[2], Fatima Arifi[2]*, Eman Abdelkareem Aly[3], Margo Van Gurp[4], Lisa White[5], Siyu Chen[5], Amal Barkat[6], Giti Azim[1], Bahara Rasoly[1], Soraya Safi[1], Jennifer Flegg[7], Nasar U. Ahmed[8], Mohammad Jamaluddin Ahadi[1], Niaz M. Achakzai[9], Alaa Abouzeid[10]

[1]General Directorate of Monitoring, Evaluation and Health Information System (M&E-HIS), Ministry of Public Health, Kabul, Afghanistan

[2]WHO Health Emergencies Programme, WHO, Kabul, Afghanistan

[3] Information Systems for Health Unit, WHO Eastern Mediterranean Regional Office, Cairo, Egypt

[4] KIT Royal Tropical Institute, Amsterdam, the Netherlands

[5] Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, United Kingdom

[6] Infectious Hazard Preparedness Unit, WHO Eastern Mediterranean Regional Office, Cairo, Egypt

[7] School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia

[8] Department of Epidemiology, Robert Stempel College of Public Health, Florida International University, Miami, FL

[9] Department of Molecular Biology, Forensic Medicine Directorate. Ministry of Public Health, Kabul, Afghanistan

[10] Department of Public Health, Faculty of Medicine, Cairo University

*Corresponding author:

Fatima Arifi (Fatima.arefi@gmail.com, arififf@who.int)

Word count: 4403 words

List of Abbreviations

CoMo Consortium – COVID-19 International Modelling Consortium

EA – enumeration area

ELISA – enzyme-linked immunosorbent assay

IFR – infection fatality ratio

IgG – immunoglobulin G

IgM – immunoglobulin M

MCMC – Markov chain Monte Carlo

MoPH – Ministry of Public Health

NPI – nonpharmaceutical intervention

NSIA – National Statistics and Information Authority

R0 – basic reproduction number

RDT – rapid diagnostic test

WHO – World Health Organization

Abstract (257 words)

Introduction

The ongoing COVID-19 pandemic continues to result in considerable morbidity and mortality around the world. However, in many countries it is difficult to estimate the true burden of COVID-19 infection in a population due to gaps in surveillance coverage and limited testing capacity.

Methods

Here, we describe a population-based, cross-sectional, age-stratified sero-epidemiological study conducted throughout Afghanistan during June/July 2020. Participants were interviewed to

complete a questionnaire, and rapid diagnostic tests were used to test for SARS-CoV-2 antibodies. The primary objectives were 1) to determine the magnitude of COVID-19 infections in the general population and age-specific cumulative incidence, as determined by seropositivity and clinical symptoms of COVID-19; and 2) to determine the magnitude of asymptomatic or subclinical infections. To adjust the seroprevalence for test sensitivity and specificity, as well as seroreversion, Bernoulli model methodology was used to infer the population exposure in Afghanistan.

Results

The survey revealed that, to July 2020, around 10 million people in Afghanistan (31.5% of the population) had either current or previous COVID-19 infection. This implies that the herd immunity threshold had not been reached and most of the population of Afghanistan remained at risk of infection. However, the herd immunity threshold may have been crossed in some localities, such as Kabul province, where more than half of the population had been infected with COVID-19, which exceeds the lowest reported herd immunity threshold.

Conclusion

As most of the population remained at risk of infection at the time of the study, any lifting of public health and social measures needed to be considered gradually.

Article Summary: Strengths and Limitations

Strengths:

- This is the first nationwide, population-based, large sample size, seroepidemiological study conducted in Afghanistan
- The study provides evidence on the high burden of COVID-19 morbidity in resource limited and conflict affected country with limited mitigation measures and limited evidence on burden
- The study highlights the limited surveillance capacity and under-reporting of COVID-19 cases in Afghanistan

Limitations:

- Due to security concerns, not all areas could be surveyed; the inability to conduct proper household listing and create maps for enumeration areas in those areas where the government lacked control may have affected the findings
- The findings may not reflect the current situation with regards to the new SARS-CoV-2 delta and omicron variants of concern, as the R0 for these variants is not well-established
- The data were entered in the DHIS2 database, which created many challenges for data verification, household matching and the subsequent analysis.

Summary box

What is already known?

- To 5 November 2021, the COVID-19 pandemic has resulted in more than 248 million cases and more than 5 million deaths worldwide.
- As in other countries, Afghanistan has introduced nonpharmaceutical interventions to control the spread of COVID-19.
- Seroepidemiological surveys can provide useful data to help inform public health policies.

What are the new findings?

- This national survey of COVID-19 morbidity and mortality in Afghanistan revealed that around 10 million people (31.5% of the population) had either current or previous COVID-19 infection.
- There was regional heterogeneity in the burden of COVID-19 disease, with urban areas such as Kabul showing higher cumulative rates of COVID-19.

What do the new findings imply?

- The cumulative number of COVID-19 cases across the country means Afghanistan had yet to reach the herd immunity threshold at the time of study, which is reported to range from 43% to 85%.
- The cumulative number of COVID-19 cases in Kabul (53%) suggests that this region appeared to might have reached the herd immunity threshold, if lower estimates for the herd immunity threshold (43%) are used.
- Seroprevalence represents a low estimate of herd immunity, while predicted exposure represents a higher estimate.

INTRODUCTION

The COVID-19 pandemic has resulted in more than 248 million confirmed cases and in excess of 5 million deaths globally to November 2021.[1] Many countries are continuing to experience epidemic waves of COVID-19, including Brazil, India and Nepal.[2-4] The first reported case of COVID-19 in Afghanistan was in Herat province on 24 February 2020; as of 20 July 2021, Afghanistan has reported 156 363 confirmed cases of COVID-19 and 7284 deaths from the disease.[5]

When the COVID-19 pandemic began, there were no vaccines or specific treatments available for COVID-19, so nonpharmaceutical interventions (NPIs) were recommended, including social distancing, home quarantine, closure of schools and universities, and bans on public gatherings. Afghanistan introduced NPIs as soon as the first case of COVID-19 was detected in the country. Case detection and isolation were seen as key features in helping to reduce the spread of COVID-19. With the recent political transition in the country and disruption of the health system, public health and social measures to tackle COVID-19 have been completely neglected, which may pose a major risk for increasing the spread of COVID-19 in Afghanistan.

The initial focus of the Afghanistan Ministry of Public Health (MoPH) was on patients with severe COVID-19 disease and ways to decrease mortality associated with the disease. Serological testing of patients can be used to provide useful information about an individual's status in terms of a current or previous COVID-19 infection. Immunoglobulin M (IgM) and G (IgG) antibodies arise at around the same time, between 1 to 3 weeks after infection; however, IgM antibodies decay more rapidly than IgG antibodies.[6] Therefore, for public health studies, IgM is used as a marker of current infection while IgG is used as a marker of previous infection, i.e. within the previous few months. There are various rapid diagnostic tests (RDTs) available that can be used to simultaneously test blood samples for IgM/IgG antibodies against COVID-19.

Due to the limited testing and surveillance capacity in Afghanistan, it seemed likely there was considerable under-reporting of cases and deaths; therefore, robust scientific studies are required to determine the actual burden of COVID-19 in the country. Serological studies can be used to estimate levels of past exposure and thus position a population in their epidemic timeline. However, serology results might underestimate the total exposure in a population[7] because of decaying antibody titres over time.[8-10] Here, we describe a national seroepidemiological survey initiated by the MoPH and conducted throughout Afghanistan between June and July 2020, involving a questionnaire survey and antibody testing of participants for COVID-19 infection using RDTs. The primary objectives of the study were: 1) to determine the magnitude of COVID-19 infection in the general population and age-specific cumulative incidence, as determined by seropositivity and clinical symptoms of COVID-19; and 2) to determine the magnitude of asymptomatic or subclinical infections. The World Health Organization (WHO) protocol for population-based age-stratified seroepidemiological investigations for COVID-19 infection was adapted for the Afghanistan context to obtain seroprevalence estimates.[11] To adjust the seroprevalence for test sensitivity and specificity, as well as seroreversion, we further adapted a methodology[12] that was originally developed for the England setting and used this to infer the population exposure and undocumented mortality associated with COVID-19 in Afghanistan.

METHODS

Patient and Public Involvement statement

As the study was not clinical trial and it did not involve patients, the public and patients were not involved directly during study design or dissemination. The study results were disseminated through public workshops in universities, seminars and workshops, and media for the general public information. Their consent was obtained for being included in the study and any personal identifier information was excluded during study design and results.

Ethical considerations

Ethical and technical clearance to conduct the survey was obtained from the Institutional Review Board of the Afghanistan MoPH with **Reference #: A.0321.0278**. Informed consent was obtained from participants aged ≥18 years, and assent from family members was obtained for those aged 5–17 years. Individuals who did not provide consent were excluded. Survey team members provided advice about home isolation to participants who tested IgM-positive for COVID-19 during the survey.

Study design

This was a population-based, cross-sectional, age-stratified seroepidemiological study. Participants were interviewed to complete a questionnaire, and RDTs were used to test for SARS-CoV-2 antibodies. The survey was conducted during June and July 2020.

Population and sampling

This was a national study conducted in the eight regions of Afghanistan plus Kabul province, which was considered as a separate region, making nine regions in total (online supplemental figure S1). The total sample size was 9514 and the number of participants required in each region was estimated proportionate to the population size of each region (online supplemental Table S1). Two-stage cluster sampling was used. In the first stage, an updated list of enumeration areas (EAs) was used as the study sampling frame, with 31 to 44 EAs (clusters) randomly selected per region, resulting in a total of 360 clusters. Due to time constraints and to ensure data validity, insecure or inaccessible EAs were excluded from the study.

In the second stage, all households in an EA were listed and 16 households per EA were selected using a random sampling table). For the age-stratification, two individuals from each household were randomly selected for testing: one aged 5–17 years and one aged ≥18 years.

Serological testing

Finger-prick blood samples were collected from the randomly selected household members in each age category. The antibody RDTs for COVID-19 were performed in the presence of the participant, and the results were shared with them. The COVID-19 RDT used was the COVID-19 IgG/IgM Rapid Test Cassette developed by Healgen Scientific LLC, USA. The RDT is US Food and Drug Administration (FDA)-authorised, with IgM relative sensitivity and specificity of 95.7% and 97.3%, respectively; IgG relative sensitivity and specificity of 91.8% and 96.4%, respectively; and both IgG-positive and/or IgM-positive specificity of 97.5%.

Data collection and analysis

The survey used a validated questionnaire that was initially piloted in Kabul province. All participants were interviewed by the survey team members, who completed a questionnaire that included questions about the demographics of each participant and their household members, their history of exposure to COVID-19, and deaths in the family during the 15-month period beginning in March 2019.

Data collection teams comprised two members, one male and one female; there were 191 teams in total. Due to the need for blood-drawing for samples, the team members were either nurses, midwives or laboratory technicians.

Regional COVID-19 data were entered into DHIS2 (District Health Information Software-2) by disease surveillance officers in the provinces. DHIS2 is the national data warehouse for Afghanistan's health information and includes data that inform the country's COVID-19 dashboard.[5] Various steps were taken for data quality assurance at both regional and central levels within the MoPH; data collection teams were monitored by the master trainers in the regional capitals and by disease surveillance staff in the provinces. Prior to being entered into the system, questionnaires were quality checked

and some participants whose phone numbers were available in the questionnaire were contacted at random by phone call to confirm that their details were correct.

Data were imported into STATA version 15[13] for the statistical analyses. To ensure a representative sample and results, weighted analysis was applied to adjust for the complex survey design. Sample weighting, non-response weighting and post-stratification weighting were performed. The proportions of infections and 95% confidence intervals were calculated and adjusted to take the survey design into account. To determine the overall levels of current and past infection of COVID-19, individuals who tested positive for IgG, IgM or both were summed. To determine the incidence of COVID-19 during the survey period, IgM positivity alone was used.

Adjustment of seroprevalence and exposure inference

We first used a simple Bernoulli model to estimate the regional SARS-CoV-2 (the virus that causes COVID-19) seroprevalence, after adjusting the proportion of individuals in each region with current or past COVID-19 infection according to the sensitivity and specificity of the serology test.[14] (The term 'seroprevalence' below denotes the serology positive ratio already adjusted by the test.) Further details of the method used can be found in online supplemental method, appendix 1. We revised the mathematical model[12] to estimate the total exposure in the population by region after taking into account waning antibody levels. Further details of the method used can be found in online supplemental method, appendix 1.

RESULTS

Demographic details

This seroepidemiological study has provided estimates of the prevalence of SARS-CoV-2 antibodies across Afghanistan, in urban and rural areas, and in the nine regions of the country. Of the 360 clusters identified for participation in the study, 338 (94%) were included; the remainder were

excluded due to insecure or inaccessible EAs and time constraints. A total of 9514 household members from these 338 clusters were interviewed and tested for COVID-19. The mean age of respondents was 27 years, 53.9% were male and 46.1% were female, 73% were from rural areas (online supplemental table S2), and most participants (79.2%) were married.

COVID-19 infections in Afghanistan

The total proportion of COVID-19 infections (including all positive results, the average of both current and past infection) for the whole of Afghanistan was 31.5%. By region, Kabul had the highest proportion of COVID-19 infections (53%), while the Central highlands region had the lowest proportion, at 21.1% (figure 1).

Based on further analysis, the adjusted seroprevalence by region was consistent with the serosurvey results. Kabul still had the highest adjusted seroprevalence (51.8%) (table 1 and figure 2).

RDT results for participants aged 18 years or more

In total, 5618 participants aged ≥18 years were interviewed and tested for this survey. Among this age group, 2056 (35.1%) of individuals tested positive for antibodies against SARS-CoV-2 (table 2). There were 885 (37.2%) females and 1170 (33.9%) males who tested positive, and there was a higher proportion of positive tests in individuals who lived in urban areas compared with the proportion in people who lived in rural areas (773, 42.3% versus 1323, 31.7%, respectively) (table 2). Kabul region had the highest proportion of participants aged ≥18 years who tested positive for antibodies against SARS-CoV-2 (357, 56.8%) (table 2). The survey results revealed that 164 (2.6%) of participants aged ≥18 years were IgM-positive for COVID-19, i.e. they had a current infection, with the highest proportion of current infections in the South-east region (37, 7.0%) (table 1).

Table 1. Seroprevalence of SARS-CoV-2 antibodies and proportion of IgM-seropositive in participants aged ≥18 years by region, area of residence and sex

	Number of positive COVID-19 tests[#]	Seroprevalence % [95% CI]	p-value	Adjusted seroprevalence [95% CI]	Number of IgM-positive COVID-19 tests	IgM-seropositive % COVID-19 tests [95% CI]	p-value
National	2056	35.1 [31–39.5]		29.8 [28.8, 30.7]	164	2.6 [2.0–3.5]	
Region							
Central	254	45.5 [37.8–53.4]	***	34.6 [31.6, 37.6]	28	4.3 [2.4–7.6]	
Central highlands	105	24.9 [17.9–33.7]		19.0 [16.4, 21.8]	5	1.0 [0.4–2.3]	***
East	294	49.1 [41.5–56.8]		41.5 [38.6, 44.4]	16	2.5 [1.4–4.5]	
Kabul	357	56.8 [52.0-62.0]		51.8 [48.8, 54.8]	17	2.7 [1.4–5.0]	
North	212	35.3 [28.1–43.4]		28.9 [26.3, 31.8]	7	1.4 [0.6–3.4]	
North-east	263	39.3 [31.9–47.4]		30.7 [28.1, 33.3]	26	4.0 [2.1–7.8]	
South	115	26.6 [19.0–36.0]		23.9 [20.7, 27.1]	8	1.6 [0.7–3.4]	

South-east	221	40.9 [34.4–47.9]	30.5 [27.4. 33.6]	37	7.0 [3.7–12.9]
West	235	39.8 [34.8–45.1]	32.4 [29.7, 35.2]	20	3.4 [1.8–6.3]
Area of residence					
Rural		31.7 [26.5–37.4]		121	3.7 [1.7–7.9]
Urban		42.3 [35.7–49.2]		43	2.3 [1.2–4.2]
Sex					
Male	1170	33.9 [29–39.2]		104	2.4 [1.4–4.0]
Female	885	37.2 [32–42.6]		60	4.1 [1.8–9.2]
Age (years)					
18–39	1109	33.7 [28.5–39.2]		96	2.7 [1.9–3.9]
40–59	657	36.5 [31.9–41.3]		50	2.4 [1.6–3.7]
60+	290	40.0 [31.8–48.2]		18	2.1 [1.1–4.2]

[#]The total number of positive COVID-19 tests includes all positive results: both current and past infections i.e. IgG-positive, IgM-positive or both. *p<.05, **p<.01, *** p<.001

CI, confidence interval

RDT results for participants aged 5 to 17 years

There were 4346 participants aged 5–17 years interviewed and tested for this survey. Among this age group, a total of 850 (25.3%) individuals tested positive for antibodies against SARS-CoV-2 (table 2), 401 (27.8%) females and 446 (24.2%) males. Again, there was a higher proportion of positive tests in individuals who lived in urban areas compared with the proportion among people who lived in rural areas (322, 30.8% versus 528, 23.4%, respectively) (table 2). Kabul region had the highest proportion of participants aged 5–17 years who tested positive for antibodies against SARS-CoV-2 (177, 46.4%) (table 2). There were 89 (3.3%) participants aged 5–17 years who were IgM-positive for COVID-19, with the highest proportion of current infections in the South region (7, 4.7%) (table 2).

Table 2. Seroprevalence of SARS-CoV-2 antibodies and proportion of IgM-seropositive results in participants aged 5–17 years by region, area of residence and sex.

	Number of positive COVID-19 tests[#]	Seroprevalence % [95% CI]	p-value	Number of IgM-positive COVID-19 tests	IgM-seropositive % COVID-19 tests [95% CI]	p-value
National	850	25.3 [20.5–30.8]		89	3.3 [1.8–6.3]	
Region						
Central	79	21.0 [14.5–29.3]*	*	10	2.8 [1.2–6.3]	**
Central highlands	42	14.6 [8.6–23.8]		3	1.6 [0.4–6.6]	
East	172	32.4 [26.8–38.6]		10	1.4 [0.7–3.1]	
Kabul	177	46.4 [40.8–52.1]		14	3.5 [1.6–7.3]	
North	96	23.0 [16.8–30.8]		6	1.2 [0.4–3.7]	
North-east	108	20.9 [15.1–28.2]		18	2.8 [1.0–7.6]	
South	55	24.4 [14.5–38.0]		7	4.7 [1.6–13.1]	
South-east	42	17.6 [10.6–27.6]		9	2.4 [0.8-6.8]	

West	79	24.5 [18.4–31.8]		12	3.2 [1.7–6.0]	
Area of residence						
Rural	528	23.4 [17.5–30.6]		60	3.7 [1.7–7.9]	
Urban	322	30.8 [24.8–37.5]		29	2.3 [1.2–4.2]	
Sex						
Male	446	24.2 [18.5–31]		47	2.4 [1.4–4.0]	
Female	401	27.8 [21.3–33]		42	4.1 [1.8–9.2]	
Age (years)						
5–9	175	[13.4–26.2]	**	20	3.3 [1.1–9.5]	**
10–14	365	[20.8–33.8]		40	3.7 [1.7–7.9]	
15–17	310	[23.5–35.6]		29	2.8 [1.5–5.2]	

[#]The total number of positive COVID-19 tests includes all positive results: both current and past infections i.e. IgG-positive, IgM-positive or both. *p<.05, **p<.01, *** p<.001

CI, confidence interval

Predictions for cumulative exposure in the population up to 21 July 2020 in the nine regions of Afghanistan are shown in figure 3. The method used for the modelling analysis, which was developed by the COVID-19 International Modelling Consortium (CoMo Consortium), is detailed in the online supplemental method, appendix 1.

The solid orange circles and black error bars in the panel for each region represent the observed seroprevalence data and the associated credible interval (CrI) after adjusting for the sensitivity and specificity of the antibody test. The green and orange lines show the median predictions for exposure and seroprevalence, respectively, while the shaded areas correspond to 95% CrI. The median predicted exposure levels by region (expressed as the proportion of the population that has been infected) as of 21 July 2020 are shown on the map of Afghanistan.

DISCUSSION

This national survey of COVID-19 morbidity in Afghanistan, which was conducted during June and July 2021, revealed that around 10 million people (31.5% of the population) were seropositive for antibodies against SARS-CoV-2, reflecting either current or previous COVID-19 infection. The population of Afghanistan is estimated to comprise approximately 33.6 million people.[15] This finding is reasonably consistent with the results of another telephone survey conducted before July 2020 with a randomly selected sample of 713 healthcare workers to estimate COVID-19 morbidity in the

country. The estimated proportion of individuals who had experienced COVID-19 signs and symptoms was 49.6%, which is close to the value for total infections for most regions reported in the present study, however, no laboratory testing was conducted for the phone survey, which only collected clinical information about symptoms. There is a discrepancy between the serosurvey results and the detected number of COVID-19 infections reported to the surveillance system in the country (36 710 cases reported by the surveillance system as of 30 July 2020 and 156 363 cases as of 5 November 2021) in Afghanistan. The under-reporting of COVID-19 cases is a problem globally due to limited testing availability, flawed test sensitivity, poor surveillance and the indeterminate proportion of asymptomatic infections.[16]

A modelling exercise was performed using the CoMo model to estimate the peak incidence of COVID-19 in Afghanistan. The CoMo model was developed by the CoMo Consortium.[17] The CoMo Consortium adopts a participatory modelling approach,[18] which places in-country subject matter experts at the forefront of model development to ensure that contextual considerations, such as local infrastructure, human resources and sociocultural considerations, are fully taken into account. The CoMo model was used to estimate the peak incidence of COVID-19 in Afghanistan under four scenarios: good, bad, very bad and appropriate, depending on the coverage of and adherence to the NPIs. If the use of NPIs (in a very bad scenario) is not considered, then the COVID-19 peak was predicted to occur in June 2020, with an estimated 69.6% of the population infected and 20 509 deaths by the end of 2020.

In communicable disease epidemiology, one of the key parameters used in decision-making is the estimate of herd immunity in a population. Herd immunity occurs when a certain proportion of the population is immune to a given infectious disease, reducing the probability that the disease will be transmitted from one individual to another, thus helping to protect the entire population from that disease.[19] Herd immunity can be achieved either through individuals being exposed or vaccinated. Determining a country's herd immunity threshold to a given disease is directly related to estimates

of the basic reproductive number, R0, of that disease. R0 indicates the average number of individuals one infected individual can go on to infect in a fully susceptible population. Different herd immunity thresholds in different contexts have been estimated for COVID-19, ranging from 43% to 85%.[19-24] For example, one study indicated that if R0=3, i.e. one infected individual can infect up to three others, meaning 67% of the population must be immune to achieve herd immunity.[20] Estimates by Johns Hopkins University suggest that 70% of the population must be immune to achieve herd immunity and end restrictions on people's day to day lives[19], while another study suggested that R0 values of 1–2, 2–4 and >4 would require herd immunity thresholds of 50%, 56.1–74.8% and 77.9–85%, respectively.[21] In addition to R0 and the herd immunity threshold, the rate of antibody decline post-infection must also be considered, with one study suggesting that antibodies to COVID-19 decline within 94 days of infection.[10]

A study conducted by Eckerle and Meyer revealed that by mid-2020, an insufficient proportion of the population had been infected globally to achieve herd immunity, and these findings were confirmed by reports of low COVID-19 morbidity levels from countries such as Sweden, where an infection rate of 7% was reported by the end of April despite no lockdown; the mentioned study also states that obtaining herd immunity by exposing the population to the disease results in the simultaneous infection of the majority of the population and paves the way for a second wave of the disease.

These estimates of herd immunity thresholds suggest that the present survey findings, of a SARS-CoV-2 antibody seroprevalence of approximately 32% among the population in Afghanistan, mean that the herd immunity threshold had not been reached by the time of the study and most of the country's population remained at risk of infection. However, the herd immunity threshold may have been crossed in some local areas. In Kabul province, for example, more than half of the population has been infected, which exceeds the lowest reported herd immunity threshold of 43%. However, as the majority of the population remains at risk of infection, NPIs should be lifted gradually, as per

WHO guidelines.[25] It should also be noted that this survey was conducted at a time when the SARS-CoV-2 alpha variant was the most prevalent variant in Afghanistan; it is unclear what effect the arrival of new variants, such as the delta and omicron variants, will have on herd immunity thresholds.

Based on evidence from countries with a similar context to Afghanistan, and if we assume an R0 in the country of 2–3, the herd immunity threshold would be between 56% and 75%. Kabul province, with a SARS-CoV-2 seroprevalence of 53%, was within range of this threshold. The Eastern and Central regions, with SARS-CoV-2 seroprevalences of 34% to 42%, were in a relatively good position, but the remaining regions with SARS-CoV-2 seroprevalence of less than 35% were in a worse position and not yet close to the herd immunity threshold at the time of the study (Online supplementary Table S3). As in many low- and middle-income countries, COVID-19 vaccination rates in Afghanistan are low, with just 12% of the population currently fully vaccinated.[5] With the disruptions to the health system as a result of the evolving political situation in the country, the COVID-19 response may deteriorate if control measures are not implemented and vigilantly maintained.

Based on the evidence outlined above, the NPIs currently in place in Afghanistan should not have been lifted, as the herd immunity threshold for the nation has yet to be reached either through natural infection or vaccination. If the NPIs are lifted, the rates of hospitalisation will increase, as will the number of patients requiring ventilation; this will place the health system under considerable pressure. However, after July 2021, the restrictions were reduced and since then the country has only focused on school closures as a mitigation measure to balance the economy, social life and the impact of COVID-19 on the health system. It is worth mentioning that with the recent transition of government in Afghanistan and decreased funding for the country's health system, there are evolving challenges that will ultimately lead to the increased spread of COVID-19 and other infectious diseases. Greater levels of poverty, a displaced population and poor sanitation will further

exacerbate this problem. The influx of refugees from Afghanistan to other countries might also facilitate the cross-border spread of disease. Particularly with the emergence of new variants and low vaccination coverage, it is crucial to have continued public health and social measures to mitigate the impact of COVID-19 in a conflict-affected and unstable country. For the continuation of health services, functional hospitals, surveillance systems and laboratories, as well as a skilled healthcare workforce, are needed to mitigate the spread of COVID-19 and other infections within Afghanistan and prevent the regional and even global spread of disease.

This study had some limitations. First, the time available to conduct the survey was limited. Second, security concerns meant that not all areas could be surveyed; the inability to conduct proper household listing and create maps for enumeration areas in those areas where the government lacked control may have affected the findings. Third, the findings may not reflect the current situation with regards to the new SARS-CoV-2 delta and omicron variants of concern, as the R0 for these variants is not well-established. Once a stable estimate of the R0 for these variants has been established then our findings can be adjusted accordingly to assist with programme planning. Fourth, the data were entered in the DHIS2 database, which created many challenges for data verification, household matching and the subsequent analysis. In future surveys, it would be preferable to collect data by entering them directly via a tablet or similar appropriate research data entry tool to improve the data quality.

CONCLUSION

Although the immunity threshold may have been reached in some localities within Afghanistan, specifically Kabul, this threshold has not yet been reached among the country's entire population. In particular, the proportion of the population that is seropositive for antibodies against SARS-CoV-2 is much lower in rural areas than urban areas. The seroprevalence represents a lower estimate of herd

immunity and the predicted exposure represents an upper limit. Given the large proportion of the population that remains susceptible to COVID-19 infection, and limited COVID-19 vaccination coverage, NPIs and vigilance should remain in place to protect the health system from an unmanageable burden of hospitalisations. The link between the presence of antibodies and immunity has yet to be established, as is the link between prior exposure and immunity. As antibody levels wane, then seroprevalence may provide an underestimate of immunity but, conversely, if immunity wanes, then prior exposure would provide a higher estimate.

Data availability statement

Survey serology data are stored in the Ministry of Public Health national database and are available upon reasonable request. All data, code and materials used in the analyses can be accessed at: https://github.com/SiyuChenOxf/AfghanistanSerologyStudy/tree/master. All parameter estimates and figures 3 and 4 can be reproduced using the code provided. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Acknowledgments

This SARS-CoV-2 Seroprevalence in Afghanistan: A Population-based Study has been supported by WHO Unity Studies, a global sero-epidemiological standardization initiative, with funding to WHO by the COVID-19 Solidarity Response Fund and the German Federal Ministry of Health (BMG) COVID-19 Research and development Fund. We would like to acknowledge the Monitoring and Evaluation and Health Information System General Directorate, which led the study in Afghanistan. The study was funded by Ministry of Public Health. We would also like to acknowledge the regional master trainers and the National Statistics and Information Authority (NSIA) team for their support in conducting the survey. We would additionally like to thank the following for their efforts during the survey: the WHO

team, USAID, World Bank, CDC Atlanta, Performance Management Office (PMO), Grants and Contracts Management Unit (GCMU), Afghanistan National Public Health Institute (ANPHI), all Provincial Public Health Directorates (PPHDs), Health Promotion Department (HPD), all Basic Package of Health Services (BPHS)- and Essential Package of Hospital Service (EPHS)-implementing NGOs, the survey steering committee team, and the data management team. We acknowledge the national and provincial surveillance teams for their extensive support in conducting the survey. We also appreciate all those individuals who supported the data collection, data entry, and listing of households for the survey frame. Medical writing assistance and editorial support was provided by Adam Bodley, according to Good Publication Practice guidelines. The views expressed in this manuscript are those of the authors and do not necessarily reflect the views of the authors' affiliated organisations or the organisations listed here.

Funding Statement:

There are no funders to report for this submission.

Authorship & Contributorship statement

SA. Seedzai, planned the study, lead the study design and tools in the country, coordinated with national authorities, provided ToT training, monitored process of data collection, analysed the data and contributed to writing and finalizing the report and findings of the study and organized dissemination workshops, seminars in the country.

MN. Sahak, supported in planning the study, supported in scientific study design, coordinated internally for sharing scientific WHO tools and protocol, supported in data analysis, contributed to writing and finalizing the study report.

F.Arifi, supported in the design and tools of the study, contributed in analysing and interpreting the results, supported in drafting the report and communicated with all authors for compiling and incorporating the feedback in the final version.

E. Aly, supported in the study design and tools, contributed to data cleaning, analysis & interpretation, and writing the report and finalization of the study report.

M.V.Gurp, supported in study design, complex and multi-level analysis, reviewed and revised the study findings and report.

L.J. White, supported in study method, analysis, interpretation of findings and modelling based on the serosurvey data, reviewed and revised the final study report.

S. Chen, supported in study method, analysis, interpretation of findings and modelling based on the serosurvey data, reviewed and revised the final study report.

A. Barakat, supported in proper laboratory testing of the study method, selection of testing kits, contributed in interpretation of findings, reviewed the draft report and provided inputs to the revision and finalization of the study report.

G. Azim, supported in planning the study, facilitated internal coordination within MoPH, provided ToT training, supervised the data collection in a cluster, supported in drafting the study report and contributed in finalizing the study findings.

B. Rasoly, supported in planning the study, provided ToT training, supervised the data collection in a cluster, supported in data entry and cleaning, drafting the study report and contributed in finalizing the study findings.

S. Safi, supported in planning the study, provided ToT training, supervised the data collection in a clusters and collected data where needed, supported in data entry and cleaning, drafting the study report and contributed in finalizing the study findings.

J. Flegg, contributed in study method, interpretation of serosurvey findings for the context, analysing the data, reviewed the draft report and provided inputs to the finalization of the study report.

N. Ahmed, contributed in study method, interpretation of serosurvey findings for the context, analysing the data, reviewed the draft report and provided inputs to the finalization of the study report.

M.J. Ahadi, supported in planning the study, provided ToT training, supervised the data collection in a clusters and collected data where needed, supported in data entry and cleaning, drafting the study report and contributed in finalizing the study findings.

N.M, Achakzai, supported in proper laboratory testing of the study method, selection of testing kits, contributed in interpretation of findings, reviewed the draft report and provided inputs to the revision and finalization of the study report.

A. AbouZeid, supported in scientific study design, internal WHO coordination, supported in data analysis, contributed to writing and finalizing the study report

Competing interests

No, there are no competing interests for any author

REFERENCES

1. COVID-19 dashboard. Johns Hopkins COVID-19 dashboard Johns Hopkins University; 2021 [Available from: https://coronavirus.jhu.edu/map.html.
2. Burki T. No end in sight for the Brazilian COVID-19 crisis. *Lancet Microbe* 2021;2(5):e180. doi: 10.1016/s2666-5247(21)00095-1 [published Online First: 2021/05/11]

3. Thiagarajan K. Why is India having a covid-19 surge? *BMJ* 2021;373:n1124. doi: 10.1136/bmj.n1124
4. ReliefWeb. Second wave of Covid-19 causes alarm: ReliefWeb; 2021 [Available from: https://reliefweb.int/report/nepal/second-wave-covid-19-causes-alarm.
5. Ministry of Public Health. COVID-19 Dashboard 2021 [Available from: http://covid.moph-dw.org/ accessed 11/06/2021 2021.
6. CDC. Interim Guidelines for COVID-19 Antibody Testing Atlanta: Centers for Disease Control and Prevention; 2021 [Available from: https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html2021.
7. Burgess S, Ponsford MJ, Gill D. Are we underestimating seroprevalence of SARS-CoV-2? *BMJ* 2020;370:m3364. doi: 10.1136/bmj.m3364
8. Ibarrondo FJ, Fulcher JA, Goodman-Meza D, et al. Rapid Decay of Anti–SARS-CoV-2 Antibodies in Persons with Mild Covid-19. *New England Journal of Medicine* 2020;383(11):1085-87. doi: 10.1056/NEJMc2025179
9. Harris RJ, Whitaker HJ, Andrews NJ, et al. Serological surveillance of SARS-CoV-2: Six-month trends and antibody response in a cohort of public health workers. *The Journal of infection* 2021;82(5):162-69. doi: 10.1016/j.jinf.2021.03.015 [published Online First: 2021/03/22]
10. Seow J, Graham C, Merrick B, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nature Microbiology* 2020;5(12):1598-607. doi: 10.1038/s41564-020-00813-8
11. WHO. Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection, 17 March 2020. Geneva: WHO, 2020.
12. Chen S, Flegg JA, White LJ, et al. Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys. *medRxiv* 2021:2021.01.08.21249432. doi: 10.1101/2021.01.08.21249432
13. StataCorp. Stata Statistical Software: Release 15 College Station, TX: StataCorp LLC; 2017 [Available from: https://www.stata.com/stata15/.
14. Healgen Scientific LLC. COVID-19 IgG/IgM Rapid Test Cassette (Whole Blood/Serum/Plasma) Instruction for Use 2020 [Available from: https://www.fda.gov/media/138438/download.
15. NSIA. Estimated Population of Afghanistan 2021-22: NSIA; 2021 [Available from: https://www.nsia.gov.af:8080/wp-content/uploads/2021/06/Estimated-Population-2-of-AFG-New-Design-14.pdf.
16. Wu SL, Mertens AN, Crider YS, et al. Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications* 2020;11(1):4507. doi: 10.1038/s41467-020-18272-4
17. CoMo Consortium. CoMo Consortium: BMJ; 2021 [Available from: https://como.bmj.com/2021.
18. Aguas R, White L, Hupert N, et al. Modelling the COVID-19 pandemic in context: an international participatory approach. *BMJ Glob Health* 2020;5(12) doi: 10.1136/bmjgh-2020-003126 [published Online First: 2020/12/29]
19. D'Souza G, Dowdy D. What is Herd Immunity and How Can We Achieve It With COVID-19? Maryland: Johns Hopkins Bloomberg School of Public Health; 2021 [Available from: https://www.jhsph.edu/covid-19/articles/achieving-herd-immunity-with-covid19.html2021.
20. Randolph HE, Barreiro LB. Herd Immunity: Understanding COVID-19. *Immunity* 2020;52(5):737-41. doi: 10.1016/j.immuni.2020.04.012 [published Online First: 2020/05/21]
21. Kwok KO, Lai F, Wei WI, et al. Herd immunity - estimating the level required to halt the COVID-19 epidemics in affected countries. *J Infect* 2020;80(6):e32-e33. doi: 10.1016/j.jinf.2020.03.027 [published Online First: 2020/03/27]
22. Gomes MGM, Corder RM, King JG, et al. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *medRxiv* 2020:2020.04.27.20081893. doi: 10.1101/2020.04.27.20081893

23. Altmann DM, Douek DC, Boyton RJ. What policy makers need to know about COVID-19 protective immunity. *Lancet* 2020;395(10236):1527-29. doi: 10.1016/s0140-6736(20)30985-5 [published Online First: 2020/05/01]

24. Britton T, Ball F, Trapman P. The disease-induced herd immunity level for Covid-19 is substantially lower than the classical herd immunity level. *arXivorg* 2020: arXiv:2005.03085v1.

25. World Health Organization. Considerations for implementing and adjusting public health and social measures in the context of COVID-19. Geneva: World Health Organization, 2021.

Figure Caption/Legend

Figure 1. Seroprevalence of SARS-CoV-2 antibodies (including all positive results: IgG-positive, IgM-positive or both) among all age groups by region in Afghanistan.

Figure 2. Adjusted seroprevalence by region by the sensitivity and specificity of the serology test for IgG-positive and/or IgM-positive.

Figure 3. Time course of the COVID-19 pandemic up to 21 July 2020 for the nine regions in Afghanistan, for all age groups.

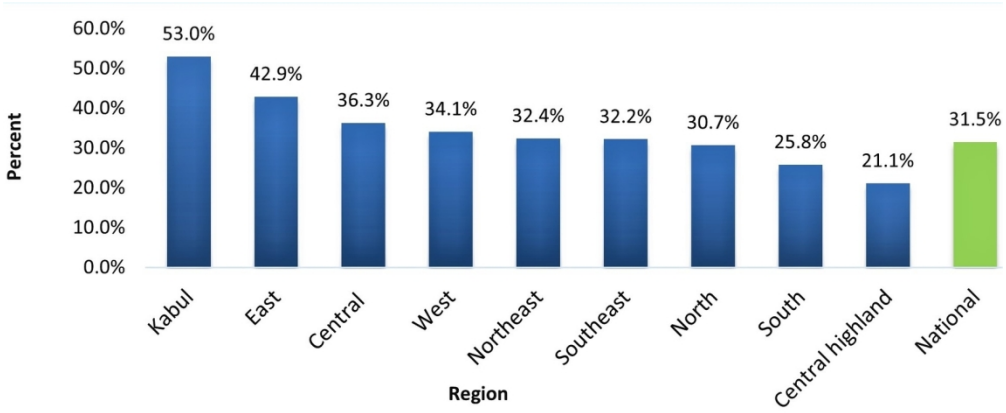


Figure 1. Seroprevalence of SARS-CoV-2 antibodies (including all positive results: IgG-positive, IgM-positive or both) among all age groups by region in Afghanistan.

149x63mm (300 x 300 DPI)

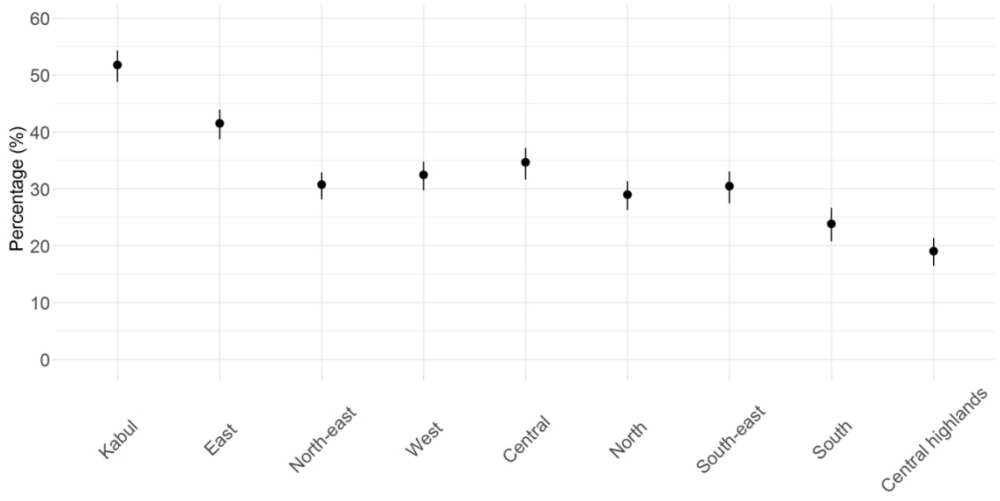


Figure 2. Adjusted seroprevalence by region by the sensitivity and specificity of the serology test for IgG-positive and/or IgM-positive.

158x82mm (220 x 220 DPI)

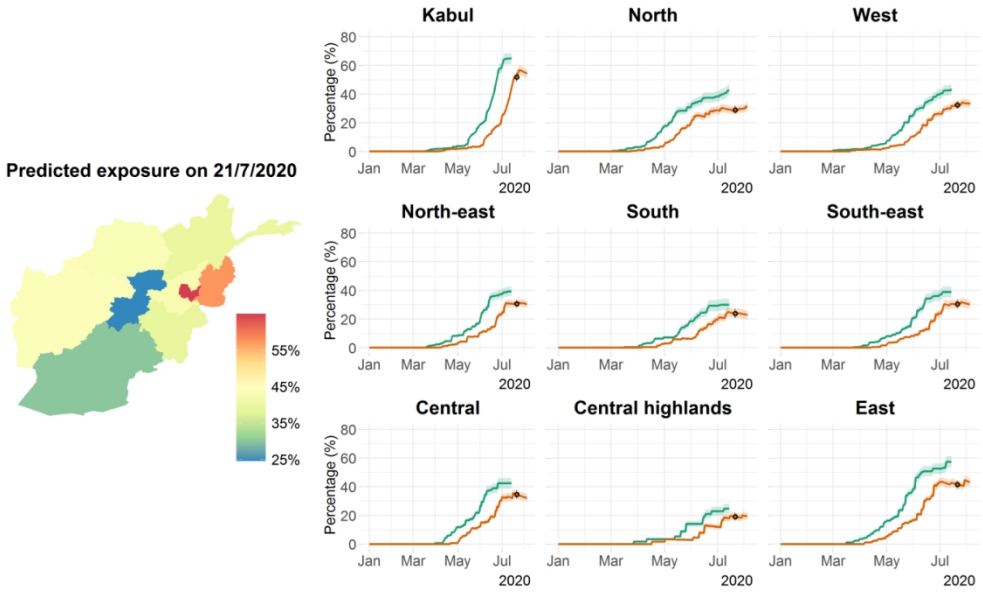


Figure 3. Time course of the COVID-19 pandemic up to 21 July 2020 for the nine regions in Afghanistan, for all age groups.

159x93mm (220 x 220 DPI)

Supplementary material

Supplementary appendix 1 – Methods

Data sources

Regional daily deaths

The documented daily mortality data associated with COVID-19, which might be subject to underreporting, for each of the nine Afghanistan regions (Kabul, East, West, North, South, North-east, South-east, Central, and Central highlands) from January 1 to August 4, 2020, were extracted from the Afghanistan Ministry of Health DHIS2 database.

Regional serology data

The proportion of individuals with current or past COVID-19 infection in each region were obtained from the seroepidemiological study data (table 3, main text). The serology survey provided a result for both IgM and IgG antibodies for each participant, using the COVID-19 IgG/IgM Rapid Test Cassette.[1] The dynamics of IgM and IgG antibodies within an infected individual are complicated.[2] Here, we take the simplified view that an individual who is either IgG positive and/or IgM positive has been exposed to COVID-19 (either past or current infection). Therefore, in the following modelling, the sensitivity and specificity provided by the manufacturer of the imperfect serology test for IgG+ and/or IgM+ was employed.

Adjustment of seroprevalence

We first used a simple Bernoulli model to estimate the regional seroprevalence, after adjusting the proportion of individuals in each region with current or past COVID-19 infection (table 3, main text) according to the sensitivity and specificity of the serology test.[1] (The term 'seroprevalence' below denotes the serology positive ratio already adjusted by the test used.) The Bayesian framework was as follows:

$$x_i(t_0) \sim Beta(1,1)$$

$$w_{ij} \sim bernoulli(k_{se} \times x_i(t_0) + (1 - k_{sp}) \times (1 - x_i(t_0)))$$

$$w_{ij} = \begin{cases} 0, IgG+ \text{ or } IgM+ \\ 1, IgG- \text{ and } IgM- \end{cases}, j = \{1,\ldots,N_i\} \quad (1)$$

where in the first equation given above we have specified a uniform prior for $x_i(t_0)$, which is the proportion of the population in region i that is serology positive, either for IgM or IgG, on July 21,

2020; w_{ij} is the serology survey result for the j-th participant in the serology study from region i; N_i is the total number of participants in the serology survey for region i (table S1); and k_{se} (k_{sp}) is the median of the serology test cassette sensitivity (specificity) reported by the manufacturer.[1] The posterior for seroprevalence on the date the serology survey was conducted, $t = t_0$, was estimated using a Markov chain Monte Carlo (MCMC) implemented in Rstan[3] and denoted as $\tilde{x}_i(t_0)$. The code associated with this work is publicly available, see the Data availability statement of the paper.

Mechanistic model

We revised the mathematical model[4] to account for the underreporting of mortality in the Afghanistan setting according to the varying serology status of the population, $X_i(t)$, of each regional population (for each region $i = 1,\ldots,9$, corresponding to Kabul, East, West, North, South, North-east, South-east, Central, and Central highlands, respectively). The population that has positive serology status increased with exposure of the population to COVID-19 and decreased due to the waning of antibodies.

Given that the constant age-averaged infection fatality rate by region is β_i, the documented mortality over time by region is $m_i(t)$, and the reporting rate of mortality associated with COVID-19 by region is q_i, which is assumed to be constant over time, then, at each time step, of the $\frac{1}{q_i\beta_i}m_i(t)$ individuals who were exposed, $\frac{1}{q_i}m(t)$ die and the remaining number of individuals, $\frac{1-\beta_i}{q_i\beta_i}m_i(t)$, seroconvert from negative to positive. Then, assuming that positive individuals convert to negative at a rate of α, the equation for the rate of change of the number of seropositive individuals is given by:

$$\frac{dX_i(t)}{dt} = \frac{(1-\beta_i)}{q_i\beta_i}m_i(t) - \alpha X_i(t) \quad (2)$$

Solving Equation (2), subject to the initial condition $X_i(t=0) = 0$ where $t = 0$ is time since January 1, 2020, gives:

$$X_i(t) = \frac{(1-\beta_i)e^{-\alpha t}}{q_i\beta_i}\int_0^t e^{\alpha r}m_i(r)dr \quad (3)$$

Discretising Equation **Error! Reference source not found.** with daily intervals ($\Delta r = 1$) gives:

$$X_i(t) = \frac{(1-\beta_i)e^{-\alpha t}}{q_i\beta_i}\sum_{r=0}^{t} e^{\alpha r}m_i(r) \quad (4)$$

Then, the proportion of the population that is serology positive over time, $x_i(t)$, is

$$x_i(t) = \frac{X_i(t)}{P_i - \frac{\sum_{r=0}^{t} m_i(r)}{q_i}} \quad (5)$$

Where P_i is the reported population in region i before the COVID-19 outbreak, and the total proportion of the population that has been exposed over time, $\varepsilon(t - \delta_\epsilon)$, is

$$\varepsilon_i(t - \delta_\epsilon) = \frac{\frac{1-\beta_i}{q_i \beta_i} \sum_{r=0}^{t} m_i(r)}{P_i - \frac{\sum_{r=0}^{t} m_i(r)}{q_i}} \quad (6)$$

where δ_ϵ is the time lag between exposure and seroconversion and is fixed at 21 days.[4]

Exposure inference

We use the posterior samples of seroprevalence at t_0, $\tilde{x}_i(t_0)$, from the MCMC and combine it with Equations (4) and (5) to calculate the posterior samples of reporting rate for mortality, $\tilde{q}_i$:

$$\tilde{q}_i = \frac{(1-\beta_i) e^{-\alpha t_0} \sum_{t=0}^{t_0} e^{\alpha t} m_i(t)}{x_i(t_0) \beta_i P_i} + \frac{\sum_{t=0}^{t_0} m_i(t)}{P_i} \quad (7)$$

Compared with the total population in Afghanistan prior to 2020 (approximately 38 million people), the cumulative mortality associated with COVID-19 by the date of serology survey, $\sum_{t=0}^{t_0} m(t)$, is small. Therefore, it is reasonable to neglect it from Equation (7), which then gives:

$$\tilde{q}_i \approx \frac{(1-\beta_i) e^{-\alpha t_0} \sum_{t=0}^{t_0} e^{\alpha t} m_i(t)}{x_i(t_0) \beta_i P_i} \quad (8)$$

Combining Equations (4), (5) and (8) we can obtain samples of seroprevalence over time, $\tilde{x}_i(t)$:

$$\tilde{x}_i(t) \approx \frac{x_i(t_0) \sum_{r=0}^{t} e^{\alpha r} m_i(r)}{e^{\alpha(t-t_0)} \sum_{t=0}^{t_0} e^{\alpha t} m_i(t)} \quad (9)$$

From Equations (6) and (8) we can obtain samples of the total proportion of the population that has been exposed over time, $\tilde{\varepsilon}_i(t - \delta_\epsilon)$:

$$\tilde{\varepsilon}_i(t - \delta_\epsilon) \approx x_i(t_0) e^{\alpha t_0} \frac{\sum_{r=0}^{t} m_i(r)}{\sum_{t=0}^{t_0} e^{\alpha t} m_i(t)} \quad (10)$$

Note that the seroprevalence (9) and exposure (10) over time are not dependent on β. We use the median estimation of α from the constant infection fatality ratio (IFR) model from Chen et al[4] as an input to Equations (9) and (10).

81

References

1. Healgen Scientific LLC. COVID-19 IgG/IgM Rapid Test Cassette (Whole Blood/Serum/Plasma) Instruction for Use 2020 [Available from: https://www.fda.gov/media/138438/download.

2. Seow J, Graham C, Merrick B, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nature Microbiology* 2020;5(12):1598-607. doi: 10.1038/s41564-020-00813-8

3. Stan Development Team. RStan: the R interface to Stan; R package version 2.21.2. 2020 [Available from: http://mc-stan.org/.

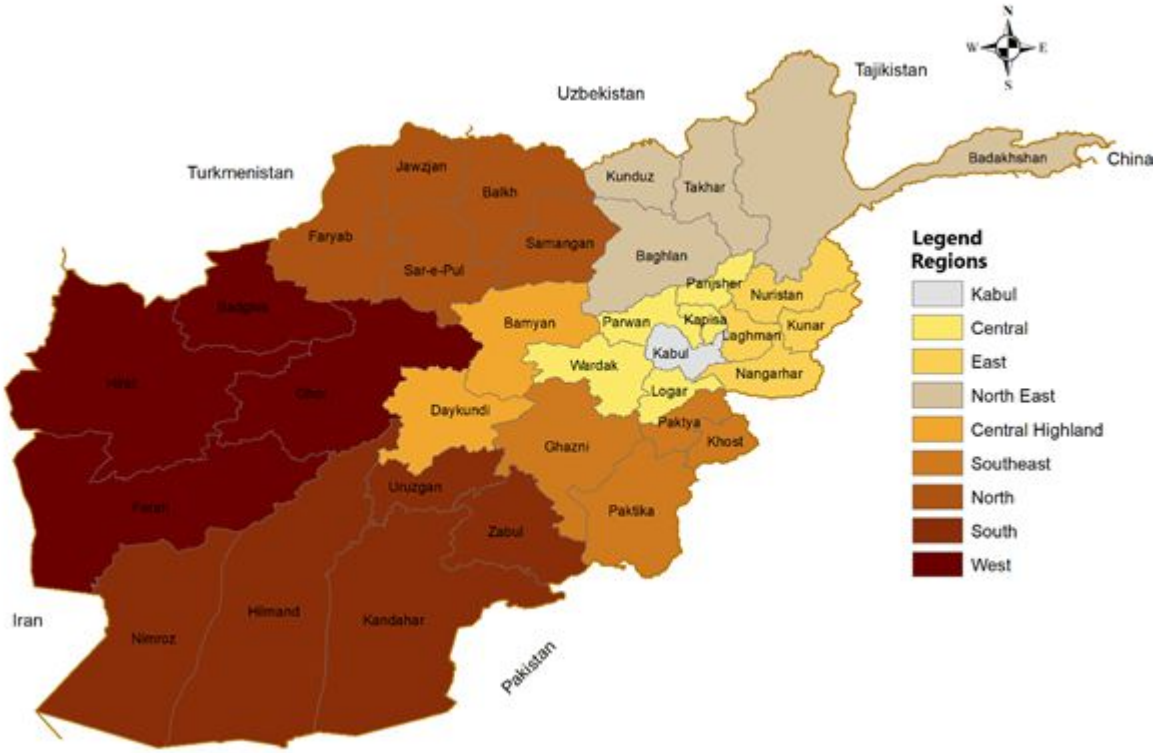
4. Chen S, Flegg JA, White LJ, et al. Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys. *medRxiv* 2021:2021.01.08.21249432. doi: 10.1101/2021.01.08.21249432

Supplementary Table S2. Sample size for the regional serology survey

Region	Sample size
Kabul	1104
Central	1056
Central highlands	902
East	1233
North	1071
South	738
North-east	1265
South-east	969
West	1176

Supplementary Table S2

Respondents' characteristics	Number	Percentage
Total respondents	9514	100%
Sex		
Male	5128	53.9%
Female	4386	46.1%
Age		
5–17 years	4346	45.7%
18 years or more	5168	54.3%
Geographical area		
Urban	2574	27%
Rural	6940	73%
Region		
Kabul	1104	11.6%
Central	1056	11.1%
Central highlands	902	9.4%
East	1233	13.0%
North	1071	11.2%
South	738	7.8%
North-east	1265	13.3%
South-east	969	10.2%
West	1176	12.4%



Supplementary Figure S1. A map showing the nine regions in Afghanistan where the study was conducted (the eight regions of Afghanistan plus Kabul province).

Supplementary Table S3. Seroprevalence of SARS-CoV-2 antibodies in Afghanistan by region, indicating whether herd immunity has been reached

Region	Seroprevalence	Herd immunity reached (based on minimum to maximum of 43% and 85%, respectively)	Number of individuals at risk of infection based on the average of all reported herd immunity thresholds
National	31.5%	No	8 462 611
Kabul	53%	Yes, if based on a herd immunity threshold of 43%; otherwise, no	352 090
East	42.9%	No	479 674
Central	36.3%	No	579 968
West	34.1%	No	1 020 314
North-east	32.4%	No	1 164 297
South-east	32.2%	No	925 019
North	30.7%	No	1 227 256
South	25.8%	No	925 019
Central highlands	21.1%	No	386 875

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5-8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*6*
Bias	9	Describe any efforts to address potential sources of bias	8
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-8
		(*b*) Describe any methods used to examine subgroups and interactions	
		(*c*) Explain how missing data were addressed	7
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	7
		(*e*) Describe any sensitivity analyses	8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	8
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8-9
		(b) Indicate number of participants with missing data for each variable of interest	8
Outcome data	15*	Report numbers of outcome events or summary measures	9
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	9-13

		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	13
Discussion			
Key results	18	Summarise key results with reference to study objectives	13-17
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	17
Generalisability	21	Discuss the generalisability (external validity) of the study results	17
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	18

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

COVID-19 morbidity in Afghanistan: a nationwide, population-based seroepidemiological study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-060739.R1
Article Type:	Original research
Date Submitted by the Author:	30-May-2022
Complete List of Authors:	Saeedzai, Sayed Ataullah; Ministry of Public Health, Monitoring, Evaluation and Health Information System Sahak, Mohammad; World Health Organization, WHE Arifi, Fatima; WHO Aly, Eman; World Health Organisation Regional Office for the Eastern Mediterranean Van Gurp, Margo; KIT Royal Tropical Institute White, Lisa; University of Oxford, Nuffield Department of Medicine; University of Oxford Chen, Siyu; University of Oxford, Discovery, Nuffield Department of Medicine Barakat, Amal; World Health Organisation Regional Office for the Eastern Mediterranean Azim, Giti; Ministry of Public Health Rasoly, Bahara; Ministry of Public Health Safi, Soraya; Ministry of Public Health Flegg, Jennifer; University of Melbourne Ahmed, Nasar; Florida International University, Department of Epidemiology Ahadi, Mohmmad Jamaluddin; Ministry of Public Health Achakzai, Niaz; Ministry of Public Health AbouZeid, Alaa; Cairo University Kasr Alainy Faculty of Medicine
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Infectious diseases, Public health, Epidemiology, Evidence based practice
Keywords:	Epidemiology < INFECTIOUS DISEASES, COVID-19, Public health < INFECTIOUS DISEASES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

COVID-19 morbidity in Afghanistan: a nationwide, population-based seroepidemiological study

Authors & affiliations

Sayed Ataullah Saeedzai[1], Mohammad Nadir Sahak[2], Fatima Arifi[2]*, Eman Abdelkareem Aly[3], Margo Van Gurp[4], Lisa White[5], Siyu Chen[5], Amal Barkat[6], Giti Azim[1], Bahara Rasoly[1], Soraya Safi[1], Jennifer Flegg[7], Nasar U. Ahmed[8], Mohammad Jamaluddin Ahadi[1], Niaz M. Achakzai[9], Alaa Abouzeid[10]

[1]General Directorate of Monitoring, Evaluation and Health Information System (M&E-HIS), Ministry of Public Health, Kabul, Afghanistan

[2]WHO Health Emergencies Programme, WHO, Kabul, Afghanistan

[3] Information Systems for Health Unit, WHO Eastern Mediterranean Regional Office, Cairo, Egypt

[4] KIT Royal Tropical Institute, Amsterdam, the Netherlands

[5] Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, United Kingdom

[6] Infectious Hazard Preparedness Unit, WHO Eastern Mediterranean Regional Office, Cairo, Egypt

[7] School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia

[8] Department of Epidemiology, Robert Stempel College of Public Health, Florida International University, Miami, FL

[9] Department of Molecular Biology, Forensic Medicine Directorate. Ministry of Public Health, Kabul, Afghanistan

[10] Department of Public Health, Faculty of Medicine, Cairo University

*Corresponding author:

Fatima Arifi (Fatima.arefi@gmail.com, arifif@who.int)

Word count: 4135 words

List of Abbreviations

CoMo Consortium – COVID-19 International Modelling Consortium

EA – enumeration area

ELISA – enzyme-linked immunosorbent assay

IFR – infection fatality ratio

IgG – immunoglobulin G

IgM – immunoglobulin M

MCMC – Markov chain Monte Carlo

MoPH – Ministry of Public Health

NPI – nonpharmaceutical intervention

NSIA – National Statistics and Information Authority

R0 – basic reproduction number

RDT – rapid diagnostic test

WHO – World Health Organization

Abstract (300 words)

Objective

The primary objectives were to determine the magnitude of COVID-19 infections in the general population and age-specific cumulative incidence, as determined by seropositivity and clinical symptoms of COVID-19, and to determine the magnitude of asymptomatic or subclinical infections.

Design, setting and participants

We describe a population-based, cross-sectional, age-stratified sero-epidemiological study conducted throughout Afghanistan during June/July 2020. Participants were interviewed to

complete a questionnaire, and rapid diagnostic tests were used to test for SARS-CoV-2 antibodies. This national study was conducted in eight regions of Afghanistan plus Kabul province, considered a separate region. The total sample size was 9514, and the number of participants required in each region was estimated proportionally to the population size of each region. For each region, 31 to 44 enumeration areas (EAs) were randomly selected, and a total of 360 clusters and 16 households per EA were selected using random sampling. To adjust the seroprevalence for test sensitivity and specificity, and seroreversion, Bernoulli's model methodology was used to infer the population exposure in Afghanistan.

Outcome measures

The main outcome was to determine the prevalence of current or past COVID-19 infection

Results

The survey revealed that, to July 2020, around 10 million people in Afghanistan (31.5% of the population) had either current or previous COVID-19 infection. By age group, COVID-19 seroprevalence was reported to be 35.1% and 25.3% among participants aged ≥18 and 5–17 years, respectively. This implies that most of the population remained at risk of infection. However, a large proportion of the population had been infected in some localities, e.g. Kabul province, where more than half of the population had been infected with COVID-19.

Conclusion

As most of the population remained at risk of infection at the time of the study, any lifting of public health and social measures needed to be considered gradually.

Strengths and Limitations

- This is a large-scale, large sample-size, nationwide, population-based, sero-epidemiological study conducted in Afghanistan
- Further analysis is conducted to adjust the seroprevalence for test sensitivity and specificity
- Due to security concerns, not all areas could be surveyed where the government lacked control, and this may have affected the findings

- The findings may not reflect the current situation with regards to the new SARS-CoV-2 delta and omicron variants of concern
- The data were entered in the DHIS2 database, which created many challenges for data verification, household matching and the subsequent analysis.

INTRODUCTION

The COVID-19 pandemic has resulted in more than 248 million confirmed cases and in excess of 5 million deaths globally to November 2021.[1] Many countries are continuing to experience epidemic waves of COVID-19, including Brazil, India and Nepal.[2-4] The first reported case of COVID-19 in Afghanistan was in Herat province on 24 February 2020; as of 20 July 2021, Afghanistan has reported 156 363 confirmed cases of COVID-19 and 7284 deaths from the disease.[5]

When the COVID-19 pandemic began, there were no vaccines or specific treatments available for COVID-19, so nonpharmaceutical interventions (NPIs) were recommended, including social distancing, home quarantine, closure of schools and universities, and bans on public gatherings. Afghanistan introduced NPIs as soon as the first case of COVID-19 was detected in the country. Case detection and isolation were seen as key features in helping to reduce the spread of COVID-19. With the recent political transition in the country and disruption of the health system, public health and social measures to tackle COVID-19 have been completely neglected, which may pose a major risk for increasing the spread of COVID-19 in Afghanistan.

The initial focus of the Afghanistan Ministry of Public Health (MoPH) was on patients with severe COVID-19 disease and ways to decrease mortality associated with the disease. Serological testing of patients can be used to provide useful information about an individual's status in terms of a current or previous COVID-19 infection. Immunoglobulin M (IgM) and G (IgG) antibodies arise at around the same time, between 1 to 3 weeks after infection; however, IgM antibodies decay more rapidly than IgG antibodies.[6] Therefore, for public health studies, IgM is used as a marker of current infection while IgG is used as a marker of previous infection, i.e. within the previous few months. There are

various rapid diagnostic tests (RDTs) available that can be used to simultaneously test blood samples for IgM/IgG antibodies against COVID-19.

Due to the limited testing and surveillance capacity in Afghanistan, it seemed likely there was considerable under-reporting of cases and deaths; therefore, robust scientific studies are required to determine the actual burden of COVID-19 in the country. Serological studies can be used to estimate levels of past exposure and thus position a population in their epidemic timeline. However, serology results might underestimate the total exposure in a population[7] because of decaying antibody titres over time.[8-10] Here, we describe a national seroepidemiological survey initiated by the MoPH and conducted throughout Afghanistan between June and July 2020, involving a questionnaire survey and antibody testing of participants for COVID-19 infection using RDTs. The primary objectives of the study were: 1) to determine the magnitude of COVID-19 infection in the general population and age-specific cumulative incidence, as determined by seropositivity and clinical symptoms of COVID-19; and 2) to determine the magnitude of asymptomatic or subclinical infections. The World Health Organization (WHO) protocol for population-based age-stratified seroepidemiological investigations for COVID-19 infection was adapted for the Afghanistan context to obtain seroprevalence estimates.[11] To adjust the seroprevalence for test sensitivity and specificity, as well as seroreversion, we further adapted a methodology[12] that was originally developed for the England setting and used this to infer the population exposure and undocumented mortality associated with COVID-19 in Afghanistan.

METHODS

Patient and Public Involvement statement

As this study was not a clinical trial and it did not involve patients, no members of the public or patients were directly involved . The study results were disseminated through public workshops in

universities, seminars and workshops, and through the media for the general public. Consent was obtained to be included in the study, and any personal identifier information was excluded during data processing and analysis.

Ethical considerations

Ethical and technical clearance to conduct the survey was obtained from the Institutional Review Board of the Afghanistan MoPH, **Reference number: A.0321.0278**. Informed consent was obtained from participants aged ≥18 years, and assent from family members was obtained for those aged 5–17 years. Individuals who did not provide consent were excluded. Survey team members provided advice about home isolation to participants who tested IgM-positive for COVID-19 during the survey.

Study design

This was a population-based, cross-sectional, age-stratified seroepidemiological study. Participants were interviewed to complete a questionnaire, and RDTs were used to test for SARS-CoV-2 antibodies. The survey was conducted during June and July 2020.

Population and sampling

This was a national study conducted in the eight regions of Afghanistan plus Kabul province, which was considered as a separate region, making nine regions in total (online supplemental figure S1). The total sample size was 9514 and the number of participants required in each region was estimated proportionate to the population size of each region (online supplemental Table S1). Two-stage cluster sampling was used. In the first stage, an updated list of enumeration areas (EAs) was used as the study sampling frame, with 31 to 44 EAs (clusters) randomly selected per region, resulting in a total of 360 clusters. Due to time constraints and to ensure data validity, insecure or inaccessible EAs were excluded from the study.

In the second stage, all households in an EA were listed and 16 households per EA were selected using a random sampling table). For the age-stratification, two individuals from each household were randomly selected for testing: one aged 5–17 years and one aged ≥18 years.

Serological testing

Finger-prick blood samples were collected from the randomly selected household members in each age category. The antibody RDTs for COVID-19 were performed in the presence of the participant, and the results were shared with them. The COVID-19 RDT used was the COVID-19 IgG/IgM Rapid Test Cassette developed by Healgen Scientific LLC, USA. The RDT is US Food and Drug Administration (FDA)-authorised, with IgM relative sensitivity and specificity of 95.7% and 97.3%, respectively; IgG relative sensitivity and specificity of 91.8% and 96.4%, respectively; and both IgG-positive and/or IgM-positive specificity of 97.5%.

Data collection and analysis

The survey used a validated questionnaire that was initially piloted in Kabul province. All participants were interviewed by the survey team members, who completed a questionnaire that included questions about the demographics of each participant and their household members, their history of exposure to COVID-19, and deaths in the family during the 15-month period beginning in March 2019.

Data collection teams comprised two members, one male and one female; there were 191 teams in total. Due to the need for blood-drawing for samples, the team members were either nurses, midwives or laboratory technicians.

Regional COVID-19 data were entered into DHIS2 (District Health Information Software-2) by disease surveillance officers in the provinces. DHIS2 is the national data warehouse for Afghanistan's health information and includes data that inform the country's COVID-19 dashboard.[5] Various steps were

taken for data quality assurance at both regional and central levels within the MoPH; data collection teams were monitored by master trainers in the regional capitals and by disease surveillance staff in the provinces. Prior to being entered into the system, questionnaires were quality checked and some participants whose phone numbers were available in the questionnaire were contacted at random by phone call to confirm that their details were correct.

Data were imported into STATA version 15[13] for the statistical analyses. To ensure a representative sample and results, weighted analysis was applied to adjust for the complex survey design. Sample weighting, non-response weighting and post-stratification weighting were performed. The proportions of infections and 95% confidence intervals were calculated and adjusted to take the survey design into account. The H0 was tested against alternative/research hypothesis at there are differences in prevalence COVID among social demographic and regional characteristics. To determine the overall levels of current and past infection of COVID-19, individuals who tested positive for IgG, IgM or both were summed. To determine the incidence of COVID-19 during the survey period, IgM positivity alone was used.

Adjustment of seroprevalence and exposure inference

We first used a simple Bernoulli model to estimate the regional SARS-CoV-2 (the virus that causes COVID-19) seroprevalence, after adjusting the proportion of individuals in each region with current or past COVID-19 infection according to the sensitivity and specificity of the serology test.[14] (The term 'seroprevalence' below denotes the serology-positive ratio already adjusted by the test.) Further details of the method used can be found in the online supplemental method, appendix 1. We revised the mathematical model[12] to estimate the total exposure in the population by region after taking into account waning antibody levels. Further details of the method used can be found in the online supplemental method, appendix 1.

RESULTS

Demographic details

This seroepidemiological study has provided estimates of the prevalence of SARS-CoV-2 antibodies across Afghanistan, in urban and rural areas, and in the nine regions of the country. Of the 360 clusters identified for participation in the study, 338 (94%) were included; the remainder were excluded due to insecure or inaccessible EAs and time constraints. Similarly, of the total planned 5408 households in 338 clusters, 5177 (96%) households completed the survey. A total of 9514 household members from these 338 clusters were interviewed and tested for COVID-19. The mean age of respondents was 27 years, 53.9% were male and 46.1% were female, 73% were from rural areas (online supplemental table S2), and most participants (79.2%) were married.

COVID-19 infections in Afghanistan

The total proportion of COVID-19 infections (including all positive results, the average of both current and past infection) for the whole of Afghanistan was 31.5%. By region, Kabul had the highest proportion of COVID-19 infections (53%), while the Central highlands region had the lowest proportion, at 21.1% (figure 1).

Based on further analysis, the adjusted seroprevalence by region was consistent with the serosurvey results. Kabul still had the highest adjusted seroprevalence (51.8%) (table 1 and figure 2).

RDT results for participants aged 18 years or more

In total, 5618 participants aged ≥18 years were interviewed and tested for this survey. Among this age group, 2056 (35.1%) of individuals tested positive for antibodies against SARS-CoV-2 (table 2). There were 885 (37.2%) females and 1170 (33.9%) males who tested positive, and there was a higher proportion of positive tests in individuals who lived in urban areas compared with the

proportion in people who lived in rural areas (773, 42.3% versus 1323, 31.7%, respectively) (table 2). Kabul region had the highest proportion of participants aged ≥18 years who tested positive for antibodies against SARS-CoV-2 (357, 56.8%) (table 2). The survey results revealed that 164 (2.6%) of participants aged ≥18 years were IgM-positive for COVID-19, i.e. they had a current infection, with the highest proportion of current infections in the South-east region (37, 7.0%) (table 1).

Table 1. Seroprevalence of SARS-CoV-2 antibodies and proportion of IgM-seropositive in participants aged ≥18 years by region, area of residence and sex

	Number of positive COVID-19 tests[#]	Seroprevalence % [95% CI]	Adjusted seroprevalence [95% CI]	Number of IgM-positive COVID-19 tests	IgM-seropositive % COVID-19 tests [95% CI]
National	2056	35.1 [31–39.5]	29.8 [28.8, 30.7]	164	2.6 [2.0–3.5]
Region* **				Region* **	
Central	254	45.5 [37.8–53.4]	34.6 [31.6, 37.6]	28	4.3 [2.4–7.6]
Central highlands	105	24.9 [17.9–33.7]	19.0 [16.4, 21.8]	5	1.0 [0.4–2.3]
East	294	49.1 [41.5–56.8]	41.5 [38.6, 44.4]	16	2.5 [1.4–4.5]
Kabul	357	56.8 [52.0-62.0]	51.8 [48.8, 54.8]	17	2.7 [1.4–5.0]
North	212	35.3 [28.1–43.4]	28.9 [26.3, 31.8]	7	1.4 [0.6–3.4]
North-east	263	39.3 [31.9–47.4]	30.7 [28.1, 33.3]	26	4.0 [2.1–7.8]
South	115	26.6 [19.0–36.0]	23.9 [20.7, 27.1]	8	1.6 [0.7–3.4]
South-east	221	40.9 [34.4–47.9]	30.5 [27.4. 33.6]	37	7.0 [3.7–12.9]
West	235	39.8 [34.8–45.1]	32.4 [29.7, 35.2]	20	3.4 [1.8–6.3]
Area of residence					
Rural		31.7 [26.5–37.4]		121	3.7 [1.7–7.9]
Urban		42.3 [35.7–49.2]		43	2.3 [1.2–4.2]
Sex					
Male	1170	33.9 [29–39.2]		104	2.4 [1.4–4.0]
Female	885	37.2 [32–42.6]		60	4.1 [1.8–9.2]
Age (years)					
18–39	1109	33.7 [28.5–39.2]		96	2.7 [1.9–3.9]
40–59	657	36.5 [31.9–41.3]		50	2.4 [1.6–3.7]
60+	290	40.0 [31.8–48.2]		18	2.1 [1.1–4.2]

[#]The total number of positive COVID-19 tests includes all positive results: both current and past infections i.e. IgG-positive, IgM-positive or both. *p<.05, **p<.01, *** p<.001

CI, confidence interval

RDT results for participants aged 5 to 17 years

There were 4346 participants aged 5–17 years interviewed and tested for this survey. Among this age group, a total of 850 (25.3%) individuals tested positive for antibodies against SARS-CoV-2 (table 2), 401 (27.8%) females and 446 (24.2%) males. Again, there was a higher proportion of positive tests in individuals who lived in urban areas compared with the proportion among people who lived

in rural areas (322, 30.8% versus 528, 23.4%, respectively) (table 2). Kabul region had the highest proportion of participants aged 5–17 years who tested positive for antibodies against SARS-CoV-2 (177, 46.4%) (table 2). There were 89 (3.3%) participants aged 5–17 years who were IgM-positive for COVID-19, with the highest proportion of current infections in the South region (7, 4.7%) (table 2).

Table 2. Seroprevalence of SARS-CoV-2 antibodies and proportion of IgM-seropositive results in participants aged 5–17 years by region, area of residence and sex.

	Number of positive COVID-19 tests[#]	Seroprevalence % [95% CI]	Number of IgM-positive COVID-19 tests	IgM-seropositive % COVID-19 tests [95% CI]
National	850	25.3 [20.5–30.8]	89	3.3 [1.8–6.3]
Region*			Region**	
Central	79	21.0 [14.5–29.3]*	10	2.8 [1.2–6.3]
Central highlands	42	14.6 [8.6–23.8]	3	1.6 [0.4–6.6]
East	172	32.4 [26.8–38.6]	10	1.4 [0.7–3.1]
Kabul	177	46.4 [40.8–52.1]	14	3.5 [1.6–7.3]
North	96	23.0 [16.8–30.8]	6	1.2 [0.4–3.7]
North-east	108	20.9 [15.1–28.2]	18	2.8 [1.0–7.6]
South	55	24.4 [14.5–38.0]	7	4.7 [1.6–13.1]
South-east	42	17.6 [10.6–27.6]	9	2.4 [0.8-6.8]
West	79	24.5 [18.4–31.8]	12	3.2 [1.7–6.0]
Area of residence				
Rural	528	23.4 [17.5–30.6]	60	3.7 [1.7–7.9]
Urban	322	30.8 [24.8–37.5]	29	2.3 [1.2–4.2]
Sex				
Male	446	24.2 [18.5–31]	47	2.4 [1.4–4.0]
Female	401	27.8 [21.3–33]	42	4.1 [1.8–9.2]
**Age (years) ** **			Age **	
5–9	175	[13.4–26.2]	20	3.3 [1.1–9.5]
10–14	365	[20.8–33.8]	40	3.7 [1.7–7.9]
15–17	310	[23.5–35.6]	29	2.8 [1.5–5.2]

[#]The total number of positive COVID-19 tests includes all positive results: both current and past infections i.e. IgG-positive, IgM-positive or both. *p<.05, **p<.01, *** p<.001

CI, confidence interval

Predictions for cumulative exposure in the population up to 21 July 2020 in the nine regions of Afghanistan are shown in figure 3. The method used for the modelling analysis, which was developed by the COVID-19 International Modelling Consortium (CoMo Consortium), is detailed in the online supplemental method, appendix 1.

The solid orange circles and black error bars in the panel for each region represent the observed seroprevalence data and the associated credible interval (CrI) after adjusting for the sensitivity and specificity of the antibody test. The green and orange lines show the median predictions for exposure and seroprevalence, respectively, while the shaded areas correspond to 95% CrI. The median predicted exposure levels by region (expressed as the proportion of the population that has been infected) as of 21 July 2020 are shown on the map of Afghanistan.

DISCUSSION

This national survey of COVID-19 morbidity in Afghanistan, which was conducted during June and July 2021, revealed that around 10 million people (31.5% of the population) were seropositive for antibodies against SARS-CoV-2, reflecting either current or previous COVID-19 infection. The population of Afghanistan is estimated to comprise approximately 33.6 million people.[15] Our finding is reasonably consistent with the results of a telephone survey conducted before July 2020 with a randomly selected sample of 713 healthcare workers to estimate COVID-19 morbidity in the country. The estimated proportion of individuals who had experienced COVID-19 signs and symptoms was 49.6%, which is close to the value for total infections for most regions reported in the present study, however, no laboratory testing was conducted for the phone survey, which only collected clinical information about symptoms. There is a discrepancy between our serosurvey results and the detected number of COVID-19 infections reported to the surveillance system in the country (36 710 cases reported by the surveillance system as of 30 July 2020 and 156 363 cases as of 5 November 2021). The under-reporting of COVID-19 cases is a problem globally due to limited testing availability, flawed test sensitivity, poor surveillance and the indeterminate proportion of asymptomatic infections.[16] However, some studies have suggested a lower prevalence of COVID-19 in countries during a similar period.[17] For example, the upper bound of COVID-19 prevalence was estimated to be 8.2% in Spain, 6.8% in Italy and 6.1% in the UK. However, the contexts, social mixing and other factors for the demographic scaling model vary across countries, particularly in resource-

limited countries. In such contexts, there are close contacts at home due to large family sizes, while social mixing in schools, communities and society might be more frequent as people rely on daily wages, and the adopted COVID-19 control measures might be less enforced and effective in such settings. Population-based seroprevalence studies are helpful to identify the true burden of disease, which might be higher compared with the burden estimated by modelling studies.

A modelling exercise was performed using the CoMo model to estimate the peak incidence of COVID-19 in Afghanistan. The CoMo model was developed by the CoMo Consortium.[18] The CoMo Consortium adopts a participatory modelling approach,[19] which places in-country subject matter experts at the forefront of model development to ensure that contextual considerations, such as local infrastructure, human resources and sociocultural considerations, are fully taken into account. The CoMo model was used to estimate the peak incidence of COVID-19 in Afghanistan under four scenarios: good, bad, very bad and appropriate, depending on the coverage of and adherence to the NPIs. If the use of NPIs (in a very bad scenario) is not considered, then the COVID-19 peak was predicted to occur in June 2020, with an estimated 69.6% of the population infected and 20 509 deaths by the end of 2020.

In communicable disease epidemiology, one of the key parameters used in decision-making is the estimate of herd immunity in a population. Herd immunity occurs when a certain proportion of the population is immune to a given infectious disease, reducing the probability that the disease will be transmitted from one individual to another, thus helping to protect the entire population from that disease.[20] Herd immunity can be achieved either through individuals being exposed or vaccinated. Determining a country's herd immunity threshold to a given disease is directly related to estimates of the basic reproductive number, R0, of that disease. R0 indicates the average number of individuals one infected individual can go on to infect in a fully susceptible population. Different herd immunity thresholds in different contexts have been estimated for COVID-19, ranging from 43% to 85%.[20-25] For example, one study indicated that if R0=3, i.e. one infected individual can infect up to

three others, 67% of the population must be immune to achieve herd immunity.[21] Estimates by Johns Hopkins University suggest that 70% of the population must be immune to achieve herd immunity and end restrictions on people's day to day lives[20], while another study suggested that R0 values of 1–2, 2–4 and >4 would require herd immunity thresholds of 50%, 56.1–74.8% and 77.9–85%, respectively.[22] In addition to R0 and the herd immunity threshold, the rate of antibody decline post-infection must also be considered, with one study suggesting that antibodies to COVID-19 decline within 94 days of infection.[10]

A study conducted by Eckerle and Meyer revealed that by mid-2020, an insufficient proportion of the population had been infected globally to achieve herd immunity, and these findings were confirmed by reports of low COVID-19 morbidity levels from countries such as Sweden, where an infection rate of 7% was reported by the end of April despite no lockdown; the mentioned study also states that obtaining herd immunity by exposing the population to the disease results in the simultaneous infection of the majority of the population and paves the way for a second wave of the disease.

These estimates of herd immunity thresholds suggest that the present survey findings, of a SARS-CoV-2 antibody seroprevalence of approximately 32% among the population in Afghanistan, mean that less than half of population was infected and most of the country's population remained at risk of infection. However, in some provinces, large numbers of individuals have been infected and recovered from COVID-19. In Kabul province, for example, more than half of the population has been infected. However, as the majority of the population remains at risk of infection, preventive measures and NPIs should be lifted gradually, as per WHO guidelines.[26] It should also be noted that this survey was conducted at a time when the SARS-CoV-2 alpha variant was the most prevalent variant in Afghanistan; it is unclear what effect the arrival of new variants, such as the delta and omicron variants, and vaccination will have on population immunity.

As in many low- and middle-income countries, COVID-19 vaccination rates in Afghanistan are low, with just 12% of the population currently fully vaccinated.[5] With the disruptions to the health system as a result of the evolving political situation in the country, the COVID-19 response may deteriorate if control measures are not implemented and vigilantly maintained.

Based on the evidence outlined above, the NPIs currently in place in Afghanistan should not have been lifted, as large numbers of the population are yet to become immune through natural infection or vaccination. If the NPIs are lifted, the rates of hospitalisation will increase, as will the number of patients requiring ventilation; this will place the already fragile health system under considerable pressure. However, after July 2021, the restrictions were reduced and since then the country has focused on school closures alone as a mitigation measure to balance the economy, social life and the impact of COVID-19 on the health system. It is worth mentioning that with the recent transition of government in Afghanistan and decreased funding for the country's health system, there are evolving challenges that will ultimately lead to the increased spread of COVID-19 and other infectious diseases. Greater levels of poverty, a displaced population and poor sanitation will further exacerbate this problem. The influx of refugees from Afghanistan to other countries might also facilitate the cross-border spread of disease. Particularly with the emergence of new variants and low vaccination coverage, it is crucial to have continued public health and social measures to mitigate the impact of COVID-19 in a conflict-affected and unstable country. For health services to continue, functional hospitals, surveillance systems and laboratories, as well as a skilled healthcare workforce, are needed to mitigate the spread of COVID-19 and other infections within Afghanistan and prevent the regional and even global spread of disease.

This study had some limitations. First, the time available to conduct the survey was limited. Second, security concerns meant that not all areas could be surveyed; the inability to conduct proper household listing and create maps for enumeration areas in those areas where the government lacked control may have affected the findings. Third, the findings may not reflect the current

situation with regards to the new SARS-CoV-2 delta and omicron variants of concern, as the R0 for these variants is not well-established. Once a stable estimate of the R0 for these variants has been established then our findings can be adjusted accordingly to assist with programme planning. Fourth, the data were entered in the DHIS2 database, which created many challenges for data verification, household matching and the subsequent analysis. All data were re-entered in DHIS2 at the central level to ensure data quality and to match the households for reliable and valid analysis. In future surveys, it would be preferable to collect data by entering them directly via a tablet or similar appropriate research data entry tool to improve the data quality.

CONCLUSION

Although the immunity threshold may have been reached in some localities within Afghanistan, specifically Kabul, this threshold has not yet been reached among the country's entire population. In particular, the proportion of the population that is seropositive for antibodies against SARS-CoV-2 is much lower in rural areas than urban areas. The seroprevalence represents a lower estimate of herd immunity and the predicted exposure represents an upper limit. Given the large proportion of the population that remains susceptible to COVID-19 infection, and the limited COVID-19 vaccination coverage, NPIs and vigilance should remain in place to protect the health system from an unmanageable burden of hospitalisations. The link between the presence of antibodies and immunity has yet to be established, as has the link between prior exposure and immunity. As antibody levels wane, seroprevalence may provide an underestimate of immunity but, conversely, if immunity wanes, then prior exposure would provide a higher estimate of immunity.

Data availability statement

Survey serology data are stored in the Ministry of Public Health national database and are available upon reasonable request. All data, code and materials used in the analyses can be accessed at: https://github.com/SiyuChenOxf/AfghanistanSerologyStudy/tree/master. All parameter estimates and figures 3 and 4 can be reproduced using the code provided. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Acknowledgments

This SARS-CoV-2 Seroprevalence in Afghanistan: A Population-based Study has been supported by WHO Unity Studies, a global sero-epidemiological standardization initiative, with funding to WHO by the COVID-19 Solidarity Response Fund and the German Federal Ministry of Health (BMG) COVID-19 Research and Development Fund. We would like to acknowledge the Monitoring and Evaluation and Health Information System General Directorate, which led the study in Afghanistan. The study was funded by Ministry of Public Health. We would also like to acknowledge the regional master trainers and the National Statistics and Information Authority (NSIA) team for their support in conducting the survey. We would additionally like to thank the following for their efforts during the survey: the WHO team, USAID, World Bank, CDC Atlanta, Performance Management Office (PMO), Grants and Contracts Management Unit (GCMU), Afghanistan National Public Health Institute (ANPHI), all Provincial Public Health Directorates (PPHDs), Health Promotion Department (HPD), all Basic Package of Health Services (BPHS)- and Essential Package of Hospital Service (EPHS)-implementing NGOs, the survey steering committee team, and the data management team. We acknowledge the national and provincial surveillance teams for their extensive support in conducting the survey. We also appreciate all those individuals who supported the data collection, data entry, and listing of households for the

survey frame. Medical writing assistance and editorial support was provided by Adam Bodley, according to Good Publication Practice guidelines. The views expressed in this manuscript are those of the authors and do not necessarily reflect the views of the authors' affiliated organisations or the organisations listed here.

Funding Statement:

There are no funders to report for this submission.

Authorship & Contributorship statement

S.A. Seedzai, planned the study, lead the study design and tools in the country, coordinated with national authorities, provided Training of Trainers (ToT), monitored process of data collection, analysed the data and contributed to writing and finalising the report and findings of the study and organised dissemination workshops, seminars in the country.

M.N. Sahak, supported in planning the study, supported in scientific study design, coordinated internally for sharing scientific WHO tools and protocol, supported in data analysis, contributed to writing and finalising the study report.

F. Arifi, supported in the design and tools of the study, contributed in analysing and interpreting the results, supported in drafting the report and communicated with all authors for compiling and incorporating the feedback in the final version.

E. Aly, supported in the study design and tools, contributed to data cleaning, analysis & interpretation, and writing the report and finalisation of the study report.

M.V. Gurp, supported in study design, complex and multi-level analysis, reviewed and revised the study findings and report.

L.J. White, supported in study method, analysis, interpretation of findings and modelling based on the serosurvey data, reviewed and revised the final study report.

S. Chen, supported in study method, analysis, interpretation of findings and modelling based on the serosurvey data, reviewed and revised the final study report.

A. Barakat, supported in proper laboratory testing of the study method, selection of testing kits, contributed to the interpretation of findings, reviewed the draft report and provided inputs to the revision and finalisation of the study report.

G. Azim, supported in planning the study, facilitated internal coordination within MoPH, provided ToT training, supervised the data collection in a cluster, supported in drafting the study report and contributed to finalising the study findings.

B. Rasoly, supported in planning the study, provided ToT training, supervised the data collection in a cluster, supported data entry and cleaning, drafting the study report and contributed to finalising the study findings.

S. Safi, supported in planning the study, provided ToT training, supervised the data collection in a cluster and collected data where needed, supported data entry and cleaning, drafting the study report and contributed to finalising the study findings.

J. Flegg, contributed to the study method, interpretation of serosurvey findings for the context, analysing the data, reviewed the draft report and provided inputs to the finalisation of the study report.

N. Ahmed, contributed to the study method, interpretation of serosurvey findings for the context, analysing the data, reviewed the draft report and provided inputs to the finalisation of the study report.

M.J. Ahadi, supported the planning of the study, provided ToT training, supervised the data collection in clusters and collected data where needed, supported data entry and cleaning, drafting the study report and contributed to finalising the study findings.

N.M. Achakzai, supported correct laboratory testing protocol of the study method, selection of testing kits, contributed to interpretation of findings, reviewed the draft report and provided inputs to the revision and finalisation of the study report.

A. AbouZeid, supported the scientific study design, internal WHO coordination, supported data analysis, and contributed to writing and finalising the study report

Competing interests

No authors declare any competing interests.

REFERENCES

1. COVID-19 dashboard. Johns Hopkins COVID-19 dashboard Johns Hopkins University; 2021 [Available from: https://coronavirus.jhu.edu/map.html.
2. Burki T. No end in sight for the Brazilian COVID-19 crisis. *Lancet Microbe* 2021;2(5):e180. doi: 10.1016/s2666-5247(21)00095-1 [published Online First: 2021/05/11]
3. Thiagarajan K. Why is India having a covid-19 surge? *BMJ* 2021;373:n1124. doi: 10.1136/bmj.n1124
4. ReliefWeb. Second wave of Covid-19 causes alarm: ReliefWeb; 2021 [Available from: https://reliefweb.int/report/nepal/second-wave-covid-19-causes-alarm.
5. Ministry of Public Health. COVID-19 Dashboard 2021 [Available from: http://covid.moph-dw.org/ accessed 11/06/2021 2021.
6. CDC. Interim Guidelines for COVID-19 Antibody Testing Atlanta: Centers for Disease Control and Prevention; 2021 [Available from: https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html2021.
7. Burgess S, Ponsford MJ, Gill D. Are we underestimating seroprevalence of SARS-CoV-2? *BMJ* 2020;370:m3364. doi: 10.1136/bmj.m3364

8. Ibarrondo FJ, Fulcher JA, Goodman-Meza D, et al. Rapid Decay of Anti–SARS-CoV-2 Antibodies in Persons with Mild Covid-19. *New England Journal of Medicine* 2020;383(11):1085-87. doi: 10.1056/NEJMc2025179
9. Harris RJ, Whitaker HJ, Andrews NJ, et al. Serological surveillance of SARS-CoV-2: Six-month trends and antibody response in a cohort of public health workers. *The Journal of infection* 2021;82(5):162-69. doi: 10.1016/j.jinf.2021.03.015 [published Online First: 2021/03/22]
10. Seow J, Graham C, Merrick B, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nature Microbiology* 2020;5(12):1598-607. doi: 10.1038/s41564-020-00813-8
11. WHO. Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection, 17 March 2020. Geneva: WHO, 2020.
12. Chen S, Flegg JA, White LJ, et al. Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys. *medRxiv* 2021:2021.01.08.21249432. doi: 10.1101/2021.01.08.21249432
13. StataCorp. Stata Statistical Software: Release 15 College Station, TX: StataCorp LLC; 2017 [Available from: https://www.stata.com/stata15/.
14. Healgen Scientific LLC. COVID-19 IgG/IgM Rapid Test Cassette (Whole Blood/Serum/Plasma) Instruction for Use 2020 [Available from: https://www.fda.gov/media/138438/download.
15. NSIA. Estimated Population of Afghanistan 2021-22: NSIA; 2021 [Available from: https://www.nsia.gov.af:8080/wp-content/uploads/2021/06/Estimated-Population-2-of-AFG-New-Design-14.pdf.
16. Wu SL, Mertens AN, Crider YS, et al. Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications* 2020;11(1):4507. doi: 10.1038/s41467-020-18272-4
17. Bohk-Ewald C, Dudel C, Myrskylä M. A demographic scaling model for estimating the total number of COVID-19 infections. *Int J Epidemiol* 2021;49(6):1963-71. doi: 10.1093/ije/dyaa198 [published Online First: 2020/12/23]
18. CoMo Consortium. CoMo Consortium: BMJ; 2021 [Available from: https://como.bmj.com/2021.
19. Aguas R, White L, Hupert N, et al. Modelling the COVID-19 pandemic in context: an international participatory approach. *BMJ Glob Health* 2020;5(12) doi: 10.1136/bmjgh-2020-003126 [published Online First: 2020/12/29]
20. D'Souza G, Dowdy D. What is Herd Immunity and How Can We Achieve It With COVID-19? Maryland: Johns Hopkins Bloomberg School of Public Health; 2021 [Available from: https://www.jhsph.edu/covid-19/articles/achieving-herd-immunity-with-covid19.html2021.
21. Randolph HE, Barreiro LB. Herd Immunity: Understanding COVID-19. *Immunity* 2020;52(5):737-41. doi: 10.1016/j.immuni.2020.04.012 [published Online First: 2020/05/21]
22. Kwok KO, Lai F, Wei WI, et al. Herd immunity - estimating the level required to halt the COVID-19 epidemics in affected countries. *J Infect* 2020;80(6):e32-e33. doi: 10.1016/j.jinf.2020.03.027 [published Online First: 2020/03/27]
23. Gomes MGM, Corder RM, King JG, et al. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *medRxiv* 2020:2020.04.27.20081893. doi: 10.1101/2020.04.27.20081893
24. Altmann DM, Douek DC, Boyton RJ. What policy makers need to know about COVID-19 protective immunity. *Lancet* 2020;395(10236):1527-29. doi: 10.1016/s0140-6736(20)30985-5 [published Online First: 2020/05/01]
25. Britton T, Ball F, Trapman P. The disease-induced herd immunity level for Covid-19 is substantially lower than the classical herd immunity level. *arXivorg* 2020: arXiv:2005.03085v1.
26. World Health Organization. Considerations for implementing and adjusting public health and social measures in the context of COVID-19. Geneva: World Health Organization, 2021.

Figure Captions/Legends

Figure 1. Seroprevalence of SARS-CoV-2 antibodies (including all positive results: IgG-positive, IgM-positive or both) among all age groups by region in Afghanistan.

Figure 2. Adjusted seroprevalence by region by the sensitivity and specificity of the serology test for IgG-positive and/or IgM-positive.

Figure 3. Time course of the COVID-19 pandemic up to 21 July 2020 for the nine regions in Afghanistan, for all age groups.

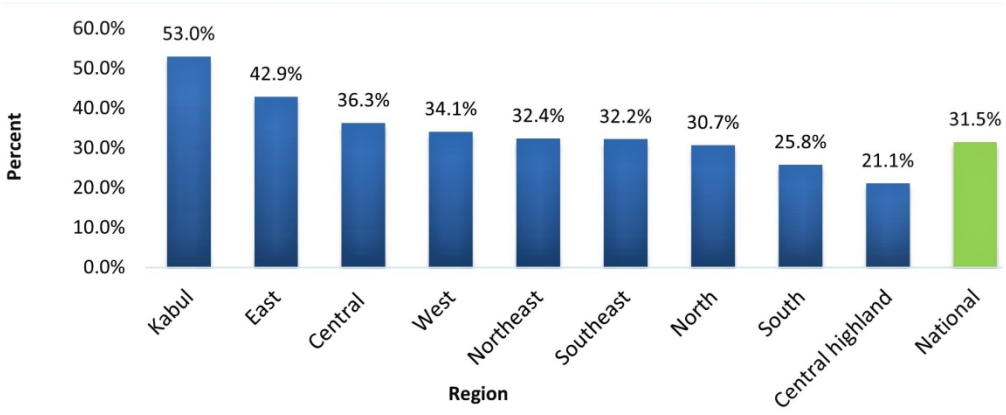


Figure 1. Seroprevalence of SARS-CoV-2 antibodies (including all positive results: IgG-positive, IgM-positive or both) among all age groups by region in Afghanistan.

149x63mm (330 x 330 DPI)

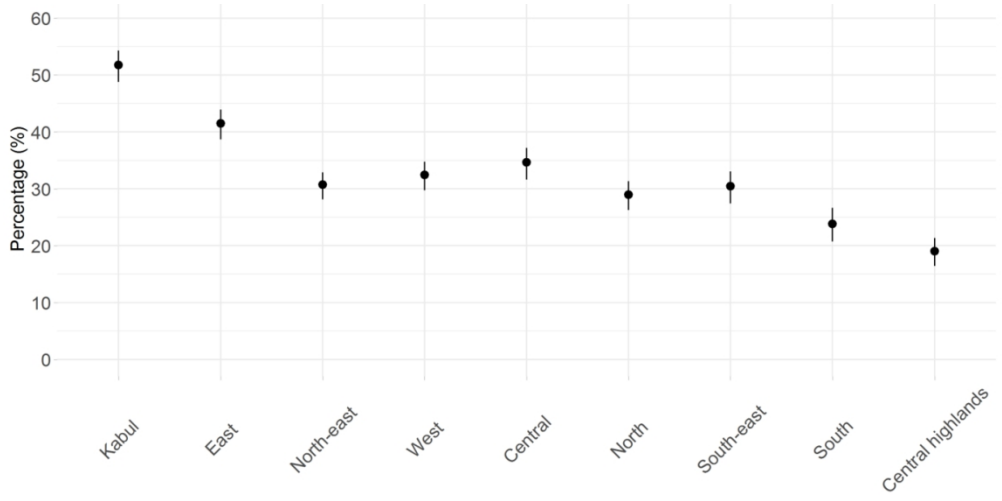


Figure 2. Adjusted seroprevalence by region by the sensitivity and specificity of the serology test for IgG-positive and/or IgM-positive.

158x82mm (220 x 220 DPI)

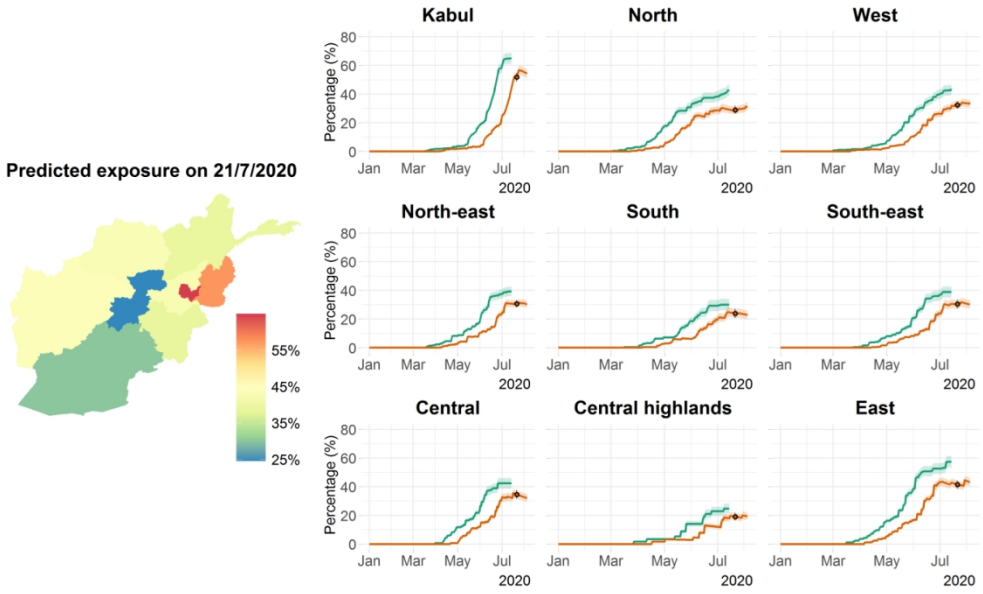


Figure 3. Time course of the COVID-19 pandemic up to 21 July 2020 for the nine regions in Afghanistan, for all age groups.

159x93mm (220 x 220 DPI)

Supplementary material

Supplementary appendix 1 – Methods

Data sources

Regional daily deaths

The documented daily mortality data associated with COVID-19, which might be subject to underreporting, for each of the nine Afghanistan regions (Kabul, East, West, North, South, North-east, South-east, Central, and Central highlands) from January 1 to August 4, 2020, were extracted from the Afghanistan Ministry of Health DHIS2 database.

Regional serology data

The proportion of individuals with current or past COVID-19 infection in each region were obtained from the seroepidemiological study data (table 3, main text). The serology survey provided a result for both IgM and IgG antibodies for each participant, using the COVID-19 IgG/IgM Rapid Test Cassette.[1] The dynamics of IgM and IgG antibodies within an infected individual are complicated.[2] Here, we take the simplified view that an individual who is either IgG positive and/or IgM positive has been exposed to COVID-19 (either past or current infection). Therefore, in the following modelling, the sensitivity and specificity provided by the manufacturer of the imperfect serology test for IgG+ and/or IgM+ was employed.

Adjustment of seroprevalence

We first used a simple Bernoulli model to estimate the regional seroprevalence, after adjusting the proportion of individuals in each region with current or past COVID-19 infection (table 3, main text) according to the sensitivity and specificity of the serology test.[1] (The term 'seroprevalence' below denotes the serology positive ratio already adjusted by the test used.) The Bayesian framework was as follows:

$$x_i(t_0) \sim Beta(1,1)$$

$$w_{ij} \sim bernoulli\left(k_{se} \times x_i(t_0) + \left(1 - k_{sp}\right) \times \left(1 - x_i(t_0)\right)\right)$$

$$w_{ij} = \begin{cases} 0, IgG + or\ IgM + \\ 1, IgG - and\ IgM - \end{cases}, j = \{1, \ldots, N_i\} \qquad (1)$$

where in the first equation given above we have specified a uniform prior for $x_i(t_0)$, which is the proportion of the population in region i that is serology positive, either for IgM or IgG, on July 21, 2020; w_{ij} is the serology survey result for the j-th participant in the serology study from region i; N_i is the total number of participants in the serology survey for region i (table S1); and k_{se} (k_{sp}) is the

median of the serology test cassette sensitivity (specificity) reported by the manufacturer.[1] The posterior for seroprevalence on the date the serology survey was conducted, $t = t_0$, was estimated using a Markov chain Monte Carlo (MCMC) implemented in Rstan[3] and denoted as $\tilde{x}_i(t_0)$. The code associated with this work is publicly available, see the Data availability statement of the paper.

Mechanistic model

We revised the mathematical model[4] to account for the underreporting of mortality in the Afghanistan setting according to the varying serology status of the population, $X_i(t)$, of each regional population (for each region $i = 1,\ldots,9$, corresponding to Kabul, East, West, North, South, North-east, South-east, Central, and Central highlands, respectively). The population that has positive serology status increased with exposure of the population to COVID-19 and decreased due to the waning of antibodies.

Given that the constant age-averaged infection fatality rate by region is β_i, the documented mortality over time by region is $m_i(t)$, and the reporting rate of mortality associated with COVID-19 by region is q_i, which is assumed to be constant over time, then, at each time step, of the $\frac{1}{q_i\beta_i}m_i(t)$ individuals who were exposed, $\frac{1}{q_i}m(t)$ die and the remaining number of individuals, $\frac{1-\beta_i}{q_i\beta_i}m_i(t)$, seroconvert from negative to positive. Then, assuming that positive individuals convert to negative at a rate of α, the equation for the rate of change of the number of seropositive individuals is given by:

$$\frac{dX_i(t)}{dt} = \frac{(1-\beta_i)}{q_i\beta_i}m_i(t) - \alpha X_i(t) \quad (2)$$

Solving Equation (2), subject to the initial condition $X_i(t=0) = 0$ where $t = 0$ is time since January 1, 2020, gives:

$$X_i(t) = \frac{(1-\beta_i)e^{-\alpha t}}{q_i\beta_i}\int_0^t e^{\alpha r}m_i(r)dr \quad (3)$$

Discretising Equation **Error! Reference source not found.** with daily intervals ($\Delta r = 1$) gives:

$$X_i(t) = \frac{(1-\beta_i)e^{-\alpha t}}{q_i\beta_i}\sum_{r=0}^{t} e^{\alpha r}m_i(r) \quad (4)$$

Then, the proportion of the population that is serology positive over time, $x_i(t)$, is

$$x_i(t) = \frac{X_i(t)}{P_i - \frac{\sum_{r=0}^{t} m_i(r)}{q_i}} \quad (5)$$

Where P_i is the reported population in region i before the COVID-19 outbreak, and the total proportion of the population that has been exposed over time, $\varepsilon(t-\delta_\epsilon)$, is

$$\varepsilon_i(t-\delta_\epsilon) = \frac{\frac{1-\beta_i}{q_i\beta_i}\sum_{r=0}^{t} m_i(r)}{P_i - \frac{\sum_{r=0}^{t} m_i(r)}{q_i}} \quad (6)$$

where δ_ϵ is the time lag between exposure and seroconversion and is fixed at 21 days.[4]

Exposure inference

We use the posterior samples of seroprevalence at t_0, $\tilde{x}_i(t_0)$, from the MCMC and combine it with Equations (4) and (5) to calculate the posterior samples of reporting rate for mortality, $\tilde{q}_i$:

$$\tilde{q}_i = \frac{(1-\beta_i)e^{-\alpha t_0}\sum_{t=0}^{t_0} e^{\alpha t} m_i(t)}{x_i(t_0)\beta_i P_i} + \frac{\sum_{t=0}^{t_0} m_i(t)}{P_i} \quad (7)$$

Compared with the total population in Afghanistan prior to 2020 (approximately 38 million people), the cumulative mortality associated with COVID-19 by the date of serology survey, $\sum_{t=0}^{t_0} m(t)$, is small. Therefore, it is reasonable to neglect it from Equation (7), which then gives:

$$\tilde{q}_i \approx \frac{(1-\beta_i)e^{-\alpha t_0}\sum_{t=0}^{t_0} e^{\alpha t} m_i(t)}{x_i(t_0)\beta_i P_i} \quad (8)$$

Combining Equations (4), (5) and (8) we can obtain samples of seroprevalence over time, $\tilde{x}_i(t)$:

$$\tilde{x}_i(t) \approx \frac{x_i(t_0)}{e^{\alpha(t-t_0)}} \frac{\sum_{r=0}^{t} e^{\alpha r} m_i(r)}{\sum_{t=0}^{t_0} e^{\alpha t} m_i(t)} \quad (9)$$

From Equations (6) and (8) we can obtain samples of the total proportion of the population that has been exposed over time, $\tilde{\varepsilon}_i(t-\delta_\epsilon)$:

$$\tilde{\varepsilon}_i(t-\delta_\epsilon) \approx x_i(t_0)e^{\alpha t_0} \frac{\sum_{r=0}^{t} m_i(r)}{\sum_{t=0}^{t_0} e^{\alpha t} m_i(t)} \quad (10)$$

Note that the seroprevalence (9) and exposure (10) over time are not dependent on β. We use the median estimation of α from the constant infection fatality ratio (IFR) model from Chen et al[4] as an input to Equations (9) and (10).

References

1. Healgen Scientific LLC. COVID-19 IgG/IgM Rapid Test Cassette (Whole Blood/Serum/Plasma) Instruction for Use 2020 [Available from: https://www.fda.gov/media/138438/download.

2. Seow J, Graham C, Merrick B, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nature Microbiology* 2020;5(12):1598-607. doi: 10.1038/s41564-020-00813-8

3. Stan Development Team. RStan: the R interface to Stan; R package version 2.21.2. 2020 [Available from: http://mc-stan.org/.

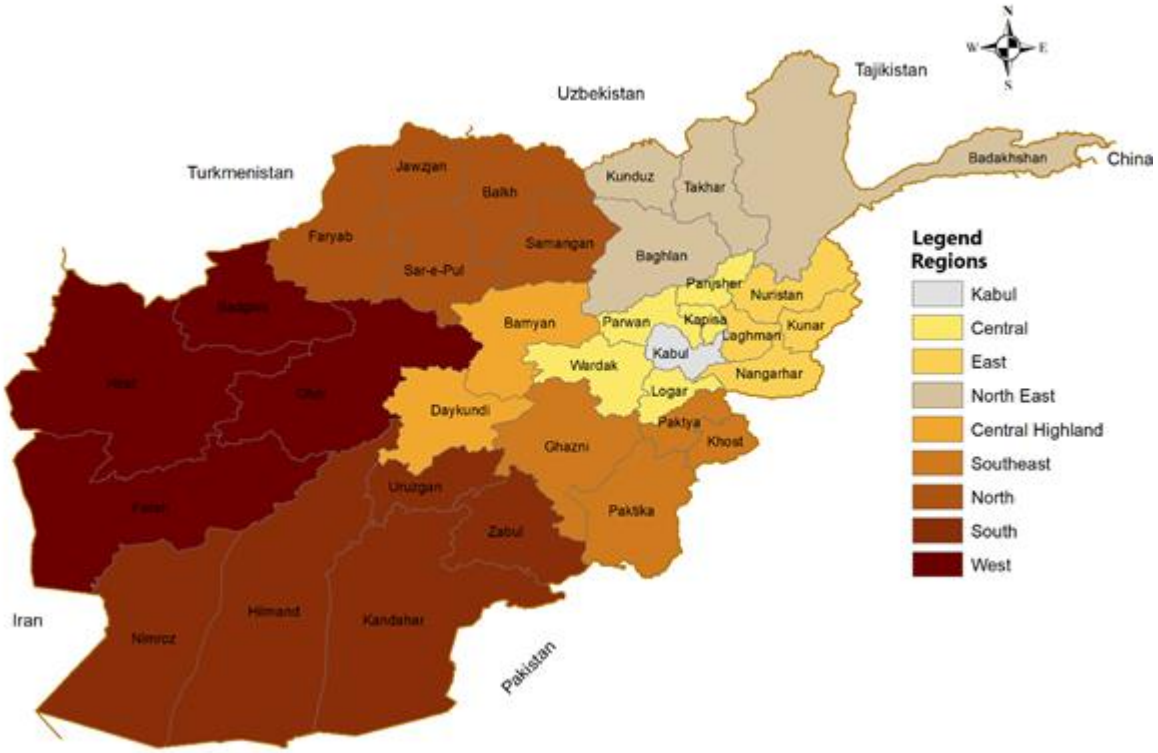
4. Chen S, Flegg JA, White LJ, et al. Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys. *medRxiv* 2021:2021.01.08.21249432. doi: 10.1101/2021.01.08.21249432

Supplementary Table S1. Sample size for the regional serology survey

Region	Sample size
Kabul	1104
Central	1056
Central highlands	902
East	1233
North	1071
South	738
North-east	1265
South-east	969
West	1176

Supplementary Table S2

Respondents' characteristics	Number	Percentage
Total respondents	9514	100%
Sex		
Male	5128	53.9%
Female	4386	46.1%
Age		
5–17 years	4346	45.7%
18 years or more	5168	54.3%
Geographical area		
Urban	2574	27%
Rural	6940	73%
Region		
Kabul	1104	11.6%
Central	1056	11.1%
Central highlands	902	9.4%
East	1233	13.0%
North	1071	11.2%
South	738	7.8%
North-east	1265	13.3%
South-east	969	10.2%
West	1176	12.4%



Supplementary Figure S1. A map showing the nine regions in Afghanistan where the study was conducted (the eight regions of Afghanistan plus Kabul province).

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5-8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*6*
Bias	9	Describe any efforts to address potential sources of bias	8
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-8
		(*b*) Describe any methods used to examine subgroups and interactions	
		(*c*) Explain how missing data were addressed	7
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	7
		(*e*) Describe any sensitivity analyses	8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	8
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8-9
		(b) Indicate number of participants with missing data for each variable of interest	8
Outcome data	15*	Report numbers of outcome events or summary measures	9
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	9-13

		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	13
Discussion			
Key results	18	Summarise key results with reference to study objectives	13-17
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	17
Generalisability	21	Discuss the generalisability (external validity) of the study results	17
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	18

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.