A rapid risk analysis tool to prioritise response to infectious disease outbreaks

Dyah A S Lesmanawati,1,2 Patrick Veenstra,3 Aye Moa,4 Dillon C Adam ☑,4 Chandini Raina MacIntyre5

ABSTRACT
Epidemics are influenced by both disease and societal factors and can grow exponentially over short time periods. Epidemic risk analysis can help in rapidly predicting potentially serious outcomes and flagging the need for rapid response. We developed a multifactorial risk analysis tool ‘EpiRisk’ to provide rapid insight into the potential severity of emerging epidemics by combining disease-related parameters and country-related risk parameters. An initial set of 18 disease and country-related risk parameters was reduced to 14 following qualitative discussions and the removal of highly correlated parameters by a correlation and clustering analysis. Of the remaining parameters, three risk levels were assigned ranging from low (1) moderate (2) and high (3). The total risk score for an outbreak of a given disease in a particular country is calculated by summing these 14 risk scores, and this sum is subsequently classified into one of four risk categories: low risk (≤21), moderate risk (21–29), high risk (30–37) and extreme risk (>37). Total risk scores were calculated for nine retrospective outbreaks demonstrating an association with the actual impact of those outbreaks. We also evaluated to what extent the risk scores correlate with the number of cases and deaths in 61 additional outbreaks between 2002 and 2018, demonstrating positive associations with outbreak severity as measured by the number of deaths. Using EpiRisk, timely intervention can be implemented by predicting the risk of emerging outbreaks in real time, which may help government and public health professionals prevent catastrophic epidemic outcomes.

INTRODUCTION
Since 2000, there has been an escalation in the number of serious infectious disease outbreaks worldwide.1-2 For example, the 2002–2003 severe acute respiratory syndrome coronavirus (SARS-CoV) epidemic spread to 37 countries, resulting in over 8000 cases and more than 700 deaths.3 Five years later in 2009, a new influenza subtype, A/H1N1pdm09 emerged in North America before circulating around the world affecting an estimated 24% of the global population.4-5 The 2013–2016 Ebola outbreak in West Africa was an unprecedented catastrophe, and Ebola continues to pose a global threat.6 Zika virus spread from Brazil and infected an estimated 1.5 million people, resulting in upwards of 4000 cases of microcephaly among infants.7-8 Outbreaks of acute flaccid myelitis have also occurred across the USA, Europe and Asia, associated with EV-D68.9 There has also been an increase in the number of vaccine-preventable disease outbreaks in recent years including polio,10 measles11 and diphtheria12. Lessons learnt from recent serious epidemics show that control measures may not be sufficient or timely enough,13 and even short delays to outbreak response result in preventable morbidity and mortality.14-15 The response to the 2013–2014 West African

Key questions

What is already known?
► There has been an increase in the number of serious infectious disease outbreaks worldwide in recent years.
► Rapid response time is critical for epidemic-prone diseases, and short delays to outbreak response result in preventable morbidity and mortality.
► Risk analysis can be used to prioritise a response to an epidemic by examining the potential impact of the infectious disease outbreak.

What are the new findings?
► EpiRisk provides rapid risk prediction of outbreaks that will assist decision makers in epidemic management.
► Both pathogen and country parameters have a significant impact on the risk of infectious disease outbreaks in particular countries or regions.
► EpiRisk provides an individualised approach where specific input data for a country and disease can be used, rather than a standard ‘one size fits all’ approach that would be less generalisable.

What do the new findings imply?
► The development of a simple risk analysis tool will be useful for global epidemic control. This tool can be used to rapidly predict the risk of outbreaks, which is useful when planning and prioritising interventions or for epidemic preparedness.
► An appropriate and timely intervention can help governments and public health professionals prevent catastrophic epidemic outcomes.

Handling editor Seye Abimbola

Accepted 29 April 2020
Revised 28 April 2020
Received 20 January 2020

Original research


BMJ Glob Health: first published as 10.1136/bmjgh-2020-002327 on 7 June 2020. Downloaded from http://gh.bmj.com/ on December 18, 2021 by guest. Protected by copyright.
Ebola outbreak is an example of a delayed intervention. Traditional laboratory and surveillance systems misdiagnosed Ebola victims as Lassa fever and cholera. The first confirmed laboratory case detected by scientists in France occurred approximately 2.5 months after illness in the index case, and intervention by WHO was delayed for at least 5 months after that. This failure in the early stages of the outbreak had significant follow-on effects, contributing substantially to an exponential increase in epidemic size to over 28,000 cases and 11,000 deaths. In 2015, 21 polio-free countries reported re-emergence of the virus with a total of over 1000 confirmed cases, with a slow response implicated as a factor in the impact of these outbreaks. Rapid response time is critical for epidemic-prone diseases, and decision support tools to prompt rapid response may be useful.

When looking at global patterns, there is diversity in the severity and impact of outbreaks, which means in the early stages, tools that can predict risk may help identify those with catastrophic potential. Both pathogen and country parameters have a significant impact on the risk of infectious disease outbreaks in a particular country or region. For example, innate characteristics of the pathogen such as the type of pathogen, transmission mode, basic reproductive number, case fatality rate, as well as the availability of effective therapy and vaccine for that particular disease all have a role in predicting the risk of an epidemic. Country factors including the social, economic and cultural characteristics of a country may also contribute to the risk of an outbreak. Risk assessment frameworks exist for polio, measles and dengue, which are used in endemic countries such as the Philippines and Romania. The availability, however, of a simple, universal epidemic risk scoring framework is valuable to predict the risk of an infectious disease outbreak and to prioritise response.

There are limited studies assessing the risk of infectious disease outbreaks based on country features. In 2017, Ajisegiri et al. used such an approach to predict the risk and outcome of the Ebola epidemic worldwide by considering country-specific parameters based on socioeconomic, health system, cultural and geographical factors. The risk framework by Ajisegiri successfully stratified the risk of Ebola outbreaks by country and predicted a much higher risk in West Africa compared with the USA or UK. Had such a tool been available in March 2014, it may have helped prompt a more rapid response. This study aimed to develop a generalised epidemic risk analysis framework that incorporates both pathogen and country-specific parameters to classify the risk of epidemics and provides an early warning system for the management of global outbreaks.

Methods
We developed a risk analysis framework to predict the risk of an epidemic of any cause modifying the approach used in the Ebola-specific tool developed by Ajisegiri et al. In 2018, Ajisegiri et al. identified the role of sociodemographic features of the Ebola-affected countries to the magnitude of the outbreak. The authors developed a framework that assigned risk scores (from 1 to 3) to country-specific characteristics such as socioeconomic, health systems and geographical factors, as well as cultural beliefs and traditional practices. These risk scores were then added into a simple summation model to produce a single risk figure for a given country. Our modified approach differs from that by Ajisegiri et al. to produce a single risk figure for a given country. Our modified approach differs from that by Ajisegiri et al. in two key ways. First, we expanded the scope of the framework developed by Ajisegiri et al. to be suitable for almost any outbreak in any country (pending data availability) by including both disease-specific and country-specific parameters. Second, we used a subset of the country parameters used by Ajisegiri et al. A key objective of the framework presented in this paper is to allow for a quick and simple evaluation of outbreak severity risk in countries around the world. Some factors used by Ajisegiri et al. require data that are not easily or quickly available (eg, screening at borders and bush meat consumption), and while those factors were suitable for evaluating a local region such as that in the Ebola outbreak, it does not meet the needs of our global framework. Data were collected on outbreak parameters that fall into two different categories: disease-related parameters and country-related parameters, as outlined in table 1.

Selection of initial risk parameters and data collection
We searched the literature to identify common risk factors of infectious disease outbreaks using three online databases: PubMed/medline, Scopus and CINAHL, using keywords ‘outbreak* OR epidemic* OR pandemic* OR emerging disease* OR re-emerged disease*’. We identified numerous initial factors associated with the severity of the disease outbreaks, which we categorised as either disease-related parameters or country-related parameters. Additional disease-related parameters were selected based on a modified Disease Attribute Intelligence System risk assessment tool by the Institute of Environmental Science and Research Limited, while additional country-related parameters were identified based on the risk analysis framework tool by Ajisegiri et al. The financial capability and the level of investment of a country plays a significant role in the early response to outbreaks. Sociopolitical factors also contribute to the length and outcome of outbreaks. For example, war and conflict provide ideal conditions for outbreaks of infection diseases. In conflict areas, health professionals flee, infrastructure is destroyed and the supply of medical equipment is halted. In some instances, hampering of immunisation efforts has also contributed to the spread of vaccine-preventable diseases, making a country more vulnerable to outbreaks. The majority of severe disease outbreaks also originate from densely populated regions. Overcrowding coupled with lower living standards can lead to the efficient transmission of diseases and positively correlate with the risk of an outbreak. The existing health system within the country also plays...
### Table 1  Risk analysis framework scoring criteria

<table>
<thead>
<tr>
<th>No</th>
<th>Risk factors</th>
<th>Risk score</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disease identification</td>
<td>1</td>
<td>Clinical syndrome is diagnostic.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>A simple laboratory test is diagnostic.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>Advance or prolonged investigation is required for confirmatory diagnosis.</td>
</tr>
<tr>
<td>2</td>
<td>Pathogen</td>
<td>Others</td>
<td>Bacterial</td>
</tr>
<tr>
<td>3</td>
<td>Reservoir</td>
<td>Animal</td>
<td>Environmental</td>
</tr>
<tr>
<td>4</td>
<td>Basic reproductive number</td>
<td>&lt;1</td>
<td>1.0–2.0</td>
</tr>
<tr>
<td>5</td>
<td>Mode of transmission</td>
<td>Vector borne or other animal borne</td>
<td>Foodborne, waterborne and direct contact</td>
</tr>
<tr>
<td>6</td>
<td>Asymptomatic transmission</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>Case fatality rate</td>
<td>&lt;1%</td>
<td>1.0%–5.0%</td>
</tr>
<tr>
<td>8</td>
<td>Therapy/drug availability</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>Vaccine availability</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>Income</td>
<td>High-income countries (&gt;12 056)</td>
<td>Middle-income countries ($995–12 056)</td>
</tr>
</tbody>
</table>
Table 1

<table>
<thead>
<tr>
<th>No</th>
<th>Risk factors</th>
<th>Risk score</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>HE total (% of GDP)</td>
<td>Upper third (&gt;10.0%)</td>
<td>Health expenditure (HE) indicates the proportion of country GDP that is assigned to the health sector expenditure. The proportion of health expenditure in a country has a strong association with the healthcare services provided to the population. A country’s health expenditure was divided into upper, middle and lower third.</td>
</tr>
<tr>
<td>12</td>
<td>Peace index</td>
<td>High peace (&lt;2.0)</td>
<td>The state of peace may impact the country's health system. In a conflict-affected country, access to essential services is poor. As a consequence, the population in that country become more vulnerable to infectious disease transmission. In addition, detection and control of the outbreak is a challenge in the conflict-affected populations. The Institute for Economics and Peace divided state of peace into five different groups: very high, high, medium, low and very low. However, for this framework, we categorised the state of peace into three groups: high (combination of very high and high group), medium and low peace (combination of very low and low).</td>
</tr>
<tr>
<td>13</td>
<td>Land border</td>
<td>Maritime only (island nation)</td>
<td>The country border is associated with the mobility of people, accessibility of transportation modes as well as the time required to travel. A land border has a higher likelihood of interborder disease transmission than water border because of easier access.</td>
</tr>
<tr>
<td>14</td>
<td>Transport network</td>
<td>&lt;2</td>
<td>The transportation network is an important determinant of health. It affects mobility from one area to another. In an outbreak event, the transportation network has a major role in the possible disease transmission. The better transportation network that the country has also increases the likelihood of disease transmission. The World Economic Forum measures the quality of transportation infrastructure based on a country’s roads, railways and air transport infrastructure data. We combined the country’s score of roads, railways and air transport infrastructure and divided them into three different categories of low, medium and high.</td>
</tr>
<tr>
<td>15</td>
<td>Population density</td>
<td>&lt;100/km²</td>
<td>Overcrowding is one of the major factors in the disease transmission risk. Epidemics of disease may be more severe in high-density populations than low-density populations.</td>
</tr>
<tr>
<td>16</td>
<td>Physician density</td>
<td>&gt;2.9/1000 populations</td>
<td>Adequate availability and skill of physicians are critical for a country to attain population health goals through the provision of sufficient basic medical care. Lack of availability and accessibility to the physician service could aggravate the impact of an outbreak because of delay in medical treatment. WHO report indicates the range of physician density per 1000 population; &lt;0.8 as low, 0.8–2.9 as middle and &gt;2.9 as high.</td>
</tr>
</tbody>
</table>
a crucial role in the risk of outbreaks. An epidemic can spread wider and faster in a country with a weak health system.33

The infectiousness of pathogens also contribute to the risk level of an outbreak.33 Diseases caused by viral pathogens are likely to transmit more rapidly due to the nature of viral replication and mode of transmission in contrast to many bacteria;35 pathogens such as measles are more infectious partly because they spread by the respiratory route and have a high R0.36 Cholera, although spread via the faecal–oral route, is not as rapid as airborne transmission; however, it has a high reproductive number.36 During the South Sudan epidemic in 2016 cholera spread rapidly, resulting in over 20,000 cases and 436 deaths. Besides infectiousness, the availability of control measures for prevention and treatment also has an impact on the overall risk of an outbreak. These factors make it easier to prevent and control epidemics. Vaccination reduces the risk of infections by reducing the number of susceptible people in the population.37 Treatment availability affects the risk of disease outbreak by reducing the duration, infectious period and severity of disease.38

Initially, 18 parameters were selected to calculate overall epidemic risk (table 1). Explanations of each parameter can also be seen in table 1. Sources of data for country-related parameters included the World Bank, WHO Global Health Observatory and the Peace Institute. For disease-related parameters, a mixture of grey literature and official sources including Centers for Disease Control and Prevention and WHO were used.

The preparation of country-specific data involved collection of both current and historical data points for 198 countries from the World Bank database. The historical data were required to allow for a valid evaluation of the tool on historical outbreaks. That is, we used country-specific data points (eg, gross domestic product and physicians’ numbers) that were relevant at the time of the outbreaks used for evaluation.

As part of the collection of pathogen-specific data, we performed a review of three outbreak databases: ProMED-mail,39 Healthmap40 and EpiWatch41 to collect a line-list of diseases from recent outbreak events for evaluation of the tool.

Criteria for allocating risk scores
We applied values ranging from one to three for each parameter to indicate the level of risk, where ‘low risk’ was denoted by ‘1’, ‘moderate risk’ was denoted by ‘2’ and ‘high risk’ was denoted by ‘3’. We applied minimum and maximum values, score 1 as ‘yes’ and 3 as ‘no’ for binary parameters that had only two different risk groups, for example, asymptomatic transmission, vaccine availability and drug availability. Criteria for assigning risk scores to selected parameters were used from the relevant studies.19 20 The values applied were based on specific risk factor criteria and are detailed in table 1. Where there was no available data for a parameter (N/A=not available), the highest risk score was assigned.

Selection of final parameters
The initial set of 18 parameters were reduced to 14 as a result of qualitative discussions and a correlation and clustering analysis on the country parameters. The country parameter ‘roadways/transport network’ was removed due to conflicting interpretations of risk i.e. (1) it may indicate better access to healthcare services, or (2) increase contact between people and hence risk of infection. The disease-factor ‘Disease identification method’ was removed as it provided little variance with most diseases requiring further investigation.

### Table 1

<table>
<thead>
<tr>
<th>No</th>
<th>Risk factors</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>Nurses and midwife density</td>
<td>&gt;7.1/1000 populations</td>
<td>1.7–7.1/1000 populations</td>
<td>&lt;1.7/1000 populations</td>
<td>Nurses and midwives play a crucial role in healthcare services, both in the hospital and community settings. Sufficient number and capacity of nurses and midwives would be associated with outbreak impact by reducing potential disease transmission in the population. WHO report categorises population-weighted density of nurse and midwife in three groups: low, moderate and high.</td>
</tr>
<tr>
<td>18</td>
<td>Hospital beds</td>
<td>&gt;4 beds/1000 populations</td>
<td>&gt;2–4 beds/1000 populations</td>
<td>0.1–2 beds/1000 populations</td>
<td>Hospital beds are an indicator of available resources to deliver healthcare services. Without sufficient hospital beds during an outbreak event, the likelihood of uncontrolled transmission increases. When insufficient resources force patients to stay longer in the community, the risk of disease transmission in the community increases.</td>
</tr>
</tbody>
</table>

Factors in bold were included in the final EpiRisk tool after qualitative analysis and a correlation analysis.

GDP, gross domestic product.
for confirmatory diagnosis. The ‘reservoir’ factor was removed as its mapping to risk is not consistent and straightforward.

After the qualitative determination to remove the three aforementioned factors, leaving 15 factors (eight county, seven disease), we performed a quantitative analysis to compute the correlations between all remaining country parameters for the 198 countries in our dataset. As some of the country factors (see table 1) consist of ordinal level data, we used the Kendall correlation coefficient to compute the correlations. See figure 1 for the results between the eight country factors.

The correlation results demonstrated that income, hospital bed density, physician density, and nurses and midwife density were highly correlated factors. We decided that income and hospital bed density were both useful information to show in the framework, and we include them both, despite the Kendall correlation of 0.53. This will lead to a bias towards this in the model, which we address in our conclusion. Regarding the physician density and nurse/midwife density factors, their correlation is 0.66, and we decided they represent similar information about the healthcare system of a country; we therefore chose to retain physician density as a factor over nurse and midwife density, because of physician diagnostic and therapy skill and due to slightly more limited historical availability of nurse and midwife density data for some countries. The initial 18 parameters were reduced to 14 after the reduction process described above (the remaining factors are bold in table 1).

In our framework, an outbreak in a given country is given risk scores in each of the seven disease factors and seven country factors. These scores are then summed to produce a disease score and a country score for the outbreak, which are then in turn summed to produce a final risk score. The minimum possible risk score is 14 (all 1s) and the maximum is 42 (all 3s). The total score was classified into four risk categories to rank the priority. The categories are ‘low risk’ for a score less than 21, ‘moderate risk’ for a score between 21 and 29, ‘high risk’ for score 30–37 and ‘Extreme risk’ for total score beyond 37. The cut-offs were determined based on expert knowledge of historical outbreaks.

The historical and current country data, together with the disease data, are stored in csv files and managed in the R statistical software package for analysis. The initial data collection was managed in Excel.

**Patient and public involvement statement**

Involvement of patients or public in this research was not applicable.

**Model evaluation**

We evaluated our risk framework in two ways. First, to demonstrate the use of the framework, we applied it to a number of past outbreaks and showed that our computed risk scores provided insight into the severity of nine historic outbreaks. Second, we collected data on 61 different outbreaks between 2002 and 2018 and showed to what extent the risk scores computed in our model correlate with the number of cases and deaths in those 61 outbreaks.

First, to demonstrate the use of the framework, we selected nine different outbreaks with varied outcomes, from different countries with low, medium and high incomes from the last 5 years (2015–2019) in order to cover a range of disease-related and country-related parameters. Table 2 presents these outbreaks together with the country and disease risk scores computed using our framework. The outbreaks consist of the following: hepatitis A in Australia (2018), Ebola in the USA (2018), measles in Japan (2018), diptheria in Bangladesh (2017), Zika in Brazil (2015), hepatitis A in Italy (2013), Ebola in Sierra Leone (2014), cholera in South Sudan (2016) and Lassa fever in Nigeria (2018).

These nine historical outbreaks were divided into three groups based on similarity of duration, cases/deaths and international aid:

- **Group 1:** hepatitis A outbreak in Australia, Ebola outbreak in the USA and the measles outbreak in Japan. These three epidemics lasted for less than 3 months, had only a few cases, few fatalities and received no international support.
- **Group 2:** diptheria in Bangladesh, Zika virus in Brazil and hepatitis A in Italy all shared similar characteristics: the number of cases were high, but the case–fatality rate was low. There were 218,931 people affected with Zika outbreak in Brazil with only 11 deaths. The outbreak duration of this group varied from 6 weeks to 18 months. Both Bangladesh and Brazil obtained international aid, while Italy managed the outbreak with its own resources.
- **Group 3:** Ebola in Sierra Leone, cholera in South Sudan and Lassa fever in Nigeria include large epidemics
### Table 2: Calibration of EpiRisk tool with nine historic outbreaks

<table>
<thead>
<tr>
<th>Outbreak characteristics</th>
<th>Hepatitis A (Australia)</th>
<th>Measles (Japan)</th>
<th>Hepatitis A (Italy)</th>
<th>Ebola (USA)</th>
<th>Zika (Brazil)</th>
<th>Diphtheria (Bangladesh)</th>
<th>Cholera (South Sudan)</th>
<th>Ebola (Sierra Leone)</th>
<th>Lassa fever (Nigeria)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>10 weeks</td>
<td>8 weeks</td>
<td>18 months</td>
<td>12 weeks</td>
<td>&gt;10 months</td>
<td>6 weeks</td>
<td>18 months</td>
<td>19 months</td>
<td>13 months</td>
</tr>
<tr>
<td>Cases (suspected and confirmed)</td>
<td>30</td>
<td>161</td>
<td>1803</td>
<td>4</td>
<td>218931</td>
<td>804</td>
<td>20438</td>
<td>13683</td>
<td>4466</td>
</tr>
<tr>
<td>Deaths (probable and confirmed)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>15</td>
<td>436</td>
<td>3953</td>
<td>142</td>
</tr>
<tr>
<td>International aid received</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Reference</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td>60</td>
<td>59</td>
<td>61</td>
<td>62</td>
</tr>
</tbody>
</table>

#### Country parameters

<table>
<thead>
<tr>
<th>Income</th>
<th>HE total (% of GDP)</th>
<th>The state of peace</th>
<th>Country’s border</th>
<th>Physician density (per 1000)</th>
<th>Hospital bed density (per 1000)</th>
<th>Population density</th>
<th>Total country score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>18</td>
</tr>
</tbody>
</table>

#### Disease parameters

<table>
<thead>
<tr>
<th>Pathogen</th>
<th>Basic reproductive number</th>
<th>Mode of transmission</th>
<th>Asymptomatic stage</th>
<th>Case fatality rate</th>
<th>Therapy/drug availability</th>
<th>Vaccine availability</th>
<th>Total disease score</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>18</td>
</tr>
</tbody>
</table>

#### Overall score

<table>
<thead>
<tr>
<th>Moderate</th>
<th>Moderate</th>
<th>Moderate</th>
<th>Moderate</th>
<th>High</th>
<th>High</th>
<th>High</th>
<th>High</th>
<th>Extreme</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>29</td>
<td>27</td>
<td>29</td>
<td>31</td>
<td>34</td>
<td>33</td>
<td>33</td>
<td>39</td>
</tr>
</tbody>
</table>

with a long duration and high case fatality and morbidity rates. All group 3 outbreaks required international aid to control the outbreak.

To demonstrate the extent to which the model risk scores correlate with outbreak severity, we collected data on 61 different (non-endemic) outbreaks between 2002 and 2018 of 18 different pathogens in 43 different countries with a wide range of number of cases (mean 25691, median 804, first quantile 100, third quantile 2734) and deaths (mean 347, median 14, first quantile 1, third quantile 76) per outbreak. Figure 2 shows the distribution of the total risk score computed for the 61 outbreaks in the dataset.

For our evaluation, we split the outbreak risk scores into four quantiles, and we split the number of deaths into four quantiles. That is, Q1 of risk scores represents the 15 outbreaks in our evaluation set with the lowest risk scores and Q4 represents the 15 outbreaks in our
evaluation set with the highest risk scores. Similarly, Q1 of deaths represents the 15 outbreaks that resulted in the lowest number of deaths and Q4 represents the 15 outbreaks that resulted in the highest number of deaths. We then tabulated these two quantile groups against each other to evaluate to what extent outbreaks with the highest number of deaths were also the outbreaks with the highest risk scores in our framework and similarly for the lower quantiles. See table 3 in the results section for the result.

RESULTS

Table 1 presents the sourced parameters for each outbreak corresponding to the nine outbreaks listed in table 2 with the corresponding risk score and risk classification. The total risk scores in table 2 is the sum of the total country score and total disease score. Australia has the lowest country score, and the overall risk score for Australia is the lowest. Likewise, Japan had a low country score. In contrast, Nigeria had the highest total country score, while Bangladesh also had a high country score; however, diphtheria’s diseases-related score is the lowest among all diseases tested, so the overall risk score is not high enough to characterise the outbreak as extreme risk but was characterised as high-risk overall.

DISCUSSION

For our quantitative evaluation of risk scores computed on an evaluation dataset of 61 outbreaks, we first split the outbreaks into quantiles. Each of the 61 outbreaks was placed in one of four quantiles for the number of deaths associated with the outbreak. An outbreak in Q1 was an outbreak where the number of deaths was in the lowest 25% of the full dataset of 61 outbreaks. An outbreak in Q4 for the number of deaths has a number of deaths associated with it that is within the highest 25% of the full dataset. The same method is used to place the outbreaks into four quantiles of risk scores as computed by our risk framework. Table 3 shows the result. What is promising is that of the 16 outbreaks that resulted in the highest number of deaths (Q4), 12 were in the top 2 quantiles in terms of risk score (Q3/Q4). However, the other four were in the lower two quantiles of risk score. Of the 15 outbreaks in our dataset that resulted in the least number of deaths (Q1), 11 were in the bottom two quantiles in terms of risk (Q1/Q2) and four were in the top quantiles of risk (Q3/Q4). An equally promising pattern is shown for outbreaks in Q2 and Q3.

There is a clear association between the risk scores computed by our framework and the severity of outbreaks as measured by the number of deaths, but there are also outbreaks that fall outside of the risk categories one would wish to see. For example, one of the most severe outbreaks was in the bottom 25% of risk scores, and three of the least severe outbreaks were in the highest 25% of risk scores. Nonetheless, the overall association exists, and in future work, we intend to make the relationship stronger.
score. For example, an outbreak of Lassa fever occurring in Nigeria at the same time as other infectious diseases such as polio and cholera could also assist the government in prioritising urgent interventions to Lassa fever because it has higher risk disease and therefore minimise the overall impact. Such a tool may assist with prioritisation and early decision making. It may also help international organisations to prioritise countries in most need of urgent assistance.28

A limitation of this study is the dependence on the quality of the data inputs, such as the validity of the data sources, and the use of fixed disease parameters of known pathogens. The data used to calculate the risk scores were collected from various sources and from different points in time, which in some cases are out of date or contain biases. On an evaluation set of 61 outbreaks, our results show that the risk score computed using our framework has a positive association with outbreak severity as measured by number of deaths associated with an outbreak. In future work, we intend to improve the model to make this association stronger. At the moment, our model is a simple linear summation of risk factors. We intend to develop a larger dataset of historical outbreaks, at which point it will become viable to develop individual weights for the factors in the model that we hope will improve the strength of the association between risk score and outbreak severity. It should also be noted that the risk categories (and their cutoffs) do lead to a situation where a unit difference in risk can change the risk category entirely. The framework is also currently limited to known pathogens. This limits the tools application during the early moments of an outbreak where the pathogen maybe unknown and awaiting diagnostic confirmation. This also includes the emergence of novel pathogens, where many disease parameters required for input such as the basic reproductive number \( R_0 \), mode of transmission and case fatality rate are likely to be uncertain. However, in the case of COVID-19 as an example, as the pandemic progressed, these parameters quickly became known. Several of the high-scoring features outlined in table 1, such as asymptomatic transmission, respiratory spread, viral infection and high case fatality rate, were present.44 45 Future iterations of the tool could allow the input of these disease parameters manually, and various potential scenarios could be tested by varying the parameters individually, such as variations of \( R_0 \) to account for uncertainty early in an epidemic. The framework is meant to be used as a whole, where the user ought to look at the risk score, the disease score, the country score and then the individual risk factor scores for deeper insight. Despite these limitations, EpiRisk provides an individualised approach where specific input data for a country and disease can be used, rather than a standard ‘one size fits all’ approach that would be even less generalisable. We believe this country-level approach is more rigorous and can predict outbreak risk more accurately than relying on disease or country factors in isolation.

The development of a simple risk analysis tool will be useful for global epidemic control. We have demonstrated in principle that EpiRisk can assess the level of epidemic risk for individual epidemics and performed well when tested against an initial set of real-world epidemics. This tool can be used to rapidly predict the risk of outbreaks that is useful when planning and prioritising interventions or for epidemic preparedness. An appropriate and timely intervention can help governments and public health professionals prevent catastrophic outcomes. To improve the relevancy for current or future prediction, the development of a real-time tool that provides the most current data for risk prediction is crucial. The involvement of local government and other health organisation to improve the data is also important. In future work, we will evaluate the generalisability of EpiRisk through the use of a more comprehensive test set of outbreaks.

**Contributors** DAS: data analysis, interpretation, study design, drafting and revision of the manuscript; PV: data analysis, interpretation, study design and revision of the manuscript; AM and DCA: interpretation, drafting and revision of the manuscript; CRM: conception of the study, interpretation, study design and revision of the manuscript.

**Funding** This work was supported by a grant from the National Health and Medical Research Council (NHMRC) Centre for Research Excellence in Integrated Systems for Epidemic Response (grant number 1107393). CRM is supported by a NHMRC Principal Research Fellowship, grant number 1137582.

**Competing interests** CRM has received funding for investigator-driven research from Sequris and Sanofi unrelated to this study. CRM has also been on advisory boards for the same companies.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data collected for the purpose of this study are publicly available. The EpiRisk Tool itself will be made available via an online system in the future.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of changes were made. See: https://creativecommons.org/licenses/by/4.0/.

**ORCID iD** Dillon C Adam http://orcid.org/0000-0002-7485-9905

**REFERENCES**

1 Davey B. Public health approaches to infectious disease, 2016.
10


9 Dyda A, Stelzer-Braides A, Adam D, et al. The association between acute flaccid myelitis (AFM) and enterovirus D68 (EV-D68) - what is the evidence for causation? Euro Surveill 2018;23.


26 Health Nlo. NIH guidelines for research involving recombinant DNA molecules (NIH guidelines), 2002.


39 International Society for Infectious Diseases. About ProMED-mail, 2010.

40 HealMap. About HealMap: Boston Children’s Hospital, 2015.


54 World Health Organization. Health workforce requirements for universal health coverage and the sustainable development goals (human resources for health observer), 2016.


