The threat of serious, cross-border communicable disease outbreaks in Europe poses a significant challenge to public health and emergency preparedness because the relative likelihood of these threats and the pathogens involved are constantly shifting in response to a range of changing disease drivers. To inform strategic planning by enabling effective resource allocation to manage the consequences of communicable disease outbreaks, it is useful to be able to rank and prioritise pathogens. This paper reports on a literature review which identifies and evaluates the range of methods used for risk ranking. Searches were performed across biomedical and grey literature databases, supplemented by reference harvesting and citation tracking. Studies were selected using transparent inclusion criteria and underwent quality appraisal using a bespoke checklist based on the AGREE II criteria. Seventeen studies were included in the review, covering five methodologies. A narrative analysis of the selected studies suggests that no single methodology was superior. However, many of the methods shared common components, around which a ‘best-practice’ framework was formulated. This approach is intended to help inform decision makers’ choice of an appropriate risk-ranking study design.

Introduction

Communicable disease outbreaks can pose a significant challenge to public health and to emergency preparedness. Types of threats and the pathogens involved shift in relation to changing factors such as climate change [1,2], global travel and trade [3,4], immigration patterns, urban sprawl, social inequalities [5,6] and other disease drivers [7-10]. An increasingly interconnected world means that diseases emerging in one part of the world, such as Zika, Middle East respiratory syndrome coronavirus or Ebola [11,12] can spread globally. Similarly, diseases once considered tropical can transmit in Europe under the right circumstances [8,13-15].

It is essential for public health agencies to be able to account for and assess the rapidly changing global context surrounding communicable disease. One of the Core Capacity Indicators of the International Health Regulations relates to mapping and using priority health risks and resources [16]. This includes conducting national risk assessments for identifying potential ‘urgent public health events’ as well as the most likely source of these events [16]. At the European level, Article 4 of the European Parliament and Council Decision 1082/2013/EU on serious cross-border threats to health focuses on preparedness and response planning, calling for ‘efforts to develop, strengthen and maintain…capacities for the monitoring, early warning and assessment of, and response to, serious cross-border threats to health’ [17].

Identifying and prioritising risks are a necessary first phase for informing the public health response to infectious disease risks, and an effective tool to guide strategic planning and ensure the efficient allocation of resources [18]. The need for methodologies to assist national efforts in this area was highlighted at a Joint European Centre for Disease Prevention and Control (ECDC)-World Health Organization (WHO) Consultation on Pandemic and All-Hazard Preparedness, held in Bratislava in November 2013 [19]. Elsewhere, the development of risk-ranking ‘toolboxes’ has been advocated, which could enable organisations to decide on the best methodologies that are commensurate with ranking exercises [20].

ECDC aims to develop a comprehensive risk-ranking tool for use in strategic prioritisation exercises. There is, however, no current consensus on the best methodology for such risk-ranking exercises, with different organisations proposing different methods. WHO, for example, has produced practical guidance on setting priorities in infectious disease surveillance, advocating a Delphi methodology [21]. Other studies have
The citation pearl-growing method [25] was used to identify search terms using an initial sample of relevant articles (identified in a scoping search [22]). Searching relevant articles (identified in a scoping search [22]). Searches were performed across biomedical databases (Medline, Embase, Cochrane Library and Centre for Reviews and Dissemination), grey literature (i.e. official documents, non-peer reviewed reports, etc.) and specialist databases (Google Advanced Search, WHO, the World Bank). Subject headings (where available) and variations on search terms related to prioritisation or ranking, were combined with ‘communicable’ or ‘infectious’ or ‘zoonoses’ to search the various sources. Supplemental search techniques of reference harvesting and citation tracking were performed for the initial sample of relevant articles and again for all articles included in the analysis [26].

The searches are not fully exhaustive, although the three-pronged approach is designed to capture the most relevant literature. Studies included in the analysis are presented in Table 1.

Quality appraisal
The aim of the quality appraisal was to evaluate the validity and reliability of individual studies, to enable comparison between individual studies and across different methodologies. No existing checklist was suitable for assessing quality across the different methodologies used in the studies included, and so a quality appraisal checklist was developed [22]. The bespoke appraisal checklist was based on the Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument criteria [27], which evaluates the methodology and reporting of guidelines. The checklist assessed the validity (how well the method measured the important facets of communicable disease) and reliability (internal consistency, inter-rater consistency and precision of the method) of the risk-ranking studies. A sample of quality appraisals was separately appraised by two reviewers to test the checklist and establish rating definitions. Studies were rated according to this set of criteria, and then given an overall rating (Table 2). The qualitative Likert assessments, which are based upon scales that typically range from ‘strongly disagree’ to ‘strongly agree’, are represented using a red-amber-green ‘traffic light rating system’ (with red indicating a high risk of bias likely). Where multiple articles described the same risk-ranking exercise, articles were appraised and extracted as one study, but counted individually within the flowchart (Figure 1) [28-32].

Methods
The project methodology comprised two key phases. First, a literature review to identify the relevant literature on risk ranking for communicable diseases was conducted. Second, the findings from this review were analysed through a narrative review, which enabled the development of a best-practices framework for ranking infectious disease threats.

Literature review
The scope of this literature review included all communicable diseases, which are defined according to the European Union (EU) list of communicable diseases for surveillance [23]. For the purposes of this review, risk was defined according to the International Organization for Standardization (ISO) standards with risk being the product of impact and likelihood [24].

Searching
The citation pearl-growing method [25] was used to identify search terms using an initial sample of relevant articles (identified in a scoping search [22]). Searches were performed across biomedical databases (Medline, Embase, Cochrane Library and Centre for Reviews and Dissemination), grey literature (i.e. official documents, non-peer reviewed reports, etc.) and specialist databases (Google Advanced Search, WHO, the World Bank). Subject headings (where available) and variations on search terms related to prioritisation or ranking, were combined with ‘communicable’ or ‘infectious’ or ‘zoonoses’ to search the various sources. Supplemental search techniques of reference harvesting and citation tracking were performed for the initial sample of relevant articles and again for all articles included in the analysis [26].

The search and sift process is presented in Figure 1. The searches are not fully exhaustive, although the three-pronged approach is designed to capture the most relevant literature. Studies included in the analysis are presented in Table 1.

Quality appraisal
The aim of the quality appraisal was to evaluate the validity and reliability of individual studies, to enable comparison between individual studies and across different methodologies. No existing checklist was suitable for assessing quality across the different methodologies used in the studies included, and so a quality appraisal checklist was developed [22]. The bespoke appraisal checklist was based on the Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument criteria [27], which evaluates the methodology and reporting of guidelines. The checklist assessed the validity (how well the method measured the important facets of communicable disease) and reliability (internal consistency, inter-rater consistency and precision of the method) of the risk-ranking studies. A sample of quality appraisals was separately appraised by two reviewers to test the checklist and establish rating definitions. Studies were rated according to this set of criteria, and then given an overall rating (Table 2). The qualitative Likert assessments, which are based upon scales that typically range from ‘strongly disagree’ to ‘strongly agree’, are represented using a red-amber-green ‘traffic light rating system’ (with red indicating a high risk of bias likely). Where multiple articles described the same risk-ranking exercise, articles were appraised and extracted as one study, but counted individually within the flowchart (Figure 1) [28-32].

In order to identify best practices in risk ranking, and to guide further ECDC work in this area, a literature review was initiated to identify and evaluate the range of methods used [22]. The findings from this review were then used to develop a best-practice framework for ranking infectious disease threats.

Figure 1
Flowchart of search and sifting process, literature review on best practices in ranking communicable disease threats, 2015

The searches are not fully exhaustive, although the three-pronged approach is designed to capture the most relevant literature. Studies included in the analysis are presented in Table 1.

Quality appraisal
The aim of the quality appraisal was to evaluate the validity and reliability of individual studies, to enable comparison between individual studies and across different methodologies. No existing checklist was suitable for assessing quality across the different methodologies used in the studies included, and so a quality appraisal checklist was developed [22]. The bespoke appraisal checklist was based on the Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument criteria [27], which evaluates the methodology and reporting of guidelines. The checklist assessed the validity (how well the method measured the important facets of communicable disease) and reliability (internal consistency, inter-rater consistency and precision of the method) of the risk-ranking studies. A sample of quality appraisals was separately appraised by two reviewers to test the checklist and establish rating definitions. Studies were rated according to this set of criteria, and then given an overall rating (Table 2). The qualitative Likert assessments, which are based upon scales that typically range from ‘strongly disagree’ to ‘strongly agree’, are represented using a red-amber-green ‘traffic light rating system’ (with red indicating a high risk of bias likely). Where multiple articles described the same risk-ranking exercise, articles were appraised and extracted as one study, but counted individually within the flowchart (Figure 1) [28-32].

In order to identify best practices in risk ranking, and to guide further ECDC work in this area, a literature review was initiated to identify and evaluate the range of methods used [22]. The findings from this review were then used to develop a best-practice framework for ranking infectious disease threats.

Methods
The project methodology comprised two key phases. First, a literature review to identify the relevant literature on risk ranking for communicable diseases was conducted. Second, the findings from this review were analysed through a narrative review, which enabled the development of a best-practices framework for ranking infectious disease threats.

Literature review
The scope of this literature review included all communicable diseases, which are defined according to the European Union (EU) list of communicable diseases for surveillance [23]. For the purposes of this review, risk was defined according to the International Organization for Standardization (ISO) standards with risk being the product of impact and likelihood [24].

Searching
The citation pearl-growing method [25] was used to identify search terms using an initial sample of relevant articles (identified in a scoping search [22]). Searches were performed across biomedical databases (Medline, Embase, Cochrane Library and Centre for Reviews and Dissemination), grey literature (i.e. official documents, non-peer reviewed reports, etc.) and specialist databases (Google Advanced Search, WHO, the World Bank). Subject headings (where available) and variations on search terms related to prioritisation or ranking, were combined with ‘communicable’ or ‘infectious’ or ‘zoonoses’ to search the various sources. Supplemental search techniques of reference harvesting and citation tracking were performed for the initial sample of relevant articles and again for all articles included in the analysis [26].

The search and sift process is presented in Figure 1. The searches are not fully exhaustive, although the three-pronged approach is designed to capture the most relevant literature. Studies included in the analysis are presented in Table 1.

Quality appraisal
The aim of the quality appraisal was to evaluate the validity and reliability of individual studies, to enable comparison between individual studies and across different methodologies. No existing checklist was suitable for assessing quality across the different methodologies used in the studies included, and so a quality appraisal checklist was developed [22]. The bespoke appraisal checklist was based on the Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument criteria [27], which evaluates the methodology and reporting of guidelines. The checklist assessed the validity (how well the method measured the important facets of communicable disease) and reliability (internal consistency, inter-rater consistency and precision of the method) of the risk-ranking studies. A sample of quality appraisals was separately appraised by two reviewers to test the checklist and establish rating definitions. Studies were rated according to this set of criteria, and then given an overall rating (Table 2). The qualitative Likert assessments, which are based upon scales that typically range from ‘strongly disagree’ to ‘strongly agree’, are represented using a red-amber-green ‘traffic light rating system’ (with red indicating a high risk of bias likely). Where multiple articles described the same risk-ranking exercise, articles were appraised and extracted as one study, but counted individually within the flowchart (Figure 1) [28-32].
Analysis of best practices in risk ranking
A standardised data-extraction form was used to extract key methodological information. Data extraction was performed in duplicate by two researchers. A narrative synthesis was performed by clustering the studies according to methodology, to compare studies within and across methodologies. The narrative review indicated that no single methodology was superior, but many of the methods shared common components. Therefore a best-practice framework was formulated, structured around the common components identified in the narrative review, which worked across the reviewed methodologies (Figure 2). The best-practice framework is designed to inform decision makers’ choice of an appropriate risk-ranking method and ensure that methodologies are carried out according to best practice.

Results

Results from the literature review
Fourteen studies, reported in 17 articles, were selected for inclusion in the review. The studies used one of five methodologies to rank communicable disease risks: bibliometric index [33,34], the Delphi technique [35-38], Multi-Criteria Decision Analysis (MCDA) [31,32,39-41], qualitative algorithms [42,43], and questionnaires [29-31,45] In general, risk-ranking exercises begin with identifying diseases to consider for prioritisation, formulating a list of criteria to assess diseases against, then weighting the criteria according to importance, and scoring diseases against the criteria to create a ranking based on the scores.

Analysis of best practices
Based on the analysis of the studies reviewed, it was possible to comment upon best practice in conducting risk-ranking exercises independent of the methodology selected and based on the steps within this generic process.

This paper focuses on the best-practice framework (Figure 2), which has the overall aim of reducing bias and strengthening the credibility and reproducibility of findings, whichever methodology is used. Some aspects of best practice run across the different steps in the framework, such as using a multidisciplinary team.

Planning
WHO guidance on priority setting in communicable disease surveillance states that planning is an essential step in the process [21]. Establishing the objectives of the exercise enables the selection of an appropriate
<table>
<thead>
<tr>
<th>Study</th>
<th>Methodology</th>
<th>Summary</th>
</tr>
</thead>
</table>
| Cox et al. [33] | Bibliometrics (h-index) | 651 diseases ranked  
Primary source: Web of Science  
Validating source: Pubmed |
| McIntyre et al. [34] | Bibliometrics (h-index) | 1,414 diseases ranked  
Primary source: Web of Science  
Validating sources: Google Scholar, Scopus |
| Balabanova et al. [35] | Delphi study | 127 diseases ranked  
10 criteria used  
Criteria weighted  
86 participants weighted criteria  
20 participants scored diseases  
3 point scale used to score diseases  
1 round of Delphi scoring |
| Economopoulou et al. [36] | Delphi study | 71 diseases ranked  
2 criteria used  
Criteria not weighted  
3 participants scored diseases  
56 participants selectively scored diseases  
5 point scale used to score diseases  
2 rounds of Delphi scoring |
| Krause et al. [37] | Delphi study | 85 diseases ranked  
12 criteria used  
Criteria weighted  
11 participants weighted criteria  
11 participants scored diseases  
3 point scale used to score diseases  
1 round of Delphi scoring |
| WHO et al. [38] | Delphi study | 53 diseases ranked  
8 criteria used  
Criteria not weighted  
24 participants scored diseases  
5 point scale used to score diseases  
1 round of Delphi scoring |
| Cardoen et al. [39] | Multi-criteria decision analysis | 51 diseases ranked  
5 criteria used  
Criteria weighted using Las Vegas method  
7 participants weighted criteria  
35 participants scored diseases  
Scores of 0–4 points allocated to each disease (based on occurrence and severity) |
| Cox et al. [31,32] | Multi-criteria decision analysis | 9 diseases ranked  
40 criteria used  
Criteria weighted using a qualitative Likert scale (based on likelihood or importance)  
64 participants weighted criteria  
47 participants scored diseases  
Likert scale used to score diseases |
| Havelaar et al. [40] | Multi-criteria decision analysis | 86 diseases ranked  
7 criteria used  
Criteria weighted using relative ranking  
29 participants  
Quantitative, scaled values used to score diseases |
| Humblet et al. [41] | Multi-criteria decision analysis | 100 diseases ranked  
57 criteria (in 5 categories)  
40 participants  
Criteria weighted using the Las Vegas method  
Co-efficients of 0–7 points assigned to each option |
| Morgan et al. [42] | Qualitative algorithm | 1 disease ranked (a worked example)  
1 participant |
| Palmer et al. [43] | Qualitative algorithm | 5 diseases ranked  
Number of participants unclear |
| Horby et al. [44] | Questionnaire studies | 61 diseases ranked  
5 criteria used  
Criteria not weighted  
518 participants |
| Ng et al. [28-30] | Questionnaire studies | 62 diseases ranked  
21 criteria used  
Criteria weighted using conjoint analysis  
4,161 participants |
<table>
<thead>
<tr>
<th>Study</th>
<th>Methodology</th>
<th>Overall score</th>
<th>Individual domain scores</th>
<th>Reviewer comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balabanova et al.</td>
<td>Delphi</td>
<td></td>
<td></td>
<td>Sources of bias were identified and mitigated where possible. Implementation issues were not discussed. The criteria used in the study did not meet all of the content validity criteria. Unclear what measures were in place to ensure internal consistency and whether any tests of validity were used.</td>
</tr>
<tr>
<td>Cardoen et al. [39]</td>
<td>Semi-quantitative methodology (analysed as multi-criteria decision analysis)</td>
<td></td>
<td></td>
<td>Unclear how criteria were developed. Implementation issues were not discussed. Either did not meet or only partly met several of the key communicable disease facets. No measures of internal consistency.</td>
</tr>
<tr>
<td>Cox et al. [31,32]</td>
<td>Multi-criteria decision analysis</td>
<td></td>
<td></td>
<td>Unclear precisely how criteria were developed. Implementation issues were not discussed. Criteria met most of the key communicable disease facets. Sensitivity analyses were used to test validity.</td>
</tr>
<tr>
<td>Cox et al.</td>
<td>Bibliometric index</td>
<td></td>
<td></td>
<td>Assessment is based on applicable criteria. This paper did not address any of the key communicable disease facets due to its design. The quality of evidence was not considered. Tested validity by comparing two data sources using Spearman’s rank test.</td>
</tr>
<tr>
<td>Havelaar et al.</td>
<td>Multi-criteria decision analysis</td>
<td></td>
<td></td>
<td>Unclear how criteria were chosen. Implementation issues were not fully discussed. Did not meet all of the key communicable disease facets, in particular it did not address mitigation. Participants were sent a repeated exercise to test internal consistency. A sensitivity analysis tested the validity of assumptions made in the different models.</td>
</tr>
<tr>
<td>Horby et al. [44]</td>
<td>Questionnaire</td>
<td></td>
<td></td>
<td>Unclear exactly how criteria were chosen, but they are compared against similar studies. Implementation issues were not discussed. Did not meet all of the key communicable disease facets, across likelihood, impact and mitigation. No tests for internal consistency, although tests to measure variation between professional groups were undertaken.</td>
</tr>
<tr>
<td>Humblet et al.</td>
<td>Multi-criteria decision analysis</td>
<td></td>
<td></td>
<td>Addresses some practical issues by stating that their intended methodology was Delphi but they did not have sufficient time. Did not meet all of the key communicable disease facets, but did consider the cost of prevention. No measures of internal consistency, but criteria definitions included to reduce inter-rater variation. Used a probabilistic method to account for variability in scores.</td>
</tr>
<tr>
<td>Krause et al. [37]</td>
<td>Delphi</td>
<td></td>
<td></td>
<td>Implementation issues were not discussed, although practical considerations were included. Did not meet all key communicable disease criteria. Did not measure internal consistency, but results were reviewed by all participants for plausibility. Criteria and scoring definitions were provided to reduce inter-rater variation.</td>
</tr>
<tr>
<td>McIntyre et al.</td>
<td>Bibliometric index</td>
<td></td>
<td></td>
<td>Assessment is based on applicable criteria. This paper did not address any of the key communicable disease facets due to its design. The quality of evidence was not considered. Tested validity by comparing two data sources using Spearman’s rank test. Authors acknowledge the limitations of the methodology.</td>
</tr>
<tr>
<td>Morgan et al. [42]</td>
<td>Qualitative algorithm</td>
<td></td>
<td></td>
<td>It is unclear how this qualitative algorithm was developed, therefore judging the risk of bias was challenging. Implementation issues were not discussed. Questions within the algorithm addressed some of the key communicable disease facets. There were no measures of internal consistency. The algorithm was completed by a single scientist.</td>
</tr>
<tr>
<td>Ng et al. [28-30]</td>
<td>Questionnaire</td>
<td></td>
<td></td>
<td>Implementation issues were not specifically discussed, but practical considerations were discussed which would assist implementation. Most of the key communicable disease criteria were not applicable as this is an early-stage risk assessment.</td>
</tr>
<tr>
<td>Palmer et al.</td>
<td>Qualitative algorithm</td>
<td></td>
<td></td>
<td>It is unclear how this qualitative algorithm was developed, with most validity criteria partly met or not met. Implementation issues were not discussed. Many key communicable disease criteria were not applicable as this is an early-stage risk assessment.</td>
</tr>
<tr>
<td>WHO et al. [38]</td>
<td>Delphi</td>
<td></td>
<td></td>
<td>Reporting lacking detail, as it was a report of a meeting to give participants experience of such an exercise. Unclear how criteria were developed. Potential sources of bias and mitigations are not reported. The publication was not peer-reviewed and it is unclear if any other review took place. Implementation issues were not discussed but Delphi scoring was limited to one round. Did not meet all of the key communicable disease facets, 95% confidence intervals used to aid discussion of discrepancies in scoring.</td>
</tr>
</tbody>
</table>

Green: criteria met, information related to that item has been clearly reported and all relevant considerations have been made.
Amber: criteria partly met, information related to that item is incomplete, or not all aspects have been considered.
Red: criteria not met, no information provided in the study that is relevant to that item, or information related to that item is very poorly reported.
NA: criteria are not applicable.
methodology that is fit for purpose. All of the methodologies reviewed can be adapted to suit the particular context and requirements of a risk-ranking exercise. Although many of the studies described the objectives of the risk-ranking exercise, they did not provide details of the planning process. Table 3 describes some scenarios in which a risk-ranking exercise might take place, with suggestions for which methodology may be most suited to meet those needs, with a rationale based on the full comparison between and across methodologies from the ECDC technical report [22].

The decision about whether to use qualitative, quantitative or mixed methods should be based on the scope and purpose of the exercise as established during the planning phase. The included studies often provided explanations for their choice of methodology in terms of overcoming or balancing the potential limitations of alternative methodologies, but rarely explained their choice of methods with regards to the specific objectives of their risk-ranking exercise. Five of the reviewed studies used a quantitative methodology [33,35,37,40,41], three used qualitative approaches [36,42,43], and six studies used semiquantitative, mixed methods [28,32,35,38,39,44]. Only four studies used either entirely qualitative or quantitative methods [33,34,42,43], however, these studies were considered by their authors to be most useful as part of a wider risk-ranking exercise rather than as a stand-alone methodology. No comprehensive methodology using only qualitative or quantitative methods was identified in this review.

There are advantages and disadvantages to using quantitative or qualitative methods in different scenarios. For example, in areas where there is little evidence (and what does exist is of poor quality) it may be preferable to use semiquantitative methods (to make best use of the evidence available [39]) or qualitative methods (in recognition that the evidence is not of much help and uncertainty remains [31]). Qualitative data generally takes longer to collect and analyse than quantitative data, although it provides a richness and context to responses that quantitative data cannot. Semiquantitative methods where quantitative data is used allow for documentation of uncertainty (and what does exist is of poor quality) it may be preferable to use semiquantitative methods (to make best use of the evidence available [39]) or qualitative methods (in recognition that the evidence is not of much help and uncertainty remains [31]). Qualitative data generally takes longer to collect and analyse than quantitative data, although it provides a richness and context to responses that quantitative data cannot. Semiquantitative methods where quantitative data is used allow for documentation of uncertainty.

WHO guidance on priority setting in communicable disease surveillance recommends that the planning process includes budgeting, covering all resources required for the ranking exercise [21]. An assessment of the resources required for any of these methods is an important part of the decision-making process. Methods requiring greater resources should not necessarily be disregarded, but the resources required for a risk-ranking exercise affects its feasibility and potentially creates barriers to the study’s application by practitioners. Thus, detailed plans should consider resources required at all stages, from the commissioners of the ranking and the deadline for delivery, to the time requirement for each participant in the process.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Methodology</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapid or large-scale risk ranking for large number of pathogens</td>
<td>H-index or qualitative algorithm</td>
<td>Both methods are suitable for ranking a large volume of pathogens within a short time period or with limited resources.</td>
</tr>
<tr>
<td>Scoping exercise to generate an initial ranking for further study</td>
<td>H-index or qualitative algorithm</td>
<td>As both methods can quickly rank a large volume of pathogens, they can be used to provide a short list for risk ranking using a more comprehensive technique.</td>
</tr>
<tr>
<td>Comprehensive risk ranking including novel, emerging and established infections</td>
<td>Multi-criteria decision analysis or Delphi, H-index</td>
<td>Both methods provide a comprehensive method for risk ranking. Where resource is restricted, consider limiting the number of criteria or the number of diseases for ranking.</td>
</tr>
<tr>
<td>Emerging infections with little published data about them</td>
<td>Qualitative algorithm</td>
<td>In qualitative methodologies, including a mechanism for respondents to identify gaps in knowledge or areas for further work could lead to improved evidence upon which to base future decisions.</td>
</tr>
<tr>
<td></td>
<td>Qualitative algorithm or questionnaires</td>
<td>In qualitative methodologies, including a mechanism for respondents to identify gaps in knowledge or areas for further work could lead to improved evidence upon which to base future decisions.</td>
</tr>
<tr>
<td></td>
<td>Multi-criteria decision analysis</td>
<td>This method can incorporate information from a variety of sources, which is useful in emerging infections where information is sparse. Ranking the risk of alternative scenarios is suitable for situations where there is less certainty about the potential course of the disease. Additionally new information can be incorporated as it emerges, without needing to re-run the entire ranking exercise</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Methodology</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapid or large-scale risk ranking for large number of pathogens</td>
<td>H-index or qualitative algorithm</td>
<td>Both methods are suitable for ranking a large volume of pathogens within a short time period or with limited resources.</td>
</tr>
<tr>
<td>Scoping exercise to generate an initial ranking for further study</td>
<td>H-index or qualitative algorithm</td>
<td>As both methods can quickly rank a large volume of pathogens, they can be used to provide a short list for risk ranking using a more comprehensive technique.</td>
</tr>
<tr>
<td>Comprehensive risk ranking including novel, emerging and established infections</td>
<td>Multi-criteria decision analysis or Delphi, H-index</td>
<td>Both methods provide a comprehensive method for risk ranking. Where resource is restricted, consider limiting the number of criteria or the number of diseases for ranking.</td>
</tr>
<tr>
<td>Emerging infections with little published data about them</td>
<td>Qualitative algorithm</td>
<td>In qualitative methodologies, including a mechanism for respondents to identify gaps in knowledge or areas for further work could lead to improved evidence upon which to base future decisions.</td>
</tr>
<tr>
<td></td>
<td>Qualitative algorithm or questionnaires</td>
<td>In qualitative methodologies, including a mechanism for respondents to identify gaps in knowledge or areas for further work could lead to improved evidence upon which to base future decisions.</td>
</tr>
<tr>
<td></td>
<td>Multi-criteria decision analysis</td>
<td>This method can incorporate information from a variety of sources, which is useful in emerging infections where information is sparse. Ranking the risk of alternative scenarios is suitable for situations where there is less certainty about the potential course of the disease. Additionally new information can be incorporated as it emerges, without needing to re-run the entire ranking exercise</td>
</tr>
</tbody>
</table>

---

www.eurosurveillance.org
to deliver the ranking. The methods used in any risk-ranking exercise can be adapted to the resource available. For example, where resources are limited the number of criteria can be limited to increase the number of pathogens that can be assessed [37]. There is always the need to balance methodological rigour and real-world practicalities. However, the reliability and validity of the methodology affects the reliability and validity of the output, and therefore whether it will be taken heed of [37].

**Identify diseases for prioritisation**

Most of the included studies (14 out of 17) described methods used for identifying and selecting diseases for risk ranking. Studies generally used existing surveillance systems to identify diseases and many used notifiable status as one of their selection criteria. Some studies also asked experts to contribute to the list of diseases for ranking, either by suggesting diseases or by commenting on a pre-formulated list. While the reviewed studies reported the method of disease selection, the rationale was generally not detailed and the potential limitations of the method were not explored. For example, using sources such as notifiable disease lists that are based on clinical and laboratory data, combined with suspected risk, would not necessarily be suitable for identifying emerging threats.

**Formulate a list of criteria to assess diseases against and weight criteria according to importance**

The criteria considered in the studies varied. However, there was a common core of key communicable disease concepts such as how easily the disease could be spread, how reliable diagnostic testing is, the treatability of the disease, impact on school and work absenteeism, and on-going illness resulting from infection. The average number of criteria was 17. The selected criteria should be specific to the context of the exercise (e.g. specific to the purpose of the exercise, the country where it is taking place): for example some studies considered the role of public concern/perception whereas many did not. Preventive measures currently in place (e.g. vaccinations) should also be considered as criteria so that diseases with low incidence due to effective control measures are not deprioritised and risk resources being allocated elsewhere [35]. The studies that weighted criteria according to importance did so using expert opinion, which creates potential subjectivity and inconsistency in weightings. Including clear definitions of criteria can help to reduce this potential bias [37,41]. Weighting can be assigned to criteria using different methods such as the Las Vegas method [45], allocating differing numbers of points to criteria, or simple relative ranking.

One study engaged members of the public, but they were included only in the initial focus groups to identify and weight criteria [28-30].

**Score diseases against the criteria**

Most studies scored diseases based on expert opinion, except for the qualitative algorithms [42,44], which provided a relative ranking, and the studies using h-index scores to rank diseases [33,34]. The incorporation of expert opinion in 13 out of 17 studies suggests that it provides a unique input that would be otherwise missing from risk-ranking exercises. The average number of experts included was 231; however, there were some outliers and so the median value of 59 (interquartile ratio: 45) may be a more useful indication. None of the included studies described how they assessed whether sufficient numbers of participants were included, and therefore it would be helpful if future studies indicated how their sample sizes were determined. Most of the studies reported how their participants were selected, which provides useful information for those seeking to apply these methods to their own setting. Multidisciplinary input based on expertise and experience can help to inform decisions where standard data are not available, such as in the case of emerging disease threats or areas with great evidential uncertainty. The variability and subjectivity of scoring decisions between individuals and between different professional groups is a potential source of bias in risk ranking. While expert input introduces potential bias, it is needed where clear quantitative metrics are not available or where they are not easily comparable. Measures can be put in place to mitigate these risks, such as clear explanations of criteria and definitions of scores to reduce inter-rater variation and interdisciplinary discussion of scores [35-37]. Formal statistical methods, such as Kappa scores, can be used to measure variation between individuals and professional groups, and appropriate adjustments can be made if the variation is considered too high. Alternatively, allowing participants to qualitatively explain their scores could be useful to assess potential causes of variation. Incorporating a method whereby participants can express uncertainty in their scoring can help to understand the rationale behind responses, identify where expert opinion disagrees with current evidence or identify areas for further research [39,44]. When incorporating expert opinion into any methodology, it is necessary to consider the representativeness of the people whose opinion is sought and, as the reviewed studies did, engage a range of multidisciplinary specialists to cover the different aspects of communicable disease risk ranking. There can be conflict between the desire to engage a variety of participants and the need to ensure that those participants are making informed decisions. This risk can be mitigated, for example by allowing respondents to acknowledge the limits of their knowledge [39], or using qualitative scales or visual representations to aid participants in interpreting otherwise abstract scores [31,32].

Five studies provided participants with evidence to support their decision-making. This evidence was collated from reliable sources such as national governments, supranational organisations (such as the EU), NGOs
(such as WHO), and charities. Providing such evidence could be interpreted as prejudicing the impartiality of the decision-making by providing information to help steer responses. However, providing evidence may help to reduce subjectivity, reduce bias (individual or professional), correct misconceptions and ensure that participants are making decisions based on reliable, up-to-date information that is relevant to the purpose of the exercise. All tools, regardless of methodology, are reliant on the quality and availability of evidence upon which to base judgments. Morgan incorporated references of the evidence used in decision-making into their qualitative algorithm [42], so that the basis of the decision could be understood and scrutinised. Decision-making should record the evidence upon which it is based, the quality of that evidence and whether any evidence gaps exist.

Rank diseases based on relative scores
Some studies reported that an indication of overall trend [44] or relative ranking was more informative than the raw individual scores of pathogens [37,40]. Various mathematical techniques were used to combine scores, depending on the methodology. As with other steps in the process it is necessary to clearly communicate the process from scoring pathogens against weighted criteria to ensure transparency and reproducibility of the method.

Evaluation
The studies included did not provide information on an evaluation of the effectiveness of the process and its output. Krause stated that the current exercise was based on experience of a previous exercise and would be further refined in future [37]. WHO guidance emphasises the role of risk-ranking exercises in the evaluation of surveillance measures and places it within a process cycle, which includes evaluation [21]. Evaluation is included in the best-practice framework, despite not being explicitly included in the reviewed studies, because it is recommended in WHO guidance [21] and is generally considered central to implementing and improving new processes. Using a process improvement cycle such as ‘Plan, Do, Study, Act’ (PDSA) [46] provides a framework for evaluating the process, comparing the rankings with actual events and enabling process improvements that can be implemented when the exercise is repeated.

Re-run the risk-ranking exercise
Placing risk-ranking exercises within a process-improvement cycle such as PDSA [46] assists in the evaluation of the process and its outcomes, but also emphasises the need to repeat the risk-ranking exercise. Krause et al. state that the experience of the current risk-ranking exercise will inform future exercises [37]. However, none of the studies lay out specific timescales or triggers for the risk-ranking exercise to be repeated. As such it is not possible to derive specific best practice in this area. However, as part of a cycle of activities, risk-ranking exercises should be re-run periodically (every five years), depending on an assessment of the extent to which the various disease drivers have changed. It is also necessary to consider triggers – such as evidence of emerging threats, the development of new interventions or new surveillance intelligence for current threats [42] – that could cue a re-ranking of diseases. In such cases it may be possible to perform an interim and rapid assessment before the next scheduled risk-ranking exercise is due.

Resource requirements
Not all studies reviewed here included information about the time and human and financial resources involved in the risk-ranking exercise. Such practical information would inform the choice of methodology and also any pragmatic modifications (such as reducing the number of pathogen included) that might be made to make the exercise viable. Some studies alluded to their method being time-consuming [35,37] or that time constraints required them to adapt their methodology or switch to another method [28-30,38,41]. General discussion of how methods can be adapted to suit time or resource constraints were discussed in some papers [31,32,36,37,40], such as reducing the number of diseases considered to allow for a larger Delphi panel [37]. One study provided data on how long the survey, which was one part of the exercise, took for participants to complete (27 min in Canada and 28 min in the US) [28-30]. In addition to the time of staff and participants, resources such as specialist software [28-30], staff training (e.g. in software or statistical methods) or outside costs (e.g. using a firm to recruit participants, hiring external skills such as focus group facilitators) were not reported.

Discussion
Predicting the future risk of communicable diseases is challenging as there are many changing factors and unknowns [1,7,18,47]. This literature review aimed to identify and evaluate the range of methods available for risk ranking of communicable diseases. The study characteristics are summarised in Table 1 and quality appraisal results are provided in Table 2. Given the diversity of methods available, it was not possible to recommend a single methodology for use in risk-ranking exercises. This finding was echoed by a scan of systematic reviews of risk ranking in other sectors including biological agents [20], pathogens, pests and weeds [48], and bioterrorism agents [49].

A best-practice framework was therefore developed using a process based on the common components identified in the studies included our literature review. It is an adaptable framework that can be applied to a variety of specific methodologies and provides best-practice recommendations to promote best practice across the various methodologies identified. We validated it by cross-checking it against the common themes of good practice identified in a systematic review of health research priority setting [50], and a
It is noteworthy that periodic evaluations of risk ranking were not explicitly considered in many of the studies reviewed here. Ultimately, risk ranking is best viewed as an initial part of the process of strategic public health planning, with the key objective being strengthened strategies to mitigate communicable disease spread. Given the rapidly changing public health landscape, it is advisable to repeat risk-ranking exercises at regular intervals. In addition, as has been observed elsewhere, there is value in the risk-ranking process itself, which has the potential to bring together stakeholders and practitioners from diverse fields to promote interdisciplinary working [52].

**Limitations**

This review focused only on ranking exercises conducted for communicable diseases. Methodologies from other sectors might also be relevant, but were not considered here. A limitation of the review is that the search, sift, quality appraisal and analysis was undertaken by a single researcher. However, quality assurance measures were put in place to mitigate any potential bias. The search strategy and approach were peer reviewed. Sifting decisions were made according to pre-defined criteria to ensure consistent decision-making. A sample of quality appraisals were duplicated to inform the development and refinement of the quality appraisal checklist, and establish scoring definitions to ensure consistent ratings. Data extractions were duplicated to ensure consistency and to check that the table captured the information required for analysis. The use of a single quality appraisal checklist across different methodologies means that the appraisal was not as deep as if method-specific appraisal tools had been used. However, the use of a single appraisal checklist enabled comparisons to be made across studies based on the principles of validity and reliability, regardless of the precise methodology. As with all quality appraisals based on published reports, the quality appraisal was affected by the reporting quality. Therefore criteria being 'not met' means that this detail was not reported in the study, however, there may be some discrepancy between the actual methodology and what was reported. Although most of the studies included in this review reported their findings clearly, there were some instances where there were gaps in reporting, which affected quality appraisals and analysis. Clear reporting ensures that processes are transparent, a stated aim of most of the included studies, so that the process can be understood and assessed by multiple stakeholders. Furthermore, it enables others to replicate, develop and improve upon previous practice, leading to improvements in methodologies.

**Conclusions**

The methodologies identified in this review mostly followed common approaches to risk ranking. The choice of methodology should reflect the purpose of the risk-ranking exercise. Common best-practice approaches, such as engaging diverse panels of stakeholders, and clearly delineating ranking criteria and criteria weights, were identified. The insights from this study will inform subsequent ECDC work on risk ranking, and should be relevant to any audience interested in ranking risks.

**Conflict of interest**

None declared.

**Authors’ contributions**

EOB: literature searches; study selection; writing. EOB and RT: generated study design; designed and performed quality appraisals; data extractions; editorial contribution. KG: expert input into study design and methodology; editorial contribution. JES and MC: initiating the study, study design input; editorial contribution. All authors have reviewed and approved the final article.

**References**


52. Suk JE, Van Cangh T, Ciotti M. Enhancing public health preparedness: towards an integrated process. eHealth Int. 2015;21(3):3.

License and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence, and indicate if changes were made.

This article is copyright of the authors, 2016.