TECHNICAL REPORT

Best practices in ranking emerging infectious disease threats

A literature review

www.ecdc.europa.eu

ECDC TECHNICAL REPORT

Best practices in ranking emerging infectious disease threats

A literature review

This report was commissioned by the European Centre for Disease Prevention and Control (ECDC), coordinated by Jonathan Suk, and produced by Bazian Ltd, London.

Declarations of interest were received from the contractor in accordance with ECDC's Independence Policy and no conflict was identified.

In accordance with the Staff Regulations for Officials and Conditions of Employment of Other Servants of the European Union and the ECDC Independence Policy, ECDC staff members shall not, in the performance of their duties, deal with a matter in which, directly or indirectly, they have any personal interest such as to impair their independence.

Suggested citation: European Centre for Disease Prevention and Control. Best practices in ranking emerging infectious disease threats. Stockholm: ECDC; 2015.

Stockholm, February 2015

ISBN 978-92-9193-621-2

doi 10.2900/653753

Catalogue number TQ-04-15-110-EN-N

Cover photo: Thomas Mathie, licenced under a Creative Commons (BY-NC-ND 2.0) license

Contents

Figures

Tables

Glossary of key terms

Bibliometrics
Bibliometrics describes a range of techniques used to analyse academic literature, such as the impact factor for journals.

Delphi
The Delphi technique is widely used to generate consensus amongst experts. It involves experts scoring various criteria, then aggregating and discussing these scores. This traditional version of this method is that the process is repeated until consensus is reached.

Hirsch index
A bibliometric measurement of the impact of an individual scientist's or team's work by calculating the number of articles and citations for a topic. It can also be used to gauge interest in a topic.

Multicriteria decision analysis (MCDA)
MCDA involves identifying and weighting criteria according to importance, identifying all possible courses of action, scoring each alternative then calculating weighted scores from the criteria weights and criteria scores.

Qualitative algorithm
Qualitative algorithms take assessors through a range of questions, which lead to different subsequent questions depending on the response; their responses then lead to a final decision.

Questionnaire
Questionnaires ask respondents to answer a number of questions. These can be quantitative, qualitative or a mixture of both.

Risk
Risk is the product of the impact multiplied by the likelihood of an outcome (based on the ISO definition).

Executive summary

Background

The threat of serious, cross-border infectious disease outbreaks in Europe is a significant challenge in terms of emergency preparedness. Types of threats and the pathogens involved shift in response to changing factors such as climate change, global travel, immigration patterns, environmental degradation, and social inequalities. In order to effectively target the use of resources to manage the risks of outbreak, it is necessary to formulate rankings or prioritisation of human and/or animal pathogens.

Methods

A literature review was conducted to identify the range of methods used to prioritise communicable disease threats for the purposes of emergency preparedness planning, and an evaluation undertaken to identify which are the most robust methodologies. Searches were undertaken across various biomedical and grey literature databases, supplemented by search techniques including reference harvesting and citation tracking. Studies were selected using transparent inclusion criteria and underwent quality appraisal by means of a bespoke checklist based on the AGREE II criteria. Due to the diversity of ranking methods identified, a narrative synthesis was performed, with studies clustered by methodology.

Results

Seventeen studies were selected for inclusion in the review. The included studies used one of five methodologies to prioritise communicable disease risks: bibliometric index, the Delphi technique, multi-criteria decision analysis (MCDA), qualitative algorithms, and questionnaires. The analysis includes an assessment of the individual studies (clustered by methodology), a discussion of the strengths and limitations of the individual methodologies, and a comparison of all of the methodologies. We also derived suggestions for good practice in risk-ranking exercises. Most of the studies included in this review followed a broadly similar approach to risk ranking: identifying diseases for ranking, identifying assessment criteria, weighting criteria, scoring diseases against criteria, and producing a ranked list of diseases. The studies that used a different approach were early-stage risk assessments, which aimed to narrow down a long list of diseases into a shorter list for use in further prioritisation exercises such as resource allocation.

Conclusions

The choice of methodology should reflect the objectives of the exercise. Instead of recommending a single definitive approach to risk ranking of communicable diseases for the purpose of preparedness planning, this review provides an evaluation of the strengths and limitations of the available methods, with a framework of best practice suggestions specific to individual methodologies and general points. This approach is intended to help inform decision-makers' choice of an appropriate risk-ranking method and ensure that these methods are carried out according to best practice.

1 Introduction

1.1 Context

The threat of serious, cross-border infectious disease outbreaks in Europe is a significant challenge in terms of emergency preparedness. Types of threats and the pathogens involved shift in relation to changing factors such as climate change [1], global travel, immigration patterns, environmental degradation, urban sprawl and other disease drivers [2, 3]. Novel infectious disease risks, such as SARS, MERS-CoV, A(H1N1), and even Ebola can arise in one part of the world and have the potential to spread globally, necessitating the implementation of a wide range of preparedness measures [4, 5]. An additional development in recent years is the emergence of what were previously considered to be tropical diseases in parts of Europe, such as West Nile fever in south-eastern Europe, outbreaks of chikungunya virus infection in Italy, France and Croatia, and indigenous cases of *Plasmodium vivax* transmission in Greece since 2009 [6-8].

In order to effectively target the use of resources to manage the risks of outbreak, it is necessary to form some sort of ranking or prioritisation of human and/or animal pathogens [9, 10]. In particular, the ranking must take account of the infectious nature, the ease and mode of transmission of the disease, identification, and specific early mitigation or preparedness measures that will positively influence outcomes. Any such method needs to be applicable across a range of pathogens and be easily updateable from year to year to reflect changing threats. Methods that try to incorporate risk have an element of estimating the likelihood and the impact of the event, but there are numerous ways to estimate these two parameters. Previous resource prioritisation processes range from qualitative, expert-led approaches, such as those that may use the Delphi method of consensus, through to a range of multi-criteria decision analysis tools including predictive modelling approaches that capture the multiple, interacting and non-linear responses of host, pathogen and vector to climate change and other variables.

1.2 Aim and objectives

This literature review aims to identify the range of methods used to rank communicable disease threats for the purposes of preparedness planning and evaluate which are the most robust methodologies. Communicable diseases are defined according to the EU's list of communicable diseases for surveillance [11]. This wide definition was adopted to reflect ECDC's full remit, and in recognition of the overlap between infectious and communicable diseases. For the purposes of this review, risk was defined according to ISO standards: risk = impact x likelihood [12]. The findings from this literature review will be incorporated into the development of an ECDC handbook on risk ranking for infectious diseases.

2 Methods

2.1. Step 1: search and sifting

Searching

A three-pronged approach to searching was used to identify both academic, peer-reviewed articles, and grey literature documents describing risk-ranking methodologies for communicable diseases:

- Biomedical databases (Medline, Embase, Cochrane Library and CRD)
- Grey literature and speciality databases (Google Advanced Search, WHO, World Bank)
- Supplemental (forensic) search techniques ('backward chaining' by scanning reference lists and 'forward chaining' using Google Scholar to identify citations) [13].

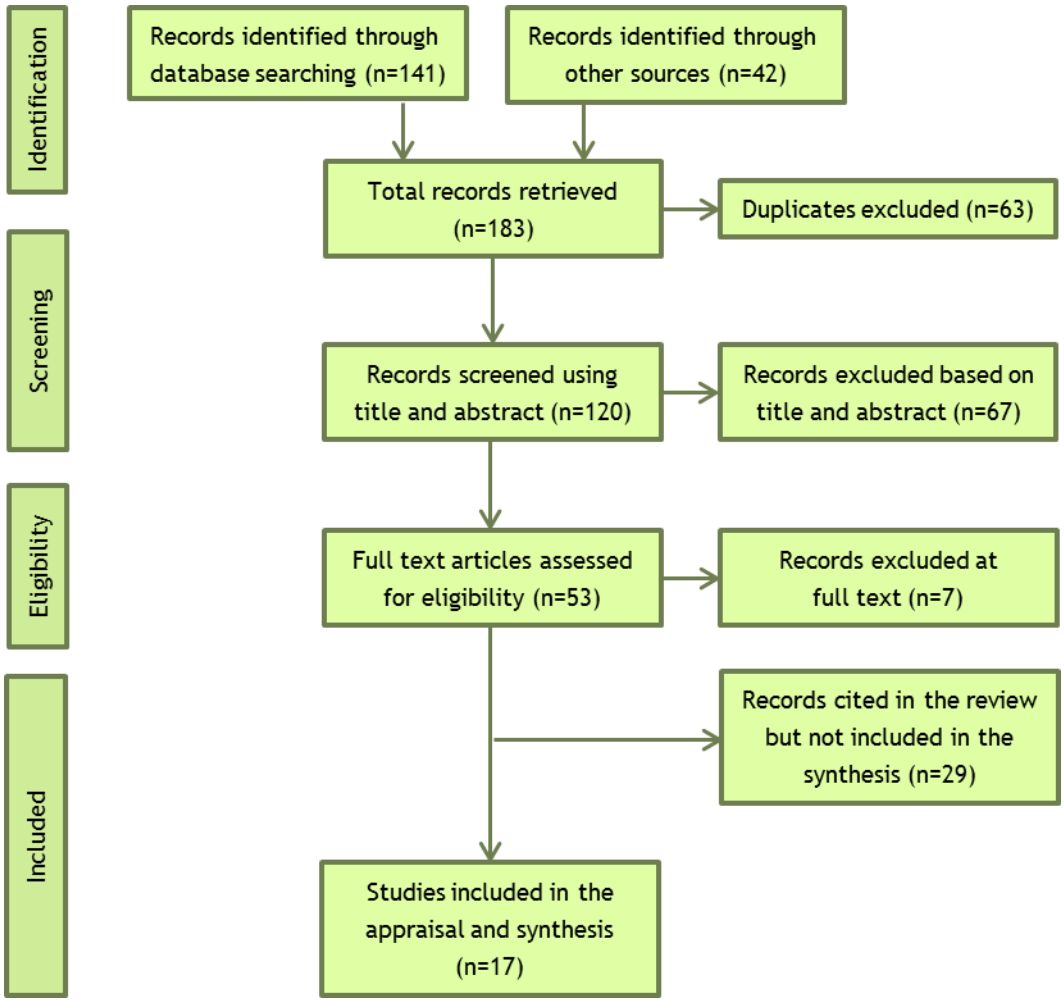
Free text and thesaurus search terms were identified by means of the citation pearl growing method [14], using an initial sample of relevant articles as well as the results of a brief scoping search (see Appendix 1 for the articles used in this process). A focused search strategy was devised for use in biomedical databases and grey literature sources. Backward and forward chaining was performed for all of these articles, and then a second round was performed for all articles included in the analysis.

Sifting

Criteria for inclusion in the review were studies:

- describing a method of prioritisation/ranking
- published in a peer-reviewed journal or by a national or supra-national government, charity, NGO or other authoritative institution
- within the geographic scope of the literature review (the EU, Australia, Canada, New Zealand and the United States)
- published in English
- published from 2000 to present for literature specifically relating to communicable diseases, and 2008 to present for literature on risk ranking/prioritisation methodologies in other sectors
- The search and sift process is presented in a PRISMA diagram in Figure 2. The searches are not fully exhaustive, though the three-pronged approach described above is both effective and efficient in capturing the most relevant literature.

Figure 1. PRISMA diagram



2.2. Step 2: evaluating validity and reliability of risk-ranking methods

Studies were evaluated based on their validity and reliability, defined thus:

- Validity: How well does the risk-ranking method measure the important facets of communicable disease emergency preparedness planning as described in the wider literature? This includes the methods used to determine a ranking/prioritisation (e.g. Delphi technique vs. survey), whether reasonable assumptions have been made, and sources of bias identified and considered.
- Reliability: What is the degree of internal consistency in the method, how precise is the methodology and what measures are in place to ensure consistency of scoring regardless of who is implementing the method?

As this project was evaluating a diverse range of studies using different methodologies, there was not a pre-existing single checklist suitable for assessing quality across the different studies. Therefore a quality appraisal checklist was developed, based on the AGREE criteria for the appraisal of clinical guidelines [15], to assess the validity and reliability of the risk ranking studies (see Appendix 3 for checklist and explanation of assessment criteria). A sample of quality appraisals were duplicated to test the validity of the instrument and to establish rating definitions. Studies were assessed against individual criteria within the three domains of 'assessment of validity', 'content validity' and 'assessment of reliability'. The appraisal form used a qualitative Likert scale of 'met', 'partly met', 'not met' or 'not applicable'. Definitions for ratings were included to ensure consistent assessment. Studies were given a rating against individual criteria, for each domain, and an overall rating based on the performance of the study against the criteria. A qualitative rather than quantitative assessment approach was used because the diversity of the included studies meant that quantitative scores would potentially be too arbitrary. For example, where assessment criteria were not applicable to a given study, a study would score a zero. This, however, could skew results as a study's score may not be an accurate reflection of the overall relative quality of that study. Using qualitative assessment allowed the assessors discretion to judge the impact of criteria ratings on the overall quality

of the study. The qualitative assessments are represented using a red, amber, green traffic light rating system (where red means that a high risk of bias is likely).

2.3. Step 3: analysis

A standardised data extraction form was used to extract key methodological information (see Appendix 4 for evidence tables). Due to the diversity of the studies, a narrative synthesis was performed. Studies were clustered by methodology to allow comparisons between studies using the same methodology and to derive strengths and weaknesses of the models. Then a narrative discussion of the different methodologies was undertaken to inform good practice points for individual methodologies and principles of good practice. There were two instances where more than one article describes the same prioritisation exercise by the same authors (Cox et al. [16, 17], Ng et al. [18-20]); in both cases, the articles were appraised and extracted as one in order to capture all of the necessary detail about the methods used. The 'et al.' suffix is used to reflect that the discussion refers to multiple publications and to distinguish Cox's single article on the Hirsch index (referenced as Cox [21]) from the two articles about multi-criteria decision analysis (referenced as Cox et al.). Included studies are counted individually within the PRISMA diagram (see Figure 1). The core set of communicable disease ranking papers were appraised using the quality appraisal tool developed for this review. A separate quality appraisal form was completed for each study (see Appendix 3). The domain ratings and the overall rating for each study were then collated into a table to produce a quality dashboard, thus facilitating quality comparisons within and across methodologies. Articles were not excluded on the basis of the 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.

3 Results

Seventeen studies were selected for inclusion in the review (see Appendix 6 for excluded studies). The studies used one of five methodologies to rank communicable disease risks: bibliometric index, the Delphi technique, multi-criteria decision analysis (MCDA), qualitative algorithms, and questionnaires. In this section, studies are clustered by methodology, with a comparison of the included studies, an overall assessment of the strengths and limitations of each methodology, and some best practice points. Then comparisons are made across the different methodologies, incorporating information from other sectors, to draw out good practice points for risk-ranking exercises.

3.1 Quality appraisals

The quality appraisals aimed to evaluate the validity and reliability of the studies. As with all quality appraisals based on published reports, the quality appraisal was affected by the reporting quality and detail. Therefore a criterion being 'not met' meant that this detail was not reported in the study; however, there may be some discrepancy between the actual methodology and what was reported.

A summary table containing the rating of each study against the three domains and the overall rating is included in Appendix 6. All included studies were rated as either having mostly met or partly met the criteria overall. The quality appraisals were used to inform the comparison between studies and across methodologies; therefore the results are incorporated into the discussion paragraphs below.

3.2 Bibliometric index (Hirsch index)

The Hirsch index (h-index) is a bibliometric measurement of the 'importance, significance and broad impact' of an individual scientist's work by calculating the number of articles and citations for a topic [22]. It can also be used to measure the impact of research units, organisations and the general 'interest' in a particular topic.

Table 1. Bibliometric studies overview

Study	Number of diseases ranked	Primary source	Validating source
Cox [21]	651	Web of Science	PubMed
McIntyre [23]	1414	Web of Science	Google Scholar, Scopus

Two studies, Cox [21] and McIntyre [23], used the h-index to rank emerging and established pathogens. Cox investigated 651 pathogens, whereas McIntyre investigated 1 414 pathogens. Web of Science (WOS) was used as the primary source of the h-index scores in both studies. Other sources were also searched to check the validity of the WOS results. Cox searched PubMed to compare the number of publications against the WOS results. McIntyre compared their WOS h-index scores with those generated by Google Scholar and Scopus. Both studies found that their WOS h-index scores were significantly correlated with the results from alternative sources, suggesting that WOS is a reliable source of h-index scores.

To test the validity of h-index scores, both studies compared the h-index results with established measures of burden of disease. Cox compared regional health-adjusted life year (HALY) data for 41 pathogens with their h-index scores. McIntyre compared the h-index scores of a subsample of 27 pathogens with World Health Organization (WHO) disability-adjusted life year (DALY) estimates. Both studies found that the h-index scores were significantly correlated with these alternative measures, suggesting that the h-index provides a reliable method of measuring the burden of disease.

Cox also undertook an analysis of h-index trends over time for seven pathogens. This indicated that h-index scores increased for all pathogens over time, even when adjusting for the general increase in the volume of publications, suggesting that older, more established pathogens may have higher h-index scores. However, it is possible to calculate the 'm-quotient' (Hirsch) which adjusts scores by dividing the h-index by the number of years since the first publication.

The use of the h-index to measure scientific interest in a subject and how this can be used in priority-setting is still an emerging methodology. Both studies found that h-index scores were highly over dispersed and most pathogens scored low. Therefore further work to understand the mechanics and potential biases of h-index scores is needed.

Although the two studies have different primary authors, there are some authors in common across the two studies.

Strengths

The main advantage of using the h-index is that it allows the rapid assessment of a large number of pathogens. McIntyre state that it took one person two weeks to obtain h-indices for the 1 414 pathogens considered in their paper. Authors speculate that it may be possible to automate this process. Furthermore, h-index is a user-objective measure so it does not share some of the biases that other semi-quantitative or qualitative measures may have.

As uncertainties remain about the methodology, the h-index may be best used as part of a wider risk-ranking exercise, for example to rapidly screen a large number of pathogens to provide a relative ranking in order to inform the next stage of risk ranking. Although it is useful in ranking pathogens already 'of interest' (Cox), it could also be used to assess those that are of international interest, but not necessarily yet affecting the region concerned. Emerging pathogens had significantly higher h-index scores than non-emerging pathogens, and h-index scores over time could be used to indicate emergence. This is especially valuable as standard metrics, such as incidence, may not yet be available in emerging infections. Therefore h-index scores could provide an alternative measure of burden of disease. As it measures the number of publications and citations, the h-index provides a good indication of both emerging and non-emerging disease threats that have a higher-than-average publication impact.

Pathogens that have been established for a longer time will have a greater number of publications and citations, which may skew an h-index. However, this bias could be mitigated by either using the m-quotient which can help to adjust for older pathogens or restricting database searches to a range of years.

Limitations

Because the h-index is reliant on publications and citations, it is subject to a lot of the same biases and issues that affect all scientific publishing. The h-index does not take account of the quality of evidence; citations are not endorsements of the quality or accuracy of research and therefore the number of citations is not necessarily positive. It may be further affected by trends in interest and funding for particular diseases, which is acknowledged as a limitation by both Cox and McIntyre. The h-index may be subject to a lag in time from research to publication, which may affect its usefulness in emerging infections, and Cox recommends that it should not be used for real-time tracking.

Both studies found that the accuracy of h-index scores can be affected by the use of pathogens in other settings, such as in laboratory settings. Search strategies could be adapted to reduce these false positives, but this would affect the rapidity of the process and may not completely eliminate the effect. There is still development work required to understand how best to search for h-index scores.

Best practice

- The h-index is best used as part of a wider risk-ranking methodology.
- H-index scores obtained via one source should be cross-checked against scores from other sources.
- When identifying relevant synonyms and acronyms to search for h-index scores, care should be taken to identify alternative uses of those acronyms and steps taken to mitigate these false positives.
- It is an emerging methodology and its limitations should be clearly reported.

3.3 Delphi studies

The Delphi technique is widely used to generate consensus amongst experts [24]. In the context of communicable diseases, Delphi discussions could be used at various stages of the risk ranking process through identifying the diseases for prioritisation, formulating criteria for assessment, deciding whether/how criteria should be weighted, independently scoring diseases, and discussing aggregated results. In most cases, Delphi discussion is restricted to the independent scoring of diseases and subsequent discussion. If there is not agreement after one cycle of discussion, independent scoring and discussion is repeated until consensus is reached.

Table 2. Delphi studies

Study	Number of participants	Number of diseases ranked	Number of criteria	Criteria weighting	Scoring method	Number of Delphi rounds
Balabanova [25]	20 (scoring) 86 (weighting)	127	10	Y	3-point scale	1
Economopoulou [26]	3 (scoring) 56 (selective scoring)	71	2	N	5-point scale	2
Krause [27]	11 (weighting and scoring)	85	12	Y	3-point scale	1
WHO [28]	24 (scoring)	53	8	N	5-point scale	1

Four studies used a Delphi technique (Balabanova [25], Economopoulou [26], Krause [27], and WHO [28]). The number of participants and their role in the Delphi process varied between studies. In Balabanova, 10 internal and 10 external experts scored the communicable diseases, while another 14 internal and 72 external experts weighted criteria. Economopoulou included a core team of three who scored diseases; then a panel of 56 internal experts from specific disease programmes scored diseases against some of the criteria. Krause had a Delphi panel of 11 who weighted the criteria and allocated scores. The WHO study involved 24 workshop participants who scored diseases against the criteria. The number of pathogens considered for ranking in these studies ranged between 53 and 127. Pathogens were identified using methods such as the current list of communicable diseases in the EU (WHO), or existing surveillance and published literature (Balabanova, Economopoulou, Krause).

The number of criteria used for risk ranking ranged from 8 to 12 (Balabanova, Krause, WHO). However, Economopoulou used only two criteria, which were highly specific to the setting for the risk-ranking exercise (the 2012 London Olympic Games). Both Balabanova and Krause used a 3-point scoring system (−1, 0, +1) to assess diseases; both state that this simple system was used to reduce inter-rater variation and provide clear definitions of scores to panellists. Economopoulou and WHO both used a 5-point scoring system (1-5, low to highest). However, Economopoulou had a two-phase scoring process where first the core team scored diseases, then the panel of disease experts scored the diseases within their specialties against selected criteria as a team. Economopoulou and WHO did not weight criteria. Both Balabanova and Krause weighted criteria by asking experts to rank criteria from the most to least important. Krause used the same group of experts to weight criteria, but at different time points; whereas Balabanova used a different group of experts to weight the criteria. Economopoulou collated evidence from a number of reliable sources to provide information on key infectious disease facets. It was unclear whether Balabanova, Krause and WHO provided any such evidence (although Balabanova does discuss the challenges in collating such evidence).

The normal process for Delphi is that rounds of discussion are repeated until consensus is reached. Balabanova, Krause and WHO only included one round of scoring. In the WHO workshop the results of the ranking exercise were discussed amongst participants but re-scoring was not possible due to time constraints. It is unclear in Balabanova and Krause whether results of the exercise were communicated with panellists and whether there was any discussion of the results. Economopoulou's core team discussed scores with the relevant disease expert teams only when there was a divergence between their scores. This Delphi discussion led to the re-ranking and prioritisation of two infections not considered relevant in the first round of scoring.

Although none of the studies specifically discussed organisational barriers to undertaking a risk-ranking exercise, they all discussed practical considerations. Krause acknowledged their small and homogenous Delphi panel, but stated that the Delphi process was time-consuming for the individuals taking part, which limited the size of their Delphi group. Krause suggest that limiting the number of pathogens for ranking could allow for a larger Delphi panel without increasing the amount of resource required. Balabanova also stated that the Delphi process required 'intensive preparation' by participants and required multiple and varied expert input (which is potentially expensive and logistically challenging). In order to engage a wide range of stakeholders, Balabanova suggest that such exercises be carried out by national organisations with sufficient power and influence to implement them. WHO also had to restrict scoring to one round due to time constraints. Economopoulou stated that their risk-ranking exercise took one week to complete and required approximately half a working day per expert.

Balabanova found that the weighting of criteria varied between participants, particularly by professional group, and speculated that it would also vary between different places based on societal values. The translation of quantitative scores into a qualitative risk matrix in Economopoulou allowed for a more nuanced and context-specific consideration of such factors; a formal weighting might have been restrictive. Therefore in risk-ranking exercises for specific settings it may be useful to use a qualitative methodology to retain that flexibility. Balabanova argues that the collocation in the final ranking of high and low incidence pathogens supports the approach of using multiple criteria upon which to base decisions, because metrics such as incidence or prevalence are not sufficient. Krause states that the relative ranking of diseases is more important than individual scores. Both Balabanova and

Krause were undertaken for the Robert Koch Institute (RKI) in Germany, and both exercises included many of the same researchers. In the interests of process improvement, Krause published a call for comments from fellow professionals [29] and carried out a survey to systematically gather feedback on their process, the results of which they implemented in subsequent prioritisation exercises [30].

Strengths

Although Delphi is a subjective process and variation in opinion is inevitable, especially between different professional groups, these can be mitigated by including evidence for experts to base their decisions on and by discussing scoring differences in multiple rounds of scoring. Using a large and multi-disciplinary panel can further mitigate subjectivity and professional bias. Providing definitions of criteria and scores can reduce inter-rater variation and subjectivity.

Delphi can allow for a nuanced and flexible, context-specific scoring and weighting. Additionally using a qualitative risk assessment may be easier for a multi-disciplinary team and audience to interpret the risk score.

Limitations

Delphi is a subjective process that can be skewed by professional bias or personal interest. Qualitative risk scores may be more open to interpretation and variation in scoring and interpretation by audiences, even when definitions are provided. It is potentially a resource-intensive process as it requires the (potentially expensive) input from a large number of experts, is time-consuming (for the experts and the organisers), and is potentially logistically challenging if done face-to-face. In order to make Delphi an affordable process, it may be necessary to adapt the process, which may in turn undermine the rigour of the process. Humblet [31] (an MCDA study described below) stated that they had intended to use Delphi, but opted for MCDA because they had limited resources and using MCDA allowed them to rank a larger number of diseases.

Best practice

- Power calculations can be used to determine the size of Delphi panel needed.
- Delphi panels should represent all relevant stakeholders.
- Criteria should reflect the aims of the risk-ranking exercise.
- Weighting of criteria should ideally be done at a separate time or by a separate group to reduce bias.
- Delphi participants should be provided with definitions of weights, scores and criteria to reduce inter-rater variation.
- Include multiple rounds of scoring, discussion and re-scoring until consensus emerges.
- The WHO has produced guidance on using Delphi to set priorities in communicable disease surveillance [32].

3.4 Multicriteria decision analysis (MCDA)

Multicriteria decision analysis (MCDA) is used in decision-making in various disciplines and settings. For example, it is widely used in environmental decision-making and policy [33, 34]. In healthcare it has been used in the decision-making process within health technology appraisals (HTA)[35] to rank and/or allocate resources, and in policy decision-making such as to prioritise investment in public health [36].

The Canadian Agency for Drugs and Technologies in Health (CADTH) describes the four steps within MCDA [37] as:

- identifying and weighting criteria according to importance
- identifying all possible courses of action (or alternatives to the planned action)
- scoring each alternative based on information collated for each criterion
- calculating weighted scores from the criteria weights and criteria scores.

These basic steps vary depending on the type of decision being made as a result of the MCDA exercise. In the case of risk-ranking exercises in communicable diseases, the ‘alternative courses of action’ are generally interpreted as the different diseases that can be ranked.

MCDA studies overview

Table 3. MCDA studies

Study	Number of participants	Number of diseases ranked	Number of criteria	Criteria weighting	Weighting method	Scoring method
Cardoen [38]	35 (scoring) 7 (weighting)	51	5	Yes	Las Vegas method	0–4 points (occurrence and severity)
Cox et al. [16, 17]	64	9	40	Yes	Qualitative Likert scale	Likert scale
Havelaar [39]	29	86	7	Yes	Relative ranking	Quantitative, scaled values
Humblet [31]	74	100	57 (5 categories)	Yes	Las Vegas method	0–7 points

Four studies used MCDA for disease risk ranking: Cardoen [38], Cox et al. [16, 17], Havelaar [39] and Humblet [31]. Cardoen self-identifies as using a 'semiquantitative' methodology; it has been discussed here alongside other MCDA studies as that is the methodology it shares most characteristics with. Cox et al. used two MCDA tools: an Excel spreadsheet, and MACBETH ('measuring attractiveness by a categorical-based evaluation technique'). Cox et al. selected the MACBETH tool because of its ability to translate qualitative responses into quantitative measures, and that it offers a variety of visual ways to compare pathogens (e.g. XY maps and difference profiles). However, there is a cost for the software and possible need for user training. By comparison, Excel is widely used, and the criteria, weighting and scoring can be easily altered to suit local context.

Cardoen, Cox et al. and Havelaar identified diseases based on the literature and expert opinion, though it was unclear how Humblet selected the diseases for ranking. The studies ranked between 9 and 100 diseases. The number of criteria varied between studies, with Humblet using 57 criteria, and Cardoen considering only five. Cardoen and Havelaar both deliberately limited the number of criteria; Cardoen did so in order to make scoring clear and reduce variation. Meanwhile Havelaar speculated that the expert decisions would only be based on a finite number of criteria, and therefore restricting the number of criteria would produce more meaningful results.

All studies weighted criteria. Cardoen and Humblet both used the Las Vegas method [40] to allow 7 and 40 experts, respectively, to score criteria based on their importance. The Las Vegas method involves the distribution of a finite number of points to criteria to weight them against a particular facet, such as importance. Cox et al. asked 64 experts to rate the impact of criteria on the likelihood of a pathogen emerging in Canada; their qualitative responses corresponding to a five-point Likert scale. A fixed weighting method and a model of weight as a probability distribution were both used to test the validity of the tool. Havelaar asked 29 experts to rank criteria by importance relative to each other. Additionally, Humblet asked six experts to use the Las Vegas method to distribute a total of 100 points to their five broad categories according to their importance to provide an inter-category weighting. Cardoen and Humblet provided experts with evidence, gathered using reliable sources, to support their decision-making. Where such information was not available Cardoen sought relevant expert opinion.

Cardoen calculated average scores per criteria using a clustered bootstrap method to estimate variance. Total scores for each disease were the sum of the bootstrapped criteria score. Havelaar and Humblet calculated uncertainty using a Monte Carlo simulation. Additionally Havelaar performed sensitivity analyses to assess the impact of the different assumptions contained in the different scenarios. Cox et al. used a sensitivity analyses to test the spreadsheet and MACBETH tools. The spreadsheet and the MACBETH tool used by Cox et al. produced comparable results overall. Cardoen and Humblet both used Classification and Regression Tree (CART) to obtain subgroups based on importance, with minimal variance using cross-validation. Havelaar's method demonstrated a good correlation between original scores and scores repeated two weeks later.

Cardoen observed small confidence intervals, indicating that variation between experts was minimal. Authors suggest this is due to the supporting information, which made decisions less subjective and more evidence-based. Similarly Havelaar stated that using quantitative scores forced participants to consider the evidence, leading to less arbitrary decision-making. However, Havelaar and Humblet still observed a lot of variation in expert opinion. Cox et al. compared their results to current literature, and presented to stakeholders and interested parties for discussion and feedback; as they were looking at emerging pathogens, there was no definitive benchmark against which to compare the model results. Cardoen gave experts the option to use 'ND' if there was insufficient data or '?' to indicate where evidence was not available or if they did not agree with the available data. Thus the study could identify areas of uncertainties, in particular knowledge gaps, upon which to recommend future research.

Strengths

MCDA incorporates expert opinion and empirical data. It is a flexible methodology that can be adapted to suit the context of the risk-ranking exercise. The number of criteria used can be varied according to need. Weighting can be assigned to criteria by using different methods such as the Las Vegas method or simple relative ranking. An MCDA tool has the potential to be developed centrally at a European level, and then adapted to suit local context within Member States. MCDA tools in Excel have the advantage that it is software that is widely used. However, the visual representations available in the MACBETH tool could allow wider stakeholder engagement because they are easier to interpret than quantitative values, so it would enable the public or other non-experts to be included in the exercise.

Limitations

The MCDA process requires the input of a range of experts, which can be resource-intensive. It is unclear how long the MCDA process takes, which is important practical information when considering implementing the process and setting re-evaluation schedules. MCDA may also require the purchase of software and expertise (software-related and statistical) not available within current teams.

Best practice

- Criteria should reflect the aims of the risk-ranking exercise.
- Weighting of criteria should ideally be done at a separate time or by a separate group to reduce bias.
- All relevant stakeholders should be included in the process.
- The UK government has devised a detailed inter-disciplinary manual for undertaking MCDA in decision-making [41].
- The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) has also launched a taskforce responsible for raising awareness of the variety of methods available within MCDA and to identify best practice [42].

3.5 Qualitative algorithm

Qualitative algorithms are a well-established tool in supporting decision-making in a variety of contexts and are often used in risk assessment. ECDC's current guidance on rapid risk assessments recommends the use of such algorithms [43], either a single-risk algorithm (as in Palmer [44]) or separate probability and impact algorithms (as in Morgan [45]). Algorithms take assessors through a range of questions, which lead to different subsequent questions depending on the response. Professor Morgan is a co-author on the Palmer paper, and Morgan, Kirkbride and Said (co-authors on the Morgan paper) are also co-authors of the ECDC guidance on risk assessment. Both studies used a qualitative approach when using a risk assessment algorithm. In both studies the algorithm was used as part of an early stage risk assessment of emerging infections to inform further research or assessment.

Qualitative algorithm studies overview

Table 4. Qualitative algorithm studies

Study	Number of participants	Number of diseases ranked
Morgan [45]	1	1
Palmer [44]	unclear	5

Both Morgan and Palmer used a qualitative algorithm to assess the risk of emerging infections. The algorithms contained a series of decision questions with binary yes/no answers. The 'stopping point' at which the assessment finishes aligns with a level of risk. Palmer used a single 'risk' algorithm, whereas Morgan separated risk into its constituent elements of probability and impact. Morgan felt that the amalgamation of the risk facets risked losing important detail.

Palmer used an expert group – the Human and Animal Infection and Risk Surveillance Group, HAIRS – to assess five pathogens, which had been referred to national authorities by medical practitioners and the expert opinion of the advisory committee. Whereas Morgan piloted the algorithm using a single pathogen which was initially assessed by a single scientist, then the results were passed to a multidisciplinary team for comment prior to sign off. Palmer does not describe the evidence, if any, that was used to support decisions; whereas Morgan state referencing any evidence used so that both the evidence and the decision can be scrutinised. Gaps in knowledge are also acknowledged, although it is not clear whether this leads to joined-up thinking in the allocation of research or surveillance funding to generate that data.

Both algorithms were designed to provide early stage risk assessment of emerging diseases, hence neither contains many of the key communicable disease facets included in the quality appraisal checklist. Detail was

lacking in the reporting of some areas of methodology in both studies; hence neither performed well on the quality appraisal checklist. For example, it is not clear how the algorithms themselves were developed, how the questions were identified or sequenced into the algorithm. Both studies state that the process attempts to distinguish a lack of evidence from a lack of evidence of zoonotic potential, which is an important distinction. However, due to the brevity of reporting it is not clear how this is achieved practically. The lack of detail about methods would make it harder to reproduce these studies; for instance, the composition of the group performing the ranking, and the quality assurance measures that should be implemented in order to ensure rigour and reduce bias.

Strengths

This process enables decision-makers to perform a rapid initial assessment of risk, prior to a full risk assessment or risk-ranking exercise. Both Morgan and Palmer claim that the process can differentiate a lack of evidence of risk from a lack of evidence (although how this is achieved is unclear), which would be useful in directing research funding or surveillance activities to generate data to inform more detailed risk assessment. A multi-disciplinary team is required to participate in the decision-making process (whether that be assessing against the algorithm or checking the results) to ensure that all relevant aspects are considered, especially given that this approach is particularly suited to emerging and zoonotic diseases. Although the qualitative approach can be said to be more subjective, algorithm-based decision-making can be evidence-informed and transparent when supporting information is clearly referenced. A qualitative approach also allows for more subtle distinctions to inform decision-making, such as balancing high-risk/low-probability diseases against low-risk/high-probability diseases. It is an early stage risk assessment which could be used to narrow down a longer list of diseases, or specifically to identify emerging diseases as part of a wider and more rigorous risk ranking process.

Limitations

It is unclear how the algorithms in these studies were formulated, making assessment of the potential bias in the process more challenging and affecting the reproducibility. As the process is qualitative it is potentially subjective, but the use of groups to undertake assessment could mitigate this potential bias. In the studies described, only one person went through the qualitative algorithm, therefore it is unclear how multiple perspectives would practically be incorporated into the process. Although this process is described as 'rapid', it is unclear what resource is required to develop and implement the algorithm.

Best practice

- Algorithms are best used as part of a wider risk-ranking methodology.
- The algorithm should cover criteria relevant to the aims of the risk-ranking exercise.
- All relevant stakeholders should be included in the process.
- ECDC has produced guidance on using qualitative algorithms to set priorities in rapid risk assessments [43].

3.6 Questionnaires

Questionnaires are widely used in many disciplines. The questions can be quantitative, qualitative or a mixture of both, which allows great flexibility in their use and the richness of data they are able to collect. Questionnaires are often used to gather quantitative data because they can be distributed to a large number of people, and the analysis of the results can be automated. Questionnaires can be used as part of a wider methodology; for example to gather large amounts of data and identify themes for richer data collection and analysis through interviews.

Questionnaire studies overview

Table 5. Questionnaire studies

Study	Number of participants	Response rate	Number of diseases ranked	Number of criteria	Criteria weighting	Weighting method
Horby [46]	518	46%	61	5	No	n/a
Ng et al. [18-20]	1539 (professionals) 2622 (public)	Not available due to recruitment method	62	21	Yes	Conjoint analysis

Two study designs, Horby [46] and Ng et al. [18-20] both used surveys to rank communicable disease risks, but their methods differed. Horby used a questionnaire to rank 58 diseases and three 'generic disease groups' (e.g. infections in the immunocompromised) based on five 'importance' criteria (e.g. burden of disease) and to identify priority areas for work (e.g. diagnosis, surveillance). Diseases were ranked based on their mean score across all

five criteria. The survey was distributed by post to 1130 various professionals including communicable disease experts, microbiologists, and genitourinary medicine practitioners. A total of 518 surveys were returned (46%).

Ng et al. first ran six focus groups with 54 multidisciplinary participants (including members of the public), using the nominal group technique, to identify criteria. A total of 59 criteria were identified, 21 of which were selected for conjoint analysis. Ng et al. used the selected criteria to develop an electronic survey using a partial-profile choice-based conjoint (CBC) analysis, which is a method usually used in market research. CBC was used to create 300 iterations of the survey, the responses provided authors with a score for the importance of the criteria. The survey was completed by members of the public (1 313 Canadian, 1 309 American) and human/animal health professionals (707 Canadian, 764 American). The CA-derived part-worth utility values were combined with data from literature searches (undertaken for each disease using a range of sources to identify reliable evidence), to prioritise 62 existing and emerging zoonotic and enteric diseases.

Despite a low response rate of 46% the final sample for Horby was still large. Ng et al. did not have figures for response rate due to the varied methods for recruiting participants. Ng et al. had a diverse survey sample (including members of the public) whereas Horby acknowledged that their survey sample was limited to mainly specialists working in secondary care. Horby found that there was generally good agreement between professional groups in their ranking of diseases, with the exception of genitourinary medicine specialists where only half of their top 10 diseases were the same as those of other groups (which may indicate a professional bias).

Authors compare the results of this survey with previous iterations and found similar results, which they suggest indicates a good reliability of the survey. Ng et al. observed some differences in the specific ranking of diseases based on the CA-derived scores of participants in Canada and the US. However, 76% of diseases were ranked within ten positions of each other, indicating general consensus between the two countries. Also, although there was some variation in the criteria importance ranking between Canadian and American respondents, all criteria were ranked within two places of each other. Horby found that 95% confidence intervals overlapped between diseases neighbouring each other in the final ranking and suggest that interpretation of the overall ranking trend is more useful than focusing too closely on the relative ranking of diseases.

Horby used fewer criteria than Ng et al. Although definitions were included, the broad nature of Horby's criteria mean that they may have been open to interpretation, and decision-making was not supported by evidence as in Ng et al.'s survey. Horby used quantitative scoring, but stated that their results may require further analysis. For example, a low-ranked disease may require a greater increase in resource allocation than a higher ranked disease. The authors suggest that including some qualitative responses in the survey would help to provide additional context to answers. Horby found that using a mixture of numerical scores and ticks confused some respondents, demonstrating that surveys must be carefully designed so that respondents understand the questions and the response they should provide. The 'area where further work is required' section in Horby could be a useful way to express uncertainty in decisions and lead to the resolution/amelioration of that uncertainty.

Neither study provided much information about the resource required for the risk ranking process. Ng et al. reported that the median time for professional respondents to complete the survey was 26.9 minutes in Canada and 28.1 minutes in America, and with over 700 respondents in each country that represents a large amount of time. It is not clear whether this time was paid for directly. Neither study provides information about how long it took to devise the surveys and analyse the results. Using a focus group to identify ranking criteria, as in Ng et al., ensures that criteria reflect the interests and needs of a diverse range of stakeholders. However, no information was provided about how much of participants' and facilitators' time was required for the focus groups or other expenses such as potentially covering travelling time and costs for attending. There was only one round of scoring in the professional focus groups, due to time constraints, suggesting that the process was time-consuming.

Strengths

Response rates are usually low for surveys; however, power calculations can be used to determine the size of the sample required based on projected non-response rates. Questionnaires provide a relatively cheap way to survey a large sample, and costs can be controlled by delivering them electronically (which may also increase response rates by reducing inconvenience). Surveys allow a potentially large and diverse sample to be consulted as part of risk ranking. Depending on the precise methods, quantitative responses can be straightforward to analyse, and automated systems are available. Mixed methods surveys can provide large quantities of quantitative data, alongside qualitative to provide context and aid interpretation of quantitative data.

Conjoint analysis (CA) in particular can facilitate the ranking of a large number of pathogens using numerous criteria. As CA is an established methodology in other sectors, it means that lessons can be learned and implemented when ranking communicable disease risks. Consideration of a larger number and range of diseases means that the ranked list can be subdivided by categories (such as vector-borne diseases) to present ranked lists to groups with particular interests. The creation and administration of questionnaires can also be outsourced to 'field and tab' companies, this can be cost-effective where this activity would take up a lot of staff time.

Limitations

Response rates are generally low, and therefore the questionnaire has to be sent out to a large sample in anticipation of non-response, which potentially increases the cost of running the survey. There can be differences in response rates within the same survey, where people do not answer all questions. Responses can be made mandatory for progression; however, this may deter respondents and lead to a higher number of non-responses. Quantitative-only responses can be reductive, therefore qualitative information to elaborate on quantitative scores may be required, which increases the resources required for data analysis. While surveying a wide range of people is useful for organisations wishing to engage all stakeholders, the varying level of expertise could undermine the validity of the results, especially where results are based on opinion. As with all surveys, there is always the possibility of respondent bias, which it is hard to mitigate.

Although those involved in communicable disease risk-ranking may have experience in running surveys, CA requires specific expertise in the development and analysis of the survey which may not be available in existing teams. It may be possible to train staff in these techniques, but there is cost involved in acquiring training materials (or sending staff to training sessions) and the staff time required for training. Therefore it may be necessary to hire in such additional expertise, thus increasing the cost. CA also requires specific software, which may have to be additionally purchased.

Best practice

- Use a power calculation to ensure that the final sample of responses is large enough to produce a meaningful and reliable result.
- Consider a sampling framework or targeted sampling to ensure that the views of all relevant stakeholders are represented.
- Ensure that questions address the aims of the risk-ranking exercise.
- Choose quantitative, qualitative or mixed questions according to the richness of responses required and the resources available for data analysis.
- Evaluate whether the project team has the necessary expertise or whether outsourcing is required.
- The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) has produced a good practice checklist for those using CA [47].

4 Discussion

4.1 Planning risk-ranking exercises

Planning is an essential step in any exercise to rank the risk of communicable disease threats. Establishing the objectives of the exercise enables the selection of an appropriate methodology that is fit for purpose: the planning of a ranking exercise should be matched to its objectives. A summary of the different methodologies, their strengths and weaknesses, and their most suitable uses is presented in Table 6.

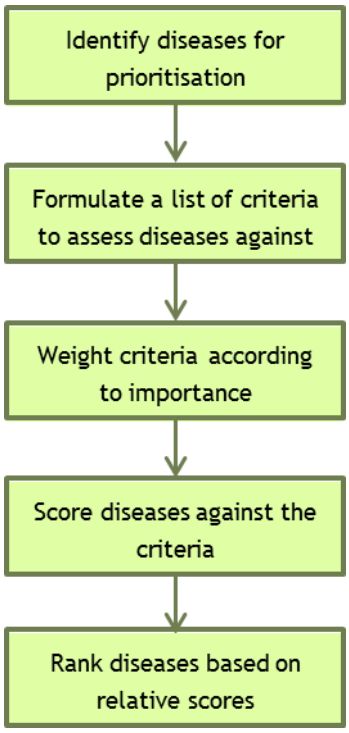
All of the methodologies described in this report can be adapted to suit the particular context and requirements of a risk-ranking exercise. For example, where resources are limited the number of criteria can be limited to increase the number of pathogens that can be assessed. There is always the need 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 [27]. Regardless of methodology, most of the included studies stated in their objectives that they wanted to rank communicable diseases using a transparent, systematic, valid and reliable method. In order to achieve this, the process for selecting diseases and criteria must be clearly stated.

An assessment of the resource required for any of these methods is an important part of the decision-making process. Methods requiring greater resource should not necessarily be disregarded, but the resource 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 resource 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 to deliver the ranking. Current WHO guidance on priority setting in communicable disease surveillance recommends that the planning process include budgeting, covering all resources required for the Delphi exercise [32].

Table 6. Scenarios and suggested methodologies for risk-ranking exercises

Scenarios	Methodology	Comments
Rapid or large-scale risk ranking for large number of pathogens	H-index or qualitative algorithm	Both methods are suitable for ranking a large volume of pathogens within a short time period or with limited resources.
Scoping exercise to generate an initial ranking for further study	H-index or qualitative algorithm	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.
Comprehensive risk ranking including novel, emerging and established infections	MCDA or Delphi	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.
Emerging infections with little published data about them	H-index	In lieu of standard data, such as burden of disease, h-index can indicate a level of professional interest/concern which may be used as an informal proxy measure of disease impact.
	Qualitative algorithm	This method combines expert opinion and evidence (where available). The qualitative nature allows for greater flexibility in decision-making and for the detailed recording of that rationale. This is particularly useful in emerging infections where decisions may be more based on expert opinion than epidemiological data.
	Qualitative algorithm or questionnaires	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.
	MCDA	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 rerun the entire ranking exercise

Figure 2. Generic components of risk ranking



4.2 Best practices in risk ranking studies

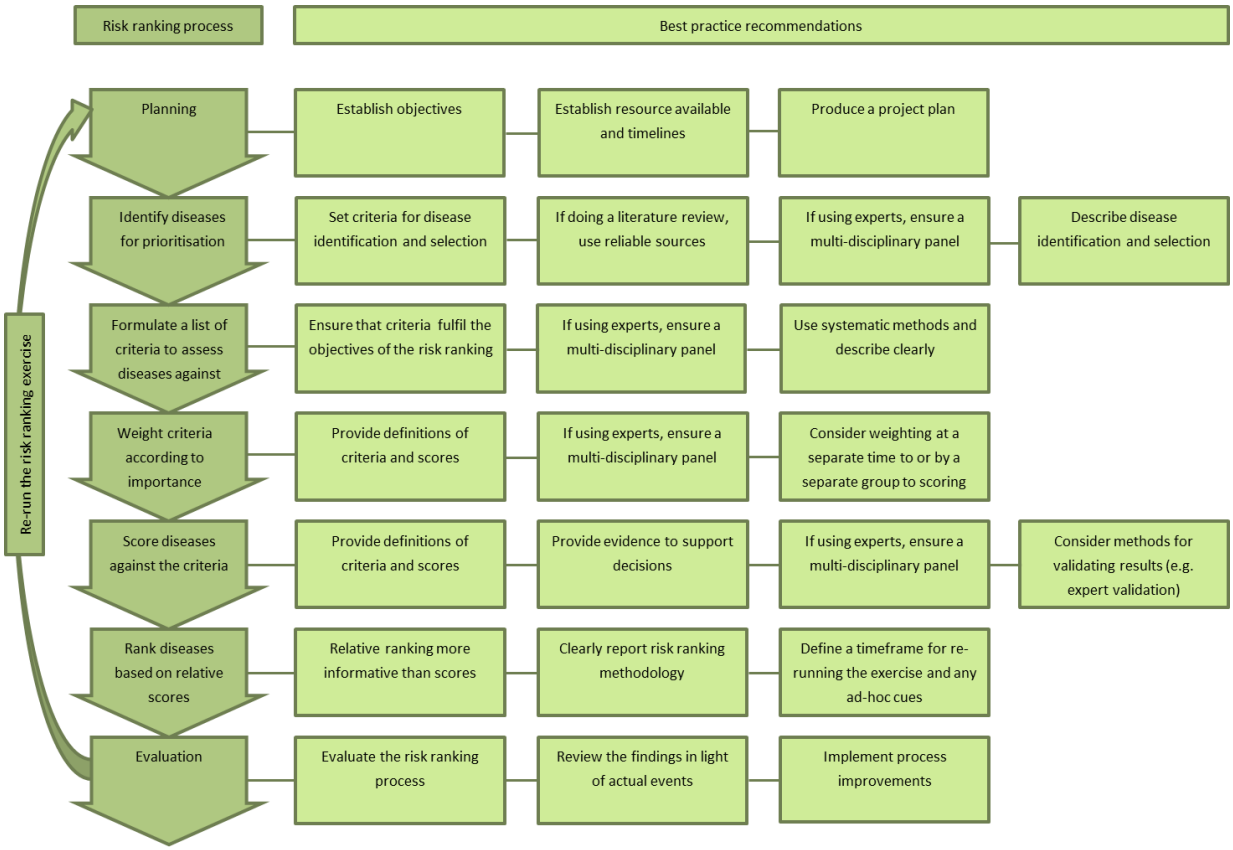
Based on the analysis of the studies reviewed in this report, it is possible to comment upon the best practices in conducting risk-ranking exercises, which is to some degree independent of the methodology selected.

In general, risk ranking infectious diseases is a process beginning with identifying diseases to consider, and ending with a ranking based on a scoring system (Figure 2). For each of the steps within this process, there are steps which can be taken to reduce bias and strengthen the credibility and repeatability of the findings (Figure 3). Within these best practices, a few are worthy of further discussion.

4.2.1 Selection of diseases

Most of the included studies (82%) assessed here 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. When using sources such as notifiable disease lists, it is necessary to consider (and report) the potential limitations of these sources. For example, notifiable disease lists are based on clinical and laboratory data, combined with suspected risk, therefore these are generally not useful for identifying emerging threats and the underlying assumptions of the risk assessment should be known and understood. Explicit reporting of the process of disease selection is important in communicating the purpose of the risk-ranking exercise.

Figure 3. Best practices in risk-ranking exercises



4.2.2 Criteria selection and weighting

The criteria considered in the studies we reviewed varied, although there was a common core of key communicable disease concepts. The average number of criteria was 17. The selected criteria should be specific to the context of the exercise (i.e. the purpose, the country it is taking place in); 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 so that diseases with low incidence due to effective control measures are not deprioritised and risk resources allocated elsewhere [25]. Instructions on interpreting criteria and score should be provided to participants to reduce inter-rater variation.

4.2.3 Selecting a multidisciplinary team

When incorporating expert opinion into any methodology, it is necessary to consider the representativeness of the people whose opinion is sought, which again relates to the scope and purpose of the exercise. All of the studies incorporating 'expert opinion' engaged 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 using qualitative scales or visual representations to aid participants in interpreting otherwise abstract scores (Cox et al.). Ng et al. was the only study to engage members of the public, and they were included only in the initial focus groups to identify and weight criteria. Predicting the impact of disease drivers such as climate change [1, 48], and predicting the future risk of communicable diseases, is challenging as there are many unknowns [2, 3, 9, 10]. Multidisciplinary input based on expertise and experience can help to inform decisions where standard data is not available, such as in the case of emerging disease threats or areas with great evidential uncertainty.

4.2.4 Evaluating expert opinion

Of the 17 included studies, 10 incorporated expert opinion in the risk-ranking methodology. The average number of experts included was 231; however, there were some outliers and so the median value of 48 may be a more useful indication. None of the included studies used any method such as power calculations to assess if sufficient numbers of participants were included, 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.

Subjectivity was an issue that nearly all of the included studies had to address, because of the use of expert opinion in the methodologies. Expert input introduces potential subjectivity and bias (particularly between different professional groups); however, expert input is needed where clear quantitative metrics are not available or where they are not easily comparable. The use of expert opinion in most of the studies suggests that they provide a unique input that would be otherwise missing from risk-ranking exercises. Measures can be put in place to mitigate this risk, such as clear explanations of criteria and definitions of scores to reduce inter-rater variation and interdisciplinary discussion of scores. Additionally, statistical methods can be used to measure variation in responses between individuals and professional groups, and appropriate adjustments can be made if the variation is considered too high. Formal statistical measures of uncertainty are also available; however, as a minimum it would be useful to incorporate a method by which participants could express any uncertainty they have so that this can be assessed. For example Cardoen allowed participants to respond with '?' where evidence was not available or they did not agree with the available evidence, which could be used to identify areas for further research. Horby explicitly asked participants to identify areas for further work such as surveillance.

4.2.5 Choosing between qualitative, quantitative and mixed methods

There are advantages and disadvantages to using quantitative or qualitative methods.

Five studies used a quantitative methodology (Cox, Havelaar, Humblet, Krause, McIntyre), three used qualitative approaches, (Economopoulou, Morgan and Palmer) and six studies used semi-quantitative, mixed methods (Balabanova, Cardoen, Cox et al., Horby, Ng et al., WHO).

The decision about whether to use qualitative, quantitative or mixed methods should be based on the scope and purpose of the exercise. For example, in areas where there is little evidence (and what does exist is of poor quality), it may be preferable to use semi-quantitative methods (to make best use of the evidence available) or qualitative methods (in recognition that the evidence is not much help).

Qualitative data generally takes longer to collect and to analyse than quantitative, although it provides a 'richness' and context to responses that quantitative data cannot. Semi-quantitative methods where respondents can provide quantitative scores with qualitative explanations seem to offer a good balance.

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 regard to the specific objectives of their risk-ranking exercise (Table 1). The only studies to not use mixed methods were Morgan and Palmer (qualitative algorithms) and Cox and McIntyre (quantitative h-index). However, these studies were considered by their authors to be most useful as part of a wider risk-ranking exercise. No comprehensive methodology using only qualitative or quantitative methods was identified in this review.

4.2.6 Providing evidence to risk ranking participants

Five studies provided their participants with information or evidence to support their decision-making. Where this evidence was provided, it was collated from reliable sources such as national governments, supranational organisations (such as the EU), non-governmental organisations (such as WHO), and charities. Providing such information 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 judgements. Morgan incorporated references of the evidence used in decision-making into their qualitative algorithm, 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.

4.2.7 Assessing resource requirements

The quality appraisal checklist devised for this literature review included two criteria relating to applicability. These assessed whether there was discussion of any potential organisational barriers to application of the method, and whether the method is supported by any advice or tools for implementation.

Unfortunately, few studies directly discussed practical considerations. Some studies allude to their method being time-consuming (Balabanova, Krause), or that time constraints required them to adapt their methodology or switch to another method (Humblet, Ng, WHO). General discussion of how methods can be adapted to suit time or resource constraints were discussed in some papers (Cox et al., Economopoulou, Havelaar, Krause), such as reducing the number of diseases considered to allow for a larger Delphi panel (Krause). Only Ng provided any data on how long the exercise took; they stated that the median time to complete their survey was 27 minutes in Canada and 28 minutes in the USA. Some authors also describe choosing their methodology because of its time-saving potential.

Additionally, resources such as software (e.g. MACBETH, Ng et al.), 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. In discussions of organisational barriers and other barriers to implementation, it would have been beneficial to see descriptions of practical experiences to help identify potential barriers, and descriptions of how barriers were overcome or suggestions for mitigations. However, these were rarely reported. There was general discussion of barriers or information about implementation, but rarely in sufficient detail to serve as an indicator for those trying to implement these methodologies.

4.2.8 Reporting results

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 our 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.

4.2.9 Evaluating the impact of risk ranking results

The included studies did not provide information on evaluating the effectiveness of the process and its output. WHO guidance on setting priorities in communicable diseases emphasises the role of the prioritisation exercise in the evaluation of the surveillance measures and places it within a process cycle, which includes evaluation [32]. This is important to ensure that activities are reviewed, but also where insufficient data exists, such a process can help to create that data.

4.2.10 Repetition and sustainability of risk-ranking exercises

Placing risk-ranking exercises within a process cycle assists in the evaluation of the process and its outcomes, but also emphasises the need to repeat the risk-ranking exercise. Krause describes how their experience of the current risk-ranking exercise will inform future risk-ranking exercises, as well as how they adapted their current methodology in light of previous experience. However, none of the studies lay out specific timescales or triggers for the risk-ranking exercise to be performed again. 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 (say every five years) depending on an assessment of the extent of change to which the various disease drivers have changed. It is also necessary to consider triggers (such as evidence of emerging threats) which could cue a re-ranking of diseases. In such ad-hoc cases it may be possible to perform a more rapid risk assessment in the interim before the next scheduled risk-ranking exercise is due.

4.3 Limitations of the review

This review focused only on ranking exercises conducted for infectious diseases. Methodologies from other sectors might also be relevant, but were not considered here. The search, sift, quality appraisal and analysis was undertaken by a single researcher, which is another study limitation. However, quality assurance measures were put in place to mitigate any potential bias. The search strategy and approach was peer reviewed by another information specialist. Sifting decisions were made according to criteria to ensure consistent decision-making. A sample of quality appraisals were duplicated to inform the development and refinement of the quality appraisal checklist, as well as establish scoring definitions to ensure consistent ratings. A sample of data extractions were duplicated (carried out by two researchers) to ensure consistency, and to ensure that the table captured the information required for analysis. The use of a single quality appraisal checklist across different methodologies further 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.

5 Conclusions

The methodologies identified in this review mostly followed common approaches to risk ranking. The studies that did not as closely follow this common method were early risk assessment studies, rather than definitive risk-ranking methods, so a different method was appropriate. Therefore the choice of methodology should reflect the purpose of the risk-ranking exercise.

This review was not able to recommend a single definitive approach for risk ranking. Instead, it provides an evaluation of the strengths and limitations of the available methods, with suggestions for best practice both within those methodologies, and a best practice framework across different methodologies (see Figure 3) to inform decision-makers' choice of risk-ranking methods.

6 Research recommendations

Future publications of risk-ranking methodologies should report their methods in detail to enable replication, adaptation and methodological appraisal. Practical information, such as the resource required for the exercise, should also be reported as much as possible.

References

1. Altizer S, Ostfeld RS, Johnson PT, Kutz S, Harvell CD. Climate change and infectious diseases: from evidence to a predictive framework. Science. 2013 Aug 2;341(6145):514-9. PubMed PMID: 23908230. Epub 2013/08/03.

2. Suk JE, Van Cangh T, Beaute J, Bartels C, Tsolova S, Pharris A, et al. The interconnected and cross-border nature of risks posed by infectious diseases. Global Health Action. 2014;7:25287. Epub 10 October.

3. Suk JE, Semenza JC. Future infectious disease threats to Europe. American Journal of Public Health. 2011;101(11):2069-78.

4. Bogoch II, Creatore MI, Cetron MS, Brownstein JS, Pesik N, Miniota J, et al. Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak. The Lancet. 2014.

5. Al-Tawfiq JA, Zumla A, Gautret P, Gray GC, Hui DS, Al-Rabeeah AA, et al. Surveillance for emerging respiratory viruses. The Lancet Infectious Diseases. 2014. English.

6. Suk JE, Semenza JC. From global to local: Vector-borne disease in an interconnected world. Eur J Public Health. 2014;24(4):531-2. English.

7. Rezza G, Nicoletti L, Angelini R, Romi R, Finarelli AC, Panning M, et al. Infection with chikungunya virus in Italy: an outbreak in a temperate region. Lancet. 2007 Dec 1;370(9602):1840-6. PubMed PMID: DOI: 10.1016/S0140-6736(07)61779-6.

8. Sudre B, Rossi M, Van Bortel W, Danis K, Baka A, Vakalis N, et al. Mapping environmental suitability for malaria transmission, Greece. Emerging Infectious Diseases. 2013;19(5):784-6. English.

9. Suk JE, Lyall C, Tait J. Mapping the future dynamics of disease transmission: risk analysis in the United Kingdom Foresight Programme on the detection and identification of infectious diseases. Euro Surveillance. 2008;13(44):ii=19021.

10. Woolhouse M. How to make predictions about future infectious disease risks. Philosophical Transactions of the Royal Society Biological Sciences. 2011;366(1573):2045-54.

11. Lal A, Hales S, French N, Baker MG. Seasonality in human zoonotic enteric diseases: a systematic review. PLoS One. 2012;7(4):e31883. Epub 2012 Apr 2.

12. International Organization for Standardization. ISO Guide 73:2009. Risk management – vocabulary. Geneva: ISO, 2009.

13. Bates MJ. The design of browsing and berrypicking techniques for the online search interface. Online Review. 1989;13(5):407-24.

14. Booth A. Unpacking your literature search toolbox: on search styles and tactics. Health Information and Libraries Journal. 2010;25:313-7.

15. AGREE Enterprise. Appraisal of guidelines for research and evaluation II. [No place] AGREE Enterprise, 2013.

16. Cox R, Crawford WR, Sanchez J. The Use of Expert Opinion to Assess the Risk of Emergence or Re-Emergence of Infectious Diseases in Canada Associated with Climate Change. PLoS ONE [electronic resource]. 2012;7(7):e41990.

17. Cox R, Sanchez J, Revie CW. Multi-criteria decision analysis tools for prioritising emerging or re-emerging infectious diseases associated with climate change in Canada. PLoS ONE [electronic resource]. 2013 2013;8(8):e68338.

18. Ng V, Sargeant JM. A quantitative and novel approach to the prioritization of zoonotic diseases in North America: a public perspective. PLoS ONE [electronic resource]. 2012 2012;7(11):e48519.

19. Ng V, Sargeant JM. A stakeholder-informed approach to the identification of criteria for the prioritization of zoonoses in Canada. PLoS ONE [electronic resource]. 2012 2012;7(1):e29752.

20. Ng V, Sargeant JM. A quantitative approach to the prioritization of zoonotic diseases in North America: a health professionals' perspective. PLoS ONE [electronic resource]. 2013 2013;8(8):e72172.

21. Cox R, McIntyre KM, Sanchez J, Setzkorn C, Baylis M, Revie CW. Comparison of the h-index scores among pathogens identified as emerging hazards in North America. Transboundary and Emerging Diseases. 2014;Early View.

22. Hirsch JE. An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(46):16569-72.

23. McIntyre KM, Hawkes I, Waret-Szkuta A, Morand S, Baylis M. The h-index as a quantitative indicator of the relative impact of human diseases. PLoS ONE [electronic resource]. 2011;6(5):e19558.

24. Sackman H. Delphi assessment: Expert opinion, forecasting, and group process. Santa Monica (CA): Rand Corporation, 1974.

25. Balabanova Y, Gilsdorf A, Buda S, Burger R, Eckmanns T, Gartner B, et al. Communicable diseases prioritized for surveillance and epidemiological research: results of a standardized prioritization procedure in Germany, 2011. PLoS ONE [electronic resource]. 2011 2011;6(10):e25691.

26. Economopoulou A, Kinross P, Domanovic D, Coulombier D. Infectious diseases prioritisation for event-based surveillance at the European Union level for the 2012 Olympic and Paralympic Games. Euro Surveillance. 2014 2014;19(15).

27. Krause G, Working Group on Prioritization at Robert Koch Institute. How can infectious diseases be prioritized in public health? A standardized prioritization scheme for discussion. EMBO Reports. 2008;9 Suppl 1:S22-S7.

28. WHO. The Dubrovnik pledge on surveillance and prioritization of infectious diseases. Copenhagen: World Health Organization, 2003.

29. Krause G, Working Group on Prioritisation at the Robert Koch Institute. Prioritisation of infectious diseases in public health – call for comments. Euro Surveillance. 2008;13(40).

30. Gilsdorf A, Krause G. Prioritisation of infectious diseases in public health: feedback on the prioritisation methodology, 15 July 2008 to 15 January 2009. Euro Surveillance. 2011 2011;16(18).

31. Humblet MF, Vandeputte S, Albert A, Gosset C, Kirschvink N, Haubruge E, et al. Multidisciplinary and evidence-based method for prioritizing diseases of food-producing animals and zoonoses. Emerging Infectious Diseases. 2012;18(4).

32. WHO. Setting priorities in communicable disease surveillance. Geneva: World Health Organization, 2006.

33. Steele K, Carmel Y, Cross J, Wilcox C. Uses and misuses of multicriteria decision analysis (MCDA) in environmental decision making. Risk analysis. 2009;29(1):26-33.

34. Huang IB. Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends. Science of The Total Environment. 2011;409(19):3578-94.

35. Thokala P. Multiple criteria decision analysis for health technology assessment. Sheffield: University of Sheffield, 2011.

36. Marsh K, Dolan P, Kempster J, Lugon M. Prioritizing investments in public health: a multi-criteria decision analysis. Journal of Public Health. 2012;35(3):460-6.

37. Canadian Agency for Drugs and Technologies in Health. Multi-criteria decision analysis (MCDA). Ottawa: CADTH, 2014.

38. Cardoen S, Van HX, Berkvens D, Quoilin S, Ducoffre G, Saegerman C, et al. Evidence-based semiquantitative methodology for prioritization of foodborne zoonoses. Foodborne Pathogens & Disease. 2009;6(9):1083-96.

39. Havelaar AH, van RF, Bucura C, Toetenel MA, Haagsma JA, Kurowicka D, et al. Prioritizing emerging zoonoses in the Netherlands. PLoS ONE [electronic resource]. 2010 2010;5(11):e13965.

40. Gore SM. Biostatistics and the Medical Research Council. Medical Research Council News. 1987;1987(35):19-20.

41. Multi-criteria analysis: a manual. London: Department for Communities and Local Government, 2009.

42. Ispor. Multi-criteria decision analysis in health care decision making emerging good practices task force. Lawrenceville (NJ): International Society for Pharmacoeconomics and Outcomes Research, 2014.

43. ECDC. Operational guidance on rapid risk assessment methodology. Stockholm: European Centre for Disease Prevention and Control, 2011.

44. Palmer S, Brown D, Morgan D. Early qualitative risk assessment of the emerging zoonotic potential of animal diseases. BMJ. 2005;331:1256-60.

45. Morgan D, Kirkbride H, Hewitt K, Said B, Walsh AL. Assessing the risk from emerging infections. Epidemiology and Infection. 2009;137:1521-30.

46. Horby P, Rushdy A, Graham C, O'Mahony M. PHLS overview of Communicable Diseases 1999. Communicable Disease & Public Health. 2001;4(1):8-17.

47. Bridges JFP, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health – a checklist: a report of the ispor good research practices for conjoint analysis task force. Value in Health. 2011;14:402-12.

48. Semenza JC, Sudre B, Oni T, Suk JE, Giesecke J. Linking environmental drivers to infectious diseases: the European environment and epidemiology network. PLoS Negl Trop Dis. 2013 Jul 25;7(7):e2323.

Appendix 1: Initial set of articles from the RFS and scoping search

List of articles used to derive the search strategy and used in the initial phase of backwards and forwards chaining.

1. Altizer S, Ostfeld RS, Johnson PT, Kutz S, Harvell CD. Climate change and infectious diseases: from evidence to a predictive framework. Science. 2013 Aug 2;341(6145):514-9.
2. Balabanova Y, Gilsdorf A, Buda S et al. Communicable diseases prioritized for surveillance and epidemiological research: results of a standardized prioritization procedure in Germany, 2011. PLoS ONE [electronic resource]. 2011 20111004 [Epub ahead of print];6(10):e25691.
3. Cox R, Sanchez J, Revie CW. Multi-criteria decision analysis tools for prioritising emerging or re-emerging infectious diseases associated with climate change in Canada. PLoS ONE [electronic resource]. 2013 20130807 [Epub ahead of print];8(8):e68338.
4. Havelaar AH, van RF, Bucura C et al. Prioritizing emerging zoonoses in the Netherlands. PLoS ONE [electronic resource]. 2010 20101115 [Epub ahead of print];5(11):e13965.
5. Krause G, Working Group on Prioritization at Robert Koch Institute. How can infectious diseases be prioritized in public health? A standardized prioritization scheme for discussion. EMBO Reports. 2008;9 Suppl 1:S22-S27.
6. McIntyre KM, Hawkes I, Waret-Szkuta A, Morand S, Baylis M. The h-index as a quantitative indicator of the relative impact of human diseases. PLoS One. 2011;6(5):e19558.
7. Ng V, Sargeant JM. A quantitative and novel approach to the prioritization of zoonotic diseases in North America: a public perspective. PLoS ONE [electronic resource]. 2012 20121101 [Epub ahead of print];7(11):e48519.
8. Suk JE, Semenza JC. Future infectious disease threats to Europe. American Journal of Public Health: November 2011, Vol. 101, No. 11, pp. 2068-2079.
9. Suk JE, Lyall C, Tait J. Mapping the future dynamics of disease transmission: risk analysis in the United Kingdom Foresight Programme on the detection and identification of infectious diseases. Euro Surveillance. 2008;13(44):pii=19021.
10. Woolhouse M. How to make predictions about future infectious disease risks. Philosophical Transactions of the Royal Society Biological Sciences. 2011;366(1573):2045-54.

Appendix 2: Quality appraisal checklist template

Domain	Item assessed	Rating				Reviewer comments
		Met	Partly met	Not met	N/A	
Assessment of validity						
Scope and purpose	The objective(s) is (are) clearly described					
	The context/setting of the prioritisation method is applicable to the EU and relevant to ranking on a national or international level					
Stakeholder involvement	Individuals from all relevant professional groups are involved in the development and validation					
Rigour of development	Systematic methods were used to search for evidence					
	The criteria for selecting the evidence are clearly described					
	The strengths and limitations of the body of evidence are clearly described					
	The methods used for formulating the prioritisation criteria are clearly described					
	There is an explicit link between the prioritisation method and the supporting evidence					
	Any assumptions and sources of bias are clearly described and mitigated where possible					
	The study has been externally reviewed by experts prior to its publication					
Applicability	The potential organisational barriers in applying the method have been discussed					
	The method is supported by advice or tools for application					
Editorial independence	The funding body has not influenced the study and any competing interests are declared					
Overall assessment of domain						
Content validity The criteria included in the prioritisation method measure the key facets of communicable diseases						
Likelihood of outbreak	Current incidence rates (in humans and animals for zoonotic diseases), emerging trends and duration					
	Previous outbreaks documented					
	Transmission routes and infectivity					
	Risk factors/at risk groups					
Impact	Case fatality					
	Acute phase and chronic state (severity)					
	Case attack rate					
	Socioeconomic burden (inequality slope?)					
	Impact on the healthcare system/workforce/school system					
Mitigation: opportunities for public health intervention	Control measures in place (e.g. early detection, surveillance)					
	Non-pharmacological prophylaxis: increasing public					

Domain	Item assessed	Rating				Reviewer comments
		Met	Partly met	Not met	N/A	
	awareness of behavioural measures; containment/suspension of travel					
	Pharmacological prophylaxis: vaccination, antimicrobials					
	Pharmacological treatment: vaccination, antimicrobials, including emerging resistance					
	Economic implications of prevention and control measures					
Weighted criteria	The method transparently weights disease criteria					
Overall assessment of domain						
Assessment of reliability						
Consistency	Internal consistency in the method (e.g. multi-item questions across surveys)					
	Measures to mitigate inter-rater variation in scoring					
Reproducibility	The prioritisation method is specific, unambiguous and easily reproducible (e.g. in another setting, context)					
	Appropriate statistical methods were considered and used by the researchers to test reliability of their methods (e.g. test and validation data sets, sensitivity analysis, sample size for surveys, model fit, test of variability in expert opinion, chi-squared tests)					
Overall assessment of domain						
Overall guideline assessment						

Score guide

The quality appraisal system is based on modified AGREE II criteria. The items assessed are scored on a 3-point Likert scale and relating to a Red Amber Green (RAG) grade, according to the following system. Comments are added to indicate any specific considerations related to each item:

Met	Information related to that item has been clearly reported and all relevant considerations have been made.
Partly met	The information related to that item is incomplete, or not all aspects have been considered. The score is assigned depending on the completeness and quality of reporting, with increased score as more criteria are met and more considerations have been made.
Not met	There is no information provided in the study that is relevant to that item, or information related to that item is very poorly reported.
Not applicable	Where criteria are not applicable to a particular study or methodology.

Studies should be rated against individual criteria; then a rating should be given for each domain based on the criteria ratings and overall rating based on the performance of the study against all criteria and domains.

Appendix 3: Evidence table

Study	Country or region	Objectives	Methods	Weighting method
Balabanova[19]	Germany	• Developing a rational system for prioritising infectious diseases • Ranking the most common pathogens in accordance with their importance for national surveillance and epi-demio-logical research	• Delphi consensus method including 10 senior external experts and 10 internal experts from the RKI (collective expertise in bacteriology, virology, mycology, parasitology, general infectious diseases, tropical medicine, general medicine, epidemiology, public health, veterinary health and infection control). • A list of 127 infectious diseases were compiled that fulfilled criteria of being: 1. notifiable according to German law for the control of infectious diseases 2. reportable within the EU and the WHO 3. an agent with potential for deliberate release 4. represented in dedicated chapters in an established infectious diseases manual 5. occurring in Germany • 10 criteria were established, modified from a previous prioritisation process and reviewed by the panel. Each pathogen was scored against the criterion (−1, 0 or +1) by the internal RKI experts, with supporting data recorded, which was followed by a modified two-round Delphi process (internal round and joint round with external experts) where scores were discussed. • Pathogens were ranked by multiplying each score by the weight for that criterion to give a total weighted score of 0 to 100, with 0 to 24 low priority, 26-50 medium, 51 to 75 high, and 76 to 100 highest priority.	• By 14 internal experts and an independent panel of 72 external experts. • A score of 1 (low) to 10 (high) was assigned reflecting the importance of the criterion for surveillance and epidemiological research. • The median weighting score for each criterion was calculated.
Cardoen[32]	Belgium	• To produce an evidence-based, standardise d, semi-quantitative method for prioritisa-tion of food and waterborne zoonoses to allow food safety authorities to focus on the most relevant hazards when forming control programme s for the food chain	• A semi-quantitative method was used with an independent weighting of criteria by food risk managers. • A list of 51 food- and waterborne pathogens was developed based on literature review and expert opinion. Zoonoses were defined as 'disease or infection naturally transmissible from animals to humans and vice versa'. • 35 experts (from medical, veterinary, agrochemical and biological backgrounds) scored pathogens (0 to 4 or 'not determined') against 5 main criteria: Severity of disease for humans, occurrence of disease in the Belgian population (2003 to 06), occurrence in live animals in Belgium, severity of the disease for animals and commercial and economic impact of the disease for the sector, occurrence of the agent in food or carcasses. • Help information' including national and international data from official sources was provided to experts in order to support their scoring decision. Data on severity and occurrence of disease in humans and animals was collected from the scientific literature and from official institutions and organisations. Information on economic and commercial impact came from expert opinion. • Individual scores of each expert were calculated to give an average score (0 to 4) and standard error for each criterion, reflecting heterogeneity between experts. Average total scores were calculated in R-project software using a clustered bootstrap allowing a correct estimate of the variance of the total score (including missing scores from the 'not determined' assignment). Non-bootstrapped and bootstrapped averages were compared to assess bias. A total score was given (0 to 20 points) with 95% confidence intervals to assess uncertainty. • Pathogens were ranked according to their total weighted scores in Excel. Different groups of importance were identified by Classification and Regression Tree (CART) software. • Scores were compared with existing national surveillance measures and recommendations and addressed to the food chain risk managers of the Belgian Federal Agency for Safety of the Food Chain.	• Criteria weighting was performed by 7 food chain risk managers. A total 20 points were distributed between the 5 criteria using Las Vegas Methodology (Gore 1987 This gave a relative weight to the criterion which was introduced in the calculations of total score
Cox et al.[10,11]	Canada	• To design and test a standardise d method to prioritise infectious diseases of	• Two different multi-criteria decision analysis (MCDA) approaches were tested both of which used the additive aggregation model: an Excel spreadsheet tool, and 'MACBETH' ('measuring attractiveness by a categorical based evaluation technique'). • Both methods involved steps of identification of criteria that can be used to prioritise pathogens, assigning	Excel spreadsheet tool • Two weighting methods • The first used fixed weighting values. Experts assigned weights of 0, 0.1,

Study	Country or region	Objectives	Methods	Weighting method
		humans and animals that may emerge in Canada in response to climate change	attributes to each criterion, and expert evaluation. • 40 criteria were identified from published literature and expert discussion, and divided into 5 main groups A to E: Disease epidemiology (12 criteria); ability to monitor, treat and control disease (5 criteria); influence of climate change in Canada (12 criteria); burden of disease (8 criteria); and economic and social impact (3 criteria). • Assigned attributes were based on published literature and aimed to be as quantitative as possible. • 64 experts were involved in criteria selection and weighting in phase 1 (a response rate of 74% of 86 who agreed to participate, of 121 invited). Most respondents (55) were from Canada, the remainder from the US, UK, France and Japan. 47 of 64 experts were involved in a phase 2 evaluation of criteria attributes and definitions, assessed by questionnaire (reported 72% response rate; calculated as 73%). Experts were identified through literature and internet searching and had academic, government or independent backgrounds with past or present study in infectious disease epidemiology and/or climate change. (The second publication by Cox et al. explicitly describes the methodology used to identify and involve experts in identifying criteria and determine the importance of these criteria.) • 14 pathogens were selected on the basis of taxonomic group, zoonotic potential, mode of transmission, endemicity, evidence for climate influence, and notifiable status in Canada in 2010 (those that are of 'significant importance to human or animal health or to the Canadian economy'). Nine of 14 were tested in the prioritisation tools; 5 were excluded due to lack of expertise on the pathogens. The 2 MCDA tool designs were: Excel spreadsheet tool • Criteria were listed in Excel and criteria attributes were implemented as predefined drop-down selection boxes. Criteria were weighted and attributes were assigned values so that completion of the spreadsheet calculated a pathogen score as a linear weighted sum of scores. MACBETH • Using the M-MACBETH software, criteria were organised into five groups in a decision tree. Information about each pathogen was entered into predefined drop-down menus. M-MACBETH calculated the score for each criterion using an additive aggregation model. M-MACBETH assessed the difference between each attribute in an attribute matrix. • Sensitivity of the spreadsheet and MACBETH were assessed by repeating pathogen ranking using reduced versions of each tool which only included the top 10 weighted criteria.	0.3, 0.5, 0.7 and 0.9 to the likelihood and importance of each criterion and a mean weight for each criterion was calculated. • The second weighting method took into account variation in expert opinion by modelling weight as a probability distribution, with likelihood or importance given a continuous distribution between 0.01 and 1 to give a total 10,000 iterations capturing the weight distribution for each criterion. • Criteria attributes were assigned values in ascending order from 1 to 4 or 0 if not applicable. Attribute values were normalised according to the number of possible attributes. MACBETH • The first weighting method was used where weights of all criteria were standardised to sum to 100 by dividing each weight by the sum of all weights and multiplying by 100. Weights ranged from 0.58 to 1.84. A value from 0 to 100 was assigned to each criterion attribute using the M-MACBETH generated matrix.

Study	Country or region	Objectives	Methods	Weighting method
Cox[15]	North America (USA, Canada, Mexico and Greenland)	• To identify and compare hazardous pathogens in North America, to see whether it might be a feasible method for ranking according to impact on human health • It further aimed to look at how it may rank pathogens identified as emerging hazards and to see how the h-index of a pathogen changes over time	• Hirsch index (h-index), which can be calculated using a range of bibliographic databases. • The ENHanCEd Infectious Disease (EID2) database (University of Liverpool) was used (search date October 2011) to identify 1827 human pathogens. • 651 of these pathogens had previous occurrence in North America, as identified by searching in the National Center for Biotechnology Information 2011 database (474 pathogens), and the PubMed database (177 pathogens), using a threshold of 5 references associating a pathogen with a country. • Various search terms were used to identify all papers relating to the 651 pathogens and obtain the h-index scores using Web of Science (WOS). • h-index calculated through WOS were compared with those obtained from the number of publications for each pathogen in PubMed (using Spearman rank correlation coefficient). • Pathogens identified as potential emerging threats in Canada were ranked by h-index. Emerging pathogens were those that have appeared in a human population for the first time, or occurred previously but increasing or expanding. Three data sources were used: 33 identified by the publication by Greer et al. 2008; 6 by the Zoonotics Division of the Public Health Agency of Canada; and for 36 the h-index was correlated with the healthy adjusted life year measurement (HALY) as measured by the Ontario Burden of Infectious Diseases study (Kwong 2010). • Change in h-index over time was assessed for a select set of pathogens of interest by looking at the h-index each year 1960 to 2011, and using 2 negative binomial models looking at cumulative h-index, and rate of change by year.	• Not applicable – no criteria developed
Economopoulou[20]	EU wide	• To use a reproducible, transparent, qualitative method to prioritise infectious diseases occurring worldwide and representing a risk for public health during mass gatherings • To develop a list of significant infectious disease events (SIDEs) to support event-based surveillance for the 2012 Olympic and Paralympic Games in London	• Delphi consensus method and development of a risk matrix. The panel involved a generic team of 3 ECDC experts and 56 ECDC experts from 7 disease programmes (food- and waterborne and zoonoses, vaccine-preventable diseases, emerging and vector-borne diseases, tuberculosis, airborne diseases, human immunodeficiency virus and other sexually transmitted infections, and antimicrobial resistance and healthcare-associated infections). • A list of 71 infectious diseases was compiled using criteria of: - Infectious diseases notifiable to The European Surveillance System in 2010 (TESSy) - Potential infectious threats to Europe that had been identified and monitored in the Threat Tracking Tool (the ECDC database for event-based surveillance) in June to September of 2005 to 2011 inclusive - Events reported in the Health Protection Agency's (HPA) weekly epidemiological reports from May to September 2011 - Diseases reportable to the World Health Organization (WHO) according to the International Health Regulations, infectious agents with deliberate release potential • A literature search provided supporting information on disease severity, incubation periods, transmissibility, routes of infection, geographical distribution, seasonality and distribution of vectors. • Generic team scored each disease on a risk matrix assessing: public health impact (1 lowest to 5 highest impact) and likelihood of occurrence (1 least likely to 5 most likely). • Public health impact included criteria of morbidity, case fatality rate, potential of sequelae, existence of disease-specific treatments, potential to provoke outbreaks and potential media interest. Likelihood took into account criteria of incidence, geographical distribution, seasonal trends, mode of transmission and incubation period. • Likelihood was scored according to 3 categories in the context of the Olympics: being imported into the Games; occurring at the Games (disease transmission during the games); and being exported from the Games to rest of the EU/EEA. • The 7 expert teams were asked only to assess likelihood of occurrence for diseases occurring at the games and being	• No weighting of criteria is described

Study	Country or region	Objectives	Methods	Weighting method
			exported from the Games. The teams received a list of diseases within their field of expertise, alongside data from TESSy and the Threat Tracking Tool, a summary of threats monitored in the HPA weekly epidemiological reports and supportive information the collated from literature search. The experts in each team discussed the score for the public health impact of each disease. • The scores assigned by the expert team were compared with those assigned by the generic team using the risk matrix. When there was a divergence in scores, they were revised. Delphi method was used to achieve consensus, through discussions between the generic team and each disease expert team separately.	
Havelaar[33]	The Netherlands	• To prioritise the threat from emerging zoonoses to support the developmen t of early warning and surveillance systems	• Emerging zoonoses were identified by searching the literature, including websites such as WHO, CDC, etc., and expert members of the Emzoo consortium were asked to suggest pathogens. • Five decision rules were applied to narrow down the number of pathogens for prioritisation, e.g. pathogens were excluded if the reservoir species is not present in Europe. • 86 emerging pathogens were selected as relevant to the Netherlands. • Seven criteria were used to score pathogens, these covered the probability of introduction, transmission in animal reservoirs, economic damage in animal reservoirs, animal-human transmission, transmission between humans, morbidity, mortality (case fatality). • Scenarios were then generated and ranked by a panel comprising, 7 risk managers, 11 infectious disease specialists and 11 medical and veterinary students. • The scenarios were offered in different sequences to the groups to avoid fatigue and were repeated across groups to test for inter-rater variation. The panellists were emailed a repeated scenario 2 weeks later to further test consistency. • Scoring against the criteria used evidence from published literature, where available. Where insufficient data was available, criteria were evaluated using 'simple decision rules'. Where those were unavailable, expert opinion was sought and the uncertainty this introduced was reflected in the score. • Data was aggregated across all groups and normalised to give a value between 0 and 1. • Sensitivity analyses were performed to assess the impact of the assumptions contained in the different scenarios used. • Aggregated scores were then categorised as low, middle and high importance using a classification and regression tree analysis (CART).	• Weights were assigned to criteria based on sessions with the panel
Horby[40]	UK	• To assess the relative priority of communica ble diseases and identify priority areas for work	• 58 diseases and 3 'generic disease groups' (infections in the immunocompromised, infections and chronic disease, infections in vulnerable groups) were considered for prioritisation (selection of these unclear). • The survey was distributed by post to 1130 various professionals including communicable disease experts, microbiologists, genitourinary medicine practitioners. • 518 survey were returned (46%). • 5 importance criteria were used based on present burden of ill-health, social and economic impact, potential threat to health (5 to 10 years), health gain opportunity, public concern and confidence. Respondents gave a score between 1 and 5 for each pathogen against each criteria. • 4 'areas for further work' were included in the survey covering diagnostic and specialist microbiology, surveillance, guidelines, evaluation of interventions. Respondents were asked to tick one or more area where further work is required for each disease. • Chi-squared test was used to test if response rate varied by professional group. • A mean score for each 'importance' criterion was calculated by summing all of the scores (range 1 to 5) then dividing that figure by the number of people who assigned a score (blank entries were disregarded). Diseases were ranked for each criterion based on this mean score. • Diseases were ranked based on their mean score (described above) across all 5 criteria. Respondents who	• No weighting of criteria

Study	Country or region	Objectives	Methods	Weighting method
			did not provide a score for all five criteria were excluded. A 95% confidence interval was calculated based on the standard error of the mean. • Spearman's rank was used to test for correlation between disease ranking based on each criterion and the overall ranking based on all 5 criteria. • For the 'areas for further work' the number of ticks for each area was calculated as a percentage of the total number of respondents. Respondents could tick more than one area, therefore percentages for individual diseases could exceed 100%.	
Humblet[25]	Europe	• To prioritise 100 animal diseases and zoonoses in Europe	• 100 diseases were included for prioritisation, selected because they were reportable to the World Organisation for Animal Health (OIE) and reported to the International Society for Infectious Diseases, considered an emerging threat. Salmonellosis was included because of its public health impact and prevalence. Foot-and-mouth diseases was included, despite its low effect on public health and the limited number of human cases. • 5 categories were used: epidemiology, prevention/control, effects on economy/trade, zoonotic characteristics and effect on society. These categories covered 57 individual criteria. These were based on a review by the Department for Environment, Food and Rural Affairs of previous priority setting exercises. • Criteria had coefficients on a scale of 0-7. • Information for each of the 57 criteria was obtained from reliable sources and authors report finding ≈100% of information. • 40 international experts, across a range of disciplines were selected. Experts gave a score (from 1 strong disagreement to 4 strong agreement) about the appropriateness of the criteria and were asked to explain any disagreement. • An overall weighted score was calculated. A global score for each criterion was calculated by multiplying the coefficient (0-7) for each criterion by the average intra-criteria weight allocated to that criterion by the 40 experts. The overall weighted score was calculated by multiplying the 6 multi-category experts' inter-category weighting was multiplied by the global score. • Uncertainty was estimated using a probabilistic method and combined using a Monte-Carlo simulation.	• The same group of experts weighted the criteria. The Las Vegas method was used to distribute values to each criteria according to their relative importance. Because categories had different numbers of criteria within them, a proportion of the total 100 points were available to distribute within each category to prevent categories with more criteria being scored lower. • Six additional experts were asked to apply the Las Vegas method to create inter-category weighting by distributing 100 points between the 5 categories.
Krause[21]	Germany	• To prioritise infectious diseases to inform resource allocation for research, surveillance and other activities	• A list of 85 pathogens was compiled which met criteria of being: notifiable under German law; reportable within the EU; listed as a chapter in established manuals and textbooks on infectious diseases; a causative agent of an outbreak reported to the RKI in the past 10 years; or an agent with potential for deliberate release. • A working group of 11 epidemiologists and infectious disease specialists scored the diseases (-1, 0, +1) according to 12 criteria covering burden of disease, disease trend, information need, and preventability and treatability, to give a criteria score. • The final score was calculated by multiplying the criteria score by the weighting score. • Unweighted scores were also calculated by summing the scores per pathogen.	• Criteria were ranked 1-12 according to importance, to provide a weighting score. • Weighting was done prior to and separately to the scoring.
McIntyre[17]	Not specified	• To conduct a rapid assessment of pathogen impact using the h-index	• The h-index, a bibliometric indicator of the number of articles and citations for a topic. • Used a previously generated list of 1414 infectious organisms (Taylor), including emerging pathogens. • To obtain the h-index score, authors searched for each pathogen on the Web of Science database, a subsample of these results were compared with h-index scores from Google Scholar and Scopus as a quality assurance measure. • H-index scores were compared with WHO-produced DALYs (considered a reliable measure of burden of disease) where available.	• Not applicable
Morgan[39]	United Kingdom	• To assess the threat level of emerging infections and inform the	• A qualitative risk assessment using an algorithm containing a series of decision questions. • The algorithm was developed based on two previous projects: a Department of Health funded 'HP Zone' project to develop a risk management model for managing communicable disease incidents; and a Department for Environment and Rural Affairs project as part of their	• Not applicable

Study	Country or region	Objectives	Methods	Weighting method
		prioritisation of resources	Veterinary Surveillance Strategy. • Separate algorithms were developed to consider the probability of and infectious disease threat in the UK, and the impact in terms of morbidity and mortality. • The stage on the algorithm where the assessment stops correlates with the level of risk. • Overall risk level is assessed as minimal, low, moderate or high. • The initial risk assessment is carried out by a single scientist, informed by information from published literature, formal and informal reports, and expert opinion. • This risk assessment is then passed to a multidisciplinary group of human and animal health experts for comment, prior to sign off by the National Expert Panel on New and Emerging Infections (NEPNEI). • In this study one pathogen (chikungunya) was used as a worked example to test the algorithm.	
Ng et al.[12-14]	Canada	• To identify criteria to prioritise zoonotic diseases for funding of control and prevention measures	• Three-part study consisting of: focus groups to identify prioritisation criteria; a questionnaire using conjoint analysis to determine the weights of criteria; disease ranking based on CA-derived criteria weighting • Six focus groups (54 participants) identified a list of prioritisation criteria. Participants were targeted to ensure that the groups reflected a range of professions and demographic groups and included members of the public. A nominal group technique was used to run the focus groups and all focus groups were conducted by the same facilitator. In the first half of the focus group sessions participants identified criteria for prioritisation, this was initially done individually then results were discussed as a group. In the second half of the session the group applied scores to the criteria they had identified, again this was done individually then the relative ranking of criteria was discussed as a group. In the focus group for members of the public a second round of scoring enabled participants to rescore criteria (the professional groups only had one round of scoring due to time constraints). • 300 iterations of a conjoint analysis questionnaire were distributed to professionals and members of the public in Canada and the US: - 1 313 (2.8%) of the targeted Canadian general public returned responses - 1 309 (15.8%) of the targeted US general public returned responses - 928 Canadian experts completed the survey (no data on numbers approached) - 998 US experts completed the survey (no data on numbers approached) • The CA scores were used to develop a point-scoring system for disease criteria. • Diseases were ranked based on the conjoint analysis score of each criteria, using data for each criteria from a literature search.	• Conjoint analysis was used to determine the importance of criteria and CA-derived scores were used to effectively weight criteria
Palmer[38]	United Kingdom	• To assess the zoonotic risk posed by emerging animal diseases by use of a qualitative algorithm.	• Five diseases were used as worked examples in the prioritisation algorithm. The list was compiled from a range of sources including those referred to national authorities by medical practitioners, US reports of endemic piglet neonatal enteritis and expert opinion of the advisory committee. • The algorithm is used to consider the key stages of zoonotic transmission: risk of cross-species transmission; exposure of humans to infected animals and secondary sources; human infection and subsequent human-to-human transmission. • The responses to questions in the algorithm are yes/no answers, and the 'stop' point then indicates the level of confidence for risk of zoonotic transmission ranging from level 0 – not zoonotic – to level 4 – confirmed zoonoses.	• Not applicable
WHO[22]	South-eastern Europe	• To prioritise areas of activity for investment in infectious diseases	• Modified Delphi approach • List of 53 diseases was taken from the EU list of communicable diseases (to save time). • Eight criteria for assessment of importance: disease impact, present burden of ill health, potential threat (5–10 years), necessity for immediate public health response, low incidence only maintained by current public health activities, long-term effects on communicable diseases, social and economic impact, health gain opportunity, public	• Not applicable

Study	Country or region	Objectives	Methods	Weighting method
			concern and confidence. • 24 workshop participants took part in the prioritisation exercise. • 1–5 scale (low to high) for ranking importance of criteria. • An overall mean score for each disease was calculated by summing the score for each criterion then dividing it by the number of participants. • 95% confidence intervals calculated to indicate the level of agreement showed varying levels of agreement across different conditions. • Only one round of scoring took place due to time constraints.	

Appendix 4: Quality appraisal summary table

Study	Methodology	Overall score	Individual domain scores			Reviewer comments
			Validity	Content validity	Relia-bility	
Balabanova	Delphi					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.
Cardoen	Semi-quantitative methodology (analysed as MCDA)					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.
Cox et al.	MCDA					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.
Cox	Bibliometric index					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.
Economopoulou	Delphi					Used two criteria of likelihood and impact. Assessment against content validity domain was based on the facets listed as included in the 'supportive information'; did not include many of those criteria. Implementation issues were not discussed. No measures of internal consistency.
Havelaar	MCDA					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.
Horby	Questionnaire					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.
Humblet	MCDA					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.

Study	Methodology	Overall score	Individual domain scores			Reviewer comments
			Validity	Content validity	Relia-bility	
Krause	Delphi					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.
McIntyre	Bibliometric index					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.
Morgan	Qualitative algorithm					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.
Ng et al.	Questionnaire					Implementation issues were not specifically discussed, but practical considerations were discussed which would assist implementation. Most of the key communicable disease facets were met. Internal consistency was not measured. The Delphi method reduces the effect of inter-rater variation because of discussion.
Palmer	Qualitative algorithm					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. This appeared to be a table-top exercise and it lacked tests of internal consistency and validity.
WHO	Delphi					Reporting lacked 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.

Met	Information related to that item has been clearly reported and all relevant considerations have been made.
Partly met	The information related to that item is incomplete, or not all aspects have been considered. The score is assigned depending on the completeness and quality of reporting, with increased score as more criteria are met and more considerations have been made.
Not met	There is no information provided in the study that is relevant to that item, or information related to that item is very poorly reported.
Not applicable	Where criteria are not applicable to a particular study or methodology.

Appendix 5: Excluded studies

Articles excluded at first sift

Study	Reviewer comments
European Technology Platform for Global Animal Health: Action Plan. Brussels: European Technology Platform for Global Animal Health; 2014.	not relevant – animal health
Exotic Animal Disease Risk Pathway & Countermeasures: Final Report. London: Department for Environment, Food and Rural Affairs; 2009.	not relevant – animals
Foresight. Infectious Diseases: Preparing for the future. Executive summary. London: Office of Science and Innovation; 2006.	not relevant – only mentions prioritising twice, difficulties ascertaining correct versions due to government website changes
NIAID describes research priorities for fighting drug-resistant tuberculosis. Home Healthcare Nurse. 2008;26(8):448-9.	not relevant – research priorities
Scientists' report outlines European priorities in tackling infectious diseases. Euro Surveillance. 2005 20050616 [Epub ahead of print];10(6):E050616.	meeting report
Adam-Poupart A, Labreche F, Smargiassi A et al. Climate change and Occupational Health and Safety in a temperate climate: potential impacts and research priorities in Quebec, Canada. [Review]. Industrial Health. 2013;51(1):68-78.	not relevant – research priorities
Akpogheneta O. Why HIV leads infectious disease priorities. The Lancet Infectious Diseases. 2011;11(7):502-3.	not relevant – general
Ashraf H. US infectious disease research leaders set out new priorities. The Lancet Infectious Diseases. 2002;2(11):651.	not relevant – research priorities
Boraschi D, Abebe AM, Aseffa A et al. Immunity against HIV/AIDS, malaria, and tuberculosis during co-infections with neglected infectious diseases: recommendations for the European Union research priorities. PLoS Neglected Tropical Diseases [electronic resource]. 2008 20080625 [Epub ahead of print];2(6):e255.	not relevant – research priorities
Borchardt SM, Ritger KA, Dworkin MS. Categorization, prioritization, and surveillance of potential bioterrorism agents. Infectious Disease Clinics of North America. 2006;20(2):213-25.	not relevant – Bioterrorism
Bots PWG, Hulshod JAM. Designing multi-criteria decision analysis processes for priority setting in health policy. J.Multi-Crit.Decis.Anal. 2000;9:56-75.	not relevant – general policy and did not have sufficient unique detail for discussion etc.
Brijnath B, Butler CD, McMichael AJ. In an interconnected world: joint research priorities for the environment, agriculture and infectious disease. Infectious Diseases of Poverty. 2014 20140128 [Epub ahead of print];3(1):2.	not relevant – research priorities
Brookes VJ, Hernandez-Jover M, Cowled B et al. Building a picture: Prioritisation of exotic diseases for the pig industry in Australia using multi-criteria decision analysis. Preventive Veterinary Medicine. 2014 20131022 [Epub ahead of print];113(1):103-17.	not relevant – animal health
Brookes VJ, Hernandez-Jover M, Neslo R et al. Identifying and measuring stakeholder preferences for disease prioritisation: A case study of the pig industry in Australia. Preventive Veterinary Medicine. 2014 20131024 [Epub ahead of print];113(1):118-31.	not relevant – animal health
Burnette WN, Hoke CH, Jr., Scovill J et al. Infectious diseases investment decision evaluation algorithm: a quantitative algorithm for prioritization of naturally occurring infectious disease threats to the U.S. military. Military Medicine. 2008;173(2):174-81.	not relevant – only discuss research priorities in this paper
Chisholm D, Baltussen R, Evans DB et al. What are the priorities for prevention and control of non-communicable diseases and injuries in sub-Saharan Africa and South East Asia? BMJ. 2012 20120302 [Epub ahead of print];344:e586.	not relevant – Out of geographical scope
Chowdhury FR, Bari MS, Alam MJ et al. Impact of climate change on prioritized infectious diseases in North-Eastern part of Bangladesh. International Journal of Antimicrobial Agents. 2013;Conference(var.pagings):June.	not relevant – about the impact of climate change on these diseases, no details of how they were prioritised
Del Rio Vilas VJ, Voller F, Montibeller G et al. An integrated process and management tools for ranking multiple emerging threats to animal health. Preventive Veterinary Medicine. 2013 20120903 [Epub ahead of print];108(2-3):94-102.	not relevant – animal health
Del Rio Vilas VJ, Montibeller G, Franco LA. Letter to the editor: Prioritization of infectious diseases in public halth: Feedback on the prioritization methodology, 15 july 2008 to 15 january 2009. Eurosurveillance. 2011;16(27).	letter
Del Rio Vilas VJ, Burgeno A, Montibeller G et al. Prioritization of capacities for the elimination of dog-mediated human rabies in the Americas: Building the framework. Pathogens and Global Health. 2013;107(7):340-5.	not relevant – specific to one disease
Del RV, V, Montibeller G, Franco L. Letter to the editor: Prioritisation of infectious diseases in public health: feedback on the prioritisation methodology, 15 July 2008 to 15 January 2009. Euro Surveillance. 2011 20110707 [Epub ahead of print];16(27).	letter

Study	Reviewer comments
Dujardin JC, Herrera S, do R, V et al. Research priorities for neglected infectious diseases in Latin America and the Caribbean region. [Review]. PLoS Neglected Tropical Diseases [electronic resource]. 2010 20101026 [Epub ahead of print];4(10):e780.	not relevant – research priorities
Duncan BB, Chor D, Aquino EM et al. [Chronic non-communicable diseases in Brazil: priorities for disease management and research]. [Portuguese]. Revista de Saude Publica. 2012;46 Suppl 1:126-34.	not relevant – non-communicable disease
Eaton L. Infectious diseases. Priorities set without nurses' input. Nursing Times. 2000;96(28):6-19.	not relevant – general process
Elmi M. Food safety: current situation, unaddressed issues and the emerging priorities. [Review] [30 refs]. Eastern Mediterranean Health Journal. 2004;10(6):794-800.	not relevant – not examining prioritisation methods
Fosse J, Seegers H, Magras C. [Prioritising the risk of foodborne zoonoses using a quantitative approach: application to foodborne bacterial hazards in pork and beef]. [French]. Revue Scientifique et Technique. 2008;27(3):643-55.	non-English
Fuller T, Bensch S, Muller I et al. The ecology of emerging infectious diseases in migratory birds: an assessment of the role of climate change and priorities for future research. [Review]. Ecohealth. 2012 20120225 [Epub ahead of print];9(1):80-8.	not relevant – research priorities
Garcia NA, Medina BG, Reinares Od, V. [Emerging zoonoses linked to pets in the autonomous community of Madrid: design of a method for setting public health priorities, Spain]. [Spanish]. Revista Espanola de Salud Publica. 2004;78(3):389-98.	non-English not relevant –
Gilsdorf A, Krause G. Authors reply: Prioritisation of infectious diseases in public health: feedback on the prioritisation methodology, 15 July 2008 to 15 January 2009. Euro Surveillance. 2011 20110707 [Epub ahead of print];16(27).	incorrect design – comments
Gilsdorf A, Krause G. Letter to the editor: Prioritisation of infectious diseases in public health: Feedback on the prioritisation methodology, 15 july 2008 to 15 january 2009. Eurosurveillance. 2011;16(27).	letter
Harley D, Bi P, Hall G et al. Climate change and infectious diseases in Australia: future prospects, adaptation options, and research priorities. [Review]. Asia-Pacific Journal of Public Health. 2011;23(2 Suppl):54S-66.	not relevant – research priorities
Hooker C. Health scares: Professional priorities. Health: an Interdisciplinary Journal for the Social Study of Health, Illness & Medicine. 2010;14(1):3-21.	not relevant – health scares
Horiguchi I, Kishiwagi T, Marui E. [Which infectious diseases should be prioritized in educating Japanese population?]. [Japanese]. Kansenshogaku Zasshi - Journal of the Japanese Association for Infectious Diseases. 2008;82(2):67-72.	non-English
Hotez PJ, Remme JH, Buss P et al. Combating tropical infectious diseases: report of the Disease Control Priorities in Developing Countries Project. Clinical Infectious Diseases. 2004 20040226 [Epub ahead of print];38(6):871-8.	Incorrect design - does not describe a clear and repeatable process
Jafar TH, Haaland BA, Rahman A et al. Non-communicable diseases and injuries in Pakistan: strategic priorities. Lancet. 2013 20130517 [Epub ahead of print];381(9885):2281-90.	not relevant – non-communicable disease
Jaffar S, Amberbir A, Kayuni N et al. Viewpoint: scaling up testing services for non-communicable diseases in Africa: priorities for implementation research. Tropical Medicine & International Health. 2013 20130904 [Epub ahead of print];18(11):1353-6.	not relevant – non-communicable disease
Jehu-Appiah C, Baltussen R, Acquah C et al. Balancing equity and efficiency in health priorities in Ghana: The use of multicriteria decision analysis. Value in Health. 2008;11(7):1081-7.	not relevant – overall health priorities
Joseph J, Kopnisky KL, Nunn M. NeuroAIDS research in resource-limited countries. Emerging priorities of the US National Institute of Mental Health and the National Institute of Neurological Disorders and Stroke. Journal of NeuroVirology. 2005;11(SUPPL. 1):1-3.	not relevant – not priority setting process
Karnon J, Goyder E, Tappenden P et al. A review and critique of modelling in prioritising and designing screening programmes. [Review] [285 refs]. Health Technology Assessment (Winchester, England). 2001;11(52):iii-iiv.	not relevant – screening across all disease types
Khazei A, Jarvis-Selinger S, Ho K et al. An assessment of the telehealth needs and health-care priorities of Tanna Island: A remote, under-served and vulnerable population. Journal of Telemedicine and Telecare. 2005;11(1):35-40.	not relevant – general healthcare priorities
Lopez A. Health and health-research priorities: has WHO got it right? Lancet. 2008;372(9649):1525-7.	not relevant – health priorities in general
MacDonald R. Prioritising neglected diseases related to poverty. BMJ. 2005;331(7507):12.	not relevant – overall health priorities
Maher D, Ford N, Unwin N. Priorities for developing countries in the global response to non-communicable diseases. Global Health. 2012 20120611 [Epub ahead of print];8:14.	not relevant – non-communicable disease
Malik P. Misplaced priorities. Canadian Journal of Cardiology. 2007;23(4):273.	not relevant – overall priorities

Study	Reviewer comments
Mangen MJ, Batz MB, Kasbohrer A et al. Integrated approaches for the public health prioritization of foodborne and zoonotic pathogens. Risk Analysis. 2010 20090917 [Epub ahead of print];30(5):782-97.	not relevant – General - discussion of potential methods
McAnulty J, Stewart K, NSW Public HN. Priorities for communicable disease control in New South Wales, 2003. New South Wales Public Health Bulletin. 2003;14(9-10): 200-5.	incorrect design - doesn't describe the process of prioritisation
Oei W, Neslo R, Janssen M. Prioritizing emerging infectious diseases(EIDS) for transfusion safety through expert elicitation. Transfusion. 2013;Conference(var.pagings):September.	not relevant – too specific to transfusion
Ordunez P. Cardiovascular health in the Americas: facts, priorities and the UN high-level meeting on non-communicable diseases. MEDICC review. 2011;13(4):6-10.	not relevant – non-communicable disease
Parry C, Ferreira-Borges C, Poznyak V et al. The international study on alcohol and infectious diseases: three priorities for research. Addiction. 2013 20121017 [Epub ahead of print];108(1):1-2.	not relevant – research priorities
Paterson DL, Steering Group of the Australasian Society for Infectious Diseases Clinical Research Network. Determining research priorities for clinician-initiated trials in infectious diseases. Medical Journal of Australia. 2013;198(5):270-2.	not relevant – research priorities
Ratnasingham S, Kwong J, Daneman N et al. Ranking the burden of infectious diseases in Ontario, Canada. International Journal of Infectious Diseases. 2014;Conference(ICID):Miami-asteur.	not relevant – burden of disease calculation only
Shiffman J. Donor funding priorities for communicable disease control in the developing world. Health Policy & Planning. 2006 20060918 [Epub ahead of print];21(6):411-20.	not disease priority setting
Soanes L, Gibson F, Bayliss J et al. Establishing nursing research priorities on a paediatric haematology, oncology, immunology and infectious diseases unit: a Delphi survey. European Journal of Oncology Nursing. 2000;4(2):108-17.	not relevant – research priorities
Soanes L, Gibson F, Hannan J et al. Establishing nursing research priorities on a paediatric haematology, oncology, immunology and infectious diseases unit: involving doctors and parents. European Journal of Oncology Nursing. 2003;7(2):110-9.	not relevant – research priorities
Steffen R, Connor BA. Vaccines in travel health: From risk assessment to priorities. Journal of Travel Medicine. 2005;12(1):26-35.	not relevant – vaccines
Stewart WC. Veterinary public health: Medical and veterinary general practitioner research priorities in Scotland. Veterinary Record. 2008;163(12):367-8.	not relevant – research priorities
Temajo NO, Howard N. The mosaic of environment involvement in autoimmunity: the abrogation of viral latency by stress, a non-infectious environmental agent, is an intrinsic prerequisite prelude before viruses can rank as infectious environmental agents that trigger autoimmune diseases. [Review]. Autoimmunity Reviews. 2014 20140111 [Epub ahead of print];13(6):635-40.	not relevant – autoimmune diseases
Unwin N. Commentary: Non-communicable disease and priorities for health policy in sub-Saharan Africa. Health Policy and Planning. 2001;16(4):351-2.	incorrect design - commentary
Unwin N. Non-communicable disease and priorities for health policy in sub-Saharan Africa. Health Policy & Planning. 2001;16(4):351-2.	not relevant – non-communicable disease
Valenciano M. Setting priorities for non-foodborne zoonoses. Medecine et Maladies Infectieuses. 2001;31(SUPPL. 2):302-4.	non-English
Wallace C, Marich A, Turahui J et al. Priorities for planning and prevention of infectious disease outbreaks at large international events. Australian & New Zealand Journal of Public Health. 2007;31(5):491-2.	not relevant – too specific - no abstract so excluded
Walsh PD. A rant on infectious disease and ape research priorities. American Journal of Primatology. 2008;70(8):719-21.	not relevant – research priorities
Watts J. G8 countries set priorities for infectious diseases but fail to make progress on debt relief. Bulletin of the World Health Organization. 2000;78(9):1168.	news article
World Health Organization. Research priorities for the environment, agriculture and infectious diseases of poverty. World Health Organization Technical Report Series. 2001;(976):i-xiii.	not relevant – research priorities
World Health Organization. Research priorities for zoonoses and marginalized infections. World Health Organization Technical Report Series. 2001;(971):ix-xi.	not relevant – research priorities

Articles excluded at second sift

Study	Reviewer comments
Cediel N, Villamil LC, Romero J et al. Setting priorities for surveillance, prevention, and control of zoonoses in Bogota, Colombia. Pan American Journal of Public Health. 2013;33(5):316-24.	Not relevant – outside geographic scope - Colombia
Doherty JA. Establishing priorities for national communicable disease surveillance. Canadian Journal of Infectious Diseases. 2000;11(1):21-4.	Incorrect design - insufficient detail for appraisal and analysis - based on full text
Kemmeren JM, Mangen M-J, Duynhoven AH et al. Priority setting of foodborne pathogens: disease burden and costs of selected enteric pathogens. Bilthoven, Holland: RIVM; 2006.	Not relevant – establishing disease burden and costs, as part of wider priority setting process
Kurowicka D. Probabilistic Inversion in Priority Setting of Emerging Zoonoses. Risk Analysis. 2010;30(5):715-23.	Not relevant – a theoretical model
McNulty CA, Smith GE, Graham C et al. PHLS primary care consultation--infectious disease and primary care research and service development priorities. Communicable Disease & Public Health. 2001;4(1):18-26.	Not relevant – general priorities, not ranking risk based on burden of disease and NHS potential to affect the burden of disease
Pavlin BI. A Standardized Process for Developing a National Notifiable Diseases List in a Pacific Island Setting. Asia Pac J Public Health. 2009;22(3):279-88.	Not relevant – ranking for notifiable disease listing
Walsh AL, Morgan D. Identifying hazards, assessing the risks. The Veterinary Record. 2005;157:684-7.	Not relevant – found when searching for a more detailed version another HAIRS report; brief version of a report, does not add anything

Appendix 6: Quality appraisals for studies included in the analysis

Quality appraisals are available upon request.