Original Contribution

Estimating the Severity and Subclinical Burden of Middle East Respiratory Syndrome Coronavirus Infection in the Kingdom of Saudi Arabia


* Correspondence to Dr. Derek A. T. Cummings, Department of Biology, Emerging Pathogens Institute, University of Florida, P.O. Box 118525, 220 Bartram Hall, Gainesville, FL 32611-8525 (e-mail: datc@ufl.edu); or Dr. Ali Al-Barrak, Public Health Department, Ministry of Health, Riyadh 11176, Saudi Arabia (e-mail: draalbarrak@yahoo.com).

Initially submitted August 31, 2015; accepted for publication December 22, 2015.

Not all persons infected with Middle East respiratory syndrome coronavirus (MERS-CoV) develop severe symptoms, which likely leads to an underestimation of the number of people infected and an overestimation of the severity. To estimate the number of MERS-CoV infections that have occurred in the Kingdom of Saudi Arabia, we applied a statistical model to a line list describing 721 MERS-CoV infections detected between June 7, 2012, and July 25, 2014. We estimated that 1,528 (95% confidence interval (CI): 1,327, 1,883) MERS-CoV infections occurred in this interval, which is 2.1 (95% CI: 1.8, 2.6) times the number reported. The probability of developing symptoms ranged from 11% (95% CI: 4, 25) in persons under 10 years of age to 88% (95% CI: 72, 97) in those 70 years of age or older. An estimated 22% (95% CI: 18, 25) of those infected with MERS-CoV died. MERS-CoV is deadly, but this work shows that its clinical severity differs markedly between groups and that many cases likely go undiagnosed.

burden; clinical symptoms; coronavirus; MERS; severity

Abbreviations: CI, confidence interval; KSA, Kingdom of Saudi Arabia; MERS-CoV, Middle Eastern respiratory syndrome coronavirus.

Most infectious disease surveillance systems only capture a fraction of the infections that have actually occurred. The metaphor that is often used for this phenomenon is that of an iceberg: The visible tip consists mostly of those infections severe enough to cause an individual to seek medical care, but a potentially large number of often mild infections remain unseen “below the water.” In the present study, we assessed the number of unrecognized mild or subclinical infections of Middle Eastern respiratory syndrome coronavirus (MERS-CoV) that may have occurred in the Kingdom of Saudi Arabia (KSA) between June 2012 and July 2014, particularly during the large outbreak that occurred in the country between March and June of 2014 (1, 2).

Determining the number of infections is important for understanding the extent of the public health threat posed by MERS-CoV. Despite occasional incidences in other countries, reported MERS-CoV infections have remained largely confined to the KSA (3). Since MERS-CoV was first detected, the incidence has been generally low, but in the spring of 2014, there was a large outbreak in the KSA, with 525 cases reported between March 1 and June 30 (1, 2). For the most part, surveillance fails to detect mild and subclinical infections. If undetected infections occur in large numbers, we may be significantly overestimating how deadly MERS-CoV infection is.

One approach to estimating the true number of persons infected with a disease is to estimate the probability of cases traversing each step on the pathway to detection (4, 5). If we know the probability that an infection will cause symptoms, the probability of a person seeking clinical care if symptomatic (related to the severity), the probability of being correctly diagnosed after seeking care, the probability of an accurate...
test, and the probability of the diagnosis being reported, we can then estimate the number of cases that we did not see for each case that was identified. However, as MERS-CoV is a newly emerged virus, we do not yet have good estimates of many of these probabilities. Hence, it is difficult to know what these probabilities are.

Another possible approach to identifying the unmeasured burden of disease is to compare the frequencies of severe disease and death in infections identified through active surveillance with those identified through passive surveillance (6). Passive surveillance is done by testing patients who present to hospitals or health care clinics and meet the case definition for MERS-CoV infection (7). Prior to a change in case definition that took place on May 13, 2014, active surveillance included testing the contacts of known case patients and health care workers in facilities where MERS-CoV case patients seek treatment, regardless of symptoms (8, 9). Case patients identified through active surveillance are more likely to have symptom profiles and mortality rates similar to those of individuals of the same age and health status in the entire infected population than are those identified through passive surveillance. This fact combined with the assumption that persons who die from MERS-CoV are nearly always detected through passive surveillance means that a comparison of cases detected through active surveillance and those detected through passive surveillance can give us insight into the total number of MERS-CoV infections that have occurred in the KSA and allow us to more accurately estimate the severity of MERS-CoV infection (Figure 1). In the present study, we used data on all laboratory-confirmed cases of MERS-CoV in the KSA identified by the Ministry of Health through active and passive surveillance as of July 25, 2014, to estimate the total number of subclinical cases that had occurred up to that point, the age-specific disease severity, and the true risk of death from infection.

METHODS
Data

The KSA Ministry of Health assembled a line list of laboratory-confirmed cases of MERS-CoV infection that occurred within the KSA from June 24, 2012, to July 25, 2014. This data set included information on age, sex, area of residence, the reason the case patient was tested for MERS-CoV, whether the case patient had symptoms that met the case definition, the reason the case patient was tested for MERS-CoV, and any indication of death, hospitalization, or MERS-CoV symptoms to have developed symptoms consistent with MERS-CoV infection severe enough to warrant testing at some point in their illness (hereafter referred to as “severe” symptoms).

In the present study, we considered laboratory-confirmed cases of MERS-CoV in patients who were tested because of the presence of MERS-CoV symptoms as having been detected through passive surveillance and cases in patients who reported undergoing testing for other reasons (e.g., contact with a MERS-CoV case patient) as having been detected through active surveillance. Within the KSA, active surveillance policies varied between locations and by time over the course of the period examined, and they were often poorly documented. Hence, there is no single protocol or set of criteria that led to individuals being tested for reasons other than having MERS-CoV symptoms. However, from mid-2013 to 2014, there was a general policy of testing home and hospital contacts of index case patients whether or not the contacts had symptoms (10, 11).

Symptomatic infection ratio estimation

We fit a probabilistic model to the observed distribution of symptoms and mortality in MERS-CoV infections detected through active and passive surveillance (see Web Appendix, available at http://aje.oxfordjournals.org/). The rate at which individuals infected with MERS-CoV develop severe symptoms and die is assumed to vary by age group. We assumed that infected individuals detected through active surveillance in each age class develop symptoms and die at the same rate.

![Figure 1. Estimating the unseen iceberg. In order to estimate the number of undetected Middle East respiratory syndrome coronavirus infections, we assume that active surveillance detects cases irrespective of their place on the iceberg and that nearly all deaths are detected (denoted by the dashed box at right enclosing all parts of the iceberg). Passive surveillance, by contrast, detects cases above the waterline (box at left above the waterline) but not those below (box at left below the waterline). That is, deaths, severe symptoms, and mild/asymptomatic infections occur in the same proportion in those identified by active surveillance as they do in infections overall. We then use the proportion of persons with infections “above the waterline” (i.e., observed through passive surveillance) who die or develop symptoms severe enough to trigger Middle East respiratory syndrome coronavirus testing to infer the number of unseen infections.](http://aje.oxfordjournals.org/)
as all (i.e., both detected and undetected) individuals infected with MERS-CoV who are the same age (Figure 1). The probability of being detected through passive surveillance is assumed to depend entirely on symptom status, and all cases of MERS-CoV in patients who die before being the infection is otherwise detected are assumed to have been detected by passive surveillance. Case patients without symptoms severe enough to trigger testing still have a small (estimated) probability of their infections being detected by passive surveillance. The probability of a case being detected by active surveillance is assumed to be independent of symptom status.

On the basis of the assumptions that all deceased case patients would be detected and that symptom status and detection through active surveillance were independent of each other, we derived the likelihood for the observed age-specific symptom distribution, mortality distribution, and incidence of actively and passively observed cases by week. For each age group, we estimated the symptomatic infection ratio, which is the rate at which MERS-CoV–infected individuals in that age group either develop symptoms suggestive of MERS-CoV infection or die. Combined with the estimated probabilities of detecting symptomatic and asymptomatic case patients through active and passive surveillance, this allowed us to estimate the weekly incidence of MERS-CoV infection.

All rates and probabilities were estimated using Bayesian Markov-chain Monte Carlo methods with the RStan package (12). Four chains of 1,000 iterations each were run, with the last half of each chain used to determine estimates. Convergence was assessed using visual examination of chains and Gelman and Rubin’s $\hat{R}$ statistic (13).

The model was validated using simulated data (Web Table 1). Sensitivity analyses were conducted to determine how sensitive the results were to temporal variations in reporting probabilities and the definition of active surveillance (Web Tables 2 and 3). Details of the model, validation, and sensitivity analyses are available in the Web Appendix.

MERS-CoV mortality

The case fatality ratio was calculated as the total number of deaths from MERS-CoV infection divided by the total number of MERS-CoV case patients detected using the current surveillance system whose final case status was reported as either deceased or recovered from MERS-CoV. The infection fatality ratio for MERS-CoV was calculated as the number of deceased case patients divided by the total number of individuals infected with MERS-CoV, and it was estimated as part of the probabilistic model described above. Risk factors for MERS-CoV infection were analyzed using multivariate logistic regression.

RESULTS

As of July 25, 2014, a total of 721 laboratory-confirmed MERS-CoV infections had been identified by the KSA Ministry of Health. Of these, 70% (504 of 721) were identified through passive surveillance and 27% (193 of 721) were identified through active surveillance (i.e., health care workers and contacts of confirmed and probable case patients); for 3% (24 of 721), the mode of detection was unknown. Thirty-three percent (63 of 193) of persons whose infections were detected through active surveillance died or were otherwise indicated to have developed symptoms consistent with the MERS-CoV case definition (i.e., would have eventually had the opportunity for their infection to be detected by active surveillance).

Although the naïve case fatality ratio has decreased over the course of the epidemic (from 75% in 2012 to 40% between January and July 2014 (Figure 2)), this difference was no longer significant once we adjusted for age, health care worker status, and reason for testing. Age was a significant risk factor for mortality, with health care workers (adjusted odds ratio = 0.20, 95% confidence interval (CI): 0.11, 0.38) and persons tested for reasons other than having MERS-CoV...
symptoms (adjusted odds ratio = 0.16, 95% CI: 0.08, 0.30) being significantly less likely to die. Though male sex was a significant independent risk factor for mortality (odds ratio = 1.89, 95% CI: 1.38, 2.60), it was no longer a significant risk factor (adjusted odds ratio = 1.15, 95% CI: 0.75, 1.76) after adjustment for age, reason for testing, and health care worker status.

We estimated a wide variation in the probability of infected individuals developing MERS-CoV symptoms by age, ranging from 11% (95% CI: 4, 25) in persons who were 0–9 years old to 88% (95% CI: 72, 97) in those 70 years of age or older (Figure 3). On the basis of these ratios, we estimated that, as of July 25, 2014, there were a total of 1,548 (95% CI: 1,327, 1,883) people infected with MERS-CoV (Table 1). Hence, reported cases likely comprised 47% (95% CI: 38, 54) of the total number of MERS-CoV infections that had occurred in the KSA.

We projected that nearly all (96%; 95% CI: 77, 100) of those infected who develop MERS-CoV symptoms consistent with the case definition were detected by MERS-CoV surveillance in the KSA. However, most asymptomatic or mildly symptomatic infections were likely missed; we estimated a 12% (95% CI: 10, 15) probability of an infected individual being detected through active surveillance (regardless of symptom status).

As of July 25, 2014, a total of 691 persons with laboratory-confirmed cases of MERS-CoV had been reported to the KSA Ministry of Health as having either recovered (i.e., been

Table 1. Middle East Respiratory Syndrome Coronavirus Infections and Deaths by Age Group, Kingdom of Saudi Arabia, June 1, 2012 to July 25, 2014

<table>
<thead>
<tr>
<th>Age, years</th>
<th>No. of Observed Cases</th>
<th>No. of Passively Observed Cases</th>
<th>Symptomatic Infection Ratio</th>
<th>95% CI</th>
<th>Estimated No. of Infections</th>
<th>95% CI</th>
<th>No. of Observed Deaths</th>
<th>Naïve CFR*</th>
<th>Estimated IFR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–9</td>
<td>15</td>
<td>5</td>
<td>0.11</td>
<td>0.04, 0.25</td>
<td>50</td>
<td>26, 89</td>
<td>4</td>
<td>0.27</td>
<td>0.10</td>
<td>0.03, 0.28</td>
</tr>
<tr>
<td>10–19</td>
<td>28</td>
<td>12</td>
<td>0.11</td>
<td>0.07, 0.16</td>
<td>97</td>
<td>61, 150</td>
<td>4</td>
<td>0.14</td>
<td>0.05</td>
<td>0.03, 0.08</td>
</tr>
<tr>
<td>20–29</td>
<td>97</td>
<td>41</td>
<td>0.14</td>
<td>0.10, 0.19</td>
<td>347</td>
<td>257, 469</td>
<td>17</td>
<td>0.18</td>
<td>0.05</td>
<td>0.03, 0.07</td>
</tr>
<tr>
<td>30–39</td>
<td>132</td>
<td>73</td>
<td>0.23</td>
<td>0.18, 0.29</td>
<td>384</td>
<td>296, 504</td>
<td>20</td>
<td>0.16</td>
<td>0.08</td>
<td>0.05, 0.11</td>
</tr>
<tr>
<td>40–49</td>
<td>115</td>
<td>82</td>
<td>0.39</td>
<td>0.31, 0.48</td>
<td>235</td>
<td>182, 307</td>
<td>36</td>
<td>0.32</td>
<td>0.17</td>
<td>0.13, 0.21</td>
</tr>
<tr>
<td>50–59</td>
<td>138</td>
<td>114</td>
<td>0.60</td>
<td>0.49, 0.71</td>
<td>204</td>
<td>163, 263</td>
<td>65</td>
<td>0.48</td>
<td>0.38</td>
<td>0.31, 0.45</td>
</tr>
<tr>
<td>60–69</td>
<td>78</td>
<td>72</td>
<td>0.78</td>
<td>0.67, 0.88</td>
<td>101</td>
<td>78, 133</td>
<td>56</td>
<td>0.74</td>
<td>0.63</td>
<td>0.56, 0.70</td>
</tr>
<tr>
<td>≥70</td>
<td>113</td>
<td>102</td>
<td>0.88</td>
<td>0.72, 0.97</td>
<td>127</td>
<td>101, 166</td>
<td>92</td>
<td>0.84</td>
<td>0.79</td>
<td>0.70, 0.86</td>
</tr>
<tr>
<td>Totalb</td>
<td>716</td>
<td>521</td>
<td>0.36</td>
<td>0.31, 0.42</td>
<td>1,548</td>
<td>1,327, 1,883</td>
<td>296</td>
<td>0.42</td>
<td>0.22</td>
<td>0.18, 0.25</td>
</tr>
</tbody>
</table>

Abbreviations: CFR, case fatality ratio; CI, confidence interval; IFR, infection fatality ratio.
* Only includes case patients with a resolved infection.
* Includes 5 participants of unknown age.
discharged) or died. Of these, 296 died, for naïve estimate of the case fatality ratio of 42%. After correcting for under-reporting of asymptomatic individuals, the infection fatality ratio was estimated to be 22% (95% CI: 18, 25) (Table 1), though it was 79% (95% CI: 70, 86, (Figure 3)) among those 70 years of age or older.

Several sensitivity analyses and validation exercises were performed to assess model performance (see Web Appendix). We were able to accurately estimate true detection rates under a variety of simulated scenarios, with 95% confidence intervals covering the true number of cases in all but one instance, in which the detection probability was low and death rates by age followed no functional form. Estimates of total number of infections (and hence the infection fatality ratio) were robust to allowing variations in reporting rates over the course of the epidemic.

**DISCUSSION**

We estimated that as of July 25, 2014, there were 1,528 (95% CI: 1,327, 1,883) MERS-CoV infections in KSA, approximately twice as many as had been observed. Nearly all persons with undetected infections were predicted to not have developed symptoms of MERS-CoV severe enough to trigger testing, but they may have developed mild symptoms or have had a severe but unusual clinical presentation.

Although MERS-CoV is always potentially deadly (we estimate the infection fatality ratio to be more than 5% in all age groups), the greatest risk is in individuals older than 60 years of age; persons in that age group have a higher than 60% chance of dying if infected. This pattern is similar to that seen for other respiratory viruses (e.g., influenza) (14), although the association is more extreme and likely results from increasing frailty and the prevalence of comorbid conditions (15). Unfortunately, clinical data on comorbid conditions were not generally available in the data set we analyzed.

The infectivity of subclinical MERS-CoV cases is unknown. The apparent rarity of MERS-CoV transmission except in health care settings suggests that the infectivity of such cases is low (16, 17). However, low-level infectivity of subclinical infections may help explain the occasional appearance of MERS-CoV infection in persons with no known human exposures or reported animal contact.

Ours is not the only analysis of MERS-CoV infection in which investigators have examined the frequency with which infections go undetected, case patients are asymptomatic, or case patients die. In an early study of household transmission, Drosten et al. (18) identified 12 household contacts with serologic or virologic evidence of MERS-CoV infection, none of whom developed symptoms. In a cross-sectional serosurvey, 15 of 10,009 sera collected in 2012 and 2013 from healthy individuals tested positive for anti–MERS-CoV antibodies, a rate that, if applied to the whole of the KSA, would imply far more cases than estimated here (see below) (19). In their analysis, Cauchemez et al. (20) used entirely different methods to estimate that at least 62% of symptomatic human case patients were going undetected as of 2013, a slightly higher rate than estimated here. In that paper (20) and a detailed investigation of the South Korean outbreak that resulted from a single imported case caused by exposure in the Middle East (21), fatality rates (20% and 21%, respectively) among secondary cases that were similar to the overall infection fatality rate estimated here were reported.

There are several limitations that could bias the results of the present study. The full histories of symptoms over the course of infection were not available, and some case patients who were initially reported as asymptomatic or as having mild symptoms might have developed symptoms that were not captured in our data at some point. However, because our estimates were also based on mortality data, the effect of this mischaracterization should be mitigated. Active surveillance may be more likely to detect case patients infected by known contacts with other infected persons than those infected by dromedary camels. If individuals infected by humans and those infected by camels develop severe symptoms and enter the passive surveillance system at the same rate, this should not bias our estimates; however, it is possible that the route of infection does impact the probability of symptoms (particularly camel-to-human infection that occurs through a route other than the respiratory system).

In our analysis, we did not take reporting delays into account; however, because no cases were reported between July 3 and July 25, 2015, doing so would have had little impact on our results. Under-reporting of deaths cannot be ruled out, nor can a bias towards reporting more severe cases; both would bias our estimates. We cannot be sure that the mortality rate and symptomatic rate in case patients tested for reasons other than MERS-CoV symptoms are the same as those of infections in general. We further assume 100% sensitivity in laboratory testing, but the true sensitivity may be less and vary over the course of a case patient’s illness. We assume that active case finding is independent of the development of symptoms, but it is likely that persons with some symptoms are more likely to be tested than are other contacts of or health care workers who dealt with MERS-CoV–infected patients, particularly after protocol changes in April and May of 2014 (8, 9). Likewise, asymptomatically infected individuals may shed less virus and thus be less likely to test positive for MERS-CoV infection. These biases would lead to an underestimation of the true number of MERS-CoV infections in our model, so our estimated number of total infections is conservative (e.g., if persons with cases detected by active surveillance have twice the odds of developing symptoms or dying compared with cases overall, the true number of MERS-CoV infections would be 1.3 times our estimate; see Web Table 3).

The estimates of the true number of MERS-CoV cases from our analysis are far fewer than other authors projected based on serological evidence (19). As discussed above, our method has limitations that tend to bias estimates of the true number of cases downward; serosurveys in general, and those by Müller et al. (19) in particular, have their own sets of limitations that may bias estimates upward. In particular, 1) cross-sectional serosurveys do not define a risk period and will pick up exposures that occurred before the current epidemic; 2) serological tests have limited ability to distinguish MERS-CoV from other recent and historic coronaviruses (camel and otherwise) that may be antigenically similar but not cause severe disease in humans; and 3) projections of a low attack rate to the entire population of the KSA may be problematic (as acknowledged by the authors). Given the differing direction of
potential biases in serological and case-based approaches, the truth may like somewhere in between the 2 approaches. If truly accurate estimates are desired in the future, serological and active surveillance methods must be implemented with this in mind from the start of the response to an emerging infection, providing the appropriate baselines and consistency in procedure that are needed to have early, precise, and unbiased estimates of the true burden of a disease. Such planning would increase the power of the approach taken here and lead to consistent estimates between approaches.

Data that are typically available from active case finding and passive surveillance are common in the investigation of emerging infectious diseases. The methods used here to estimate the true number of MERS-CoV infections could be applied to any disease in which this combination of data is available. Given the diverse assortment of first-line health care providers and local health systems involved in outbreak response and the enormous workload typically associated with the response, a key challenge is to collect the relevant data in a coordinated and standardized way from the early stages of the outbreak. This aspect of the response may be improved by numerous efforts to develop generic protocols for outbreak response prepared in advance (22, 23). In particular, it would be useful to have a clear record of the reasons for a case patient’s entry into the line list to facilitate analysis such as that presented here.

Although the at least 53% of MERS-CoV cases that our analysis suggested would be unobserved is not quite the 90% implied by the iceberg metaphor, the observed cases are only a fraction of the number of MERS-CoV infections observed. As is common when dealing with emerging diseases, our data on MERS-CoV are imperfect, and we must make assumptions to produce meaningful estimates of the true number of cases. If wrong, these assumptions generally bias the estimated number of infections down; thus, our estimate may be best viewed as a lower limit of the true number of MERS-CoV infections in the KSA. However, even the number of undetected infections projected by our work represents a sizeable population of individuals who were infected with the virus within the KSA but whose infections were not detected. More refined estimates will require serological surveys or modifications to the surveillance system with the specific aim of improving detection of asymptomatic cases.

ACKNOWLEDGMENTS

Author affiliations: Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Justin Lessler, Henrik Salje, Isabel Rodríguez-Barraquer, Derek A. T. Cummings); MRC Centre for Outbreak Analysis and Modelling, Faculty of Medicine, Imperial College London, London, United Kingdom (Maria D. Van Kerkhove, Neil M. Ferguson, Thibaut Jombart, Ricardo Aguas); Center for Global Health, Institut Pasteur, Paris, France (Maria D. Van Kerkhove); Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, Paris, France (Simon Cauchemez); Ministry of Health, Riyadh, Kingdom of Saudi Arabia (Rafat Hakeem, Ali Al-Barrak); Emerging Pathogens Institute, University of Florida, Gainesville, Florida (Derek A. T. Cummings); and Department of Biology, University of Florida, Gainesville, Florida (Derek A. T. Cummings).

J.L. was supported in part by the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. D.A.T.C. acknowledges funding from the US National Institute of General Medical Sciences (grant 5U54GM088491, Computational Models of Infectious Disease Threats). N.M.F. and S.C. acknowledge funding from the Medical Research Council, the National Institute of Health Research, and the Fogarty International Center, National Institutes of Health (grant 5U54GM088491, Computational Models of Infectious Disease Threats). N.M.F. and S.C. acknowledge funding from the Medical Research Council, the National Institute of Health Research, and the Fogarty International Center, National Institutes of Health (grant 5U54GM088491, Computational Models of Infectious Disease Threats).

REFERENCES

1. Gulland A. WHO voices concern over rising numbers of MERS-CoV cases. BMJ. 2014;348:g2968.
9. World Health Organization. Interim surveillance recommendations for human infection with Middle


