Underlying trend, seasonality, prediction, forecasting and the contribution of risk factors: an analysis of globally reported cases of Middle East Respiratory Syndrome Coronavirus

Omar B. Da’ar and Anwar E. Ahmed
King Abdullah International Medical Research Centre (KAIMRC)/College of Public Health & Health Informatics, King Saud bin Abdulaziz University of Health Sciences, National Guard Health Affairs, Riyadh, Saudi Arabia

Abstract
This study set out to identify and analyse trends and seasonal variations of monthly global reported cases of the Middle East respiratory syndrome coronavirus (MERS-CoV). It also made a prediction based on the reported and extrapolated into the future by forecasting the trend. Finally, the study assessed contributions of various risk factors in the reported cases. The motivation for this study is that MERS-CoV remains among the list of blueprint priority and potential pandemic diseases globally. Yet, there is a paucity of empirical literature examining trends and seasonality as the available evidence is generally descriptive and anecdotal. The study is a time series analysis using monthly global reported cases of MERS-CoV by the World Health Organisation between January 2015 and January 2018. We decomposed the series into seasonal, irregular and trend components and identified patterns, smoothened series, generated predictions and employed forecasting techniques based on linear regression. We assessed contributions of various risk factors in MERS-CoV cases over time. Successive months of the MERS-CoV cases suggest a significant decreasing trend ($P = 0.026$ for monthly series and $P = 0.047$ for Quarterly series). The MERS-CoV cases are forecast to wane by end 2018. Seasonality component of the cases oscillated below or above the baseline (the centred moving average), but no association with the series over time was noted. The results revealed contributions of risk factors such as camel contact, male, old age and being from Saudi Arabia and Middle East regions to the overall reported cases of MERS-CoV. The trend component and several risk factors for global MERS-CoV cases, including camel contact, male, age and geography/region significantly affected the series. Our statistical models appear to suggest significant predictive capacity and the findings may well inform healthcare practitioners and policymakers about the underlying dynamics that produced the globally reported MERS-CoV cases.

Introduction
This study set out to identify trends and seasonal variations; made a prediction based on the globally reported cases of the Middle East respiratory syndrome coronavirus (MERS-CoV), extrapolated into the future by forecasting the trend and assessed contributions of various risk factors for the MERS-CoV cases. Specifically, we consider the questions: (1) what is the underlying growth or trend of the globally reported MERS-CoV cases? (2) Are there seasonal variations present in globally reported MERS-CoV cases over time and how do they affect the series? (3) What are the contributions of various risk factors for MERS-CoV cases?

The motivation for this study is that to date, the World Health Organisation (WHO) places the MERS-CoV among the list of blueprint priority diseases. Although a survey of the literature shows a rapid increase in research activities related to MERS-CoV [1], yet there is still a general paucity of empirical literature examining trends and seasonality. To the best of our knowledge, there is a single study, which anecdotally examined seasonality and time series patterns of MERS-CoV to date [2]. Given MERS-CoV remains a potential pandemic disease globally, it is important to understand the dynamics of the underlying growth or trend of the globally reported cases.

The motivation for this study also comes from the aetiology of MERS-CoV, especially its causes, spread, the complexity of its diagnosis and mortality. MERS-CoV is a virus that causes severe viral pneumonia in humans, known to have a high mortality rate [3–7] and has clinical symptoms similar to severe acute respiratory syndrome coronavirus [8, 9]. It was first reported in Saudi Arabia [10] and after that, the virus exhibited outbreaks in several regions of the world, particularly Saudi Arabia and the Republic of Korea [10, 11]. Additionally, the complexity of MERS-CoV and its diagnosis of infection have been acknowledged in the literature [12, 13]. According to WHO, 2160 laboratory-confirmed
cases of MERS-CoV were reported at the end of January 2018, including 773 associated deaths (case–fatality rate: 35.8%) that were reported globally [14]. The majority of these cases were reported in Saudi Arabia.

Methods

This is a time series analysis using publicly reported MERS-CoV monthly global cases. The WHO receives confirmed MERS-CoV cases from countries across the world. To date, new cases continue to be identified and reported to the WHO, specifically from the Middle East region. These data are available at http://www.who.int/csr/don/archive/disease/coronavirus_infections/en/. The latest report included one case from Malaysia on 8 January 2018. We used time series of MERS-CoV cases reported between January 2015 and January 2018, where WHO began using standard case report. A research assistant retrieved data from WHO webpage and reviewed for quality by the second study author. The data retrieved include patient and clinical data such as age, gender, healthcare worker, comorbidity, the source of infection and geographical regions. The main outcome was the number of cases reported on a monthly basis from January 2015 to January 2018.

Statistical analysis

The analysis was performed using STATA 12 (STATA Corp., Texas, USA) and Microsoft Excel 10. Using the classical multiplicative time series model, we decomposed the original series into seasonal, irregular and trend components and examined their effects. Additionally, we identified patterns, smoothed series, generated predictions and employed forecasting techniques based on linear regression. Finally, we assessed contributions of various risk factors in MERS-CoV cases over time. P-values <0.05 were considered statistically significant.

Results

Figure 1 shows that although cases of MERS-CoV are decreasing in the range period selected, the series exhibits peaks and spikes. Figure 1 also reveals a negative trend of MERS-CoV series from January 2015 to January 2018.

We investigated direction and significance of this trend, as well as stationarity of the series. While a unit root test for non-stationarity confirmed the MERS-CoV cases had a negative and statistically significant trend, the series was found to be stationary. The negative and statistically significant trend was also confirmed by subsequent regressions.

We collapsed the monthly series into quarters and then smoothened out the series using a centred moving average in order to understand underlying growth component. We assumed the classical multiplicative time series model by decomposing original series into seasonal, irregular, and trend components (Table 1).

Our decomposition of the MERS-CoV cases shows that in 2016Q2, the seasonality and irregularity components of the series were 8% below the baseline (the centred moving average). Decomposing further, seasonality component was 11% above the baseline in 2016Q2, while it was 41% below the baseline in 2016Q4. Our analysis also shows the de-seasonalised series of the original MERS-CoV by removing seasonality and irregularity components. Using a linear regression, we then estimated the effect of time on the deseasonalised series to capture the underlying growth or trend component using a linear regression in order to make predictions. Since the last available data was in January 2018, we also made a forecast of three more quarters (2018Q2, Q3, Q4), an additional 9 months into the future. The forecast of the series revealed that MERS-CoV cases would approach zero by end of 2018 or beginning of 2019, making further extrapolation into the horizon infeasible.

Figure 2 shows decomposition of MERS-CoV series of the original series, centred moving average (smoothed series), forecast...
### Table 1. Decomposing reported MERS-CoV cases

<table>
<thead>
<tr>
<th>Time</th>
<th>Year</th>
<th>Quarter</th>
<th>MERS-CoV Cases</th>
<th>Moving average (4)</th>
<th>Centred moving average (4)</th>
<th>Seasonality with irregularity$(S_t, I_t)$</th>
<th>Seasonality, $(S_t)$</th>
<th>Deseasonalised</th>
<th>Trend</th>
<th>Predicted/Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2015</td>
<td>Q1</td>
<td>151</td>
<td></td>
<td></td>
<td>1.05</td>
<td>144</td>
<td></td>
<td>164.74</td>
<td>173</td>
</tr>
<tr>
<td>2</td>
<td>Q2</td>
<td>152</td>
<td></td>
<td></td>
<td></td>
<td>1.11</td>
<td>137</td>
<td></td>
<td>153.91</td>
<td>170</td>
</tr>
<tr>
<td>3</td>
<td>Q3</td>
<td>228</td>
<td></td>
<td></td>
<td></td>
<td>0.69</td>
<td>330</td>
<td></td>
<td>143.07</td>
<td>99</td>
</tr>
<tr>
<td>4</td>
<td>Q4</td>
<td>33</td>
<td>141</td>
<td>131</td>
<td>0.25</td>
<td>0.59</td>
<td>56</td>
<td>132.23</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2016</td>
<td>Q1</td>
<td>73</td>
<td>122</td>
<td>111</td>
<td>0.66</td>
<td>1.05</td>
<td>70</td>
<td>121.39</td>
<td>127</td>
</tr>
<tr>
<td>6</td>
<td>Q2</td>
<td>71</td>
<td>101</td>
<td>78</td>
<td>0.92</td>
<td>1.11</td>
<td>64</td>
<td>110.55</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Q3</td>
<td>38</td>
<td>54</td>
<td>57</td>
<td>0.67</td>
<td>0.69</td>
<td>55</td>
<td>99.72</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Q4</td>
<td>58</td>
<td>60</td>
<td>64</td>
<td>0.91</td>
<td>0.59</td>
<td>99</td>
<td>88.88</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2017</td>
<td>Q1</td>
<td>102</td>
<td>67</td>
<td>71</td>
<td>1.44</td>
<td>1.05</td>
<td>97</td>
<td>78.04</td>
<td>82</td>
</tr>
<tr>
<td>10</td>
<td>Q2</td>
<td>99</td>
<td>74</td>
<td>76</td>
<td>1.30</td>
<td>1.11</td>
<td>89</td>
<td>67.20</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Q3</td>
<td>54</td>
<td>78</td>
<td>76</td>
<td>0.71</td>
<td>0.69</td>
<td>78</td>
<td>56.36</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Q4</td>
<td>37</td>
<td>73</td>
<td>62</td>
<td>0.60</td>
<td>0.59</td>
<td>63</td>
<td>45.53</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2018</td>
<td>Q1</td>
<td>14</td>
<td>51</td>
<td></td>
<td>1.05</td>
<td>13</td>
<td>34.69</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Q2*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.11</td>
<td></td>
<td>23.85</td>
<td>26.41</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Q3*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.69</td>
<td></td>
<td>13.01</td>
<td>8.99</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Q4*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.59</td>
<td></td>
<td>2.17</td>
<td>1.28</td>
<td></td>
</tr>
</tbody>
</table>

Q1, Q2, Q3, Q4 are the Quarters of the year and * = implies forecast.
and linear forecast. Together, after accounting for irregular and trend components of the series, seasonality was found to range between 41% below the baseline i.e. the centred moving average (of order 4) in some Quarters and 11% above the baseline in other Quarters. The average seasonality component was found to be 14% below the baseline. However, regression estimation revealed that, unlike the trend component, seasonality was not statistically significant, a fact also backed by our statistical test, which showed the monthly global MERS-CoV cases series were stationary.

**Regressions**

Table 2 shows results of the effect of trend and seasonality on MERS-CoV cases. We compared the seasonal dummy variables, interpreted by comparing them with Quarter 4 (Q4) (base season) while holding time constant. Time, in this analysis, was represented by successive months and was interpreted as the effect of the linear trend on MERS-CoV cases over time, holding the effect of the seasons constant.

The regression results revealed that after accounting for the trend, MERS-CoV cases Quarter 1 (Q1) each year averaged about 14 cases more than Q4 cases after accounting for the trend component time. It is important to note that the effects of seasonality were not statistically significant. However, what was revealed statistically significant is the negative trend. Consistent with the negative trend shown in Figure 1, the regression results revealed that each additional quarter registered approximately an average decrease of one case, after adjusting for the season. In other words, the MERS-CoV cases decreased, on average, by four (4 × −0.867) per quarter year to year.

The model as a whole appears to suggest statistically significant predictive capacity.

We analysed fluctuations of reported MERS-CoV cases from period to another by graphing the residuals (generated from the regression of trend and seasonality on MERS-CoV cases) against time, as is the convention with time series analysis. The results indicated no clear patterns, suggesting that correlated errors are not a problem with this model. Other diagnostic tests also revealed neither violation of the classical linear assumptions nor correlation between the reported MERS-CoV cases in each month with cases reported in earlier months.

We further examined the effects of various risk factors of MERS-CoV cases such as camel contact, healthcare worker contact, exposure, gender and region. Table 3 shows regression that adjusts for these factors. The results reveal camel contact, Saudi

---

**Table 2. Effect of trend and seasonality (n = 37)**

<table>
<thead>
<tr>
<th>Dependent variable = MERS-CoV cases</th>
<th>Coefficient</th>
<th>Robust s.e.</th>
<th>t value</th>
<th>P-value</th>
<th>(95% Conf. Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>34.158</td>
<td>11.025</td>
<td>3.100</td>
<td>0.004</td>
<td>11.701</td>
</tr>
<tr>
<td>Time</td>
<td>−0.867</td>
<td>0.420</td>
<td>−2.070</td>
<td>0.047&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−1.722</td>
</tr>
<tr>
<td>Quarter_1</td>
<td>13.970</td>
<td>8.998</td>
<td>1.550</td>
<td>0.130</td>
<td>−4.358</td>
</tr>
<tr>
<td>Quarter_2</td>
<td>16.355</td>
<td>9.647</td>
<td>1.700</td>
<td>0.100</td>
<td>−3.295</td>
</tr>
<tr>
<td>Quarter_3</td>
<td>18.733</td>
<td>12.321</td>
<td>1.520</td>
<td>0.138</td>
<td>−6.363</td>
</tr>
</tbody>
</table>

<sup>a</sup>Significant at <5% level.
Table 3. Effects of risk factors of MERS-CoV (n = 36)

<table>
<thead>
<tr>
<th>MERS-CoV Cases</th>
<th>Coefficient</th>
<th>Robust s.e.</th>
<th>t value</th>
<th>P &gt; t</th>
<th>(95% Conf. Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.400</td>
<td>0.434</td>
<td>0.920</td>
<td>0.365</td>
<td>−0.049 – 1.292</td>
</tr>
<tr>
<td>Camel exposure</td>
<td>0.166</td>
<td>0.065</td>
<td>2.580</td>
<td>0.016a</td>
<td>0.034 – 0.299</td>
</tr>
<tr>
<td>Comorbidities</td>
<td>−0.114</td>
<td>0.099</td>
<td>−1.140</td>
<td>0.264</td>
<td>−0.318 – 0.091</td>
</tr>
<tr>
<td>Healthcare worker</td>
<td>0.004</td>
<td>0.070</td>
<td>0.060</td>
<td>0.951</td>
<td>−0.140 – 0.148</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>0.149</td>
<td>0.034</td>
<td>4.430</td>
<td>&lt;0.001a</td>
<td>0.080 – 0.217</td>
</tr>
<tr>
<td>Middle East</td>
<td>0.511</td>
<td>0.134</td>
<td>3.810</td>
<td>0.001a</td>
<td>0.235 – 0.788</td>
</tr>
<tr>
<td>Male</td>
<td>0.372</td>
<td>0.093</td>
<td>4.020</td>
<td>&lt;0.001a</td>
<td>0.182 – 0.563</td>
</tr>
<tr>
<td>Age ≥ 60 years</td>
<td>0.493</td>
<td>0.164</td>
<td>3.010</td>
<td>0.006a</td>
<td>0.156 – 0.831</td>
</tr>
<tr>
<td>Age 30–59 years</td>
<td>0.596</td>
<td>0.094</td>
<td>6.350</td>
<td>&lt;0.001a</td>
<td>0.403 – 0.789</td>
</tr>
<tr>
<td>Exposure to MERS-CoV cases</td>
<td>0.239</td>
<td>0.055</td>
<td>4.360</td>
<td>&lt;0.001a</td>
<td>0.126 – 0.352</td>
</tr>
</tbody>
</table>

*aSignificant at <5% level.

Arabia and the Middle East regions, as well as being male significantly contribute to the global reported MERS-CoV cases (Table 3). The model in this estimation also appears to suggest statistically significant predictive capacity.

Discussion

Using linear time series models and their application to the modelling and prediction of the globally reported MERS-CoV data, the present study identified trends, analysed seasonality, predicted and forecast evolution of MERS-CoV cases and assessed the contribution of various risk factors. The decomposition of the time series of MERS-CoV cases into trend and seasonality components and making predictions have not hitherto been studied in the context of MERS-CoV pandemic. In this study, we set out to understand the dynamics of its growth over time.

The results of our time series analysis of globally reported MERS-CoV cases suggest a significant negative trend that is forecast to be eradicated in the near future unless something unexpected happens.

Our study showed that although seasonality oscillated below or above the baseline i.e. the centred moving average (of order 4) over time, the average seasonality component was found to be 14% below the baseline. Even then, our analysis showed that, unlike the trend component, seasonality did not affect the series over time. Many risk factors are associated with MERS-CoV cases, mortalities, or complications. Our results indicate those aged under 30 years (reference category) are less likely to be a MERS-CoV case than those aged over 30, consistent with several studies that associated MERS-COV with elderly patients [15, 16]. Surprisingly, comorbidity did not show a statistically significant contribution to MERS-CoV cases. However, there are studies that showed MERS-CoV cases were associated with patients with comorbidities [17–19]. A recent systematic study, for example, suggests the prevalence of comorbidities of MERS-CoV cases, diabetes, hypertension and cardiac diseases [20]. While our analysis suggests males contribute to the global reported MERS-CoV cases, gender was reported to have a mixed effect on MERS-CoV mortalities in the literature. Some studies showed men as high risk [7, 18] and MERS-CoV infects more males than females [5, 21–23]. Other literature indicated that the frequency of deaths was less in men [24].

The literature showed that MERS-CoV can be spread via human–human [25–29], or healthcare facilities [23, 30–32]. Other studies revealed animal to human [33, 34] as the primary culprit of MERS-CoV virus transmission. Specifically, the literature showed camels act as a direct source of human MERS-CoV infection [31, 35], while healthcare workers were reported to be at higher risk [7, 16, 19, 24]. The results of our study in this regard were mixed. While our study indicated that the effect of camel cases on overall MERS-CoV reported cases are positive and significant, the contribution of healthcare workers was not. Our analysis also showed evidence of geographical contributions to MERS-CoV cases such as Saudi Arabia and greater Middle East compared with South Korea. This can be seen as somewhat consistent with earlier studies that demonstrated a link between mortality associated with MERS-CoV and geography [15]. This finding is also intuitive in that Saudi Arabia and Middle East, in general, remain the global epicentre of MERS-CoV, hence the name.

The contribution of our study is that it adduces empirical evidence by making inferences and predictions based on the globally reported cases of MERS-CoV and extrapolated into the future by forecasting the trend. Unlike previous studies that descriptively analysed seasonality patterns of MERS-CoV and influenza in the Middle East [2], our study presents statistically significant results of trends of global MERS-CoV cases, consistent with regularities underlying the empirical dynamics and classical time series analysis. However, there are limitations of this study. First, the data used for this study comprised 37 months (January 2015 to January 2018). While this was just enough for several years’ worth of monthly observations to appropriately model seasonality, time series analysis can be sensitive to the number of observations. Hence, sufficiently large number of observations might have provided a better fit and results. Additionally, the analysis utilised WHO open source globally reported data, which may lack harmonisation from the various country sources.

Conclusion

This study contributes to the time series analysis of MERS-CoV literature. In particular, our analysis of trends and seasonality components the series, the prediction based on the globally reported cases of MERS-CoV and extrapolation into the future...
by forecasting the trend is envisaged to help in understanding the dynamics of the reported cases over time. The study findings suggest a significant negative trend of the monthly and quarterly data from 2015 to 2018. However, a further extrapolation into the future reveals that the MERS-CoV cases are forecast to be zero by end 2018 or beginning of 2019 unless something unexpected happens. Seasonality component of the series oscillated below or above the baseline, i.e. the centred moving average but did not affect the series over time. The results demonstrated that camel contact, exposure, gender, age and geography/region significantly contributed to the overall global reported MERS-CoV cases. The findings may well inform healthcare practitioners and policymakers about the underlying dynamics that produced the globally reported MERS-CoV cases.

Data. These dataset used and/or analysed are available at http://www.who.int/csr/don/archive/disease/coronavirus_infections/en/.

Acknowledgements. The authors are grateful for the collegiality and research support at the College of Public Health and Health Informatics, King Saud bin Abdulaziz International Medical Research Centre (KAIMRC), King Saud bin Abdulaziz University for Health Sciences, the information and opinions contained in this work do not necessarily reflect the views or policies of these institutions. This research is supported by King Abdullah International Medical Research Centre (KAIMRC), King Saud bin Abdulaziz University for Health Sciences, National Guard Health Affairs, Riyadh, Saudi Arabia.

Author contributions. OBD analysed the data and wrote manuscript. AEA Saudi Arabia.

International Medical Research Centre (KAIMRC), King Saud bin Abdulaziz University for Health Sciences. The information and opinions contained in this work do not necessarily reflect the views or policies of these institutions. This research is supported by King Abdullah International Medical Research Centre (KAIMRC), King Saud bin Abdulaziz University for Health Sciences, National Guard Health Affairs, Riyadh, Saudi Arabia.

Conflict of interest. None declared.

Ethical standards. Not applicable.

References


Downloaded from https://www.cambridge.org/core. IP address: 54.70.40.11, on 03 May 2019 at 01:01:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1017/S0950268818001541

