PLOS ONE

COVID-19 risk score as a public health tool to guide targeted testing: A demonstration study in Qatar

Laith J. Abu-Raddad[1,2,3,4]*, **Soha Dargham**[1,2], **Hiam Chemaitelly**[1,2,3], **Peter Coyle**[5,6,7], **Zaina Al Kanaani**[5], **Einas Al Kuwari**[5], **Adeel A. Butt**[3,5], **Andrew Jeremijenko**[5], **Anvar Hassan Kaleeckal**[5], **Ali Nizar Latif**[5], **Riyazuddin Mohammad Shaik**[5], **Hanan F. Abdul Rahim**[8], **Gheyath K. Nasrallah**[7,9], **Hadi M. Yassine**[7,9], **Mohamed G. Al Kuwari**[10], **Hamad Eid Al Romaihi**[11], **Mohamed H. Al-Thani**[11], **Abdullatif Al Khal**[5], **Roberto Bertollini**[11]

1 Infectious Disease Epidemiology Group, Weill Cornell Medicine-Qatar, Cornell University, Doha, Qatar, 2 World Health Organization Collaborating Centre for Disease Epidemiology Analytics on HIV/AIDS, Sexually Transmitted Infections, and Viral Hepatitis, Weill Cornell Medicine–Qatar, Cornell University, Qatar Foundation – Education City, Doha, Qatar, 3 Department of Population Health Sciences, Weill Cornell Medicine, Cornell University, New York, New York, United States of America, 4 Department of Public Health, College of Health Sciences, QU Health, Qatar University, Doha, Qatar, 5 Hamad Medical Corporation, Doha, Qatar, 6 Wellcome-Wolfson Institute for Experimental Medicine, Queens University, Belfast, United Kingdom, 7 Biomedical Research Center, QU Health, Qatar University, Doha, Qatar, 8 College of Health Sciences, QU Health, Qatar University, Doha, Qatar, 9 Department of Biomedical Science, College of Health Sciences, QU Health, Qatar University, Doha, Qatar, 10 Primary Health Care Corporation, Doha, Qatar, 11 Ministry of Public Health, Doha, Qatar

* lja2002@qatar-med.cornell.edu

Abstract

We developed a Coronavirus Disease 2019 (COVID-19) risk score to guide targeted RT-PCR testing in Qatar. The Qatar national COVID-19 testing database, encompassing a total of 2,688,232 RT-PCR tests conducted between February 5, 2020-January 27, 2021, was analyzed. Logistic regression analyses were implemented to derive the COVID-19 risk score, as a tool to identify those at highest risk of having the infection. Score cut-off was determined using the ROC curve based on maximum sum of sensitivity and specificity. The score's performance diagnostics were assessed. Logistic regression analysis identified age, sex, and nationality as significant predictors of infection and were included in the risk score. The ROC curve was generated and the area under the curve was estimated at 0.63 (95% CI: 0.63–0.63). The score had a sensitivity of 59.4% (95% CI: 59.1%-59.7%), specificity of 61.1% (95% CI: 61.1%-61.2%), a positive predictive value of 10.9% (95% CI: 10.8%-10.9%), and a negative predictive value of 94.9% (94.9%-95.0%). The concept and utility of a COVID-19 risk score were demonstrated in Qatar. Such a public health tool can have considerable utility in optimizing testing and suppressing infection transmission, while maximizing efficiency and use of available resources.

Introduction

Suppressing the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic necessitates strategic preparedness and response [1]. The World Health Organization (WHO) has urged countries to adopt a "testing, tracing, and isolation" approach as the "backbone" of their SARS-CoV-2 national response [2]. However, to suppress the epidemic, deliver healthcare services to those in need, and ensure optimal use of resources, testing strategies need to be guided by real-time data analysis so that testing is prioritized to those at higher risk of exposure.

A risk score is an objective set of simple questions or measurements that can be used to assess the likelihood of an individual having a specific infection/disease condition [3–6]. Such scores have been useful in designing initial screening or testing strategies for a variety of diseases, as they reduce the need for more invasive, time-consuming, and expensive testing, while optimizing resource allocation by targeting individuals at higher risk of having the infection/disease [7]. The utility of developing a risk score for SARS-CoV-2 infection offers the benefits of earlier case detection, isolation of cases, and quarantine of contacts, given the disease burden associated with this infection.

Qatar is a high-income country in the Arabian Gulf with a total population of 2.8 million, the majority of whom (89%) are expatriates from over 150 countries [8–10]. The nation's rapid development resulted in a unique socio-demographic structure dominated by men, who comprise 74% of the total population [8], and by younger age cohorts (ages 20–50 years), who likewise comprise 74% of the population [8].

The country was afflicted with a large first epidemic wave of SARS-CoV-2 infection that peaked toward the end of May, 2020 [11]. As of January 29, 2021, >65,000 infections per million population had been laboratory-confirmed [11–13]. Qatar has also one of the world's most extensive databases to document this epidemic and its toll at the national level [14], such that Qatar's epidemic has been one of the most thoroughly investigated and best characterized [11, 14–24].

This study had three objectives. The first was to present a derived risk score for SARS-CoV-2 infection that was developed during the first epidemic wave in April, 2020 to inform the national response to the epidemic. The second objective was to assess the prospective performance of this risk score on epidemic data collected after its derivation. The third objective was to update this risk score to end of January, 2021, and to assess its diagnostic metrics for future use as part of the national response.

The overarching goal of this study was to demonstrate the feasibility and utility of the concept of a Coronavirus Disease 2019 (COVID-19) risk score as a public health tool in an emergent epidemic, applying it to a specific country. Building on the public health utility of risk scores for other diseases such as diabetes [3–6], we believe that this study provides the first COVID-19 risk score for any country. The score has been named "*COVID-19 risk score*", given the prevailing public use of "COVID-19", as opposed to SARS-CoV-2.

Materials and methods

Data source

We analyzed the national database for SARS-CoV-2 real-time polymerase chain reaction (RT-PCR) testing compiled by Hamad Medical Corporation (HMC), the main public healthcare provider in Qatar. The database includes results of all RT-PCR testing conducted in Qatar, regardless of whether it was for suspected SARS-CoV-2 cases, traced contacts, infection surveillance, or other purposes, between February 5, 2020 and January 27, 2021. February 5 is

the day on which the first RT-PCR positive patient was diagnosed, a traveler arriving in Qatar [16].

The database included the swab date, SARS-CoV-2 laboratory result, and demographic information, including age, sex, and nationality. Age was categorized into 10-year age brackets, except for the last category (<10, 10–19, . . ., 80+). Nationality comprised 11 classifications. Nationalities with <1% of diagnosed SARS-CoV-2 cases were grouped as "Other nationalities"; nationalities with >1% of diagnosed cases constituted individual categories.

Two risk scores were derived. The "original" Qatar COVID-19 risk score was derived in April 2020, during the expanding phase of the epidemic [11], utilizing half of the RT-PCR tests administered from February 5, 2020 to April 21, 2020. This half of the sample was chosen randomly. Performance of the risk score was subsequently assessed and validated utilizing the remaining half of the sample.

Similarly, an updated version of the Qatar COVID-19 risk score was derived utilizing half of the RT-PCR testing sample compiled from February 5, 2020 to January 27, 2021. Performance of the updated risk score was subsequently assessed and validated utilizing the remaining half of the sample.

Laboratory methods

Nasopharyngeal and/or oropharyngeal swabs (Huachenyang Technology, China) were collected for PCR testing and placed in Universal Transport Medium (UTM). Aliquots of UTM were: extracted on a QIAsymphony platform (QIAGEN, USA) and tested with RT-qPCR using TaqPath™ COVID-19 Combo Kits (100% sensitivity and specificity [25]; Thermo Fisher Scientific, USA) on an ABI 7500 FAST (ThermoFisher, USA); extracted using a custom protocol [26] on a Hamilton Microlab STAR (Hamilton, USA) and tested using AccuPower SARS-CoV-2 Real-Time RT-PCR Kits (100% sensitivity and specificity [27]; Bioneer, Korea) on an ABI 7500 FAST; or loaded directly into a Roche cobas® 6800 system and assayed with a cobas® SARS-CoV-2 Test (95% sensitivity, 100% specificity [28]; Roche, Switzerland). The first assay targets the viral S, N, and ORF1ab regions. The second targets the viral RdRp and E-gene regions, and the third targets the ORF1ab and E-gene regions.

All tests were conducted at the HMC Central Laboratory or Sidra Medicine Laboratory, following standardized protocols.

Statistical analysis

Risk score derivation. Bivariable logistic regressions were performed to identify associations between each demographic factor and SARS-CoV-2 status. Multivariable logistic regression was then conducted to identify independent predictors of SARS-CoV-2 RT-PCR positivity and to estimate adjusted odds ratios (aOR) and corresponding 95% confidence intervals (CI). A p-value ≤ 0.05 in the multivariable analysis for any predictor was considered to provide strong evidence for an association with the outcome. Predictors with p-values ≤ 0.05 were retained in deriving the Qatar COVID-19 risk score.

Each predictor level was assigned scoring points using the corresponding regression model's β-coefficient multiplied by 10 (and rounded to the nearest integer) for ease of implementation, per established methodology [3–6, 29]. An aggregate risk score for each test was then derived by summing the scoring points, given the individual's profile. No interaction terms between covariates were included, so as to keep the score simple and accessible for broad use. The score was used to determine an individual's level of risk of exposure to SARS-CoV-2 infection.

Risk score performance and validation. A receiving operating characteristics (ROC) curve was plotted to determine the capacity of the risk score to diagnose SARS-CoV-2 infection at different cut-off values resulting in a positive outcome. Sensitivity was defined as the proportion of those with a positive outcome when applying the score among tests with a positive RT-PCR result, that is, the capacity of the score to detect a true SARS-CoV-2 infection. Specificity was defined as the proportion of those with a negative outcome when applying the score among the tests with a negative RT-PCR result, that is, the capacity of the score to detect true absence of SARS-CoV-2 infection.

The optimal score cut-off/criterion to identify infected or uninfected cases was determined by selecting the value that maximized the sum of sensitivity and specificity. The area under the ROC curve (AUC) was also estimated to quantify the accuracy of the risk score, that is, how well the risk score separated infected from uninfected persons.

The risk score derived utilizing half of the sample was applied to the other half of the sample to assess and validate its performance. The risk score's predictive and diagnostic performance was assessed by estimating the sensitivity, specificity, positive predictive value (PPV; probability of being infected given a positive outcome when applying the score), and the negative predictive value (NPV; probability of being uninfected given a negative outcome when applying the score).

Performance assessment of the original risk score on prospective data. The "original" Qatar COVID-19 risk score that was derived from testing data up to April 21, 2020 was applied to all testing data from April 22, 2020 up to January 27, 2021. The diagnostic metrics described above were calculated to assess the performance of the risk score on data collected *after* its derivation. All analyses were conducted using SPSS version 27.0 (Armonk, NY, USA).

The study was approved by the Hamad Medical Corporation (HMC IRB number MRC-05-011) and Weill Cornell Medicine-Qatar (WCM-Q IRB number 20–00017) Institutional Review Boards with waiver of informed consent. All methods were carried out in accordance with relevant guidelines and regulations.

Results

Characteristics of SARS-CoV-2 RT-PCR testing conducted in Qatar

Between February 5, 2020 and April 21, 2020, there was a total of 69,820 individuals tested for SARS-CoV-2, of which 13,654 (19.6%) had more than one test. A total of 90,027 RT-PCR tests were performed for SARS-CoV-2 infection, and 10,362 were positive for an overall RT-PCR positivity of 11.5% (95% CI: 11.3%-11.7%).

Between February 5, 2020 and January 27, 2021, there was a total of 1,041,022 individuals tested for SARS-CoV-2, of which 406,048 (39.0%) had more than one test. A total of 2,688,232 RT-PCR tests were performed, and 200,646 were positive for an overall RT-PCR positivity of 7.5% (95% CI: 7.4%-7.5%). Characteristics of SARS-CoV-2 RT-PCR testing conducted in Qatar are presented in Table 1.

Original Qatar COVID-19 risk score

Risk score derivation. Bivariable logistic regression of half the testing sample from February 5, 2020-April 21, 2020 identified significant associations between individual variables, age, sex, and nationality, and RT-PCR outcome (Table 2A). All three demographic variables were retained in the multivariable logistic regression and were included in the risk score. Scoring points were lower for females than for males and higher for specific nationalities. The risk score was expressed as a mathematical formula illustrated in Box 1A.

Table 1. Characteristics of SARS-CoV-2 RT-PCR testing conducted in Qatar.

Characteristics	Original sample of Feb 5, 2020-Apr 21, 2020			Extended sample of Apr 22, 2020-Jan 27, 2021		
	Tested N = 90,027	SARS-CoV-2 RT-PCR positive N = 10,362 (11.5%)		Tested N = 2,598,205	SARS-CoV-2 RT-PCR positive N = 190,284 (7.3%)	
	N (%)	N (%)	P-value	N (%)	N (%)	P-value
Sex						
Male	69,440 (77.1)	9,022 (13.0)	<0.001	1,773,809 (68.3)	144,875 (8.2)	<0.001
Female	20,586 (22.9)	1,339 (6.5)		824,220 (31.7)	45,393 (5.5)	
Age (years)						
<10	2,266 (2.5)	175 (7.7)	<0.001	206,506 (7.9)	12,806 (6.2)	<0.001
10–19	3,493 (3.9)	221 (6.3)		184,501 (7.1)	11,606 (6.3)	
20–29	25,844 (28.7)	2,594 (10.0)		601,053 (23.1)	43,073 (7.2)	
30–39	30,823 (34.2)	3,578 (11.6)		834,766 (32.1)	64,301 (7.7)	
40–49	16,087 (17.9)	2,173 (13.5)		450,517 (17.3)	36,374 (8.1)	
50–59	7,658 (8.5)	1,085 (14.2)		219,455 (8.4)	15,715 (7.2)	
60–69	2,791 (3.1)	403 (14.4)		77,662 (3.0)	4,973 (6.4)	
70–79	737 (0.8)	115 (15.6)		18,109 (0.7)	1,090 (6.0)	
80+	328 (0.4)	18 (5.5)		5,635 (0.2)	346 (6.1)	
Nationality						
Other*	13,448	909 (6.8)	<0.001	646,379 (24.9)	26,124 (4.0)	<0.001
Bangladeshi	8,622	2,013 (23.3)		142,649 (5.5)	18,932 (13.3)	
Nepalese	8,024	1,712 (21.3)		187,623 (7.2)	26,393 (14.1)	
Indian	17,232	2,632 (15.3)		557,106 (21.4)	49,085 (8.8)	
Pakistani	4,036	643 (15.9)		135,337 (5.2)	10,736 (7.9)	
Kenyan	909	110 (12.1)		36,505 (1.4)	2,114 (5.8)	
Egyptian	2,836	298 (10.5)		140,872 (5.4)	9,808 (7.0)	
Sri Lankan	2,281	174 (7.6)		55,635 (2.1)	6,450 (11.6)	
Sudanese	2,058	166 (8.1)		83,714 (3.2)	5,222 (6.2)	
Filipino	6,408	352 (5.5)		185,033 (7.1)	13,171 (7.1)	
Qatari	24,173	1,353 (5.6)		427,352 (16.4)	22,249 (5.2)	

RT-PCR; real-time polymerase chain reaction

*These include 148 other nationalities residing in Qatar.

Risk score performance and validation. The ROC curve was generated, and the AUC was estimated at 0.67 (95% CI: 0.66–0.67) (Fig 1A). A score cut-off value of 6.5 maximized the sum of sensitivity and specificity. This indicated that individuals with a risk score ≥6.5 should be prioritized for RT-PCR testing.

To validate the risk score, it was applied to the other half of the sample and yielded the following diagnostic metrics: 66.8% (95% CI: 65.5%-68.0%) for sensitivity, 62.6% (95% CI: 62.2%-63.1%) for specificity, 19.1% (95% CI: 18.6%-19.7%) for PPV, and 93.4% (93.1%-93.7%) for NPV (Table 3).

Updated Qatar COVID-19 risk score

Risk score derivation. Bivariable logistic regression of half the testing sample from April 22, 2020-January 27, 2021 identified significant associations between individual variables, age, sex, and nationality and RT-PCR outcome (Table 2B). All three demographic variables were retained in the multivariable logistic regression and were included in the risk score. Scoring

Table 2. Results of regression analyses used to derive a) the original and b) updated Qatar COVID-19 risk scores.

A) "Original" Qatar COVID-19 risk score.			
Characteristics	**β**	**aOR (95% CI)**	**Score points**
Sex			
Male	0.000	1.00	0
Female	-0.384	0.68 (0.62–0.75)	-4
Age (years)			
<10	0.000	1.00	0
10–19	-0.028	0.97 (0.72–1.31)	0
20–29	-0.123	0.88 (0.70–1.12)	-1
30–39	-0.109	0.90 (0.71–1.13)	-1
40–49	0.100	1.11 (0.87–1.40)	1
50–59	0.295	1.34 (1.05–1.72)	3
60–69	0.561	1.75 (1.34–2.30)	6
70–79	1.030	2.80 (1.97 (3.99)	10
80+	-0.083	0.92 (0.45–1.88)	-1
Nationality			
Other*	0.000	1.00	0
Bangladeshi	1.362	3.91 (3.45–4.93)	14
Nepalese	1.358	3.89 (3.43–4.41)	14
Indian	0.897	2.45 (2.19–2.75)	9
Pakistani	0.852	2.35 (2.01–2.74)	9
Kenyan	0.902	2.47 (1.85–3.28)	9
Egyptian	0.459	1.58 (1.30–1.93)	5
Sri Lankan	0.059	1.06 (0.83–1.35)	1
Sudanese	0.318	1.37 (1.09–1.74)	3
Filipino	-0.083	0.92 (0.77–1.11)	-1
Qatari	-0.205	0.82 (0.72–0.92)	-2
B) "Updated" Qatar COVID-19 risk score			
Characteristics	**β**	**aOR (95% CI)**	**Score points**
Sex			
Male	0.000	1.00	0
Female	-0.193	0.82 (0.81–0.84)	-2
Age (years)			
<10	0.000	1.00	0
10–19	0.077	1.08 (1.04–1.12)	1
20–29	-0.160	0.85 (0.83–0.88)	-2
30–39	-0.088	0.92 (0.89–0.94)	-1
40–49	-0.017	1.02 (0.99–1.05)	0
50–59	-0.017	0.98 (0.95–1.02)	0
60–69	-0.028	0.97 (0.93–1.02)	0
70–79	0.049	1.05 (0.96–1.15)	0
80+	0.068	1.07 (0.91–1.26)	1
Nationality			
Other*	0.000	1.00	0
Bangladeshi	1.307	3.69 (3.59–3.80)	13
Nepalese	1.368	3.93 (3.83–4.03)	14
Indian	0.830	2.29 (2.26–2.35)	8
Pakistani	0.718	2.05 (1.99–2.12)	7

(Continued)

Table 2. (Continued)

Kenyan	0.478	1.61 (1.51–1.72)	5
Egyptian	0.553	1.74 (1.68–1.80)	6
Sri Lankan	1.096	2.99 (2.87–3.12)	11
Sudanese	0.456	1.58 (1.51–1.65)	5
Filipino	0.647	1.91 (1.85–1.97)	6
Qatari	0.236	1.27 (1.23–1.30)	2

β, beta coefficient; aOR, adjusted odds ratio; CI, confidence interval.

*These include 148 other nationalities residing in Qatar.

The Hosmer and Lemeshow goodness of fit test reported a p-value of 0.053, indicating a good fit. The chi-square test of the model reported a p-value <0.001, also indicating a good fit.

https://doi.org/10.1371/journal.pone.0271324.t002

points were lower for females than for males and higher for specific nationalities. The risk score was expressed as a mathematical formula illustrated in Box 1B.

Risk score performance and validation. The ROC curve was generated, and the AUC was estimated at 0.63 (95% CI: 0.63–0.63) (Fig 1B). A score cut-off value of 5.5 maximized the sum of sensitivity and specificity. This indicated that individuals with risk scores ≥5.5 should be prioritized for RT-PCR testing.

To validate the risk score, it was applied to the other half of the sample and yielded the following diagnostic metrics: 59.4% (95% CI: 59.1%-59.7%) for sensitivity, 61.1% (95% CI: 61.1%-61.2%) for specificity, 10.9% (95% CI: 10.8%-10.9%) for PPV, and 94.9% (94.9%-95.0%) for NPV (Table 3).

Performance assessment of the original derived risk score on prospective data

The original Qatar COVID-19 risk score derived from testing data until April 21, 2020 was applied to all testing data from April 22, 2020 to January 27, 2021. It yielded the following diagnostic metrics: 52.6% (95% CI: 52.4%-52.9%) for sensitivity, 65.2% (95% CI: 65.2%-65.3%) for specificity, 10.7% (95% CI: 10.6%-10.8%) for PPV, and 94.6% (94.6%-94.6%) for NPV (Table 3).

Discussion

To illustrate the concept and public health value of COVID-19 risk scores, we derived a simple COVID-19 risk score for Qatar, which to our knowledge is the first for any country. The risk score demonstrated relatively strong performance supporting the utility of using such risk scores to inform national testing strategies. A main finding is that the COVID-19 risk score performed similarly to other public health risk scores, such as those for diabetes [6, 29–34]. Indeed, this risk score, though simple to implement, demonstrated reasonably high diagnostic accuracy (Fig 1 and Table 3). The original risk, which was derived based on early epidemic data until only April, 2020 proved effective and offered comparable performance to the updated risk score based on all data until the present (Table 3). This further affirms the utility of such scores even when they are derived from a more limited set of testing data during a specific phase of the epidemic.

While our study provided a proof of concept for the use of such scores, implementation of them can be further optimized. We reported a risk score derived over one year. The score's performance could have been improved, with higher diagnostic ability, if different scores were

Box 1. Mathematical formula for the A) original and the B) updated Qatar COVID-19 risk scores.

A.

$$Qatar\ COVID - 19\ risk\ score = \begin{cases} 14\ if\ Nationality = Bangladeshi \\ 14\ if\ Nationality = Nepalese \\ 9\ if\ Nationality = Indian \\ 9\ if\ Nationality = Kenyan \\ 9\ if\ Nationality = Pakistan \\ 5\ if\ Nationality = Egyptian \\ 3\ if\ Nationality = Sudanese \\ 1\ if\ Nationality = Sri\ Lankan \\ 0\ if\ Nationality = Other \\ -1\ if\ Nationality = Filipino \\ -2\ if\ Nationality = Qatari \end{cases} + \begin{cases} -4\ if\ Sex = Female \\ 0\ if\ Sex = Male \end{cases} + \begin{cases} 0\ if\ Age < 10 \\ 0\ if\ Age\ 10-19 \\ -1\ if\ Age\ 20-29 \\ -1\ if\ Age\ 30-39 \\ 1\ if\ Age\ 40-49 \\ 3\ if\ Age = 50-59 \\ 6\ if\ Age = 60-69 \\ 10\ if\ Age = 70-79 \\ -1\ if\ Age \geq 80 \end{cases}$$

**If Qatar COVID-19 risk score is ≥ 6.5, then the individual is classified with higher risk of being SARS-CoV-2 RT-PCR positive

B.

$$Qatar\ COVID - 19\ risk\ score = \begin{cases} 14\ if\ Nationality = Nepalese \\ 13\ if\ Nationality = Bangladeshi \\ 11\ if\ Nationality = Sri\ Lankan \\ 8\ if\ Nationality = Indian \\ 7\ if\ Nationality = Pakistan \\ 6\ if\ Nationality = Egyptian \\ 6\ if\ Nationality = Filipino \\ 5\ if\ Nationality = Kenyan \\ 5\ if\ Nationality = Sudanese \\ 2\ if\ Nationality = Qatari \\ 0\ if\ Nationality = Other \end{cases} + \begin{cases} -2\ if\ Sex = Female \\ 0\ if\ Sex = Male \end{cases} + \begin{cases} 0\ if\ Age < 10 \\ 1\ if\ Age\ 10-19 \\ -2\ if\ Age\ 20-29 \\ -1\ if\ Age\ 30-39 \\ 0\ if\ Age\ 40-49 \\ 0\ if\ Age = 50-59 \\ 0\ if\ Age = 60-69 \\ 0\ if\ Age = 70-79 \\ 1\ if\ Age \geq 80 \end{cases}$$

**If Qatar COVID-19 risk score is ≥ 5.5, then the individual is classified with higher risk of being SARS-CoV-2 RT-PCR positive

derived in real-time at every phase of the epidemic and their use is updated continuously. It is remarkable that the risk score derived using a year of RT-PCR testing performed well, even though the epidemiology of the infection in Qatar has evolved immensely during this year [11, 14–24]. A month-by-month risk score, derived based on RT-PCR testing of only the previous month, would have better predicted the risk of infection month by month. With the ease of the process of deriving such risk scores, continuous updating of risk scores is feasible even in resource-limited settings, provided there is a minimal digital healthcare system to track RT-PCR testing.

A finding of this study is that there is always likely to be considerable variation in the risk of exposure to the infection based on basic demographics (such as age, sex, and nationality). This reflects the underlying dynamics of infection transmission in any country, as those delineated earlier for Qatar [11, 14–24]. Biological factors such as age [35–41], may also cause variation in susceptibility to the infection or in the likelihood of the infection's being symptomatic, which may affect the likelihood of testing or of a positive test outcome. The most affected
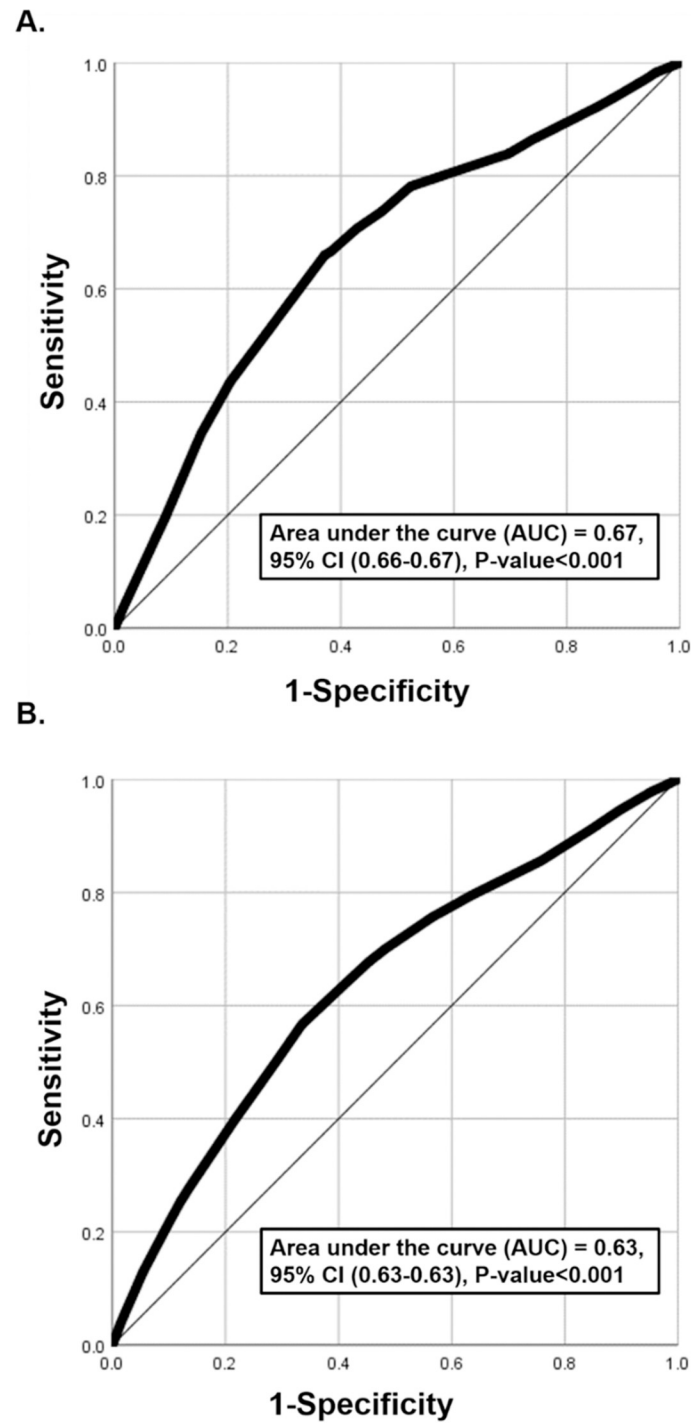
A.

B.

Fig 1. Diagnostic performance of the A) original and B) updated Qatar COVID-19 risk scores, assessed using the area under the receiver operating characteristic (ROC) curve.

https://doi.org/10.1371/journal.pone.0271324.g001

Table 3. Validation and diagnostic performance of the Qatar COVID-19 risk score, assessed using measures of sensitivity, specificity, positive predictive value, and negative predictive value.

		SARS-CoV-2 infection status using RT-PCR			Risk score metrics			
		Positive	Negative	Total	Sensitivity % (95% CI)	Specificity % (95% CI)	Positive Predictive Value % (95% CI)	Negative Predictive Value % (95% CI)
SARS-CoV-2 infection status using the "original" Qatar COVID-19 risk score applied to the original sample of Feb 5, 2020-Apr 21, 2020	Positive	3,520	14,876	18,396	66.8 (65.5–68.0)	62.6 (62.2–63.1)	19.1 (18.6–19.7)	93.4 (93.1–93.7)
	Negative	1,752	24,951	26,703				
	Total	5,272	39,827	45,099				
SARS-CoV-2 infection status using the "updated" Qatar COVID-19 risk score applied to the updated sample of Feb 5, 2020-Jan 27, 2021	Positive	58,735	482,098	540,833	59.4 (59.1–59.7)	61.1 (61.1–61.2)	10.9 (10.8–10.9)	94.9 (94.9–95.0)
	Negative	40,168	758,474	798,642				
	Total	98,903	1,240,572	1,339,475				
SARS-CoV-2 infection status using the "original" Qatar COVID-19 risk score by applying it prospectively to the extended sample of Apr 22, 2020-Jan 27, 2021	Positive	100,183	836,863	937,046	52.6 (52.4–52.9)	65.2 (65.2–65.3)	10.7 (10.6–10.8)	94.6 (94.6–94.6)
	Negative	90,101	1,571,058	1,661,159				
	Total	190,284	2,407,921	2,598,205				

*RT-PCR, real-time polymerase chain reaction.

subpopulation in Qatar by COVID-19 during the first wave was the craft and manual worker subpopulation living in shared housing accommodations; that is, similar to living in dormitories where they shared common spaces such as bathrooms and kitchen/dining areas [11, 14, 16, 18, 20, 21]. The contribution of each nationality reflected the association between nationality and occupation, as well as the differences in the social contact structure in Qatar [14, 18, 20, 21]. Social contacts are more prominent within nationality groups who share the same culture, language, and/or national background. A COVID-19 risk score can be seen as a metric that quantifies these variations in any setting, creating an opportunity for more effective public health action that addresses the needs of different segments of the population.

This study has some limitations. The COVID-19 risk score was derived using the national testing database rather than a nationally representative, probability-based survey of the total population of Qatar. Infection levels and patterns among tested individuals may not necessarily reflect actual levels and patterns in the wider population. The score used a small number of demographic variables, but its predictive power might have been enhanced if other variables had been available, such as more socio-demographic indicators.

Because of the central tracking of all PCR testing in Qatar, we were able to conduct this study on a very large sample size, which may not be available in other countries. However, we conducted a sensitivity analysis where the same regression analysis was done on only (random) 25% of the sample size. The analysis yielded similar effect sizes suggesting that sample size should not be a hindrance in applying this concept to other countries and settings (S1 Table). Of note that the impact of each variable is reflected and measured by the score points. For example, Nepalese nationality was very predictive with a score of 14 while Qatari nationality had a score of -2 and was not as highly predictive (Table 2A).

It was not possible to account for other factors, such as geography and comorbidities, as such data were not available. The study covered the duration of only one epidemic wave that was followed by a long low-incidence phase, lasting for seven months. Qatar is primarily a city state where infection was broadly distributed across the country's neighborhoods/areas; thus, geography is unlikely to have been a confounding factor. We conducted a sensitivity analysis where the multivariable logistic regression included a random effect at the PCR testing site,

and the regression generated similar results to the baseline analysis (S2 Table). While data on comorbidities were not available, adjusting for age may have served as a proxy given the association between comorbidities and older age. Nonetheless, keeping the risk score as simple as possible and the number of variables to a minimum enhance its value to be implemented broadly as part of awareness campaigns and in primary care settings.

Although the diagnostic metrics (sensitivity, specificity, and PPV) were not particularly high, the value of a risk score is not in providing a highly sensitive and specific measure for infection diagnosis. Infection diagnosis should only be done using biological testing such as by PCR. The value of a risk score is to optimize who to test by PCR resulting in reduced costs, consumption of resources, and logistics. The risk score enables identification of persons more likely to be infected shortly after infection, thereby reducing the potential severity of the infection through treatment and allowing faster isolation to reduce infection transmission.

Despite these limitations, the study had important strengths. The testing database was massive and encompassed all RT-PCR testing done in Qatar using validated commercial platforms with very high sensitivity and specificity. The database included results of over two million tests, representing a majority of the population of Qatar [8, 42]. While adding other variables to the score may have improved its predictive power, it may have reduced its accessibility and utility for broad use as a tool of public health. The score value is in providing a non-invasive tool for identification of individuals at higher risk of being infected, who should be prioritized for PCR testing, in addition to typical cases of clinical suspicion and contact tracing. Therefore, use of such scores may substantially enhance the effectiveness of the "testing, tracing, and isolation" approach that is the "backbone" of the COVID-19 national response in different countries [2]. Indeed, the present analyses have helped to guide Qatar's national COVID-19 response to control transmission and to reduce the disease burden.

In conclusion, the concept and utility of a COVID-19 risk score was demonstrated in a single country. Policy makers should consider the application of this method in streamlining PCR testing to minimize costs, consumption of resources, and logistics. Such public health tool, based on a set of non-invasive and easily captured variables, can help optimize testing and suppression of infection transmission, while maximizing efficient use of available resources.

Supporting information

S1 Table. Results of multivariable logistic regression analysis using only 10% of the sample to derive the original Qatar COVID-19 risk score.
(DOCX)

S2 Table. Results of multivariable random-effect logistic regression analysis (with random level at location of PCR testing) used to derive the original Qatar COVID-19 risk score.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Laith J. Abu-Raddad, Mohamed H. Al-Thani.

Data curation: Laith J. Abu-Raddad, Soha Dargham, Hiam Chemaitelly, Peter Coyle, Zaina Al Kanaani, Einas Al Kuwari, Adeel A. Butt, Andrew Jeremijenko, Anvar Hassan Kaleeckal, Ali Nizar Latif, Riyazuddin Mohammad Shaik, Hanan F. Abdul Rahim, Gheyath K. Nasrallah, Hadi M. Yassine, Mohamed G. Al Kuwari, Hamad Eid Al Romaihi, Abdullatif Al Khal, Roberto Bertollini.

Formal analysis: Laith J. Abu-Raddad, Soha Dargham, Hiam Chemaitelly.

Funding acquisition: Laith J. Abu-Raddad, Adeel A. Butt.

Investigation: Laith J. Abu-Raddad, Soha Dargham, Hiam Chemaitelly, Peter Coyle, Zaina Al Kanaani, Einas Al Kuwari, Adeel A. Butt, Andrew Jeremijenko, Anvar Hassan Kaleeckal, Ali Nizar Latif, Riyazuddin Mohammad Shaik, Hanan F. Abdul Rahim, Gheyath K. Nasrallah, Hadi M. Yassine, Mohamed G. Al Kuwari, Hamad Eid Al Romaihi, Mohamed H. Al-Thani, Abdullatif Al Khal, Roberto Bertollini.

Methodology: Laith J. Abu-Raddad, Soha Dargham, Hiam Chemaitelly.

Project administration: Laith J. Abu-Raddad.

Supervision: Laith J. Abu-Raddad.

Validation: Laith J. Abu-Raddad, Soha Dargham, Hiam Chemaitelly.

Writing – original draft: Laith J. Abu-Raddad, Soha Dargham.

Writing – review & editing: Laith J. Abu-Raddad, Soha Dargham, Hiam Chemaitelly, Peter Coyle, Zaina Al Kanaani, Einas Al Kuwari, Adeel A. Butt, Andrew Jeremijenko, Anvar Hassan Kaleeckal, Ali Nizar Latif, Riyazuddin Mohammad Shaik, Hanan F. Abdul Rahim, Gheyath K. Nasrallah, Hadi M. Yassine, Mohamed G. Al Kuwari, Hamad Eid Al Romaihi, Mohamed H. Al-Thani, Abdullatif Al Khal, Roberto Bertollini.

References

1. World Health Organization. 2019 Novel Coronavirus (2019-nCoV): Strategic preparedness and response plan. https://www.who.int/publications-detail/strategic-preparedness-and-response-plan-for-the-new-coronavirus. Geneva, Switzerland: World Health Organization, 2020.

2. Lacina L. WHO coronavirus briefing: Isolation, testing and tracing comprise the 'backbone' of response. https://www.weforum.org/agenda/2020/03/testing-tracing-backbone-who-coronavirus-wednesdays-briefing/. Accessed on: April 27, 2020. World Economic Forum. 2020.

3. Brown N, Critchley J, Bogowicz P, Mayige M, Unwin N. Risk scores based on self-reported or available clinical data to detect undiagnosed type 2 diabetes: a systematic review. Diabetes Res Clin Pract. 2012; 98(3):369–85. Epub 2012/09/27. https://doi.org/10.1016/j.diabres.2012.09.005 PMID: 23010559.

4. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Med. 2011; 9:103. Epub 2011/09/10. https://doi.org/10.1186/1741-7015-9-103 PMID: 21902820.

5. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. BMJ. 2011; 343:d7163. Epub 2011/11/30. https://doi.org/10.1136/bmj.d7163 PMID: 22123912.

6. Awad SF, Dargham SR, Toumi AA, Dumit EM, El-Nahas KG, Al-Hamaq AO, et al. A diabetes risk score for Qatar utilizing a novel mathematical modeling approach to identify individuals at high risk for diabetes. Sci Rep. 2021; 11(1):1811. Epub 2021/01/21. https://doi.org/10.1038/s41598-021-81385-3 PMID: 33469048.

7. International Diabetes Federation. Risk prediction tools (PREDICT—2). http://www.idf.org/risk-prediction-tools-predict-2. Accessed on April 30, 2020. 2017.

8. Planning and Statistics Authority-State of Qatar. The Simplified Census of Population, Housing & Establishments. https://www.psa.gov.qa/en/statistics/Statistical%20Releases/Population/Population/2018/Population_social_1_2018_AE.pdf Accessed on: April 2, 2020. 2019.

9. World Population Review. Qatar Population (Live). https://worldpopulationreview.com/countries/qatar-population/. Accessed on April 23, 2020. 2020.

10. Priya Dsouza Communications. Population of Qatar by nationality—2019 report. https://priyadsouza.com/population-of-qatar-by-nationality-in-2017/. 2019.

11. Ayoub HH, Chemaitelly H, Seedat S, Makhoul M, Al Kanaani Z, Al Khal A, et al. Mathematical modeling of the SARSCoV-2 epidemic in Qatar and its impact on the national response to COVID-19. Journal of Global Health. 2021; 11(05005). https://doi.org/10.7189/jogh.11.05005 PMID: 33643638

12. Hamad Medical Corporation. SARS-CoV-2 hospitalizations and care. 2020.

13. Ministry of Public Health-State of Qatar. Coronavirus Disease 2019 (COVID-19). https://covid19.moph.gov.qa/EN/Pages/default.aspx. Accessed on: May 25, 2020. 2020.

14. Abu-Raddad LJ, Chemaitelly H, Ayoub HH, Al Kanaani Z, Al Khal A, Al Kuwari E, et al. Characterizing the Qatar advanced-phase SARS-CoV-2 epidemic. Sci Rep. 2021; 11(1):6233. Epub 2021/03/20. https://doi.org/10.1038/s41598-021-85428-7 PMID: 33737535.

15. Abu-Raddad LJ, Chemaitelly H, Malek JA, Ahmed AA, Mohamoud YA, Younuskunju S, et al. Two prolonged viremic SARS-CoV-2 infections with conserved viral genome for two months. Infect Genet Evol. 2020; 88:104684. https://doi.org/10.1016/j.meegid.2020.104684 PMID: 33352320.

16. Al Kuwari HM, Abdul Rahim HF, Abu-Raddad LJ, Abou-Samra A-B, Al Kanaani Z, Al Khal A, et al. Epidemiological investigation of the first 5685 cases of SARS-CoV-2 infection in Qatar, 28 February–18 April 2020. BMJ Open. 2020; 10(10):e040428. https://doi.org/10.1136/bmjopen-2020-040428 PMID: 33033033

17. Abu-Raddad LJ, Chemaitelly H, Malek JA, Ahmed AA, Mohamoud YA, Younuskunju S, et al. Assessment of the Risk of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Reinfection in an Intense Reexposure Setting. Clin Infect Dis. 2020. https://doi.org/10.1093/cid/ciaa1846 PMID: 33315061

18. Al-Thani MH, Farag E, Bertollini R, Al Romaihi HE, Abdeen S, Abdelkarim A, et al. SARS-CoV-2 Infection Is at Herd Immunity in the Majority Segment of the Population of Qatar. Open Forum Infectious Diseases. 2021; 8(8). https://doi.org/10.1093/ofid/ofab221 PMID: 34458388

19. Ayoub HH, Chemaitelly H, Makhoul M, Al Kanaani Z, Al Kuwari E, Butt AA, et al. Epidemiological impact of prioritising SARS-CoV-2 vaccination by antibody status: mathematical modelling analyses BMJ Innovations. 2021; 7(2):327–36. https://doi.org/10.1136/bmjinnov-2021-000677 PMID: 34192020

20. Coyle PV, Chemaitelly H, Ben Hadj Kacem MA, Abdulla Al Molawi NH, El Kahlout RA, Gilliani I, et al. SARS-CoV-2 seroprevalence in the urban population of Qatar: An analysis of antibody testing on a sample of 112,941 individuals. iScience. 2021; 24(6):102646. Epub 2021/06/01. https://doi.org/10.1016/j.isci.2021.102646 PMID: 34056566.

21. Jeremijenko A, Chemaitelly H, Ayoub HH, Alishaq M, Abou-Samra AB, Al Ajmi J, et al. Herd Immunity against Severe Acute Respiratory Syndrome Coronavirus 2 Infection in 10 Communities, Qatar. Emerg Infect Dis. 2021; 27(5):1343–52. Epub 2021/04/27. https://doi.org/10.3201/eid2705.204365 PMID: 33900174.

22. Younes S, Al-Jighefee H, Shurrab F, Al-Sadeq DW, Younes N, Dargham SR, et al. Diagnostic Efficiency of Three Fully Automated Serology Assays and Their Correlation with a Novel Surrogate Virus Neutralization Test in Symptomatic and Asymptomatic SARS-COV-2 Individuals. Microorganisms. 2021; 9(2). Epub 2021/01/29. https://doi.org/10.3390/microorganisms9020245 PMID: 33504067.

23. Abu-Raddad LJ, Chemaitelly H, Coyle P, Malek JA, Ahmed AA, Mohamoud YA, et al. SARS-CoV-2 antibody-positivity protects against reinfection for at least seven months with 95% efficacy.

EClinicalMedicine. 2021; 35:100861. Epub 2021/05/04. https://doi.org/10.1016/j.eclinm.2021.100861 PMID: 33937733.

24. Seedat S, Chemaitelly H, Ayoub HH, Makhoul M, Mumtaz GR, Al Kanaani Z, et al. SARS-CoV-2 infection hospitalization, severity, criticality, and fatality rates in Qatar. Sci Rep. 2021; 11(1):18182. Epub 2021/09/16. https://doi.org/10.1038/s41598-021-97606-8 PMID: 34521903.

25. Thermo Fisher Scientific. TaqPath™ COVID-19 CE-IVD RT-PCR Kit instructions for use. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0019215_TaqPathCOVID-19_CE-IVD_RT-PCR%20Kit_IFU.pdf. Accessed on December 02, 2020. 2020.

26. Kalikiri MKR, Hasan MR, Mirza F, Xaba T, Tang P, Lorenz S. High-throughput extraction of SARS-CoV-2 RNA from nasopharyngeal swabs using solid-phase reverse immobilization beads. medRxiv. 2020:2020.04.08.20055731. https://doi.org/10.1101/2020.04.08.20055731

27. Kubina R, Dziedzic A. Molecular and Serological Tests for COVID-19 a Comparative Review of SARS-CoV-2 Coronavirus Laboratory and Point-of-Care Diagnostics. Diagnostics (Basel). 2020; 10(6). Epub 2020/07/02. https://doi.org/10.3390/diagnostics10060434 PMID: 32604919.

28. US Food and Drug Administration. Cobas® SARS-CoV-2: Qualitative assay for use on the cobas® 6800/8800 Systems. https://www.fda.gov/media/136049/download. Accessed on: December 02, 2020. 2020.

29. Al-Lawati JA, Tuomilehto J. Diabetes risk score in Oman: a tool to identify prevalent type 2 diabetes among Arabs of the Middle East. Diabetes Res Clin Pract. 2007; 77(3):438–44. Epub 2007/02/20. https://doi.org/10.1016/j.diabres.2007.01.013 PMID: 17306410.

30. Aekplakorn W, Bunnag P, Woodward M, Sritara P, Cheepudomwit S, Yamwong S, et al. A risk score for predicting incident diabetes in the Thai population. Diabetes Care. 2006; 29(8):1872–7. Epub 2006/07/29. https://doi.org/10.2337/dc05-2141 PMID: 16873795.

31. Al Khalaf MM, Eid MM, Najjar HA, Alhajry KM, Doi SA, Thalib L. Screening for diabetes in Kuwait and evaluation of risk scores. East Mediterr Health J. 2010; 16(7):725–31. Epub 2010/08/31. PMID: 20799528.

32. Glumer C, Carstensen B, Sandbaek A, Lauritzen T, Jorgensen T, Borch-Johnsen K, et al. A Danish diabetes risk score for targeted screening: the Inter99 study. Diabetes Care. 2004; 27(3):727–33. Epub 2004/02/28. https://doi.org/10.2337/diacare.27.3.727 PMID: 14988293.

33. Memish ZA, Chang JL, Saeedi MY, Al Hamid MA, Abid O, Ali MK. Screening for Type 2 Diabetes and Dysglycemia in Saudi Arabia: Development and Validation of Risk Scores. Diabetes Technol Ther. 2015; 17(10):693–700. Epub 2015/07/15. https://doi.org/10.1089/dia.2014.0267 PMID: 26154413.

34. Sulaiman N, Mahmoud I, Hussein A, Elbadawi S, Abusnana S, Zimmet P, et al. Diabetes risk score in the United Arab Emirates: a screening tool for the early detection of type 2 diabetes mellitus. BMJ Open Diabetes Res Care. 2018; 6(1):e000489. Epub 2018/04/10. https://doi.org/10.1136/bmjdrc-2017-000489 PMID: 29629178.

35. Ayoub HH, Chemaitelly H, Mumtaz GR, Seedat S, Awad SF, Makhoul M, et al. Characterizing key attributes of the epidemiology of COVID-19 in China: Model-based estimations. Global Epidemiology. 2020;100042. https://doi.org/10.1016/j.gloepi.2020.100042 PMID: 33235991

36. Ayoub HH, Chemaitelly H, Seedat S, Mumtaz GR, Makhoul M, Abu-Raddad LJ. Age could be driving variable SARS-CoV-2 epidemic trajectories worldwide. PLoS One. 2020; 15(8):e0237959. Epub 2020/08/21. https://doi.org/10.1371/journal.pone.0237959 PMID: 32817662.

37. Davies NG, Klepac P, Liu Y, Prem K, Jit M, group CC-w, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat Med. 2020. https://doi.org/10.1038/s41591-020-0962-9 PMID: 32546824.

38. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020. https://doi.org/10.1056/NEJMoa2002032 PMID: 32109013.

39. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020; 395(10223):497–506. Epub 2020/01/28. https://doi.org/10.1016/S0140-6736(20)30183-5 PMID: 31986264.

40. Nikolich-Zugich J, Knox KS, Rios CT, Natt B, Bhattacharya D, Fain MJ. SARS-CoV-2 and COVID-19 in older adults: what we may expect regarding pathogenesis, immune responses, and outcomes. Geroscience. 2020. https://doi.org/10.1007/s11357-020-00186-0 PMID: 32274617.

41. Zhu Y, Bloxham CJ, Hulme KD, Sinclair JE, Tong ZWM, Steele LE, et al. Children are unlikely to have been the primary source of household SARS-CoV-2 infections. medRxiv. 2020:2020.03.26.20044826.

42. Planning and Statistics Authority- State of Qatar. Qatar Monthly Statistics. https://www.psa.gov.qa/en/pages/default.aspx. Accessed on: may 26,2020. 2020.