Safety Science xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

A systems approach to risk analysis validation for risk management

John Lathrop [a,*], Barry Ezell [b]

[a] Innovative Decisions, Inc., 1905 Milano Way, Mountain View, CA 94040, United States
[b] Old Dominion University, Virginia Modeling, Analysis and Simulation Center, United States

ARTICLE INFO

Article history:
Received 28 October 2016
Received in revised form 6 April 2017
Accepted 18 April 2017
Available online xxxx

Keywords:
Risk analysis
Risk analysis validation
Trust
Risk management
Systems approach
Culture of analysis quality

ABSTRACT

This paper presents a logical structure to address the topic of this special issue: Risk Analysis Validation and Trust in Risk Management. We do that by presenting a systems approach that links all four of those concepts. The underlying logic: Validation should test how effectively a risk analysis supports actual, real-world implemented risk management. Our approach is based on a flowchart linking all of the elements from inputs through risk analysis, risk reporting and transparency, then how that reporting-transparency support the risk management decision making process and associated third party and stakeholder reviews (formal or informal), which in turn determine the trust and acceptance necessary for the real-world implementation of risk management actions. We take that flowchart and identify within it sixteen critical elements, then specify a validation test for each of those elements. Validation, then, consists of subjecting the risk analysis to those sixteen tests. Those tests, together, test the risk analysis for how effectively it supports implemented risk management. Another key feature: We divide the flowchart into Analysts' Domain, Users' Domain, and Analysis Community Domain. The Analysts' Domain is where the risk analysts work, then the Users' Domain stands between their work and implementation. The Analysis Community Domain is comprised of the communities of risk analysts and commissioners of risk analyses. Those two communities are where we would, as part of building our systems approach to risk analysis validation, build a "Culture of Analysis Quality," where the sixteen validation tests would be enforced by both of those communities.

1. Introduction

Risk Analysis shows up in our lives in several arenas. In many of those arenas, e.g. consumer product safety, medical treatment strategies, siting of hazmat facilities, routing of hazmat transport (rail, pipeline, truck), nuclear power and many more, risk analysis does not show up as a set of calculations, but shows up as support for arguments on one side or another (or both) of vigorous public debates over actions, regulations, laws and policies. In those cases the effectiveness of a risk analysis depends on a great deal more than what is typically covered in "Verification and Validation" (Goerlandt et al., 2016; Aven and Heide, 2009; Sargent, 2013; Petty, 2010; Department of Defense, 2008; United States Coast Guard, 2006). An analysis can be fully verified and validated in a purely analytic sense, yet still be ineffective because it is not accepted and trusted in the public debate it is to support. In particular, if one side of the debate can credibly cast doubt on the risk analysis, its role can be markedly limited. So what we have, there, are cases where the definition of "Validation" should be extended beyond a solely analytic test of the risk analysis, to concepts of validation covering the effectiveness of the risk analysis in the debate it is to support. That, in turn,

* Corresponding author.
E-mail addresses: jlathrop@innovativedecisions.com (J. Lathrop), bezell@odu.edu (B. Ezell).

http://dx.doi.org/10.1016/j.ssci.2017.04.006
0925-7535/

calls upon us to adopt a systems approach to risk analysis validation – extending to tests of achieving trust and acceptance in the applicable public debate. This paper presents such a systems approach.

That systems approach has several implications. The most important one is that the duty of the risk analyst is not only to conduct all calculations in a valid and validly scoped way, but also to design his or her analysis specifically to most effectively couple with downstream elements standing between the risk analysis and its effectiveness in the real-world risk management process. What matters, at the end of the day, is the risk management that actually occurs, and that risk management is the result of a system of elements, only some of which are the analytic elements of risk analysis.

That reasoning is based on the definition of validation presented in ISO 15288: "Confirmation, through ... objective evidence, that the requirements for a specific intended use ... have been fulfilled" (International Organization for Standardization, 2015). While that definition is not specifically concerning risk analysis/management, it applies at the more general level of validation of systems approaches, which is the perspective taken in this paper. We add to that the obvious point that in the case of risk analysis for risk management, the specific intended use is to support the risk management involved, that is, the risk management decisions involved. That scope reflects the scope of this special issue, risk analysis in risk management. As Rae et al. (2014) point out, risk assessment "is used in many domains for many different purposes." That statement clearly applies more broadly to risk analysis as discussed here. We have written this paper specifically to apply to all uses of risk assessment in its many domains, focusing not on the substantive domains but on supporting the risk management decisions involved. Furthermore, as will become clear later in this paper, we describe validation tests in terms that apply equally well to supporting any risk management decisions in any risk domain.

To extend that point to a higher level: The scope of this paper extends to all risk analyses in support of risk management decisions. Sections of this paper discuss scenarios and adversary decisions, but those sections do not have the effect of limiting the scope of this paper to risk analyses based on explicit lists of scenarios, or risk analyses involving adversary decisions. As we discuss later, even risk analyses not based on explicit lists of scenarios should be examined with validation tests that ask, at a high conceptual level, whether or not all significant scenarios and/or scenario-like processes have been adequately considered in terms of initiation, unfolding and completeness.

Shortly we will present a graphic, Fig. 1, that presents all of the elements and relationships we have mentioned above. Then after explaining that graphic we will map each of the sixteen elements in the analysts'-domain part of Fig. 1 to a validity test, worded as a question. Those sixteen elements in the analysts' domain are so central to the logic of this paper, in Fig. 1 we have colored them a distinctive green color. Each test is presented paired with a discussion of the shortfalls associated with failures to pass that test. We note, in advance, that the list of tests is long – sixteen tests, one for each analysts'-domain element. We make no apologies for that. The fact of the matter is that those sixteen analysts'-domain elements operate as a system to support real-world risk management, in ways depicted in Fig. 1. So once we define validation as we have here, in terms of how effectively it supports risk management, we are forced to recognize that validation must involve many considerations, and so many tests.

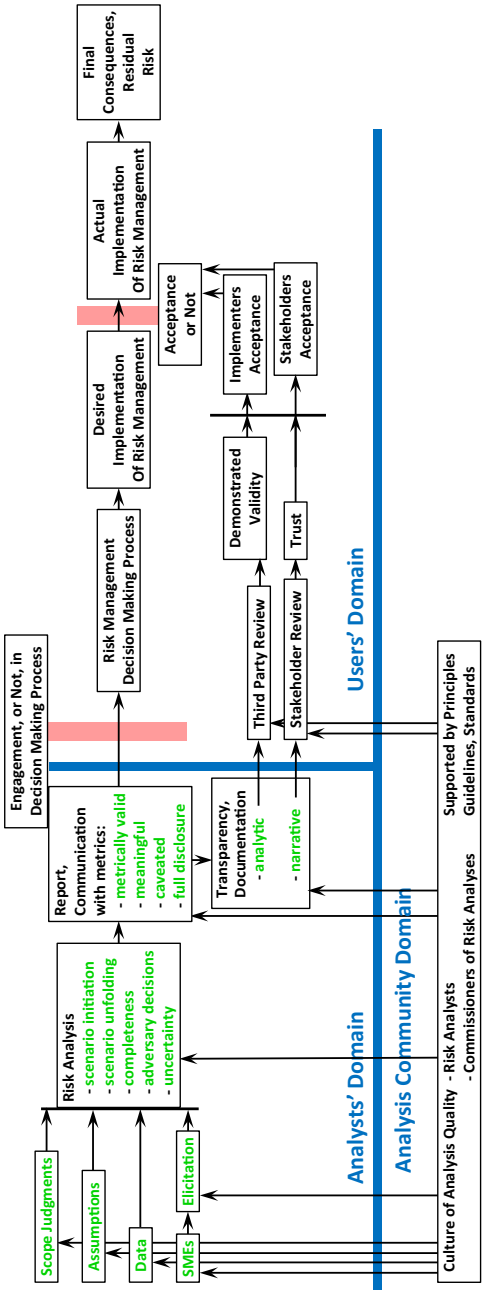

Fig. 1. Risk analysis validation – a systems perspective. Green font denotes the sixteen analysts'-domain elements that form the basis of the logic of the paper.

Please cite this article in press as: Lathrop, J., Ezell, B. A systems approach to risk analysis validation for risk management. Safety Sci. (2017), http://dx.doi.org/10.1016/j.ssci.2017.04.006

But before we present and discuss those Fig. 1 elements, we need to present the basis for our many statements. That basis is, simply, our experiences with many risk analyses. Table 1 presents eleven terrorism risk assessments as one subset of our experience, but of course, with a combined 70 years of experience between us, we have conducted, modified, added to or reviewed-critiqued many more risk assessments than that, in fact a total of 48 risk assessments (Lathrop and Ezell, 2016). In addition, one of us is coordinator and majority author of a document under development in the Society for Risk Analysis: Principles, Guidelines and Core Knowledge for Analytic Support of Risk Management. Every statement in this paper is based on some part or parts of that extensive experience. Every concept we present here could be linked back to one or more particular risk analyses in our experience base. We will not present those linkages, of course, due to considerations of confidentiality and sensitivity. This is not a paper embedded in an exhaustive survey of all available refereed literature on the subject. Rather, all concepts presented here are experience-based, as "reports from the trenches." To complement that orientation, we refer the reader to an excellent state-of-the-art review of validation of safety-related quantitative risk analysis, by Goerlandt et al. (2016).

While we present a long list of considerations and tests of validity here, we do not maintain that this is an exhaustive list. Rather, this list is gleaned from our experience as considerations that have occurred with special significance for the effectiveness (or ineffectiveness) of the real-world implemented risk management involved.

Finally, the value of this paper lies not in an extensive, deep support of each of the concepts we list, but in its system-level overview. A full treatment of each concept would result in a paper far in excess of the word count limits of this special issue. As a consequence, each concept is treated only at a very high level.

2. The overall logic of our systems approach to risk analysis validation: Fig. 1

Now we present Fig. 1. After an overview we will present and discuss each of the sixteen analysts'-domain elements, indicated in green font in Fig. 1, in turn. First, we do not maintain that Fig. 1 is the only or the best way to present the elements involved and their relationship. We do maintain that it is a sound basis for the discussion that follows. In particular there is one widely accepted structure at a level similar to Fig. 1 found in ISO 31000:2009 (International Organization for Standardization, 2009). Future work aligning Fig. 1 with ISO 31000:2009 would generate valuable insights and aid substantially in communication with risk analysis practitioners.

Fig. 1 presents the logical flow from inputs to the risk analysis on the left to the final consequences on the right. To reiterate: All that matters, in the end, are the final consequences on the right in Fig. 1, the residual risk remaining after risk management as it is actually applied in the real world. The fact that risk analysis is only one box in the larger sequence of Fig. 1 is a key message of this paper. It follows that validation of risk analysis should be framed in the larger sequence of Fig. 1. The figure is largely self-explanatory, so we won't step through it box by box here, but rather will present the overall structure. First, note the two pink boundaries, which represent the key barriers between risk analysis

Table 1
Authors' experience in terrorism risk assessment.

WMD Terrorism Risk Assessment Model (TRAM)	Customer	Description of work
Bioterrorism Risk Assessment (BTRA)[a,b,c,d]	Dept. of Homeland Security Science & Technology	Third Party Review Project Manager: Lathrop
Radiological Nuclear Terrorism Risk Assessment (RNTRA)[a]	DHS Domestic Nuclear Detection Office	Third Party Review Project Manager: Ezell
Risk Assessment Process for Informed Decision Making (RAPID)[e]	DHS Strategy, Planning, Analysis and Risk	Reviewed, then recommended methodological enhancements. Project Manager: Ezell
Global Biological Threat Priority Model (GBTP)[f]	Dept. of State, through Sandia National Laboratories	Verified and validated this multiple objective decision model. (Ezell only)
Maritime Security Risk Analysis Model (MSRAM)[g]	US Coast Guard	Third Party Review (Lathrop) Recommending enhancements
Models below developed/being developed, or analysis conducted by, the first author in his support of national laboratories.		
Modeling the Adversary for Responsive Strategy (MARS): Evaluates defensive measures vs. WMDs & IEDs on a "Common Yardstick."	Department of Energy, through Lawrence Livermore National Laboratory (LLNL)	Developed model, including Adaptive Adversary function, Dueling Portfolio framework.
Make, Buy, Steal (MBS): Compares IND risks by types, sources, to US targets. Compares defensive strategies.	" " "	Developed model, with three different adversary model elements.
Matthew Bunn: A mathematical model of the risk of nuclear terrorism[h]	" " "	Compared 4 Nuclear TRAMS: 2 listed above, RNTRA & MBS, then Bunn, Mueller, and the Lugar Survey[j].
John Mueller: IND risk model presented in his book, *Atomic Obsession*[i]	" " "	
Mosaic Effect: Evaluating IND disclosures	" " "	Developing model.
Model cannot be named here, name is sensitive information.	DOE, through three national laboratories	Supported ongoing development including adversary modeling.

[a] Pillai (2011).
[b] Ezell and von Winterfeldt (2009).
[c] Ezell and Collins (2010).
[d] National Research Council (2008).
[e] Ezell et al. (2011).
[f] Caskey et al. (2013).
[g] U.S. Coast Guard (2017).
[h] Bunn (2006).
[i] Mueller (2010).
[j] Lugar (2005).

Please cite this article in press as: Lathrop, J., Ezell, B. A systems approach to risk analysis validation for risk management. Safety Sci. (2017), http://dx.doi.org/10.1016/j.ssci.2017.04.006

and final consequences. The first boundary is between the risk analysis report, which is simply a document, and the actual risk management decision making process, which is where the actual risk management starts to happen. There is one more barrier, between the desired implementation of risk management, a product of the risk management decision making process, and what actually happens. That barrier is the one of acceptance by implementers and stakeholders. We structured Fig. 1 this way to explicitly present the role of trust in risk management, in keeping with the topic of this special issue. Those two barriers are downstream of the risk analysis, but in fact the validity of the risk analysis, as we have defined that here, depends upon how effectively the risk analysis takes into account those two downstream barriers. How much good would a Nobel Prize winning risk analysis be if it did not take into account the engagement with the decision making process, and the critical barrier of acceptance? Would such a risk analysis be valid, as we have defined validity here? Not necessarily. If it is effective across both barriers, yes. If it fails to cross either of those barriers, no.

Which brings us to the other large structural feature of Fig. 1, the blue boundaries. They separate the Analysts' Domain, the Users' Domain and the Analysis Community Domain. This paper will focus on the Analysts' Domain because that is where the risk analysis is executed, but we will also discuss the critical roles played by the Users' Domain and the Analysis Community Domain. The impacts of those different domains are captured in the arrows crossing the blue boundaries.

Consider the three arrows crossing the boundary from Analysts' Domain to Users' Domain. Analysts can only directly affect the boxes in the Analysts' Domain, but the effectiveness of the risk analysis is determined by:

- First, how effectively the risk analysis engages the risk management decision making process (top arrow), which is determined by the effectiveness of the report including all the considerations listed in the "Report" box.
- Second, that effectiveness is determined by the transparency and documentation of the analysis as that supports any third party review (by methodological experts, second arrow).
- Third, that effectiveness is determined by the transparency and documentation of the analysis as that supports any stakeholder review (by stakeholders, not methodological experts, third arrow).

The other ten, vertical, boundary-crossing arrows capture the role of the Analysis Community Domain, which we present as a single box, Culture of Analysis Quality. That culture affects all eight boxes and sixteen elements in the Analysts' Domain, and two of the boxes in the Users' Domain. With that background, we will now present tests of risk analysis validity box by box and element by element, simply moving from left to right in Fig. 1, then consider the roles of Culture of Analysis Quality.

3. Analysts' domain: Tests of risk analysis validity, element by element in Fig. 1

Fig. 1 presents sixteen elements of risk analysis in its Analyst's Domain section. Those sixteen elements are the five boxes on the left (Scope Judgments, Assumptions, Data, SMEs and Elicitation), then the five dash-list elements under Risk Analysis (scenario initiation, scenario unfolding, completeness, adversary decisions and uncertainty), the four dash-list elements under Report, Communication with metrics (metrically valid, meaningful, caveated and full disclosure) and the two dash-list elements under Transparency, Documentation (analytic, narrative). As mentioned earlier, these sixteen elements are colored a distinctive green color in Fig. 1. Our basic premise is that a risk analysis validation, at the systems level discussed here, should be conducted by critically posing and investigating sixteen questions, one for each of the sixteen elements just listed. So this section simply steps through those sixteen elements, from left to right in Fig. 1, discussing for each one: (1) A validation test, worded as one or more questions; (2) Shortfalls, one or more, corresponding to the issues associated with the test.

3.1. Scope judgments

3.1.1. Test

Was the scope set in a fully and explicitly considered, transparent and documented process, considering the implications of at least one alternative scope? Then was the scope presented, and summaries of the key issues associated with the scope presented, in immediate proximity to the results, with full disclosure of the implications of the scope for the results, including how those results would be apt to differ from results based on an assessment with a different scope?

3.1.2. Shortfall

The scope of an analysis is often simply presumed or announced, when in fact there may be alternative scopes, and which scope is chosen can have a very important effect on the results. One example in our experience: Two terrorism risk assessments, each assessing the risk of the same type of weapon of mass destruction. The two assessments made different scope assumptions, involving geographic areas considered, the materials considered, the range of effects of the weapons considered, and the adversary groups considered. The two assessments produced very different results, yet a reader of each of the two reports would not be aware, from each report itself, that the results depended upon the scope, or what the considerations of the scope were, and how the results would be apt to differ with a different scope.

3.2. Assumptions

This discussion closely parallels the discussion just presented concerning scope.

3.2.1. Test

Were the assumptions made set in a fully and explicitly considered, transparent and documented process, considering the implications of at least one alternative assumption set? Then were the assumptions presented, and summaries of the key assumptions presented in immediate proximity to the results, with full disclosure of the implications of the assumptions for the results, including how those results would be apt to differ from results based on a different assumption set?

3.2.2. Shortfall

The assumptions of an analysis are often simply presumed or announced, when in fact there may be alternative assumption sets, and which assumption set is chosen can have an important effect on the results. One example in our experience: One major terrorism risk assessment made two very key assumptions, out of necessity concerning computation time. Those assumptions had important implications for the relative risks of several alternative weapons considered, and while the sizes of the differences in those relative risks could not be assessed within the scope of the study, the directions of those differences were known from first principles. So the readers of the report could have been advised of those assumptions and the directions of differences in relative risks of

Please cite this article in press as: Lathrop, J., Ezell, B. A systems approach to risk analysis validation for risk management. Safety Sci. (2017), http://dx.doi.org/10.1016/j.ssci.2017.04.006

the alternative weapons if those assumptions were changed, but the readers were not so advised.

3.3. Data

In any risk assessment we have been aware of, there has never been an ideal data set or sets. The data sets used always have known shortfalls relative to an ideal data set or sets.

3.3.1. Test

Were the effects of the data sets on the results fully considered and disclosed? That is, were the effects of the ways in which the actual data sets used differed from ideal data sets, were those effects fully considered and disclosed? And were summaries of those effects presented in immediate proximity to the results? Those effects could include any of a number of things, including for example biases in known directions, or overstatement or understatement of the uncertainty involved.

3.3.2. Shortfall

While it is usually the case that the risk analysts are aware of the effects of the non-ideality of the data sets on the results, it is often the case that those effects are not adequately communicated to the readers of the analysis results.

3.4. SMEs, subject matter experts

The issue of SMEs is essentially the same as that of data. We separate it out here because it has important differences that call for separate attention.

3.4.1. Test

Was the process of selecting SMEs systematically conducted, and that process documented in a complete and transparent way? Were the possible distorting effects of the set of SMEs that were selected considered and disclosed in a complete and transparent way? Were summaries of those effects presented in immediate proximity to the results?

3.4.2. Shortfall

As is the case with data (see above), while it is usually the case that the risk analysts are aware of the effects of the non-ideality of the set of SMEs on the results, it is often the case that those effects are not adequately communicated to the readers of the analysis results.

3.5. SME elicitations

3.5.1. Test

Were the SME elicitations conducted in a way consistent with currently recognized best practice? Was the problem of SME overconfidence explicitly addressed? Were the results of elicitations combined across SMEs and if so, what method was used for that combination? Whether that combination included equal or differential weighting of each SME, what is the justification for that weighting? Were the possible distorting effects of the weighting (or not weighting, i.e. equal weighting) in aggregation across SMEs considered and disclosed in a complete and transparent way? Were summaries of those effects presented in immediate proximity to the results? (Hubbard, 2010).

3.5.2. Shortfalls

SME elicitation, associated known SME biases and the combination of results of the elicitations of multiple SMEs can have a very pronounced effect on the results. As is the case with data and SMEs (see above), while it is usually the case that the risk analysts are aware of the biases and effects of the SME elicitations and any combinations of SME elicitations, it is often the case that those effects are not adequately communicated to the readers of the analysis results.

3.6. Risk analysis

It may strike some readers as strange to have the topic of "Risk Analysis" consigned to one section of this paper. But in fact, in our experience the five elements (boxes) upstream of Risk Analysis in Fig. 1, and the elements downstream of Risk Analysis, are just as important in terms of effects on validity (as defined here) as variations in practice within the risk analysis itself. Then within risk analysis, the elements of concern are the five listed in the Risk Analysis box and are here considered below, in turn. That is, there are many elements involved in risk analysis, but in our experience, the five elements listed here are the elements that most significantly result in shortfalls of analysis.

3.6.1. Scenario Initiation, completeness

3.6.1.1. Test. To the extent that the risk generating mechanism can be represented in terms of lists of scenarios, are all the mechanisms for scenario initiation fully considered and accounted for? Here we note that some risk analyses are not based on explicit lists of scenarios. Examples include dose-response relations, ordinal rankings and procedural approaches. We maintain that, even in those cases, the analysis should be subjected to a version of this test since, most fundamentally, at some conceptual level any risk analysis must capture the risk generating mechanism and that mechanism, at some conceptual level, involves some version of a list of one or more scenarios and/or scenario-like processes.

3.6.1.2. Shortfall. A key challenge with this issue is one of imagination – attempting to characterize an effectively complete set of all of the scenario initiations. By effectively complete, we mean complete enough to characterize the risk adequately enough to adequately support the risk management process the analysis is to support. Kaplan (1997) gives us a clear language to use here. He defines scenario s in companionship with likelihood l and consequence x where $\langle s_i l_i x_i \rangle$ as one risk, $\{\langle s_i l_i x_i \rangle\}$ a set of risks and $\{\langle s_i l_i x_i \rangle\}_c$ the complete set. Aven (2016) provides an insightful perspective on the problem of completeness and how best to address it, including procedural measures to improve completeness, the role of specificity in scenario definitions, and systematic analyses of precursors. Striving for completeness can be quite challenging. At times the lack of completeness is not so much a lack of imagination as it is a deliberate but perhaps not fully considered limitation in scope. For example, the NRC did not address the risk of multiple reactor failures on one site, and then Fukushima happened (World Nuclear Association, 2017). This is a significant issue, in that it may cause risks in some cases to be underestimated, and it may cause estimates of the risk reductions of considered measures to be miss-assessed.

3.6.2. Scenario Unfolding, completeness

3.6.2.1. Test. To the extent that the risk generating mechanism can be represented in terms of lists of scenarios, are all the mechanisms by which a scenario can unfold, after initiation, fully considered and accounted for? The comments of Section 3.6.1 concerning analyses not based on explicit lists of scenarios apply here as well.

3.6.2.2. Shortfall. This is a concern parallel to initiation completeness, just discussed. All the points presented there apply here also, and will not be repeated. This consideration is listed separately because the issue of imagination applies differently. That is, it is

Please cite this article in press as: Lathrop, J., Ezell, B. A systems approach to risk analysis validation for risk management. Safety Sci. (2017), http://dx.doi.org/10.1016/j.ssci.2017.04.006

a different process to imagine all significant ways a scenario can unfold after initiation, than to imagine all significant ways a scenario can be initiated. Examples: A terrorist group acquiring not just one but several weapons of mass destruction, and perhaps using them, or using the second through Nth ones, for blackmail. Or a biological agent attack where, through contagion, the impacts extend to multiple cities. There may be risk assessments for those cases that account for those branchings, but at one point in the past there were risk assessments that did not account for those branchings.

3.6.3. Explicitly assessing completeness

This consideration may seem redundant with Sections 3.6.1 and 3.6.2 just presented. But those two considerations are focused on attempting to attain completeness. This consideration is focused on explicitly assessing the degree to which that completeness is attained, and assessing and reporting the implications of how much the issue of falling short of completeness may affect the risk analysis's support of the risk management decision making process. This may seem an impossible consideration. How, after all, is one to assess the importance of scenarios that haven't been imagined? But some risk spaces are more able to be completely imagined than others. For example, the set of all possible improvised nuclear devices may be more tightly constrained by the laws of physics than the set of all possible biological weapons. Note that the comments of Section 3.6.1 concerning analyses not based on explicit lists of scenarios apply here as well. The issue of completeness is a more general one than simply the completeness of any explicit list of scenarios.

3.6.3.1. Test. Is the problem of completeness explicitly considered and assessed, and then if that is a problem, are the implications adequately assessed and presented to the decision makers?

3.6.3.2. Shortfalls. Lack of explicitly assessing completeness and reporting the implications of shortfalls in completeness can have a significant impact on the effectiveness of a risk analysis. Completeness directly affects the relative attractiveness of risk management strategies centered on robustness and resilience. Risk management decision making operating without adequate understanding of the degree of completeness and its implications can make significantly suboptimal decisions, and those decision makers may not even be aware of that suboptimality. This issue is problematic, since in many cases, it could be argued that completeness is simply impossible. In fact mishaps "not on the list" even have a popular name, "Black Swans," as presented in Taleb's book of the same name (Taleb, 2010) and discussed in Haugen and Vinnem (2015) and Aven (2013). Given that common problem, it is important to communicate to decision makers the two key implications of the completeness issue: The degree of lack of completeness: (1) ... affects the degree to which the risk assessment underestimates the risk; and (2) ... affects the degree to which risk reduction measures that address Black Swans, such as measures increasing resilience and robustness, are implicitly or explicitly undervalued relative to other risk reduction measures.

Revisiting Sections 3.6.1–3.6.3: It may seem like overkill to have broken out issues of completeness into those three considerations. But in fact, in our experience we have found that completeness may often be an extremely important issue, in part because it can result in under-assessments of risk where the degree of under assessment, and the implications of that under assessment, are not communicated at all to the risk management decision making process. In addition, the three different considerations represent three significantly different ways that the issue of completeness can affect the effectiveness of a risk analysis.

3.6.4. Adversary decisions

3.6.4.1. Test. If part of the risk generating mechanism involves adversary decisions, including in some cases adversary adaptation to newly deployed defenses, are those decisions fully considered and accounted for?

3.6.4.2. Shortfall. If adversary decisions are in fact a significant part of the risk generation process, and not accounted for, that can result in a significant miss-assessment of the risk. In particular if an important part of the risk generation process involves an adversary shifting his attack planning in response to newly deployed defenses, then a risk analysis not accounting for that will overestimate the risk reduction of those newly deployed defenses. This issue was emphatically pointed out in the 2008 National Research Council's review of the Biological Terrorism Risk Assessment model, BTRA (2008).

3.6.5. Uncertainty

Uncertainty may seem an odd issue to raise as one issue of the sixteen listed here, given that the "whole point" of risk analysis is to take uncertainty into account, but we have found cases where uncertainty is simply not adequately quantified or communicated to the decision makers.

3.6.5.1. Test. Are all uncertainties adequately assessed and communicated to decision makers? That includes both aleatory and epistemic uncertainties, correlations in uncertainty when that is important, and adequately labelling metrics of uncertainty, e.g. error bars.

3.6.5.2. Shortfalls. We have found cases where there are error bars, but in fact those error bars only represented some, not all, of the uncertainty. Another common problem is not accounting for correlation among errors. For example, one assessment ranked many different weapons by overall risk, including highly overlapping error bars, but with no clear indication of the correlations among those errors. If they were highly correlated, then the ranking would be much more robust than if they were uncorrelated. So a decision maker concerned with the robustness of the ranking was not provided an important element of information.

3.7. Risk analysis report, communication with metrics

As indicated in Fig. 1, we have identified four aspects of risk analysis reports that constitute bases for significant validity (as defined here) issues. That is, in our experience we have found the four listed issues to be significant issues of concern in the validity of risk analyses in their role of advising risk management decisions. Three of the four explicitly concern risk metrics, and the fourth one also concerns the concept of risk metrics more broadly defined, i.e. mechanisms to communicate risk. Risk metrics are central to risk analysis, as discussed in Johansen and Rausand (2014). They define risk metrics: "The term risk metric is interchangeably used with risk measure and has been defined as 'a mathematical function of the probability of an event and the consequences of that event' (Jonkman et al., 2003). A risk metric should hence have a probability and consequence element and relate to the occurrence of one or more hazardous events." Though as discussed here, we take a broader view of risk metrics as mechanisms to communicate risk not limited to the probability, consequence and occurrence concepts just listed.

3.7.1. Metric validity

3.7.1.1. Test. Are the metrics metrically valid?

Please cite this article in press as: Lathrop, J., Ezell, B. A systems approach to risk analysis validation for risk management. Safety Sci. (2017), http://dx.doi.org/10.1016/j.ssci.2017.04.006

3.7.1.2. Shortfall. This may be a case where we might assume that all risk analyses of course deliver valid metrics, but we know of one significant, widely used case where the metrics include bar charts that are sums of products of ordinal measures, clearly metrically invalid. One of us tested those bar charts with the actual data, and found bar chart rank reversals upon allowable data metric transformations.

3.7.2. Meaningful metrics

3.7.2.1. Test. Are the metrics meaningful to the perhaps non-specialist decision maker, and are the metrics specifically designed to support the risk management decisions considered?

3.7.2.2. Shortfall. Some risk management decisions should be based on metrics such as individual risk and changes in that individual risk. Other risk management decisions should be based on population risk and changes in that risk. Others considerations include attribution, known vs. statistical fatalities, voluntariness of risk, media spectacle of a risk, the impact of dread, and others. In some cases a key issue is robustness and/or resilience. Section 3.6.5 discussed several issues of meaningful metrics concerning uncertainty, which we won't repeat here but that apply here also. In many if not most cases, there is a need for multiple metrics to address multiple metric issues such as those listed here. There is no space, here for a detailed discussion of those considerations, but the point here is that the risk analysis should not depend on the decision making process to be knowledgeable of all of those considerations. Rather, the risk analyst should fully consider all of those considerations and design metrics, including the visual formats for those metrics, to bring those perhaps several considerations most clearly into the minds of the perhaps non-specialist decision makers.

3.7.3. Caveats

3.7.3.1. Test. Are the metrics adequately caveated accounting for all of the concerns listed here in the upstream boxes, the preceding subsections of Section 3, and are those caveats summarized in immediate proximity to the metrics and associated graphics?

3.7.3.2. Shortfalls. In one important case in our experience, both issues of the test applied. The assumptions (and all of the other concerns listed above in Section 3) were not translated into the associated caveats to be taken into account when reading the metrics and graphics, and the identified assumptions were listed in one part of the report, but those assumptions were not summarized in immediate proximity to the associated results graphics. In the real world, most decision makers are not apt to adequately take into account the implications of assumptions (or any of the other issues of the above subsections in Section 3) and are even more not apt to adequately take into account those implications when they are several tens of pages separated from the results presented as graphics of metrics.

3.7.4. Full disclosure

This is another version of the caveats of Section 3.7.3, reworded here from another perspective:

3.7.4.1. Test. Any risk analysis must represent a list of compromises between an ideal risk analysis and what is possible given the realities of finite budgets, schedules, and all of the issues listed above in Section 3. The important point raised here is that the results should include full disclosure of all of those realities, and their implications for interpreting the results. Rae et al. (2014) address full disclosure issues from the perspective of their maturity model framework.

3.7.4.2. Shortfall. We can make this simple. We have found no case, in all of our experience, where the final report included full disclosure of all of the realities of Section 3 and their implications for interpreting the results.

3.8. Transparency, documentation

3.8.1. Transparency, documentation: Analytic

3.8.1.1. Test. Is the transparency and documentation of the risk analysis and the risk analysis process sufficient for a third party review to ask the tests of Sections 3.1–3.7.4 from the perspective of methodological validity? The assumption is that a third party review would be staffed by methodologically knowledgeable analysts. This test holds even for cases where no third party review is actually conducted, or where that review is conducted at an informal level.

3.8.1.2. Shortfall. In the third party reviews we have been part of, in some cases we found that we could only ask the questions of Sections 3.1–3.7.4 in a person to person discussion – the documentation was not adequate to support those tests. In those cases the degree of validity could only be established by a third party review and depended on the competence and completeness of that review, then finally adequate documentation could only be generated by that third party review. The net result was that there was an inadequate completion of the risk analysis in the Analysts' Domain. The risk analysis should be completed, including full documentation, within the Analysts' Domain without depending on an external third party review.

3.8.2. Transparency, documentation: Narrative

This section is the same as Section 3.8.1 with the exception that the transparency and documentation should be sufficient for a stakeholder review, where those stakeholders are representatives of stakeholder interests and may not be knowledgeable in risk assessment methodology. That is, the transparency and documentation should be sufficient for stakeholders to ask the tests of Sections 3.1–3.7.4 from the point of view of stakeholder interests. As with transparency and documentation – analytic, this test holds even for cases where no stakeholder review is actually conducted, or where that review is conducted at an informal level. To summarize: This transparency and documentation differs from the Section 3.8.1 transparency and documentation two ways:

(1) This documentation must present its results in formats understandable to stakeholders who may not be knowledgeable in risk assessment methodology.
(2) Stakeholders are typically concerned about whether or not their interests are adequately considered and considered in a balanced way. That may call for documentation of the methodology and results specifically tailored to address those concerns.

3.9. Metrics of overall risk management effectiveness of a risk analysis

The previous subsections of Section 3 have presented tests for each of the sixteen analysts'-domain elements presented in Fig. 1. In that discussion we have not raised the issue of scoring how well each element is addressed in a risk analysis. One could argue for some sort of scoring, e.g. on a 0–10 scale, for each test. But we would argue against that. We have ample experience in scoring, in particular multiattribute utility analysis (Keeney and Raiffa, 1993). Based on that experience, we don't feel that the concepts presented here should be treated in a scoring context. Two reasons: (1) Validity is not a matter of scorable degree. It is a matter of text descriptions of how well the analysis meets the test of

Please cite this article in press as: Lathrop, J., Ezell, B. A systems approach to risk analysis validation for risk management. Safety Sci. (2017), http://dx.doi.org/10.1016/j.ssci.2017.04.006

each element, to be evaluated at that text level. (2) Risk analyses, and their contexts, vary over such a wide range of considerations, it is infeasible to devise a scoring system such that, e.g., a "7" in one risk analysis is meaningfully and defensibly comparable to a "7" in another analysis. There may be confusion, here, between the concept of scoring as we discuss it here, and Rae et al.'s (2014) discussion of maturity levels. Their four maturity levels are actually levels of importance of flaws, and are meant to prioritize improvements in a risk assessment and in risk assessment research.

4. Users' Domain

This section, though quite important, is quite short. All we do, here, is explain the upper right portion of Fig. 1, the Users' Domain, in terms of what it means for requirements for the risk assessment as that is generated in the Analysts' Domain. The functions in the Users' Domain are functions of other parties and processes, other than the analysts conducting the risk analysis. Yet as explained earlier, the analysts' responsibilities are to support all of those functions, as represented by the three arrows in Fig. 1 crossing the boundary from Analysts' Domain to Users' Domain. Those functions:

(1) To support the Risk Management Decision Making Process, which takes the results of the risk analysis and uses them in its determination of risk management actions. That process typically involves many considerations not addressed in the risk analysis. Then that process seeks to implement those risk management actions, presented in Fig. 1 as Desired Implementation of Risk Management. Then that desired implementation only results in actual implementation depending on the degree to which it is accepted by implementers and stakeholders. That acceptance is structured in Fig. 1 as a process of third party review (either formal or informal), stakeholder review (either formal or informal), then the resulting demonstrated validity and degree of trust, which in turn leads to the degree of acceptance of the desired implementation of risk management, i.e. the actual implementation of risk management actions. Expressing that in terms of the other two arrows crossing the Analysts' Domain – Users' Domain boundary in Fig. 1.
(2) To support a Third Party Review, which can be either formal or informal. That is important for demonstrated validity and degree of trust, which in turn leads to implementers' and stakeholders' acceptance.
(3) To support a Stakeholder Review, which can be either formal or informal. Exactly as with the Third Party Review, that is important for demonstrated validity and degree of trust, which in turn leads to implementers' and stakeholders' acceptance.

Finally, the actual implementation of risk management actions results in what actually matters, which is the final consequences and residual risk. As explained earlier, this is a quite cursory, notional treatment of the Users' Domain. That is because it would take an entire other paper or even book to explain the Users' Domain at a more complete level, and in fact for the purposes of this paper we only need to explain the Users' Domain at this cursory level to explain the requirements for validity in the Analysts' Domain.

5. Analysis community domain: Culture of analysis quality

This section presents a different perspective than the sections on Analysts' Domain and Users' Domain. We have found more than one instance where the shortcomings of a risk analysis have been recognized, but the commissioners of the analysis have not been willing to allocate funds to correcting those shortcomings. The clear answer is that there is a need for a "Culture of Analysis Quality," that is, a culture of analysts and (perhaps more importantly) commissioners of analyses where validation standards are widely recognized and agreed upon, and where risk analysis commissioners insist that any risk analysis they commission meet those standards, and where they provide adequate funding and a schedule such that those standards can be met. In Section 3 we suggest one draft of those standards, worded as tests or questions covering the list of elements: Scope Judgments, Assumptions, Data, SMEs, SME Elicitation, Risk Analysis (in particular the five topics listed in Section 3.6), the reporting of that analysis (in particular the four topics in Section 3.7), and the analytic and narrative transparency and documentation points of Section 3.8. The point of this section is not limited to the particular standards suggested in Section 3, but simply some agreed upon set of standards. In fact other frameworks for standards can be inferred from Goerlandt et al. (2016), Rae et al. (2014) and Rouhiainen (1992). At another level, that culture should be pervasive and accepted enough such that any analysis or model should have to be tested and found in compliance with those standards if it is to be used to support a significant decision. We have one possible framework for that process in Department of Defense Instruction 5000.61, which sets policy that "Models, simulations and associated data used to support DoD processes, products and decisions shall undergo verification and validation (V&V) throughout their life cycles …. And shall be accredited for an intended use," then references Military Standard 3022 as providing "suggested templates for documenting VV&A that are tailored to the application" (Department of Defense, 2008, 2009).

6. How can we make that culture of analysis quality happen?

We include this section as a section separate from Section 5 to highlight the problem of "How can we make that culture of analysis quality happen?" and to separate out an issue where we are, frankly, unclear on the answer. We have our experience base for Sections 1–5, but we pose the question of this Section 6 as an open question for the readers of this journal. We began this paper citing ISO 15288, and there is also ISO 31000, Risk management – Principles and Guidelines (International Organization for Standardization, 2009, 2015). In addition, as mentioned, one of us is coordinator and majority author of a document under development in the Society for Risk Analysis: Principles, Guidelines and Core Knowledge for Analytic Support of Risk Management. But all of those are simply standards and standards-like documents. None of them, by themselves, can bring about the Culture of Analysis Quality we describe here. One promising idea, suggested by a reviewer, would be to perform research to show possible negative risk management outcomes if that "analysis quality" is lacking. Though one of us did exactly that, showing how a methodological flaw in one major risk management study could lead to rank reversals in assessed risks of different risk elements. But that research did not lead to changes in that study. That is, in that example, we found that research alone, while it might help, may not be sufficient. We admit that it is quite challenging to identify how to create such a culture.

7. Summary and conclusion

We began this paper presenting the logical sequence upon which our reasoning is based: Validation of a risk analysis should be based on how well the risk analysis supports risk management,

Please cite this article in press as: Lathrop, J., Ezell, B. A systems approach to risk analysis validation for risk management. Safety Sci. (2017), http://dx.doi.org/10.1016/j.ssci.2017.04.006

and assessing how well the risk analysis supports risk management should include considering how well it supports the actual, real-world decision making process, and the roles trust, third party review and stakeholder review and acceptance may play in that. Then we reworded that logic: Validation for risk management must be considered from a systems perspective, and presented that system in Fig. 1. That figure, in turn, provided the structure for a series of eight functions of risk analysis, and within those functions sixteen tests of validity, each test worded as a question. Each test is paired with a discussion of the shortfalls associated with failing that test. That Fig. 1 is divided into three domains: Analysts' Domain, Users' Domain and Analysis Community Domain. The Analysts' Domain is the one domain implicitly assumed in any discussion of risk analysis validation we have seen. We add to that perspective the fact that there is a Users' Domain separating any risk analysis from the final consequences and residual risk, which is the only thing that actually matters. That Users' Domain includes the risk management decision process, desired implementation of risk management actions, and the actual risk management actions achieved based on the acceptance or denial of those actions, where that acceptance/denial is a result of implementers and stakeholder acceptance, which in turn is a result of the risk analysis report and its transparency and documentation. Underlying both the Analysts' and Users' Domains is the Analysis Community Domain, which could provide a "Culture of Analysis Quality" where the sixteen validation tests listed here would be enforced by both risk analysts and commissioners of risk analyses. Though we end by admitting that it is quite challenging to identify how to create that culture.

In conclusion, we have here presented a very broad, insightful systems perspective on the matter of how to establish risk analysis validity, including sixteen concrete tests of validity.

The authors are grateful to the four anonymous reviewers for their useful comments and suggestions to the original version of this paper. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Aven, T., 2013. On the meaning of a black swan in a risk context. Saf. Sci., 44–51
Aven, T., 2016. Ignoring scenarios in risk assessments: understanding the issue and improving current practice. Rel. Eng. Syst. Saf. 145, 215–220.
Aven, T., Heide, B., 2009. Reliability and validity of risk analysis. Rel. Eng. Syst. Saf. 94, 1862–1868.
Bunn, M., 2006. A mathematical model of the risk of nuclear terrorism. Ann. Am. Acad. Polit. Soc. Sci. 607 (1).
Caskey, S., Ezell, B., Dillon-Merrill, R., 2013. Global, chemical, biological, and nuclear threat potential prioritization model (G-CBN-TP). Bioterror. Biodef. 4 (1).
Department of Defense, 2008. Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations, MIL-STD-3022.
Department of Defense, 2009. Instruction 5000.61, DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A).
Ezell, B., Collins, A., 2010. Response to Parnell, Smith, and Moxley, intelligent adversary risk analysis: a bioterrorism risk management model. Risk Anal. 30 (6).
Ezell, B., Dillon-Merrill, R., Parnell, G., Lathrop, J., Barrett, A., Krempley, M., 2011. Methodology Enhancements to Risk Assessment Process for Informed Decision Making (RAPID). Innovative Decisions, Inc., Report, Contract # HSHQDC-10-P-00253.
Ezell, B., von Winterfeldt, D., 2009. Probabilistic risk analysis and bioterrorism risk. Biosecur. Bioterror. 7 (1).
Goerlandt, F., Khakzad, N., Reniers, G., 2016. Validity and validation of safety-related quantitative risk analysis: a review. Saf. Sci. http://dx.doi.org/10.1016/j.ssci.2016.08.013.
Haugen, S., Vinnem, J.E., 2015. Perspectives on risk and the unforeseen. Rel. Eng. Syst. Saf. 137, 1–5.
Hubbard, D., 2010. How to Measure Anything. Wiley.
International Organization for Standardization, 2009. ISO 31000:2009, Risk Management – Principles and Guidelines. <http://www.iso.org/iso/home/standards/iso31000.htm> (Last accessed 4-4-17).
International Organization for Standardization, 2015. ISO/IEC/IEEE 15288:2015, Systems and Software Engineering – System Life Cycle Processes. <http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=63711> (Last accessed 4-4-17).
Johansen, I.L., Rausand, M., 2014. Foundations and choice of risk metrics. Saf. Sci. 62, 386–399.
Jonkman, S., van Gelder, P., Vrijling, J., 2003. An overview of quantitative risk measures for loss of life and economic damage. J. Hazard. Mater. 99, 1–30.
Kaplan, S., 1997. The words of risk analysis. Risk Anal. 17 (4).
Keeney, R., Raiffa, H., 1993. Decisions with Multiple Objectives. Cambridge University Press.
Lathrop, J., Ezell, B., 2016. Validation in the absence of observed events. Risk Anal. 36 (4).
Lugar, R., 2005. The Lugar Survey on Proliferation Threats and Responses. Senate Foreign Relations Committee.
Mueller, J., 2010. Atomic Obsession. Oxford University Press.
National Research Council, 2008. Department of Homeland Security Bioterrorism Risk Assessment: A Call for Change. National Academies Press.
Petty, M., 2010. Verification, validation, and accreditation. In: Banks, C., Sokolowski, J. (Eds.), Modeling and Simulation Fundamentals: Theoretical Underpinnings and Practical Domains. Wiley.
Pillai, S., 2011. Taking Measure of Countermeasures (Part 1). <https://www.dhs.gov/news/2011/04/14/written-testimony-st-house-homeland-security-subcommittee-emergency-preparedness> (Last accessed 4-4-17). Department of Homeland Security.
Rae, A., Alexander, R., McDermid, J., 2014. Fixing the cracks in the crystal ball: a maturity model for quantitative risk assessment. Reliab. Eng. Syst. Saf. 125, 67–81.
Rouhiainen, V., 1992. QUASA: a method for assessing the quality of safety analysis. Saf. Sci. 15, 155–172.
Sargent, R., 2013. Verification and validation of simulation models. J. Simul. 7.
Taleb, N., 2010. The Black Swan. Random House.
United States Coast Guard, 2006. Commandant Instruction 5200.40, Verification, Validation and Accreditation (VV&A) of Models and Simulations (M&S).
United States Coast Guard, Port Security Evaluation Division, 2017. Maritime Security Risk Analysis Model. <http://aapa.files.cms-plus.com/PDFs/MSRAMBrochureTrifold.pdf> (Last accessed 4-4-17).
World Nuclear Association. Fukushima Accident, 2017. <http://www.world-nuclear.org/information-library/safety-and-security/safety-of-plants/fukushima-accident.aspx> (Last accessed 4-4-17).